Defining and computing reproduction numbers to monitor the outbreak of Covid-19 or other epidemics

Paulo R. Zingano, Janaína P. Zingano,

Institute of Mathematics and Statistics
Universidade Federal do Rio Grande do Sul
Porto Alegre, RS 91509-900, Brazil

Alessandra M. Silva

Companhia de Planejamento do Distrito Federal
Governo de Brasília
Brasília, DF 70620-080, Brazil

and

Carolina P. Zingano,

School of Medicine
Universidade Federal do Rio Grande do Sul
Porto Alegre, RS 90035-003, Brazil

Abstract

We present a general approach to define *reproduction ratios* or *numbers* to monitor the outbreak of epidemics that are modeled by mathematical evolution equations. This provides a solution to an important topic that has not been completely settled in the literature, especially in the case of complex epidemiological models. We illustrate our procedure with a full implementation of a standard deterministic SEIR model that is applied to examine the Covid-19 outbreaks and the effects of intervention measures in several countries in America (Argentina, Brazil, Mexico, USA) and Europe (France, Italy, Spain and UK) in 2020. Our code is also used to investigate *herd immunity* levels for Covid-19, indicating values between 85% and 90%.

Key words: Covid-19 outbreak, SARS-Cov-2 coronavirus, reproduction numbers, SEIR models, parameter determination, herd immunity against Covid-19

**AMS Subject Classification:** 92-04 (primary); 92-08, 92-10 (secondary)

Matlab code: A complete matlab source code to compute reproduction numbers of Covid-19 or other epidemics is freely available by clicking here: find_Rt.m.

Corresponding author: Paulo R. Zingano (ORCID: 0000-0002-5074-9146)
E-mail: paulo.zingano@ufrgs.br

1. Introduction

The monitoring of the evolving state of a serious epidemic can be done during and after its outbreak by estimating the daily values of basic ratios generally known as *reproductive* or *reproduction numbers* [7, 10, 11, 16]. While not properly geared at providing serious predictions of future values of the epidemic, they are powerful indicators of the present state of transmission and the effect of intervention measures. However, their calculation may not be immediately clear, particularly in the case of complex (e.g., highly stratified) mathematical models, where open issues still remain [10, 18]. For a review of several methods presently in use, see [8, 10, 17, 20].

Common choices of reproductive ratios are the *basic reproduction number* R_0 and its time dependent variant, the *effective reproduction number* R_t, which measure the expected number of secondary infections caused by one typical infected individual during the entire transmission period [3, 5, 6, 9, 20], and popular methods to estimate them can be found in [2, 20, 22]. As with other indicators, their estimation depends on the values of various mathematical parameters, like the length of transmission and latent periods, which may impair their determination (Figure 1). Other reproductive ratios, like the equally useful *weekly reproduction number* and further examples considered here, might be preferred for their increased robustness.
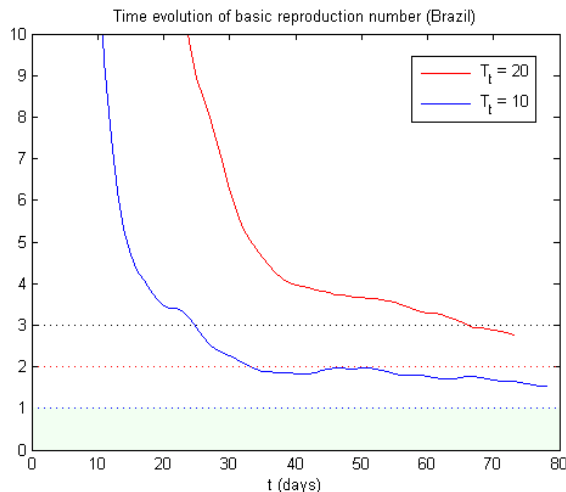
Fig. 1: Time evolution of *effective reproduction numbers* of Covid-19 in Brazil during the first few weeks since the date of 100 cases reported ($t = 0$), showing the effects of two distinct hypothetical transmission periods ($T_t = 20$ and $T_t = 10$, resp.). In this example, $t = 0$ corresponds to 03/13/2020.

(Data source: covid.saude.gov.br)

In fact, it appears to have not been sufficiently stressed in the literature that once some mathematical model has been chosen to simulate the disease dynamics and had its parameters determined from the available data, several natural reproductive numbers become automatically available at no additional computational cost. This is true of both simple and complex models, deterministic or stochastic, potentially providing new indicators that may reveal important information not available in the standard ratios R_0 or R_t above, especially for complex models.

