On COVID-19 Modelling

Robert Schaback[1]

June 2, 2020

Abstract. This contribution analyzes the COVID-19 pandemic by comparably simple mathematical and numerical methods. The final goal is to predict the peak of the epidemic outbreak per country with a reliable technique. The difference to other modelling approaches is to stay extremely close to the available data, using as few hypotheses and parameters as possible. This is done by an algorithm motivated by standard SIR models that directly works with the standard data provided by the Johns Hopkins University. But this model not sufficient to deal with the hidden part of the pandemics. To reconstruct data for the unregistered Infected, a second algorithm uses current experimental values of the infection fatality rate and a data-driven estimation of a specific form of the recovery rate. All other ingredients are data-driven as well. Various examples of predictions are provided for illustration. They show how dramatic the predictions during the uncontrolled outbreak were, how the situation changed by political and social counteraction, and what countries expecting their infection peak possibly have to face. In the beginning, the basic notions of modelling epidemics are collected, for the convenience of readers.

1 Introduction and Overview

During an epidemic outbreak like COVID-19, everybody wants to know how hard the impact will be. In particular:

- What is the health risk for me, my family, our friends, the city, the country, and the world?
- Is the health system prepared properly?
- Should households fill up their reserves in time?

[1]Address:
Prof. Dr. R. Schaback, Institut für Numerische und Angewandte Mathematik,
Universität Göttingen, Lotzestraße 16-18, 37083 Göttingen
schaback@math.uni-goettingen.de, http://num.math.uni-goettingen.de/schaback/

This is a situation that asks for mathematics, like in the old times when mathematicians were needed to predict floods or solstices. Such predictions should be based on data and arguments, and they should provide well-supported suggestions for what to do. To understand the process and to make predictions, it should be modelled, and the model should be computable. Then predictions will be possible, and reality will decide later whether the model and the predictions were useful. Many models are possible, and the approach presented here is just one of them. The specific goal is to stay as close as possible to the available data, but it turns out that the available data are not directly usable for the standard models that give the basic understanding. To this end, two extensions to the standard SIR model are developed that get closer to the available data and finally are able to make data-driven predictions.

The beginning is made in section 2 with an introduction to standard terms like *Basic Reproduction Number*, *Herd Immunity Threshold*, and *Doubling Time*, together with some critical remarks on their abuse in the media. These notions are based on the standard SIR model for epidemics. Experts can skip over this completely. Readers interested in the predictions should jump right away to section 5. For simplicity, the presentation ignores all delay-related issues like *incubation period* and *serial interval* in the theoretical model of Section 2.

To bridge the gap between model and data, Section 3 describes the Johns Hopkins data source with its limitations and flaws, and then presents a variation of a SIR model that can be applied directly to the data. It allows to estimate basic parameters, including the Basic Reproduction Number, but does not work for predictions of peaks of epidemics.

To achieve the latter goal, section 4 combines the data-compatible model of section 3 with a SIR model dealing with the unknown Susceptibles and the unregistered Infectious. This needs extra parameters that should be extracted from the literature. The first is the *infection fatality rate*, as provided e.g. by an der Heiden/Buchholz [2], Streeck et al. [21], Verity et al. [22]. Section 4.3.1 pairs it with the *case fatality rate* and shows how the latter can be deduced from the Johns Hopkins data. Like in Bommer/Vollmer [3], their combination gives a detection rate for the confirmed cases.

Section 4.4 introduces the second additional parameter: a recovery rate that can be directly used in the model and estimated from the infection fatality rate and the observable case fatality and case death rates. However, this parameter is not

needed for prediction, just for determination of the unknown variables from the known data as long as the latter are available.

Then section 5 combines all of this into an algorithm that makes predictions under the assumption that there are no further changes to the parameters by political action. It estimates the parameters of a full SIR model from the available Johns-Hopkins data by the techniques of section 4, using two additional technical parameters: the number of days used backwards for estimation of constants, and the number of days in which recovery or death must be expected, for estimation of case fatality and recovery rates. This is where time delays enter, but not into the model, only into internal estimation procedures. After the data-driven estimation of these parameters, the prediction uses only the infection fatality rate. All other ingredients are derived from the Johns Hopkins data.

Results are presented in section 5. Given the large uncertainties in the Johns-Hopkins data, the predictions are rather plausible. However, reality will have the final word on this prediction model.

The paper closes with a summary and a list of open problems.

2 Classical SIR Modeling

This contains the basic notions that are useful for modelling epidemics, and that were in use in the media during the COVID-19 outbreak. Among other things, there will be a rigid mathematical underpinning of what is precisely meant by

- *flattening the epidemic outbreak (mitigation)*,
- *basic reproduction number*,
- *Herd Immunity Threshold*, and
- *doubling time*,

pointing out certain abuses of these notions. This will not work without calculus, but things were kept as simple as possible. Readers from outside the mathematics community should take the opportunity to brush up their calculus knowledge. Experts should go over to section 3.

2.1 The Model

The simplest standard "SIR" model of epidemics (e.g. Hethcote [10], and easily retrievable in the wikipedia [23]) deals with three variables

1. Susceptible (S),
2. Infectious (I), and
3. Removed (R).

The Removed cannot infect anybody anymore, being either dead or immune. This is the viewpoint of bacteria or viruses. The difference between death and immunity of subjects is totally irrelevant for them: they cannot proliferate anymore in both cases. The SIR model cannot say anything about death rates of persons.

The Susceptible are not yet infected and not immune, while the Infectious can infect Susceptibles. The three classes S, I, and R are disjoint and add up to a fixed total population count $N = S + I + R$. All of these are ideally assumed to be smooth functions of time t, and satisfy the differential equations

$$\begin{aligned} \dot{S} &= -\beta \frac{S}{N} I, \\ \dot{I} &= +\beta \frac{S}{N} I - \gamma I, \\ \dot{R} &= \gamma I. \end{aligned} \tag{1}$$

where the dot stands for the time derivative, and where β and γ are positive parameters. The product $\frac{S}{N} I$ models the probability that an Infectious meets a Susceptible.

Managing an SIR epidemic means modifying the constants β and γ. This is why one should see the parameters as control variables, and we shall treat them even as time series from section 3 on.

Note further that the Removed of the SIR model are not the Recovered of the Johns Hopkins data that we treat later, and the SIR model does not account for the Confirmed counted there. Similarly, there is no direct relation to the data published by the Robert Koch Institute. It is a major problem to match models with the available data, and we shall explain the latter to some detail in section 3.

2.2 Other Models

In many publications concerning COVID-19 (e.g. an der Heiden/Buchholz [2], Dandekar/Barbasthatis [4], De Brouwer et al. [5], Friston et al. [8], Khailaie et al. [14], Kucharski et al. [15], Maier/Brockmann [16]), the SIR model is extended by Exposed E that are infected, but not (yet) infectious. This introduces an additional parameter and would require dealing with a latency delay properly. We avoid this complication to keep the model as simple as possible. Note that there are extensions of SIR models with 14 to 21 parameters, e.g. (Friston et al. [8], Giordano et al. [9], Khailaie et al. [14]. Fitting model parameters in the above papers is partially done numerically and partially by Bayesian approaches using Monte Carlo sampling. Here, we avoid fitting and time delays as far as possible.

Conceptionally different are the individual-based transition model for contacts of Ferguson et al. [7], and the approach of Mohring et al. [17] working consistently with time delays.

2.3 Simple Properties of the SIR Model

Since $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$, the equation $N = S + I + R$ is kept valid at all times. The term $\beta \frac{S}{N} I$ moves Susceptibles to Infectious, while γI moves Infectious to Removed. Thus β represents an *infection rate* while the *removal rate* γ accounts for either healing or fatality after infection, i.e. immunity. Political decisions about reducing contact probabilities will affect β, while γ resembles the balance between the medical aggressivity of the infection and the quality of the health care system.

As long as the Infectious I are positive, the Susceptibles S are decreasing, while the Removed R are increasing. Excluding the trivial case of zero Infectious from now on, the Removed and the Susceptible will be strictly monotonic.

The SIR model is not really dependent on the total population N. Moreover, if we scale time by $\tau := t \cdot \gamma$ and go over to *relative* quantities

$$\begin{aligned} s(\tau) &:= \frac{S(\tau/\gamma)}{N}, \\ r(\tau) &:= \frac{R(\tau/\gamma)}{N}, \\ i(\tau) &:= \frac{I(\tau/\gamma)}{N}, \end{aligned}$$

we get the new system

$$\begin{aligned} \frac{ds}{d\tau} &= -\frac{\beta}{\gamma}s(\tau)i(\tau) &&= -R_0 s(\tau)i(\tau) \\ \frac{di}{d\tau} &= \left(\frac{\beta}{\gamma}s(\tau)-1\right)i(\tau) &&= (R_0 s(\tau)-1)\,i(\tau) \\ \frac{dr}{d\tau} &= i(\tau) \end{aligned} \tag{2}$$

only containing the *Basic Reproduction Number*

$$R_0 := \frac{\beta}{\gamma} \tag{3}$$

that will turn out to be of central importance. Both β and γ vary under a change of time scale, but the basic reproduction number is invariant. Physically, β and γ have the dimension $time^{-1}$, but $R_0 = \beta/\gamma$ and the new "time" parameter τ are dimensionless. Another interpretation is that after a time scale one can assume $\gamma = 1$ and $R_0 = \beta$. The new "time" scale lets removal take one unit of "time". We shall call it *unit removal parameter*, and use a prime to denote derivatives with respect to τ. But in all sections that make real-world interpretations, we have to use real time, and then we shall go back to (1).

Another standard mathematical trick is to divide the first equation by the third to arrive at

$$\begin{aligned} \frac{ds}{dr} &= -R_0 s, \\ s(r) &= s(r(0))\exp(-R_0(r-r(0))). \end{aligned} \tag{4}$$

We shall use (4) in section 2.11 to study the long-term behaviour of solutions. The introduction of (4) is a typical pitfall for mathematics: it is a nice theoretical simplification, but it obscures the most interesting practical aspect, in this case the number I of infectious persons.

Figures 1 and 2 show some simple examples that will be explained in some detail below.

2.4 Interpretation of the Basic Reproduction Number R_0

Media often say that R_0 gives the number of persons an average Infectious infects while being infectious. This is a rather mystical statement that needs underpin-

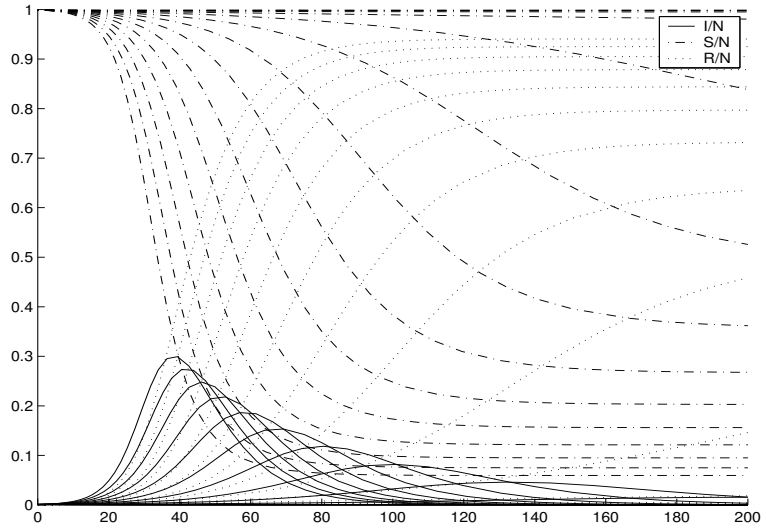

Figure 1: Some typical SIR system solutions, relative to the total population N, with small $I(0)$ and varying R_0.

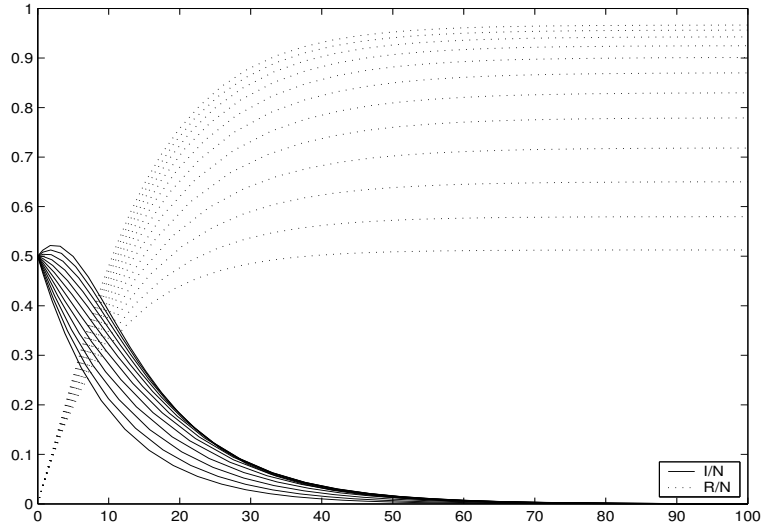

Figure 2: Some other typical SIR system solutions, now with $I(0) = N/2$.

ning. The quantity

$$\frac{1}{\gamma} = \frac{I}{\dot{R}}$$

is a value that has the physical dimension of time. It describes the ratio between current Infectious and current newly Removed, and thus can be seen as the average time needed for an Infectious to get Removed, i.e. it is the average time that an Infectious can infect others. Correspondingly,

$$\dot{I} + \gamma I = \dot{I} + \dot{R} = \beta \frac{S}{N} I$$

are the newly Infected, and therefore

$$\frac{1}{\beta} \frac{N}{S} = \frac{I}{\dot{I} + \dot{R}}$$

can be seen as the time it needs for an average Infectious to generate a new Infectious. The ratio $\frac{\beta}{\gamma} \frac{S}{N}$ then gives how many new Infectious can be generated by an Infectious while being infectious, but this is only close to R_0 if $S \approx N$, i.e. at the start of an outbreak. A correct statement is that R_0 is the average number of infections an Infectious generates while being infectious, but within an unlimited supply of Susceptibles.

The above interpretation of the Basic Reproduction Number is used as an argument to change the parameters of the epidemic towards small R_0 by administrative action. We shall see that this is correct from a mathematical viewpoint as well, ands we shall study the influence of R_0.

On the side, the above viewpoint shows two major ways to make R_0 small: reducing the number of possible contacts to Susceptibles, and reducing the time an Infectious has to infect others. The first can be done by reducing contacts of all persons, even the Susceptibles, and the second by putting all infectious persons into strict quarantine. Both were used by various countries.

SIR-based models of the COVID-19 pandemics estimate R_0 between 2 and 6 during an outbreak (see e.g. the Robert Koch-Institute [18], De Brouwer et al. [5], and Maier/Brockmann [16]), while *non-pharmaceutical interventions* (NPI) bring R_0 below 1. We shall see examples in sections 3.3.2 and 5.2.

The use of the Basic Reproduction Number R_0 in the media suggests that large R_0 are generally serious, because each Infectious infects several people. This is only

true at the beginning of an outbreak, because then there are enough Susceptibles. But it will turn out in section 2.7 that the Infectious will always finally go to zero, whatever the Basic Reproduction Number is.

2.5 Conditions for Outbreaks

The first interesting question in a beginning epidemic is:

Will there be a serious outbreak, or will the infection disappear quickly?

Therefore we start by looking at the initial conditions for the model. Since everything is invariant under an additive time shift, we can start at time 0, and since time scales are irrelevant either, we can go to the simplified system (2).

The relative Infectious i do not increase right from the start if

$$s(0) \leq \frac{1}{R_0}, \tag{5}$$

and then they decrease further since the Susceptibles σ must decrease and

$$(\log i(\tau))' = R_0 s(\tau) - 1 < R_0 \sigma(0) - 1 \leq 0. \tag{6}$$

There is no outbreak, and this must occur for all initial conditions if $R_0 \leq 1$. But if $R_0 > 1$, the outbreak depends on the initial condition (5). We require

$$1 > s(0) > \frac{1}{R_0} \tag{7}$$

for an outbreak, i.e. there must be sufficiently many Susceptibles. When we discuss an outbreak in what follows, we always assume $R_0 > 1$ and (7).

2.6 Herd Immunity Threshold

In connection with an outbreak, the Herd Immunity Threshold

$$HIT = 1 - \frac{1}{R_0}$$

is often mentioned. The background question is:

If an uninfected population is threatened by an infection with Basic Reproduction Number R_0, what is the number of immune persons needed to prevent an outbreak right from the start?

We can read this off equation (5) in the ideal situation that $i(0) = 0$ and $s(0) + r(0) = 1$, namely

$$s(0) = 1 - r(0) = \frac{1}{R_0}$$

implying

$$r(0) = 1 - \frac{1}{R_0}$$

as the threshold between outbreak and decay for the relative Removed. This does not refer to a whole epidemic scenario. It is a condition to be checked before anything happens, and useless within a developing epidemic, whatever the media say.

Furthermore, the Herd Immunity Threshold has nothing to do with the important long-term ratio of Susceptibles to Removed. We shall address this ratio in section 2.11.

2.7 The Peak

In the outbreak case (7), the main questions are:

- when will the Infectious reach their maximum, and
- how large the maximal value is.

In more generality, we ask for a unit removal parameter τ_I where the Infectious are positive and do not change. Then we have

$$0 = \frac{di}{d\tau}(\tau_I) = (R_0 s(\tau_I) - 1) i(\tau_I), \tag{8}$$

and the monotonicity of s implies uniqueness of τ_I and

$$s(\tau_I) = \frac{1}{R_0}. \tag{9}$$

If i would increase without reaching a maximum in finite time, the first equation of (2) would imply that s goes exponentially to zero, but then there is a τ_I with

(9), and (8) follows. Summarizing, this proves that whenever there is an outbreak by (7), there is a unique maximum of the relative Infectious i that we call the *peak* from now on.

Determining the peak is theoretically difficult, and in practice it requires good estimates for β and γ. We shall deal with this problem in the major part of this paper,

In real life it is highly important to avoid the peak situation, and this can only be done by administrative measures that change β and γ in (1) to the situation $\beta < \gamma$. This is what management of epidemics is all about, provided that an epidemic follows the SIR model. We shall see how countries perform.

In the peak situation of (8), the fraction

$$1 - \frac{1}{R_0} = 1 - s(\tau_I) = r(\tau_I) + i(\tau_I) \geq i(\tau_I) \tag{10}$$

of the relative Non-Susceptible at the peak is exactly the Herd Immunity Threshold. Thus it is correct to say that if the Immune of a population are below the Herd Immunity Threshold at startup, and if the Basic Reproduction Number is larger than one, the sum of the Immune and the Infectious will rise up to the Herd Immunity Threshold and then the Infectious will decay. This is often stated imprecisely in the media.

2.8 Examples

Figure 1 shows a series of test runs with $s(0) = 0.999$ and $r(0) = 0$ with R_0 varying from $1/5$ to 3. Due to the realistically small $i(0) = 0.001$, one cannot see the decaying cases near startup, but the tails of the peaked i curves are decaying examples by starting value, due to $s(\tau) < 1/R_0$ when started at time τ. Decreasing R_0 flattens the peaks of i. One can observe that i always dies out, while s and r tend to fixed positive levels. We shall prove this below. From the system, one can also infer that r has an inflection point where i has its maximum, since $r'' = i'$. If only r would be observable, one could locate the peak of i via the inflection point of r.

When countries change parameters by administrative actions like a shutdown, they jump to a more flat i curve, e.g. at an intersection point. Recall that the i values

are the most interesting ones, but they are usually not covered by the media. They usually focus on the cumulative number of confirmed Infectious, containing the Removed.

Figure 2 shows an artifical case with a large starting value $i(0) = 1/2 = s(0)$ and R_0 varying from 0.05 to 3. In contrast to Figure 1, this example shows cases with small R_0 properly. The essence is that the Infectious go down, whether they have a peak or not, and there will always be a portion of Susceptibles. Again, we shall prove this below.

2.9 Analyzing the Outbreak

In the beginning of the outbreak, s is near to one, and therefore

$$i' = R_0 s - 1 \approx R_0 - 1$$

models an exponential outbreak with exponent $R_0 - 1 > 0$ in unit removal parametrization, with a solution

$$i(\tau) \approx i(0) \exp((R_0 - 1)t).$$

If this is done in real time and discrete time steps Δt, one has

$$\frac{I(t + \Delta t)}{I(t)} \approx \exp((\beta - \gamma)\Delta t).$$

The severity of the outbreak is not controlled by $R_0 = \beta/\gamma$, but rather via $\beta - \gamma$. Publishing single values $I(t)$ does not give any information about $\beta - \gamma$. Better is the ratio of two subsequent values

$$\frac{I(t_2)}{I(t_1)} \approx \exp((\beta - \gamma)(t_2 - t_1)),$$

and if this gets smaller ove time, the outbreak gets less dramatic because $\beta - \gamma$ gets smaller. Really useful information about an outbreak does not consist of values and not of increments, but of increments of increments, i.e. some second derivative information. This is what the media rarely provided during the outbreak. On the other hand, publishing estimates of $R_0 > 1$ indicates the severity of an outbreak in unit removal parametrization.

2.10 Doubling "Time"

Another information used by media during the outbreak is the *doubling time*, i.e. how many days it takes until daily values double. It clearly is no time, but it is the number n in

$$2 = \frac{I(t+n\Delta t)}{I(t)} \approx \exp((\beta-\gamma)n\Delta t) = (\exp((\beta-\gamma)\Delta t)^n$$

or

$$n = \frac{\log 2}{(\beta-\gamma)\Delta\tau},$$

i.e. it is inversely proportional to $\beta-\gamma$. If political action doubles the "doubling time", if halves $\beta-\gamma$. If politicians do this repeatedly, they never reach $\beta<\gamma$, and they never escape an exponential outbreak if they do this any finite number of times. Extending the doubling time will never prevent a peak, it only postpones it and hopefully flattens it. When presenting a "doubling time", media should always point out that this makes only sense during an exponential outbreak. And it is not related to the basic reproduction number $R_0 = \beta/\gamma$, but rather to the difference $\beta-\gamma$. If R_0 is used in

$$\nu = \frac{\log 2}{R_0 - 1},$$

one gets the doubling number ν in removal time units.

2.11 Long-term Behaviour

Besides the peak in case $R_0 > 1$, it is interesting to know the portions of the population that get either removed (by death or immunity) or never come into contact with the infection. This concerns the long-term behaviour of the Removed and the Susceptibles. Figures 1 and 2 demonstrate how r and r level out under all circumstances shown, but is this always true, and what is the final ratio? And if one has addition information on the percentage of casualties within the Removed, what is total death toll in the long run?

So far, section 2.7 showed that the Infectious i always go to zero, independent of how large R_0 is, but we are still left with s and r. Going back to (4), we get

$$s(r) = s(0)\exp(-R_0 r) \tag{11}$$

when assuming $r(0) = 0$ at startup. Since r is increasing, it has a limit $0 < r_\infty \leq 1$ for $\tau \to \infty$, and in this limit

$$s(r_\infty) = s(0) \exp(-R_0 r_\infty)$$

holds, together with the condition $r_\infty + s(r_\infty) = 1$, because there are no more Infectious. The equation

$$s(0) \exp(-R_0 r_\infty) = 1 - r_\infty$$

has a unique solution in $(0,1)$ dependent on $s(0) < 1$ and R_0. See Figure 3 for illustration. The right-hand side $1 - r_\infty$ of the equation is the fixed straight line, while the dotted curves are the left-hand side for varying values of R_0. The intersection points are the values r_∞ of r that solve the equation. Looking at both sides of the equation as functions of r_∞, an increase of R_0 for fixed $s(0) < 1$ lets the intersection point move towards 1 on the r axis.

Qualitatively, we can use (11) in the form

$$R_0 = -\frac{\log(s(r)) - \log(s(0))}{r} = \frac{\log(s(0)/s_\infty)}{r_\infty} \tag{12}$$

to see that the ratio of Susceptibles to Removed decreases with R_0, but there is a logarithm involved.

All of this has some serious implications, if the model is correct for an epidemic situation. First, the Infectious always go to zero, but Susceptibles always remain. This means that a new infection can always arise whenever some infected person enters the sanitized population. The outbreak risk is dependent on the portion $s_\infty = 1 - r_\infty$ of the Susceptibles. This illustrates the importance of vaccination, e.g. against measles or influenza.

The above analysis shows that large values of R_0 lead to large relative values of Removed to Susceptible in the limit. The consequence is that systems with large R_0 have a dramatic outbreak and lead to a large portion of Removed. This is good news if the rate of fatalities within the Removed is low, but very bad news otherwise.

When politicians try to "flatten the curve" by bringing R_0 below 1 from some time on, this will automatically decrease the asymptotic rate of Removed and

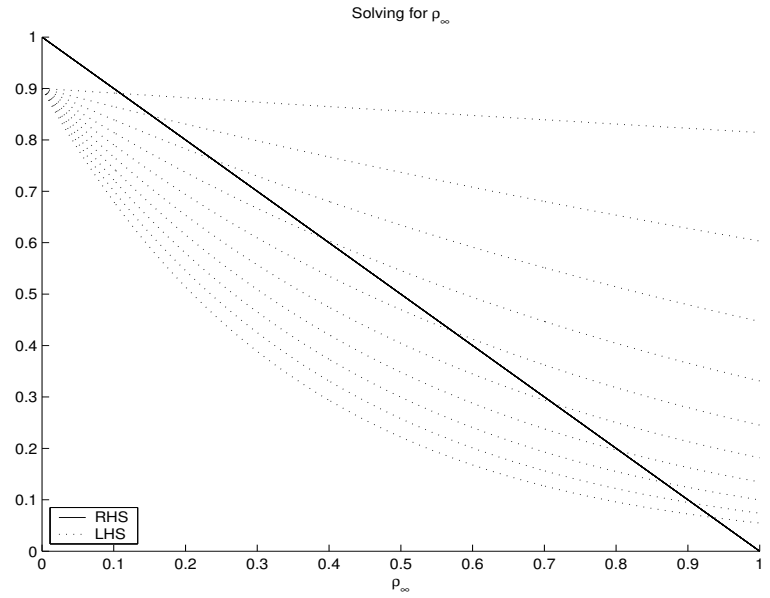

Figure 3: Solving for r_∞ for fixed $C(0) = 0.9$ and varying R_0

increase the asymptotic rate of Susceptibles in the population. This is particularly important if the rate of fatalities within the Removed is high, but by the previous argument the risk of re-infection rises due to the larger portion of Susceptibles.

The decay situation (5) implies that

$$s_\infty \leq \frac{1}{R_0}$$

and consequently

$$r_\infty = 1 - s_\infty \geq 1 - \frac{1}{R_0} = HIT.$$

Therefore the final rate of the Removed is not smaller than the Herd Immunity Threshold. This is good news for possible re-infections, but only if the death rate among the Removed is small enough.

2.12 Asymptotic Exponential Decay

If we go back to (6) for a unit removal parameter τ_D where i decreases (in an outbreak or not), we have

$$R_0 s_\infty \leq R_0 s(\tau_D) < 1$$

and then

$$i(\tau_D)\exp(-(R_0 s_\infty - 1)(\tau - \tau_D)) \leq i(\tau) \leq i(\tau_D)\exp(-(R_0 s(t_D) - 1)(\tau - \tau_D))$$

for all $\tau \geq \tau_D$. Therefore the exponential decay in unit removale parametrization is not ruled by $R_0 - 1$ as in the outbreak case with $R_0 > 1$, but rather by $R_0 s_\infty - 1$. This also holds for large R_0 because s_∞ counteracts. The bell shapes of the peaked i curves are not symmetric with respect to the peak.

2.13 Back to the Peak

If we go back to analyzing the peak of i at τ_I for $R_0 > 1$, we know

$$s(\tau_I) = \frac{1}{R_0} = s(r(\tau_I)) = s(0)\exp(-R_0 r(\tau_I))$$

and get

$$r(\tau_I) = \frac{1}{R_0}\log(s(0)R_0) \tag{13}$$

leading to

$$i(\tau_I) = 1 - s(\tau_I) - r(\tau_I) = 1 - \frac{1}{R_0} - \frac{1}{R_0}\log(s(0)R_0) \tag{14}$$

as the exact value at the maximum, improving (10). Note that the final log is positive due to the condition (7) for an outbreak.

For standard infections that have starting values $s(0) = S(0)/N$ very close to one, the maximal ratio of Infectious is

$$i(\tau_i) \approx 1 - \frac{1}{R_0} - \frac{1}{R_0}\log(R_0).$$

Figure 4 shows the behaviour of this function. A value of $R_0 = 4$ leads to a maximum of more than 40% of the population infectious at a single time. If 5%

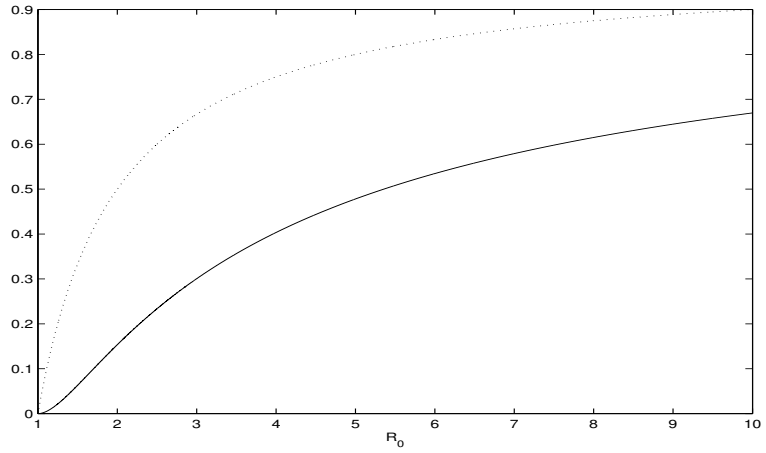

Figure 4: The effect of R_0 on the maximum rate of Infectious within the population

need hospital care, a country needs hospital beds for 2% of the population. This is what mitigation by "flattening the curve" is all about.

The dotted line leaves the log term out, i.e. is marks the rate of the Susceptibles at the peak, and by (10) the difference is the rate $r(\tau_i)$ of the Recovered at the peak. The line also marks the extreme case in (7) with $R_0 s(0) = 1$, i.e. having the smallest possible initial value of $s(0)$ for a given R_0 to generate an outbreak. Therefore all possibilities vary between the two lines.

2.14 Flattening the Curve

We first evaluate the integral

$$\int_0^\infty i(\sigma)d\sigma = \int_0^\infty r'(\sigma)d\sigma = r_\infty - r(0).$$

If we are in a peak situation (7), we can consider the *turnaround* parameter τ_T at which i comes down back to $i(0)$ behind the peak. At that point the relative number of Infectious is back to its starting value, but now the population has accumulated more Immune.

We compare areas under the i curve and get

$$i(0)\tau_T \le r_\infty - r(0) \le r_\infty \le 1, \tag{15}$$

proving that the turnaround time $t_T = \tau_T/\gamma$ has a fixed bound $t_T \le r_\infty/(i(0)\gamma)$, and this is also a crude bound on the peak time t_I. It depends on R_0 only indirectly via r_∞.

If we look at the integral left of the peak and assume $r(0) = 0$, we get

$$\int_0^{\tau_I} i(\sigma)d\sigma = \int_0^{\tau_I} r'(\sigma)d\sigma = r(\tau_I) = \frac{1}{R_0}\log(s(0)R_0).$$

from (13). This goes to zero for $R_0 \to \infty$, i.e. the left part of the full integral gets small though the value at the peak goes to 1. By additional arguments one can show that the peak position moves towards zero. The major part of the curve of the Infectious lies behind the sharp peak when R_0 is large.

Flattening the curve by bringing R_0 down is limited by (7) in the form $R_0 s(0) > 1$, if one stays in the outbreak situation. If $s(0)$ is small, one has a chance to avoid the outbreak even when $R_0 > 1$, namely by $1 \le R_0 \le 1/s(0)$.

Otherwise, when using small R_0, we are near the dotted line situation of Figure 4 near $R_0 = 1$ and know that decreasing R_0 brings the peak *value* down. But where does the peak *time* t_I or the peak parameter τ_I go?

By a detailed argument that we suppress here, there is a lower bound

$$\tau_I \ge \frac{\log(s(0)R_0)}{s(0)(R_0 - 1 - \log(s(0)R_0))} \ge \frac{\log(s(0)R_0)}{s(0)(R_0-1)}.$$

If we go into the minimal R_0 situation by setting

$$R_0 s(0) = \exp(1+\varepsilon) = 1+\delta$$

with small ε and δ, the bound turns int0

$$\tau_I \ge \frac{1+\varepsilon}{1+\delta-s(0)} = \frac{1+\varepsilon}{i(0)+\delta}$$

if $r(0) = 0$. The bound is sharp because of an additional upper bound

$$\tau_I \le \frac{\log(s(0)R_0)}{i(0)} = \frac{1+\varepsilon}{i(0)}$$

that we also suppress. This blends in with the upper bound (15) on the turnaround. If R_0 is chosen as small as possible to still guarantee an outbreak, the peak moves away from the origin at least like $1/i(0)$.

2.15 The Infection Timescale

Here is a detour that is well-known in the SIR literature. The full SIR system (1) can be written as

$$\begin{aligned} \frac{dS}{N} &= -\beta \frac{S}{N}\frac{I}{N}dt \\ \frac{dI}{N} &= \left(+\beta \frac{S}{N} - \gamma\right)\frac{I}{N}dt \\ \frac{dR}{N} &= \gamma \frac{I}{N}dt \end{aligned}$$

and in a new time variable τ with $d\tau = \frac{I}{N}dt$, one gets the system

$$\begin{aligned} \frac{ds}{d\tau} &= -\beta s(\tau), \\ \frac{dr}{d\tau} &= -\gamma \end{aligned}$$

for $s = S/N$ and $r = R/N$ as functions of the new *infection timescale* τ that one can fix as

$$\tau(t) = \int_0^t \frac{I(s)}{N} ds$$

to make sure that $\tau(0) = 0$. This implies

$$\begin{aligned} s(\tau) &= s(0)\exp(-\beta\tau), \\ r(\tau) &= r(0) + \gamma\tau. \end{aligned}$$

The beauty of this is that the roles of β and γ are perfectly split, like the roles of s and r. In the new timescale, r increases linearly and s decreases exponentially. The Basic Reproduction Number then describes the fixed ratio

$$\frac{\beta}{\gamma} = R_0 = \frac{\log S(\tau) - \log S(0)}{r(\tau) - r(0)},$$

like in (12), and the result (11) of section 2.11 comes back as

$$s(\tau) = s(0)\exp\left(-\frac{\beta}{\gamma} r(\tau)\right)$$

for the case $r(0) = 0$. This approach has the disadvantage to conceal the peak within the new timescale, and is useless for peak prediction. Of course, by a scale of the infection timescale one can set $\gamma = 1$ and $R_0 = \beta$, going to a dimensionless parametrization, as in the transition to (2).

2.16 Estimating and Varying Parameters

If real-time data for the SIR model were fully available, one could solve for

$$\begin{aligned} \gamma &= \frac{\dot{R}}{I} \\ b := \beta\frac{S}{N} &= \frac{\dot{I}+\gamma I}{I} = \frac{\dot{I}+\dot{R}}{I}, \\ \beta &= \frac{N}{N-I-R}\cdot\frac{\dot{I}+\dot{R}}{I}, \\ R_0 &= \frac{N}{N-I-R}\cdot\frac{\dot{I}+\dot{R}}{\dot{R}} = -\frac{N}{S}\frac{\dot{S}}{\dot{R}} = -\frac{1}{s}\frac{ds}{dr}, \end{aligned} \tag{16}$$

and we shall use this in section 3.3. The validity of a SIR model can be tested by checking whether the right-hand sides for β, γ and R_0 are roughly constant. If data are sampled locally, e.g. before or after a peak, the above technique should determine the parameters for the global epidemic, i.e. be useful for either prediction or backward testing.

However, in pandemics like COVID-19, the parameters β and γ change over time by administrative action. This means that they should be considered as functions in the above equations, and then their changes may be used for conclusions about the influence of such actions. This is intended when media say that "R_0 has changed". From this viewpoint, one can go back to the SIR model and consider β and γ as "control functions" that just describe the relation between the variables.

But the main argument against using (16) is that the data are hardly available. This is the concern of the next section.

3 Using Available Data

Now we want to confront the modelling of the previous section with available data. This is crucial for maneuvering countries through the epidemics (Sentker [20])[2]. Note that from now on we have work in real time and go back to (1) instead of all mathematical simplifications.

[2]Original text in German, April 16th: *Schnelle Modelle, die dem Abgleich mit der Wirklichkeit standhalten, sind eine wichtige Voraussetzung, das Land politisch durch die Seuche zu steuern.*

3.1 Johns Hopkins Data

In this text, we work with the COVID-19 data from the CSSEGISandData repository of the Johns Hopkins University [13]. They are the only source that provides comparable data on a worldwide scale.

The numbers there are

1. Confirmed (C) or *cumulative infected*
2. Dead (D), and
3. Recovered (R)

as cumulative integer valued time series in days from Jan. 22nd, 2020. All these values are absolute numbers, not relative to a total population. Note that the unconfirmed cases are not accessible at all, while the Confirmed contain the Dead and the Recovered of earlier days.

We take the data as presented, but there are many well-known flaws. In particular, the values for specific days are partly belonging to previous days, due to delays in the chains of data transmission in different countries. This is why, at some points, we shall apply some conservative smoothing of the data. Finally, there are inconsistencies that possibly need data changes. In particular, there are countries like Germany who deliver data of Recovered in a very questionable way. The law in Germany did not enforce authorities to collect data of Recovered, and the United Kingdom did not report numbers of Dead and Recovered from places outside the National Health System, e.g. from Senior's retirement homes. Both strategies have changed somewhat in the meantime, as of early May, but the data still keep these flaws.

We might assume that the Dead plus the Recovered of the Johns Hopkins data are the Removed of the SIR model, and that the Infectious $I = C - R - D$ of the Johns Hopkins data are the Infectious of the SIR model. But this is not strictly valid, because registration or confirmation come in the way.

On the other hand, one can take the radical viewpoint that facts are not interesting if they do not show up in the Johns Hopkins data. Except for the United Kingdom, the important figures concern COVID-19 casualties that are actually registered as such, others do not count, and serious cases needing hospitalization or leading to

death should not go unregistered. If they do in certain countries, using such data will not be of any help, unless other data sources are available. If SIR modelling does not work for the Johns Hopkins data, it is time to modify the SIR technique appropriately, and this will be tried in this section.

An important point for what follows is that the data come as daily values. To make this compatible with differential equations, we shall replace derivatives by differences.

3.2 Examples

To get a first impression about the Johns Hopkins data, Figure 5 shows raw data up to day 120 (May 21^{st}, as of this writing). Here, and in all plots to follow, the x axis has the days after Jan. 22^{nd}, 2020. It might be helpful to remember that day 100 is May 1^{st}. The y axis is logarithmic, because then linearly increasing or decreasing parts in the figures correspond to exponentially increasing or decreasing numbers in the real data. Many presentations in the media are non-logarithmic, and then all exponential outbreaks look similar. The real interesting data are the Infectious $I = C - R - D$ that show a peak or not. The other curves are cumulative. The data for other countries tell similar stories and are suppressed.

The media, in particular German TV, present COVID-19 data in a rather debatable way. When mentioning Johns Hopkins data, they provide C, D, and R separately without stating the most important figures, namely $I = C - D - R$, their change, and the change of their change. When mentioning data of the Infectious from the Robert Koch institute alongside, they do not say precisely that these are non-cumulative and should be compared to the $I = C - R - D$ data of the Johns Hopkins University. And, in most cases during the outbreak, they did not mention the change of the change. Quite like all other media.

One can see in Figure 5 that Germany has passed the peak of the Infectious, while France is roughly at the peak and the United States and Brazil are still in an exponential outbreak. The early figures, below day 40, are rather useless, but then an exponential outbreak is visible in all cases. This outbreak changes its slope due to political actions, and we shall analyze this later. See Dehning et al. [6] for a detailed early analysis of slope changes.

There are strange anomalies in the Recovered (green). France seems not to have delivered any data between days 40 and 58, Germany changed the data delivery

policy between days 62 and 63, and the UK data for the Recovered are a mess.

It should be noted that the available medical results on the COVID-19 disease often state that Confirmed will die or survive after a more or less fixed number of days. This would imply that the red curves for the Dead and the green curves for the Recovered should roughly follow the blue curves for the Confirmed with a fixed but measurable delay. This is partially observable, but much less accurately for the Recovered.

3.3 The Johns Hopkins Data Model

The idea is to define a model that works exclusively with the Johns Hopkins data, but comes close to a SIR model, without being able to use S. Since the SIR model does not distinguish between recoveries and deaths, we set

$$R_{SIR} \Leftrightarrow D_{JH} + R_{JH}$$

and let the Infectious be comparable, i.e.

$$I_{SIR} \Leftrightarrow I_{JH} := C_{JH} - D_{JH} - R_{JH}$$

which implies

$$(I+R)_{SIR} \Leftrightarrow C_{JH},$$

and we completely omit the Susceptibles. From now on, we shall omit the subscript JH when we use the Johns Hopkins data, but we shall use SIR when we go back to the SIR model.

Now we go back to (16) in section 2.16 and use

$$\begin{aligned} \gamma &= \frac{\dot{R}_{SIR}}{I_{SIR}} \\ &\approx \frac{(D+R)_{n+1} - (D+R)_n}{I_n} =: \gamma_n \\ b := \beta\frac{S_{SIR}}{N} &= \frac{\dot{I}_{SIR} + \gamma I_{SIR}}{I_{SIR}} = \frac{\dot{I}_{SIR} + \dot{R}_{SIR}}{I_{SIR}}, \\ &\approx \frac{C_{n+1} - C_n}{I_n} =: b_n, \end{aligned}$$

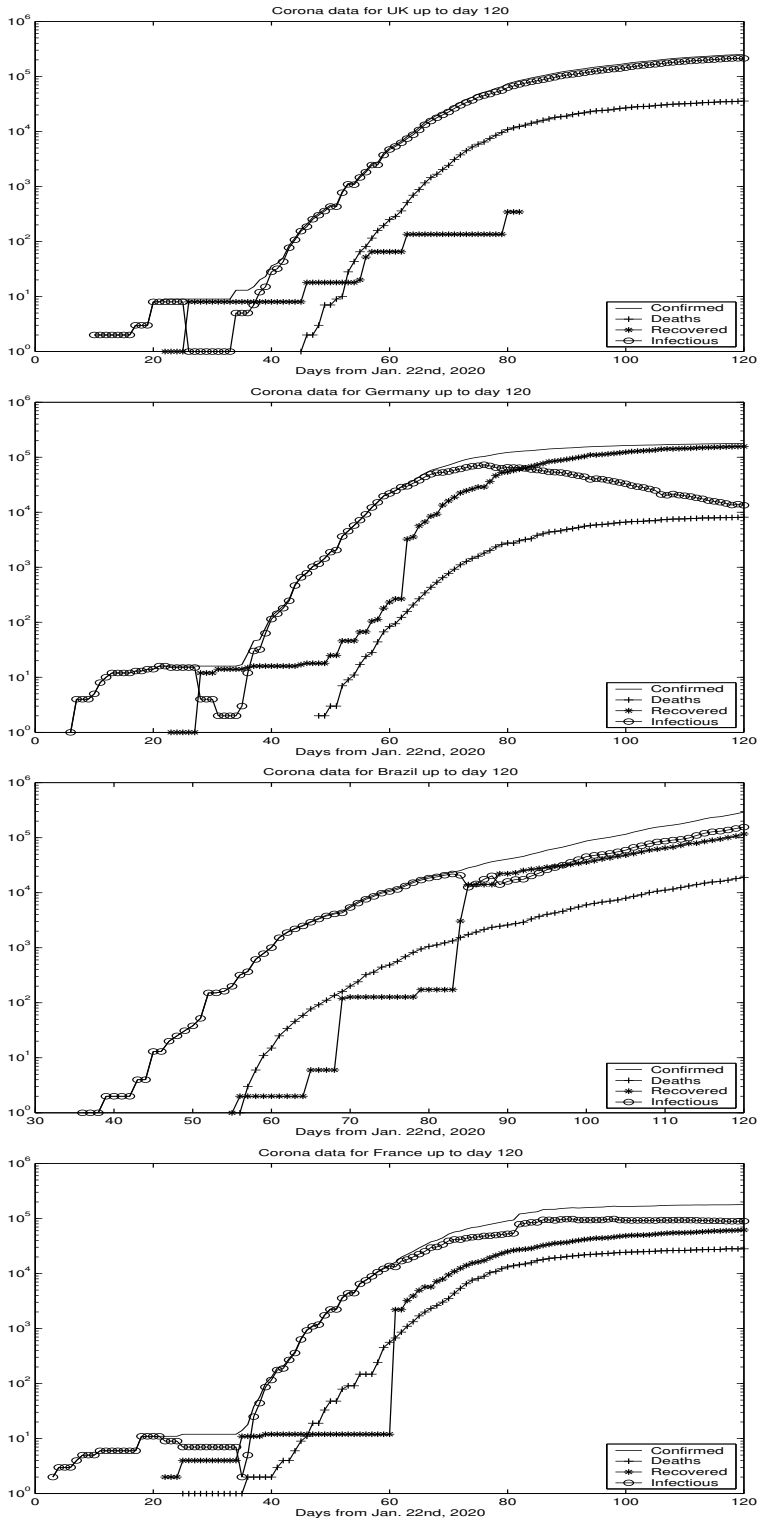

Figure 5: Raw Johns Hopkins data in logarithmic presentation up to day 120, from top: UK, Germany, Brazil, and France

defining time series γ_n and b_n that model γ and $b = \beta \cdot S_{SIR}/N$ without knowing S_{SIR}. This is equivalent to the model

$$\begin{aligned} C_{n+1} - C_n &= b_n I_n, \\ I_{n+1} - I_n &= b_n I_n - \gamma_n I_n, \\ (R+D)_{n+1} - (R+D)_n - &= \gamma_n I_n \end{aligned}$$

that maintains $C = I + R + D$, and we may call it a *Johns Hopkins Data Model*. It is very close to a SIR model if the time series b_n is not considered to be constant, but just an approximation of $\beta \cdot S_{SIR}/N$.

3.3.1 Estimating R

By brute force, one can consider

$$r_n = \frac{b_n}{\gamma_n} = \frac{C_{n+1} - C_n}{R_{n+1} + D_{n+1} - R_n - D_n} \tag{17}$$

as a data-driven substitute for

$$\frac{\beta}{\gamma} \frac{S_{SIR}}{N} = R_0 \frac{S_{SIR}}{N}.$$

Then there is a rather simple observation:

> If r_n is smaller than one, the Infectious decrease.

It follows from

$$\begin{aligned} C_{n+1} - C_n &= r_n(R_{n+1} + D_{n+1} - R_n - D_n) \\ I_{n+1} - I_n + R_{n+1} - R_n + D_{n+1} - D_n &= r_n(R_{n+1} + D_{n+1} - R_n - D_n) \\ I_{n+1} - I_n &= -(1 - r_n)(R_{n+1} - R_n + D_{n+1} - D_n), \end{aligned}$$

but this is visible in the data anyway and not of much help.

Since r_n models $R_0 \frac{S_{SIR}}{N}$, it always underestimates R_0. This underestimation gets dramatic when it must be assumed that S_{SIR} gets seriously smaller than N.

At this point, it is not intended to forecast the epidemics. The focus is on extracting parameters from the Johns Hopkins data that relate to a background SIR-type model.

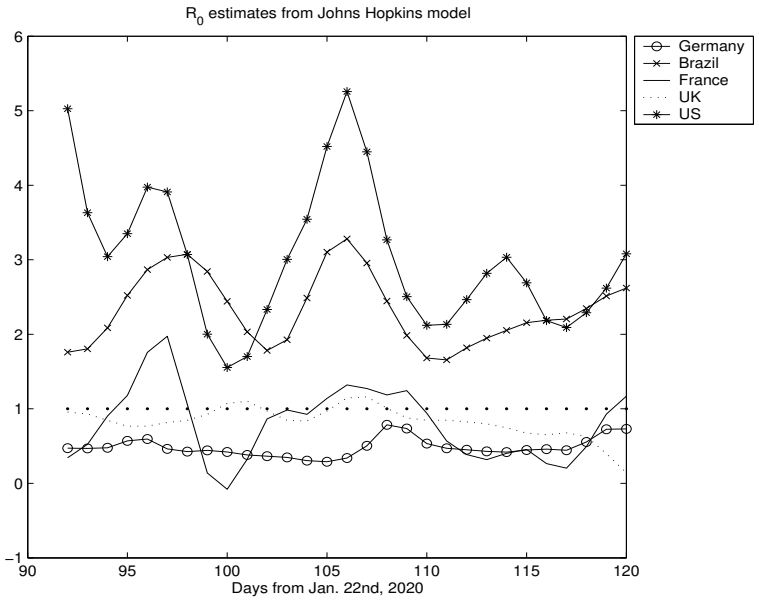

Figure 6: Estimates of R_0 via the time series r_n

3.3.2 Example

Figure 6 shows $R_0 \frac{S_{SIR}}{N}$ estimates via r_n for the last four weeks before day 120, i.e. March 21^{st}. Except for the United States and Brazil, all countries were more or less successful in pressing R_0 below one. In all cases, S_{SIR}/N is too close to one to have any influence. The variation in r_n is not due to the decrease in S_{SIR}/N, but should rather be attributed to political action. As mentioned above, the estimates for R_0 by r_n are always too optimistic.

For the figure, the raw Johns Hopkins data were smoothed by a double action of a $1/4, 1/2, 1/4$ filter on the logarithms of the data. This smoother keeps constants and linear sections of the logarithm invariant, i.e. it does not change local exponential behaviour. This smoothing was not applied to Figure 5. It was by far not strong enough to eliminate the apparent 7-day oscillations that are frequently in the Johns Hopkins data, see the figure. Data from the Robert Koch Institute in Germany have even stronger 7-day variations.

3.3.3 Properties of the Model

As long as r_n is roughly constant, the above approach will always model an exponential outbreak or decay, but never a peak, because the difference equations are linear. It can only help the user to tell if there is a peak ahead or behind, depending on $r_n \approx R_0$ being larger or smaller than 1. If r_n is kept below one, the Confirmed Infectious will not increase, causing no new threats to the health system. Then the S/N factor will not decrease substantially, and a full SIR model is not necessary.

On the other hand, if a country manages to keep r_n smaller than some $r = \frac{b}{\gamma} < 1$, it is clear that it takes

$$j = \frac{\log(I_n)}{-\log(1+b-\gamma)}$$

steps to bring the Confirmed Infectious down to 1, if we assume b and γ to be constant. Making $R_0 \approx \frac{b}{\gamma} < 1$ small is not the best strategy. Instead, one should maximize $\gamma - b$.

As long as countries keep r_n clearly below one, e.g. below 1/2, this would mean that $R_0 \approx r_n \frac{N}{S_{SIR}}$ stays below one if $S_{SIR} \geq N/2$, i.e. as long as the majority of the population has not been in contact with the SARS-CoV-2 virus. This is good news. But observing a small r_n can conceal a situation with a large R_0 if S_{SIR}/N is small. This is one reason why countries need to get a grip on the Susceptibles nationwide.

It is tempting to use the above technique for prediction in such a way that the b_n and the γ_n are fitted to a constant or a linear function, and using the values of the fit for running the system into the future. This is very close to extending the logarithmic plots of Figure 5 by lines, using pencil and ruler, and thus hardly interesting. It would extend also all effects due to political measures.

So far, the above argument cannot replace a SIR model. It only interprets the available data. However, monitoring the Johns Hopkins data in the above way will be very useful when it comes to evaluate the effectivity of certain measures taken by politicians. It will be highly interesting to see how the data of Figure 6 continue, in particular when countries relax their contact restrictions.

3.4 Extension Towards a SIR Model

For cases where one still has to expect $R_0 > 1$, e.g. US and Brazil (see Figure 6), the challenge remains to predict a possible peak. Using the estimates from

the previous section is questionable, because they concern the subpopulation of Confirmed and are systematically underestimating R_0. The "real" SIR model will have different parameters, and it needs the Susceptibles to model a peak or to make the r_n estimates realistic. So far, the Johns Hopkins data work in a range where S/N is still close to one, and the Susceptibles are considered to be abundant. But the bad news for countries with $r_n > 1$ is that r_n underestimates R_0.

Anyway, if one trusts the above time series as approximations to β and γ, one can run a SIR model, provided one is in the case $R_0 > 1$ and has reasonable starting values. But these are a problem, because the unconfirmed Infectious and the unconfirmed Recovered are not known, even if, for simplicity, one assumes that there are no unconfirmed COVID-19 deaths.

For an unrealistic scenario, consider *Total Registration*, i.e. all Infected are automatically confirmed. Then the Susceptibles in the Johns Hopkins model would be $S_n = N - C_n = N - I_n - R_n - D_n$. Now the estimate for R_0 must be corrected to

$$r_n \frac{N}{S_n} = r_n \frac{N}{N - C_n} = r_n \left(1 + \frac{C_n}{N - C_n}\right)$$

but this change will not be serious during the outbreak.

If the time series $\beta_n = b_n \frac{N}{S_n} = b_n \frac{N}{N - C_n}$ for β and γ_n for γ are boldly used as predictors for β and γ in a SIR model, and if the model is started using $S_n = N - C_n = N - I_n - D_n - R_n$ in the discretized form

$$\begin{aligned} S_{n+1} - S_n &= -\beta \frac{S_n}{N} I_n, \\ I_{n+1} - I_n &= +\beta \frac{S_n}{N} I_n - \gamma I_n, \\ (R+D)_{n+1} - (R+D)_n &= -\gamma I_n, \end{aligned}$$

one gets a crude prediction of the peak in case $R_0 = \beta/\gamma > 1$.

Figure 7 shows results for two cases. The top shows the case for the United States using data from day 109 (May 10^{th}) and estimating β and γ from the data one week before. The peak is predicted at day 473 (May 9^{th}, 2021) with a total rate of 33% Infectious, i.e. about 124 million people. With an infection fatality rate of 0.005%, this means about 600000 casualties in the two weeks around the peak. To

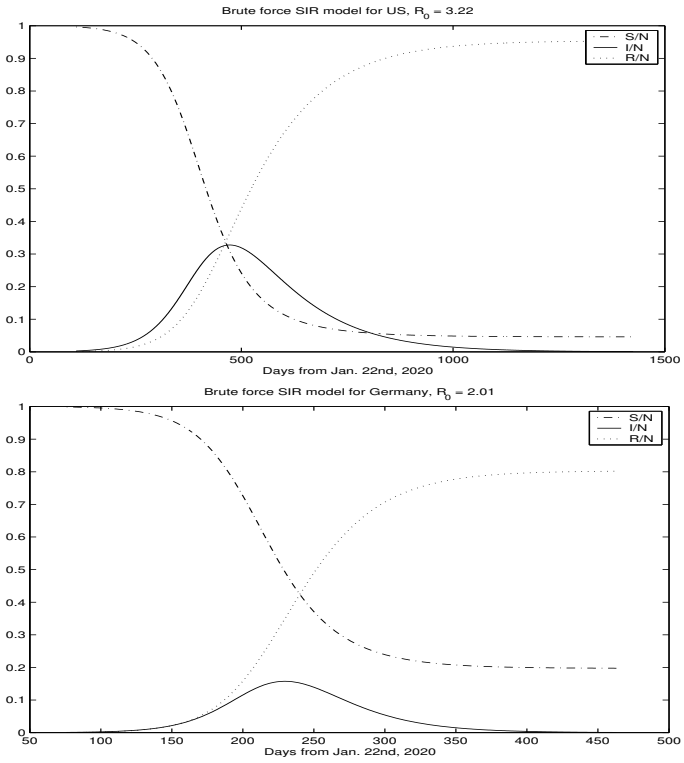

Figure 7: Brute force SIR modeling for US and Germany using last week's data, at days 109 and 75, respectively.

see how crude the technique is, the second plot shows the case of Germany using data up to day 75 (April 6^{th}, 2020), i.e. before the peak, and the peak is predicted at day 230 (Sept. 8^{th}, 2020) with about 16% Infected. This would imply about 65000 casualties around the peak. At day 75, R_0 was estimated at 2.01, but a few days later the estimate went below 1 (Figure 6) by political intervention changing b_n considerably. See Figure 11 for a much better prediction using data only up to day 67.

4 Extended SIR Model

To get closer to reality, one should combine the data-oriented Johns Hopkins Data Model with a SIR model that accounts for what happens outside of the Confirmed.

We introduce the time series

S for the Susceptibles like in the SIR model,

M for the Infectious, not yet confirmed, (M standing for *mysterious*),

H for the unconfirmed Recovered (H standing for *healed*).

This implies that all deaths occur within the Confirmed, though this is a highly debatable issue. It assumes that persons with serious symptoms get confirmed, and nobody dies of COVID-19 without prior confirmation.

4.1 The Hidden Model

The Removed from the viewpoint of a global SIR model including H and M are $H+C$, and thus the SIR model is

$$\begin{aligned} S_{n+1}-S_n &= -\beta\frac{S_n}{N}M_n, \\ M_{n+1}-M_n &= \beta\frac{S_n}{N}M_n-\gamma M_n, \\ (H+C)_{n+1}-(H+C)_n &= \gamma M_n. \end{aligned} \tag{18}$$

To run this *hidden* model with constant $N=S+M+H+C$, one needs initial values and good estimates for β and γ, which are not the ones of the Johns Hopkins Data Model of section 3.3.

4.2 The Observable Model

The Johns Hopkins variables D and R are linked to the hidden model via $C=I-R-D$. They follow an *observable* model

$$\begin{aligned} I_n &= C_n-R_n-D_n, \\ D_{n+1}-D_n &= \gamma_{iCD}I_n, \\ R_{n+1}-R_n &= \gamma_{iCR}I_n \end{aligned} \tag{19}$$

with *instantaneous case death and recovery rates* γ_{iCD} and γ_{iCR} for the Confirmed Infectious. These rates can be estimated separately from the available Johns Hopkins data, and we shall do this below. We call thes rates *instantaneous*, because they artificially attribute the new deaths or recoveries at day $n+1$ to the previous day, not to earlier days. In this sense, they are rather artificial, and we shall address this question. They are *case* rates, because they concern the Confirmed.

The observable model is coupled to the hidden model only by C_n. Any data-driven C_n from the observable model can be used to enter the $H+C$ variable of the hidden model, but in an unknown ratio. Conversely, any version of the hidden model produces $H+C$ values that do not determine the C part. Summarizing, there is no way to fit the hidden model to the data without additional assumptions.

Various possibilites were tried to connect the Hidden to the Observable. Two will be presented now.

4.3 Fatality Rates

4.3.1 Infection Fatality Rate

Recall that the parameter γ_{iCD} in the observable model (19) relates case fatalities to the confirmed Infectious of the previous day. In contrast to this, the *infection fatality rate* in the standard literature, denoted by γ_{IF} here, is relating to the infection directly, independent of the confirmation, and gives the probability to die of COVID-19 after infection with the SARS-CoV-2 virus, whatever the delay between infection and death is. It was estimated as $\gamma_{IF} = 0.56\%$ by an der Heiden/Buchholz [2] and 0.66% by Verity et al. [22], but specialized for China. Recent data from the Heinsberg study by Streeck et. al. [21] gives a value of 0.36% for the Heinsberg population in Germany. For the UK, Ferguson et al. [7] arrive at 0.9%. We shall later use 0.5% for our predictions. But it is very desirable to get more information on infection fatality rates, in particular for different countries. So far, we use thus a single value globally.

The idea to use the infection fatality rate for information about the hidden system comes from Bommer/Vollmer [3]. The way to use it will depend on how to handle delays, and it turned out to be difficult to use these rates in a stable way.

4.3.2 Estimation of Case Fatality Rates

Let us focus on probabilities to die either after an infection or after confirmation of an infection. The first is the infection fatality rate given in the literature, but what is latter, the *case fatality rate* γ_{CF} when using the Johns Hopkins data? It is clearly not the γ_{iCD} in (19), giving the ratio of new deaths at day $n+1$ as a fraction of the confirmed Infectious at day n. The deaths at day $n+1$ must be assigned to various earlier days instead.

Case fatality rates in the literature vary strongly, and they are country-dependent. Countries have different ways to detect cases, and because the mortality is age-dependent, different age structures will have a serious influence. The Robert-Koch-Institute [18] mentions 10.5% for Europe and 4.6% for Germany, while De Brouwer et al. [5] has 10.0% for Italy, 4.0% for China, 6.0% for Spain, and 4.3% worldwide. According to Streeck et al.[21], the current estimate of the case fatality rate in Germany by the World Health Organization (WHO) is between 2.2% and 3.4%.

We cannot clean up these inconsistencies. Instead, we now describe a way to estimate case fatality rates per country from the Johns Hopkins data. The basic idealistic assumption is that COVID-19 diseases end after k days from confirmation with either death or recovery. Let us call this the *k-day rule*. Suggested values for k start from 14 days for mild cases (an der Heiden/Buchholz [2] WHO [1]) and go up to 30 days, composed of an incubation time of about 5 days and various values between 11 and 25 days for hospitalization, depending on the amount of intensive care (an der Heiden/Buchholz [2], Robert Koch-Institut [18], Verity et al. [22], Mohring et al. [17]).

Following Schaback [19], one can estimate the probability to survive on day $k+1$ after confirmation, and this works in a stable way per country, based only on C and D, not on the unstable R data. In [19] this approach was used to produce R values that comply with the k-day rule, but here we use it for estimating the case fatality. The technique of [19] performs a fit

$$\begin{aligned} D_n - D_{n-1} &\approx \sum_{i=1}^{k} q_i (C_{n-i} - C_{n-i-1}), \\ q_i &= p_i \prod_{j=1}^{i-1} (1-p_j),\ 1 \le i \le k+1 \end{aligned} \tag{20}$$

i.e. it assigns all new deaths at day n to previous new infections on previous days in a hopefully consistent way, minimizing the error in the above formula under variation of the probabilities p_i to die on day i after confirmation. If the p_i are known for days 1 to k, the *case fatality rate* is

$$\gamma_{CF} = \sum_{i=1}^{k} q_i = 1 - q_{k+1}.$$

This argument can also be seen the other way round: the new Confirmed $C_n - C_{n-1}$ at day n enter into $D_{n+1} - D_n$ with probability $q_1 = p_1$, into $D_{n+2} - D_{n+1}$ with

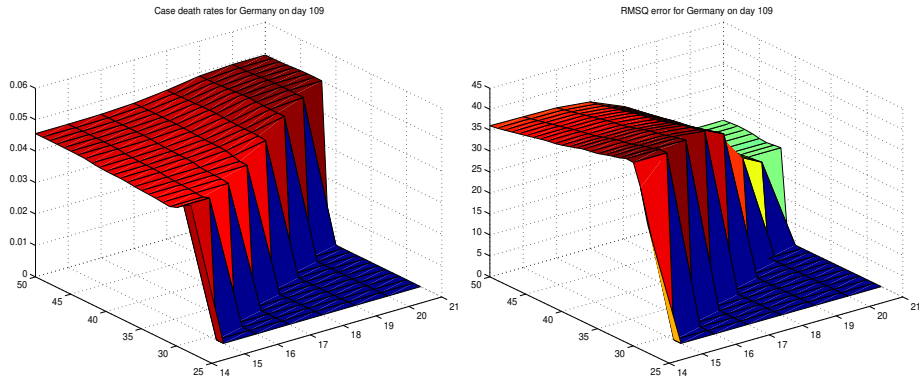

Figure 8: Left: case fatality rate for Germany based on data at day 109, as functions of k and the data backlog. Right: Root mean-square error.

probability $q_2 = p_2(1-p_1)$ and so on. The rest enters into the new Recovered at day $n+k$ with probability q_{k+1} if we set $p_{k+1} = 0$. Thus the case fatality rate can be expressed as $1-q_{k+1}$ like above.

At this point, there is a hidden assumption. Persons that are new to the Confirmed at day n are not dead and not recovered. The change $C_{n+1} - C_n$ to the Confirmed is therefore understood as the number of new registered infections. Otherwise, one might replace $C_{n-i} - C_{n-i-1}$ by $I_{n-i} - I_{n-i-1}$ in (20), but this would connect a cumulative function to a non-cumulative function. Furthermore, this would use the unsafe data of the Recovered.

In fact, equation (20) is unexpectedly reliable, provided one looks at $1-q_{k+1}$ or q_{k+1}, not at single probabilities p_j. This follows from series of experiments that we do not document fully here, except for Figure 8. In [19], data for $2k$ days backwards were used for the estimation, and results did not change much when more or less data were used or when k was modified. Here, the range $7 \le k \le 21$ was tested, and backlogs of up to 50 days from day 109. See Figure 8 below for an example. It is typical here and for many other cases that a value of $k = 14$ performs well, with a backlog of $2k = 28$ days for the fit in (20). Using larger k needs a larger backlog, but then the estimation is not time-local enough to produce current estimates, because outdated values are used. Figure 8 shows the variation of the case fatality rate estimation when k and the backlog are varied.

See the first column of Table 1 for estimates of case fatality rates for different

Country	Death rate	Detection rate
Germany	0.047	0.106
Brazil	0.094	0.053
Russia	0.012	0.407
Italy	0.138	0.036
Spain	0.085	0.059
Sweden	0.157	0.032
Austria	0.052	0.096
France	0.122	0.041
UK	0.145	0.035
US	0.067	0.075

Table 1: Case fatality and detection rates, estimated on day 109 using the 14-day rule and a backlog of 28 days

countries, calculated on day 109 (May 10^{th}) for $k = 14$ and a backlog of 28 days. They comply with the values from the literature cited above, and depend strongly on the strategy for confirmation. In particular, they are high when only serious cases are confirmed, e.g. cases that need hospital care. If many more people are tested, confirmations will contain plenty of much less serious cases, and then the case fatality rates are low.

The instantaneous case death rate γ_{iCD} of (19) for the Johns Hopkins data comes out around 0.004 for Germany on day 109 by direct inspection of the data via

$$\gamma_{iCD} \approx \frac{D_{n+1} - D_n}{I_n}, \tag{21}$$

while the Case Fatality Rate γ_{CF} in Table 1 is about 0.047. The deaths have to be attributed to different days using the k-day rule, they cannot easily be assigned to the previous day without making the rate smaller.

4.3.3 Using Fatality Rates for the Hidden Model

If the case fatality rates γ_{CF} of Table 1 are used with a known infection fatality rate γ_{IF}, one should obtain an estimate of the total Infectious. If the formula (20)

is written as

$$\sum_{i=1}^{k} q_i(C_{n-i} - C_{n-i-1}) \approx D_n - D_{n-1} \approx \sum_{i=1}^{k} \tilde{q}_i(S_{n-i-1} - S_{n-i})$$

in terms of the previous new infections $S_{n-i-1} - S_{n-i}$ with infection fatality probabilities $\tilde{q}_i$, one should maintain

$$\gamma_{CF} = \sum_{i=1}^{k} q_i \text{ and } \gamma_{IF} = \sum_{i=1}^{k} \tilde{q}_i$$

and this works by setting

$$C_n - C_{n-1} = \frac{\gamma_{IF}}{\gamma_{CF}}(S_{n-1} - S_n) \tag{22}$$

in general, without using the unstable p_i.

4.3.4 The Detection Rate

The quotient $\frac{\gamma_{IF}}{\gamma_{CF}}$ can be called the *detection rate*, stating the fraction of Infectious that is entering confirmation. See the second column of Table 1. The rate depends on good estimates of the infection fatality rate, and the new value by Streeck et al. [21] will roughly double the detection rate for Germany.

All of this is comparable to the findings of Bommer/Vollmer [3] and uses the basic idea from there, but with a somewhat different technique and different results. There, the values were 7% for March 23rd and 9% for March 30th, while Mohring et al. [17] assume 20%.

A simple way to understand the quotient $\frac{\gamma_{IF}}{\gamma_{CF}}$ as a detection rate is to ask for the probability $p(C)$ for Confirmation. If the probability to die after Confirmation is γ_{CF}, and if there are no deaths outside confirmation, then

$$p(D) = p(C) \cdot p(D|C),$$

and

$$p(C) = \frac{p(D)}{p(D|C)} = \frac{\gamma_{IF}}{\gamma_{CF}}.$$

It is tempting to replace S by M in (22), but this would make M cumulative.

4.3.5 Local Estimation of Fatality Rates

Under political changes of the parameters β and γ, and when the testing strategies are changed, the estimation of the Case Fatality Rate should be made locally, not globally. Using the experience of Schaback [19] and section 4.3.2, we shall do this using a fixed $k = 14$ for the k-day rule and data for a fixed backlog of $2k$ days.

4.4 Recovery Rates

We need another parameter to connect the hidden to the observable model. There are many choices, and after some failures we selected the constant γ_{iIR} in a model equation

$$H_{n+1} - H_n = \gamma_{iIR} M_n.$$

Following what was mentioned about *instantaneous* rates in section 4.2, this is an *instantaneous Infection Recovery Rate*, relating the new unregistered Recovered to the unregistered Infections the day before.

4.4.1 Estimation of Recovery Rates

A good value of γ_{iIR} can come out of a field experiment that produces a time series for M and H, i.e. for unregistered Infectious and unregistered Recovered. Then the instantaneous Infection Recovery rate γ_{iIR} can be obtained directly by

$$\frac{H_{n+1} - H_n}{M_n} \approx \gamma_{iIR}.$$

The Infection Recovery rate $\gamma_{IR} = 1 - \gamma_{IF}$ does not help much, because we need an instantaneous rate that has no interpretation as a probability.

With the risk of using unstable data of the Recovered, we can look at the instantaneous Case Recovery rate

$$\frac{R_{n+1} - R_n}{I_n} \approx \gamma_{iCR} \tag{23}$$

that is available from the Johns Hopkins data, and comes out experimentally to be rather stable, provided that countries have useful data for the Recovered. Otherwise, we have to use the technique of Schaback [19]. The rate γ_{iIR} must be larger, because we now are not in the subpopulation of the Confirmed, and nobody can

die without going first into the population of the Confirmed. As long as no better data are available, we shall use the formula

$$\gamma_{iIR} = \frac{1-\gamma_{IF}}{1-\gamma_{CF}}\gamma_{iCR} = \frac{\gamma_{IR}}{\gamma_{CR}}\gamma_{iCR} = \frac{\gamma_{iCR}}{\gamma_{CR}}\gamma_{IR} \tag{24}$$

that implements two meaningful arguments:

1. the value γ_{iCR} is increased by the ratio of Recovered probabilities for the Infected and the Confirmed,

2. the value γ_{IR} is multiplied by a factor for transition to immediate rates, and this factor is the transition factor for the Confirmed Recovered.

The above strategy is debatable and may be the weakest point of this approach. However, others turned out to be worse, mainly due to instability of results. On the positive side, the final prediction does not need it. It enters only in the intermediate step when S, M, and H are calculated in the time range of the available Johns Hopkins data, see section 4.5. And, finally, there is hope that there will be field experiments that yield reliable values directly.

4.4.2 Practical Approximation of Recovery Rates

In (24) the rate γ_{IR} is fixed, and the rate γ_{CR} is determined locally via section 4.3.5. The rate γ_{iCR} follows from the time series

$$\frac{R_{n+1}-R_n}{I_n} \approx \gamma_{iCR}$$

as in (19). This works for countries that provide useful data for the Recovered. In that case, and in others to follow below, we can take the time series itself as long as we have data. For prediction, we estimate the constant from the time series using a fixed backlog of m days from the current day, i.e. we take the mean of the last $m+1$ values. Since many data have a weekly oscillation, due to data being not properly delivered during weekends, the backlog should not be less than 7.

But for certain countries, like the United Kingdom, the data for the Recovered are useless. In such cases, we employ the technique of Schaback [19] to estimate the Recovered using the k-day rule and a backlog of $2k$ days, like in section 4.3.5 for the case fatality rates.

4.5 Model Completion

We now have everything to run the hidden model, but we do it first for days that have available Johns Hopkins data. This leads to estimations of S, M, and H from the observed data of the Johns Hopkins source, without any need for sophisticated fitting algorithms, With the parameters from above, we use the new relations

$$\begin{aligned} C_{n+1} - C_n &= \frac{\gamma_{IF}}{\gamma_{CF}}(S_n - S_{n+1}), \\ H_{n+1} - H_n &= \gamma_{iIR} M_n \end{aligned} \tag{25}$$

in a specific way. We set up the second model equation in (18) for M as

$$\begin{aligned} M_{n+1} - M_n &= S_n - S_{n+1} - \gamma_n M_n \\ &= \frac{\gamma_{CF}}{\gamma_{IF}}(C_{n+1} - C_n) - \gamma_n M_n \\ &= \frac{\gamma_{CF}}{\gamma_{IF}}(C_{n+1} - C_n) - (C_{n+1} - C_n + H_{n+1} - H_n) \\ &= \left(\frac{\gamma_{CF}}{\gamma_{IF}} - 1\right)(C_{n+1} - C_n) - \gamma_{iIR} M_n \end{aligned} \tag{26}$$

that can be solved if an initial value $\tilde{M}_0$ is prescribed. Then (25) is run to produce the S_n and H_n, with starting values that we describe in section 4.5.1. If β_n and γ_n are calculated by

$$\begin{aligned} \beta_n \frac{S_n}{N} M_n &= S_n - S_{n+1}, \\ \gamma_n M_n &= C_{n+1} - C_n + H_{n+1} - H_n, \end{aligned} \tag{27}$$

respectively, the balance equation $N = S + M + H + C$ follows from (26).

4.5.1 Starting Values

Since the populations are large, the starting values for S are not important. Starting at the full population N from a very early day, the S values are calculated from (25), just to get values S_j for starting at later days.

Then the first day j is taken where C_j is at least 100, and the start value for H there is set as

$$H_j = C_{j-k} \frac{\gamma_{CF}}{\gamma_{IF}}$$

using the k-day rule with $k = 14$. This divides the $C_{j-k} > I_{j-k}$ value by the detection rate, i.e. roughly all estimated undetected Infectious at time $j-k$ are assumed to be healed k days later, i.e. at day j. Then the starting value for M_j is calculated via the balance equation $N = S_j + M_j + H_j + C_j$.

The starting value for H is irrelevant for H itself, because only differences enter, but it determines the starting value for M due to the balance equation. Anyway, it turns out experimentally that the starting values do not matter, if the model is started early. The hidden model (18) depends much more strongly on C than on the starting values.

Figure 11 contains a wide variation of the starting value $H = N - S - C$ at the starting point, by multipliers between 1/32 and 32. This has hardly any effect on the results, the lines getting somewhat thicker. The variation in starting values get more visible in other cases, see the right-hand plot in Figure 11 for the United States. But the influence on predictions is negligible.

4.5.2 Examples

The figures to follow in section 5.2 show the original Johns Hopkins data together with the hidden variables S, M, and H that are calculated by the above technique. Note that the only ingredients beside the Johns Hopkins data are the number $k = 14$ for the k-day rule, the Infection Fatality Rate γ_{IF} from the literature, equations (25), and the backlog $m = 7$ for estimation of constants from time series.

5 Predictions using the Full Model

To let the combined model predict the future, or to check what it would have predicted if used at an earlier day, we take the completed model of the previous sections up to a day n and use the values $S_n, M_n, H_n, C_n, I_n, R_n$ and D_n for starting

the prediction. With the variable $HC := H + C$, we use the recursion

$$\begin{array}{rcl} S_{i+1} &=& S_i - \beta \dfrac{S_i}{N} M_i, \\ M_{i+1} &=& M_i + \beta \dfrac{S_i}{N} M_i - \gamma M_i, \\ HC_{i+1} &=& HC_i + \gamma M_i, \\ C_{i+1} &=& C_i + \gamma_{IF}(S_i - S_{i+1})/\gamma_{CF}, \\ R_{i+1} &=& R_i + \gamma_{iCR} I_i, \\ D_{i+1} &=& D_i + \gamma_{iCD} I_i, \\ I_{i+1} &=& C_{i+1} - R_{i+1} - D_{i+1}, \\ H_{i+1} &=& HC_{i+1} - C_{i+1}. \end{array} \tag{28}$$

This needs fixed values of β and γ that we estimate from the time series for β_n and γ_n by using a backlog of 7 days, following Section 4.5. The instantaneous rates γ_{iCR} and γ_{iCD} can be calculated via their time series, as in (23) and (21), using the same backlog. We do this at the starting point of the prediction, and then the model runs in what can be called a *no political change* mode. Examples will follow in section 5.2.

5.1 Properties of the Full Model

The first part of the full model (28) runs as a standard SIR model for the variables S, M and $H + C$, and inherits the properties of these as described in section 2. It does not use the γ_{iIR} parameter of the second equation in (25), and it uses the first the other way round, now determining C from S, not S from C.

The balances $N = S + M + H + C$ and $C = I + D + R$ are maintained automatically, and the time series for $S, C, R, H + C$, and D stay monotonic as long as M and I are nonnegative. To check the monotonicity of H, consider

$$\begin{array}{rcl} H_{i+1} - H_i &=& HC_{i+1} - HC_i - C_{i+1} + C_i \\ &=& \gamma M_i - \dfrac{\gamma_{IF}}{\gamma_{CF}}(S_i - S_{i+1}) \\ &=& \left(\gamma - \beta \dfrac{\gamma_{IF}}{\gamma_{CF}} \dfrac{S_i}{N}\right) M_i. \end{array}$$

The bracket is positive if

$$R_0 = \frac{\beta}{\gamma} < \frac{\gamma_{CF}}{\gamma_{IF}} \frac{N}{S_i} \geq \frac{\gamma_{CF}}{\gamma_{IF}},$$

which is enough for practical purposes as long as detection rates are high and R_0 is not excessively large.

The slopes of S and C are always connected by (22), and those of R and D are connected by

$$R_{i+1} - R_i = \frac{\gamma_{iCR}}{\gamma_{iCD}}(D_{i+1} - D_i)$$

in the prediction part. But the figures below will show logarithms, and therefore the slope parallelism will not be visible.

By section 2.11, the hidden Infectious M will always go to zero, and the variables S and $H+C$ will level out in the long run. Since C is increasing, it must level out as well, and I must go to zero because R and D level out. The asymptotic levels of S and $H+C$ follow from 2.11, but not the interesting asymptotic level of D. If the prediction is started at day n, then

$$R_\infty - R_n = \frac{\gamma_{iCR}}{\gamma_{iCD}}(D_\infty - D_n)$$

connects the asymptotic deaths and confirmed recoveries. From the connection of S and C we likewise get

$$C_\infty - C_n = \frac{\gamma_{iIF}}{\gamma_{CF}}(S_n - S_\infty).$$

With $C_\infty = R_\infty + D_\infty$ we now have three independent equations for the unknowns $C_\infty, D_\infty, R_\infty$. Because the theory of Section 2.11 yields S_∞ and $H_\infty + C_\infty$ in terms of β and γ, we can also get H_∞. But if the simulation is run long enough, one can read the asymptotic values off the plots.

5.2 Examples of Predictions

Figure 9 shows predictions on day 122, May 23rd, for Germany, Brazil, France, and US, from the top. The plots for countries behind their peak are rather similar to those for Germany and France. The other two countries are selected because they still have to face their peak, if no action is taken to change the parameters.

The plots show that Germany can expect to get away with no more than 10000 casualties in the long run, while Brazil goes for a peak of about 20 million hidden Infectious in fall 2020 (M, symbol □) and a final death toll of about 1 million

(D, symbol $+$). The United States would have to face a peak of hidden Infectious of about 25 million in mid-January 2021, and more than 1 million COVID-19 deaths in October 2021, and still rising. But of course, these predictions assume that reality follows the model and that there are no parameter changes by political action.

The estimated R_0 values are 0.65, 2.19, 0.4s, and 1.75, respectively. Note that these are not directly comparable to Figure 6, because they are the fitted constants to the backlog of a week, and using (27) instead of (17), avoiding the systematic underestimation of the latter. The hidden M and H (symbols $\square$ and $\diamond$) follow roughly the observable I and C (symbols O and x), but with a factor due to the detection rate that is different between countries, see Table 1. To enhance visibility, not all points in the plots are marked with symbols. The C, R, I and D data left of the vertical line are the original Johns Hopkins data. The S, M, H data there are calculated by section 4, while to the right the data are predictions for all variables by the full model (28).

All test runs were made for the infection fatality rate $\gamma_{IF} = 0.005$, the delay $k = 14$ for estimating case fatalities, and a backlog of 7 days when estimating constants out of recent values of time series. The choice $\gamma_{IF} = 0.005$ is somewhat between 0.56% from an der Heiden/Buchholz [2], 0.66% from Verity et al. [22], and 0.36% from Streeck et al. [21]. New information on infection fatality rates should be included as soon as they are available, and if possible different values for different countries.

When used within estimation routines, the Johns Hopkins data were smoothed by a double application of the 1/4, 1/2, 1/4 filter on the logarithms, like for Figure 6. But the plots show the original data.

5.3 Evaluation of Predictions

To evaluate the prediction quality, one should go back and start the predictions for earlier days, to compare with what happened later. Figure 10 shows overplots of predictions for days 94, 108, and 122, each a fortnight apart. The starting points of the predictions are marked by vertical lines again. For better visibility, only the death count D (symbol $+$) and the two non-cumulative variables M and I for the hidden and confirmed Infectious (symbols $\square$ and O) are shown. Each prediction has slightly different estimates for S, M, and H due to different available data, and

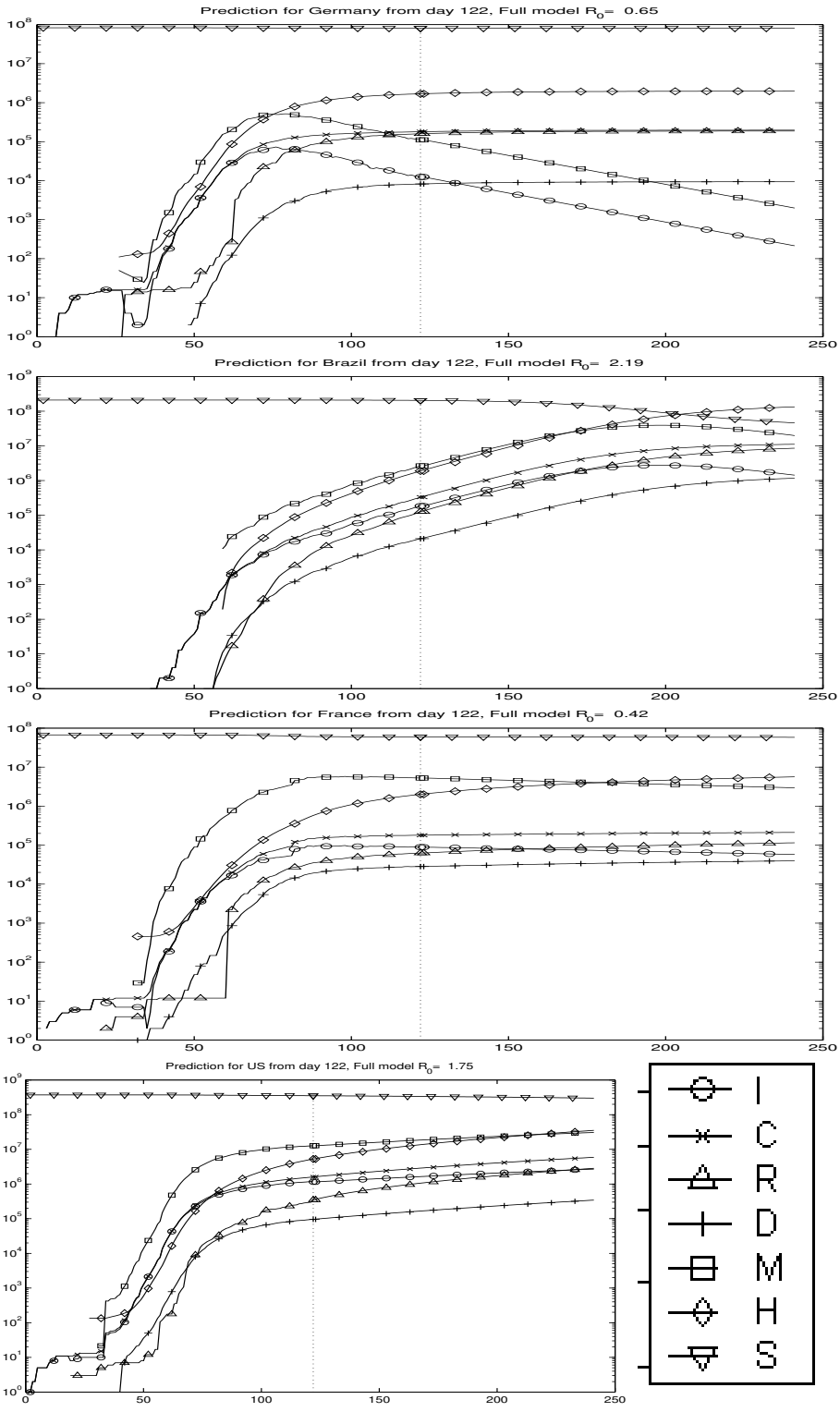

Figure 9: Predictions for countries Germany, Brazil, France, and US on day 122 marked by the vertical dotted line

therefore there are different M curves even left of the leftmost prediction day. Recall that the determination of these variables is done while there are Johns Hopkins data available, following section 4.5, and will be dependent on the data-driven estimations described there. In particular, the case fatality rates and detection rates of Table 1 change with the starting point of the prediction, and they determine S, M, and H backwards. The leftmost prediction on day 94 roughly matches the data available up to day 122 in all cases. One can see that the Brazil parameters do not change much, while the three predictions for the United States get better. Germany was slightly better off on day 94 than on days 108 and 122. This might be used to assess effectivity of administrative efforts to handle the pandemics.

For an early case in Germany, Figure 11 shows the prediction based on data of day 67, March 27 th. The peak of about 20 million Infected is predicted on day 121, May 22nd, with roughly 20 million unconfirmed Infectious and about 70,000 casualties at the peak and about 200,000 finally. A good reason to act politically. Note that the real death count is about 8300 on May 23rd, and the prediction of the day, in Figure 9, targets a final count of below 10,000.

Quantitative commitments to predictions are rare in the literature, except for rough estimations of dramatic outbreak scenarios. On April 3rd, after the last public restrictions in Germany of March 22nd, 2020, Germany had 1107 deaths and Khailaie et al. [14] predicted 10,000 deaths for early May. Roughly the same is predicted by this model using the same data, but also a final death toll of 25,000 in the long run. The true deaths were 6812 on May 3rd.

On March 16th, Ferguson et al. [7] predicted deaths in the order of 250,000 in GB, and 1.1 to 1.2 million in the USA. If the model (28) is run on data up to March 16th, it predicts 2 million deaths in the USA and 300,000 for the UK. Again, there were non-pharmaceutic interventions in between that brought the figures down.

The use of the Infection Fatality Rate is somewhat different from Streeck et al. [21] and Bommer/Vollmer [3], but results are similar. If the rate 0.0036 of [21] is used in a test run based on data of May 2nd, the estimated number of Infected comes out as 1.54 million, while [21] gets 1.8 million by the formula $(M_n + C_n) = D_n/0.0036$ for the same day.

The parameter changes by political measures turned out to be rather effective, like in many countries that applied similar strategies. But since parts of the population

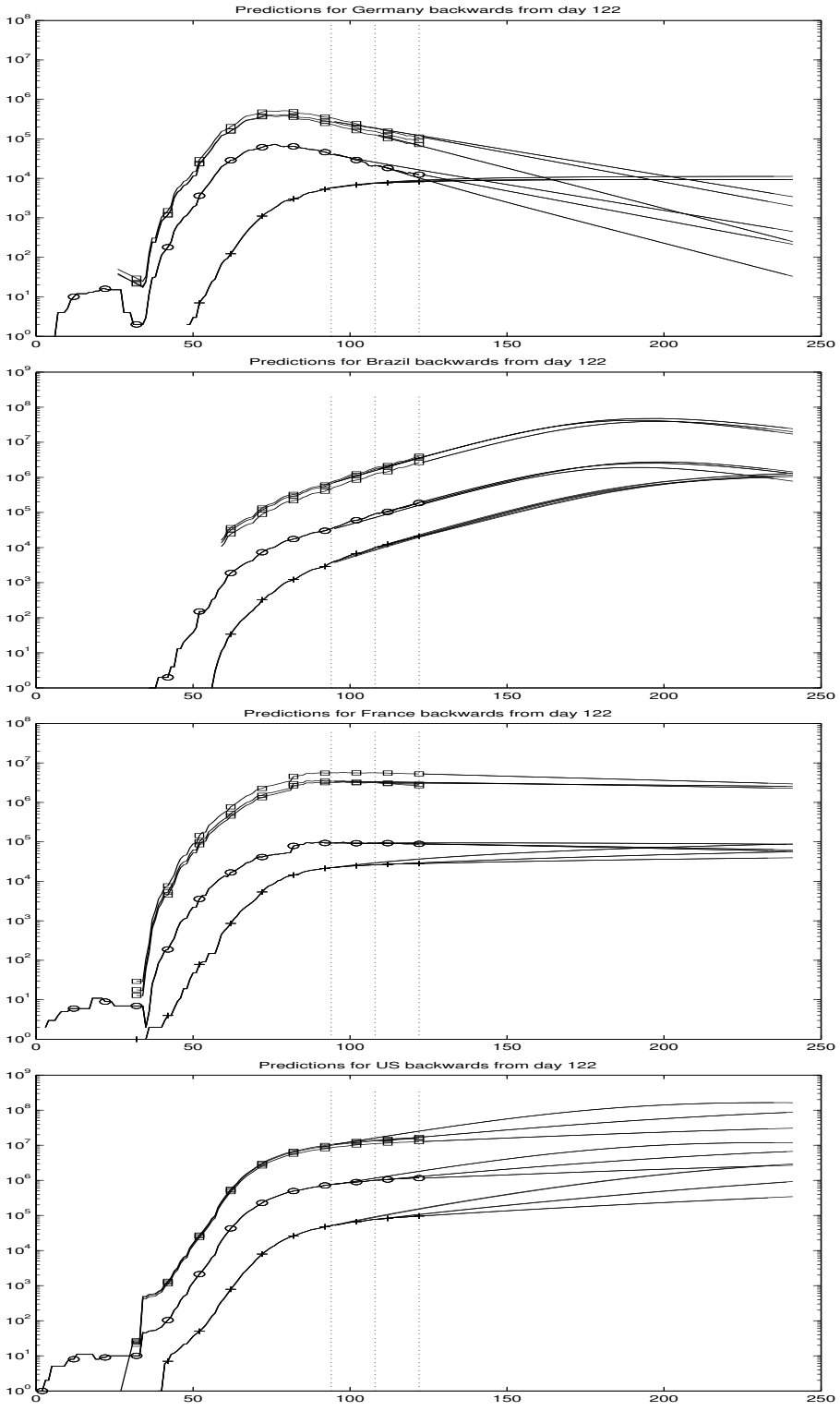

Figure 10: Predictions for countries Germany, Brazil, France, and USA on days 122, 108, and 94, marked by vertical dotted lines. Legend as in Figure 9.

want to go back to their previous lifestyle, all of this is endangered, and the figures should be monitored carefully.

Of course, all of this makes sense only under the assumption that reality follows the model, in spite of all attempts to design a model that follows reality.

6 Conclusion and Open Problems

So far, the model presented here seems to be useful, combining theory and practically available data. It is data-driven to a very large extent, using only the infection fatality rate from outside for prediction, and the approximation (24) for calibration.

On the downside, there is quite a number of shortcomings:

- Like the data themselves, the model needs regular updating. As far as the Johns Hopkins data are concerned, the model updates itself by using the latest data, but it needs changes as soon as new information on the hidden infections come in.
- There may be better ways of estimating the hidden part of the epidemics. However, it will be easy to adapt the model to other parameter choices. If time series for the unknown variables get available, the model can easily be adapted to being data-driven by the new data.
- The treatment of delays is unsatisfactory. In particular, infected persons get infectious immediately, and the k-day rule is not followed at all places in the model. But the rule is violated as well in the data (Schaback [19]).
- There is no stochastics involved, except for simple things like estimating constants by means, or for certain probabilistic arguments on the side, e.g. in section 4.3.2. But it is not at all clear whether there are enough data to do a proper probabilistic analysis.
- As long as there is no probabilistic analysis, there should be more simulations under perturbations of the data and the parameters. A few were included, e.g. for section 4.3.2 and Figures 8 and 11, but a large number was performed in the background when preparing the paper, making sure that results are stable. However, there are never too many test simulations.

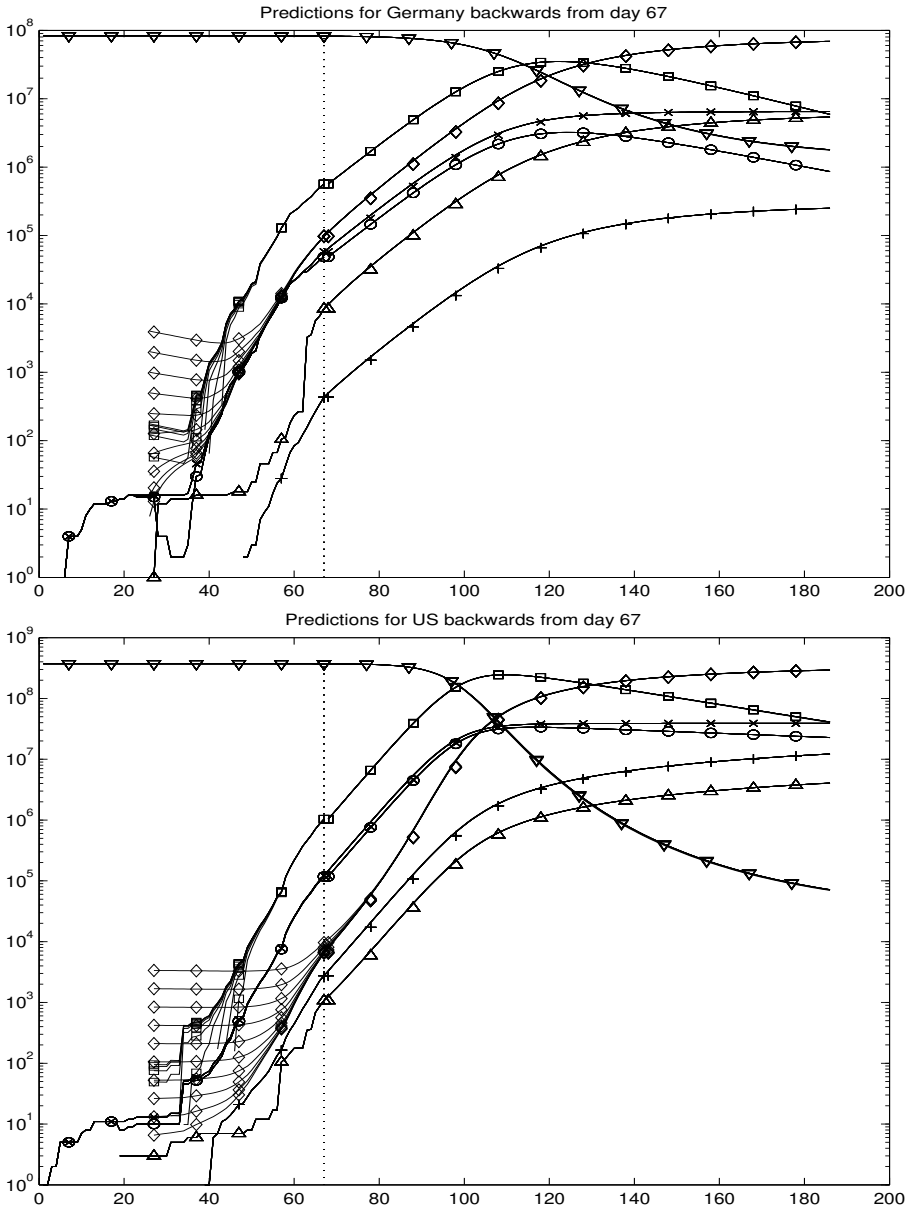

Figure 11: Predictions for Germany and USA on day 67, March 27^{th}, with varying starting values. Legend as in Figure 9.

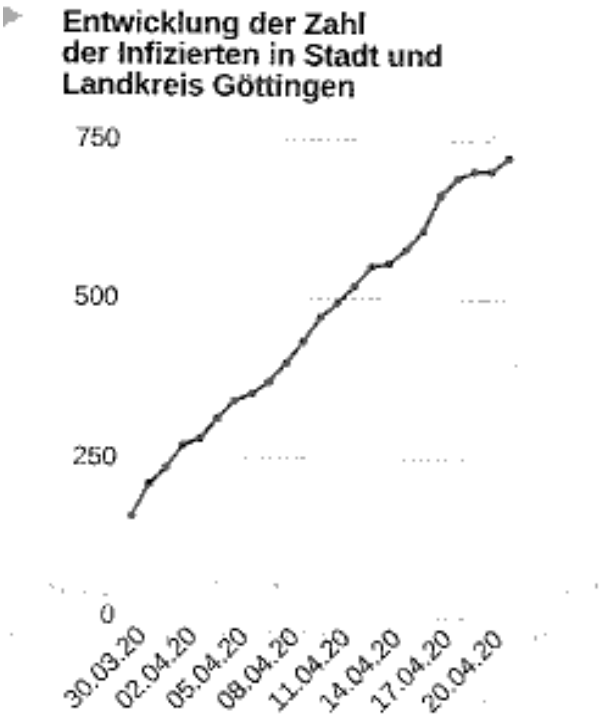

Figure 12: Infectious in Göttingen city and county, as of April 22nd, 2020 in the local newspaper "Göttinger Tageblatt". No exponential outbreak.

- Totally different models were not considered, e.g. the classical ones with delays (Hoppenstaedt/Waltman [11, 12].
- Under certain circumstances, epidemics do not show an exponential outbreak, in particular if they hit only locally and a prepared population. See Figure 12 for the COVID-19 cases in Göttingen and vicinity.

MATLAB programs and more recent predictions will be on the research website `http://num.math.uni-goettingen.de/schaback/research/group.html` of the author. Special thanks go to Tara Fickle, Viola Priesemann, Jalda Schaback, and Wolfgang Warth for comments and corrections. All links in the references were verified on June 2nd, 2020.

References

[1] Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), Feb 28th, 2020.
https://www.who.int/publications-detail/report-of-the-who-china-joint--mission-on-coronavirus-disease-2019-(covid-19)

[2] M. An der Heiden and U. Buchholz. Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland. *Bekanntmachungen des Robert Koch-Instituts*, March 3rd, 2020. DOI: 10.25646/6571.2.
https://edoc.rki.de/handle/176904/6547.2

[3] C. Bommer and S. Vollmer. Average detection rate of SARS-CoV-2 infections is estimated around six percent. April 2nd, 2020.
https://reason.com/wp-content/uploads/2020/04/Bommer-Vollmer-2020-COVID-19-detection-April-2nd.pdf

[4] R. Dandekar and G. Barbastathis. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. April 6th, 2020, medRxiv, DOI: 10.1101/2020.04.03.20052084.

[5] E. De Brouwer, D. Raimondi, and Y. Moreau. Modeling the COVID-19 outbreaks and the effectiveness of the containment measures adopted across countries. April 19th, 2020, medRxiv, DOI: 10.1101/2020.04.02.20046375

[6] J. Dehning, J. Zierenberg, P. Spitzner, M. Wibral, J. Pinheiro Neto, M. Wilczek, and V. Priesemann. Inferring Covid-19 spreading rates and potential change points for case number forecasts. Revised May 4th, 2020, arXiv:2004.01105v2

[7] N.M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L.C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. GT Walker, C. Walters, P. Winskill, C. Whittaker, C.A. Donnelly, S. Riley, and A.C. Ghani.
Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. DOI: 10.25561/77482, Imperial College London, March 16, 2020

[8] K.J. Friston, T. Parr, P. Zeidman, A. Razi, G. Flandin, J. Daunizeau, O. J. Hulme, A. J. Billig, V. Litvak, R. J. Moran, C. J. Price, and C. Lambert. Dynamic causal modelling of COVID-19, arxiv:2004.04463, April 9th, 2020.

[9] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo and M. Colaneri. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. naturemedicine, April 22nd, 2020, https://doi.org/10.1038/s41591-020-0883-7

[10] HW Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42:599–653, 2000.

[11] F.C. Hoppensteadt and P. Waltman. A problem in the theory of epidemics I. *Math. Biosci.*, 9:71–91, 1970.

[12] F.C. Hoppensteadt and P. Waltman. A problem in the theory of epidemics II. *Math. Biosci.*, 12:133–145, 1971.

[13] GitHub Inc. Covid-19 repository at GitHub. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

[14] S. Khailaie, T. Mitra, A. Bandyopadhyay, M. Schips, P. Mascheroni, P. Vanella, B. Lange, S. Binder, and M. Meyer-Hermann. Estimate of the development of the epidemic reproduction number R_t from Coronavirus SARS-CoV-2 case data and implications for political measures based on prognostics. medRxiv, April 7th, 2020, DOI: 10.1101/2020.04.04.20053637v1

[15] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R.M. Eggo. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. May 1st, 2020, Volume 20, issue 5, 553-558, DOI: 10.1016/S1473-3099(20)30144-4.

[16] B. F. Maier and D. Brockmann. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science, 368 (6492): 742–746, April 2020, DOI: 10.1126/science.abb4557

[17] J. Mohring, R. Wegener, S. Gramsch und A. Schöbel. Prognosemodelle für die Corona-Pandemie. Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM Kaiserslautern. April 29th, 2020, https://www.itwm.fraunhofer.de/content/dam/itwm/de/documents/

PressemitteilungenPDF/2020/20200429_Bericht_Prognosemodelle-für-die-Coronapandemie.pdf

[18] Robert-Koch-Institut. SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19). Technical report, Version of May 22nd,2020 https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html

[19] R. Schaback. Modelling recovered cases and death probabilities for the COVID-19 outbreak. arxiv:2003.12068, March 26nd, 2020.

[20] A . Sentker. Bloß raus hier! *DIE ZEIT*. April 16th, 2020, page 20.

[21] H. Streeck, B. Schulte, B. Kuemmerer, E. Richter, T. Hoeller, C. Fuhrmann, E. Bartok, R. Dolscheid, M. Berger, L. Wessendorf, M. Eschbach-Bludau, A. Kellings, A. Schwaiger, M. Coenen, P. Hoffmann, M. Noethen, A.-M. Eis-Huebinger, M. Exner, R. Schmithausen, M. Schmid, and B.Kuemmerer. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. medRxiv, May 8th, 2020, DOI: 10.1101/2020.05.04.20090076

[22] R. Verity, L.C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G.T. Walker, H. Fu, A. Dighe, J.T, Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A.Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H, Wang, Y. Wang X. Xi, C.A. Donnelly, A. C. Ghani, and N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases, Volume 20, issue 6, 669-677, June 01, 2020 DOI: 10.1016/S1473-3099(20)30243-7

[23] Wikipedia. Compartmental models in epidemiology. https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

Appendix, not for publication:
The suppressed argument:

Between 0 and τ_I:

$$
\begin{array}{rcccl}
s(\tau_I) & \leq & s(\tau) & \leq & s(0) \\
i(0) & \leq & i(\tau) & \leq & i(\tau_I) \\
-R_0 i(0) s(\tau_I) & \geq & -R_0 i(\tau) s(\tau) & \geq & -R_0 i(\tau_I) s(0) \\
-R_0 i(0) s(\tau_I) \tau & \geq & \log\left(\dfrac{s(\tau)}{s(0)}\right) & \geq & -R_0 i(\tau_I) s(0) \tau \\
-R_0 i(0) s(\tau_I) \tau_I & \geq & \log\left(\dfrac{s(\tau_I)}{s(0)}\right) & \geq & -R_0 i(\tau_I) s(0) \tau_I \\
+R_0 i(0) s(\tau_I) \tau_I & \leq & \log(s(0) R_0) & \leq & +R_0 i(\tau_I) s(0) \tau_I \\
i(0) s(\tau_I) & \leq & \dfrac{\log(s(0) R_0)}{R_0 \tau_I} & \leq & i(\tau_I) s(0) \\
\dfrac{i(0)}{R_0} & \leq & \dfrac{\log(s(0) R_0)}{R_0 \tau_I} & \leq & s(0)\left(1 - \dfrac{1}{R_0} - \dfrac{\log(s(0) R_0)}{R_0}\right) \\
i(0) & \leq & \dfrac{\log(s(0) R_0)}{\tau_I} & \leq & s(0)\left(R_0 - 1 - \log(s(0) R_0)\right) \\
\dfrac{\log(s(0) R_0)}{s(0)\left(R_0 - 1 - \log(s(0) R_0)\right)} & \leq & \tau_I & \leq & \dfrac{\log(s(0) R_0)}{i(0)}
\end{array}
$$