The idea behind it is easily explained by considering a simple SEIR model, defined by the equations (1.1) below, where for the sake of simplicity we have ignored effects like birth or migration rates, deaths by natural or other causes, and so forth. This model divides the entire population in question into four classes: the *susceptible* individuals (class S), those *exposed* (class E, formed by infected people who are still inactive, that is, not yet transmitting the disease), the *active infected* or *infectious* individuals (class I) and the *removed* ones. The latter class is formed by those who have *recovered* from the disease (class R) or who have *died* from it (class D). The dynamics between the various classes in this basic setting is given in the universal language of calculus by the differential equations

$$
\begin{cases}
\dfrac{dS}{dt} = -\beta \dfrac{S(t)}{N} I(t), \\[2mm]
\dfrac{dE}{dt} = \beta \dfrac{S(t)}{N} I(t) - \delta E(t), \\[2mm]
\dfrac{dI}{dt} = \delta E(t) - (r + \gamma) I(t), \\[2mm]
\dfrac{dR}{dt} = \gamma I(t), \\[2mm]
\dfrac{dD}{dt} = r I(t),
\end{cases}
\tag{1.1}
$$

see e.g. [1, 6, 11, 15] for a detailed discussion of the various terms and their meanings. The parameters β (AVERAGE TRANSMISSION RATE) and r (AVERAGE LETHALITY RATE due to the disease) vary with t (time, here measured in DAYS), but δ and γ are typically positive constants given by

$$
\gamma = \frac{1}{T_t}, \qquad \delta = \frac{1}{T_\ell},
\tag{1.2}
$$

where T_t denotes the AVERAGE TRANSMISSION TIME and T_ℓ stands for the mean length of the LATENT PERIOD (the time taken to become infectious, once infected), which for Covid-19 can be taken as 14 and 5 days, respectively [13, 14, 21]). In the system (1.1), N denotes the full size of the susceptible population initially exposed, so that we have $S(t_0) + E(t_0) + I(t_0) + R(t_0) + D(t_0) = N$, where t_0 denotes the initial time. Observing that, by the equations (1.1), the sum $S(t) + E(t) + I(t) + R(t) + D(t)$ is invariant in time, it follows the CONSERVATION LAW

$$
S(t) + E(t) + I(t) + R(t) + D(t) = N, \qquad \forall\ t > t_0,
\tag{1.3}
$$

which is a consequence of the assumptions above.

To well define the model (1.1), besides informing the functions $\beta(t)$ and $r(t)$ we need to provide the initial values $S(t_0)$, $E(t_0)$, $I(t_0)$, $R(t_0)$, $D(t_0)$, which is not a trivial task, since not all of these variables are reported, and those reported may be in error — which may well be large. It thus seems clear that predicting reasonably correct values for the variables $S(t)$, $E(t)$, $I(t)$, $R(t)$ and $D(t)$ at future times is *not* a simple task, especially in the long time range. The situation becomes even more complicated for more complex (e.g., stratified) models, which add other variables and parameters to be determined. Calibrating three or more parameters can quickly become a real nightmare. For all its simplicity, models with only a few variables and parameters like (1.1) can yield surprisingly good results and thus should not be overlooked, as will be seen in the sequel. Still, estimating the missing data and parameters for the ODE system (1.1) requires careful thinking and is a critical part of the implementation stage. The specific procedures are described in Section 2.

Once the epidemiological model has been fully implemented, it can be used in a number of ways to investigate the disease dynamics. An important application is the generation of suitable time varying *reproduction numbers* to monitor various facets of the disease transmission levels. The general idea is to look at the variables of key interest and determine the ratios of their values at the endpoints of some time frame $[t,\,t+d]$ (or, if preferred, $[t-d,\,t+d]$, etc), with $d > 0$ chosen. Complex models naturally offer in this way a realm of interesting choices, leading itself to a rich set of reproductive ratios that should all be considered. Of course, simpler models are more restricted. For (1.1), we may choose to consider

$$R_t^{(1)} = \frac{I(t+d)}{I(t-d)} \tag{1.4}$$

(a reproductive ratio related to the infectious population), or the slight variant

$$R_t^{(2)} = \frac{E(t+d) + I(t+d)}{E(t-d) + I(t-d)} \tag{1.5}$$

(related to the total infected population), with the choice $d = T_t/2$ corresponding to the *effective reproduction number R_t* already mentioned. For Covid-19, a convenient choice is to take $d = 3$, as explained in Section 3. This reproductive ratio is used in Section 4 to examine the unfolding of the epidemic in several countries in 2020.

With vaccination against Covid-19 now well under way in many countries, we close our discussion in Section 5 with one last illustration of the multiple uses of our implementation of (1.1): this time, estimating *herd immunity* values of the disease. Following (1.1) and the results obtained for the parameters β and r by our code, herd immunity is estimated to lie between 85% and 90% of the entire population, depending on the transmission levels that can be safely assumed.

2. Implementing the SEIR model

Having introduced the SEIR equations (1.1), we now describe an implementation of this model that is suitable for all our needs (and much more).

(*i*) *assigning a value to the population parameter N*

In the case of Covid-19, which was caused by a new virus (SARS-CoV-2), it is reasonable to assume that the entire population of the region under consideration is initially susceptible, which was done in the code. In any case, it turns out that the exact value of N is not so important for the short range dynamics (which accounts for the applications studied in this paper) as it proves to be for long time simulations (Figures 2*a* and 2*b*), so that for our present purposes this is not a serious issue.
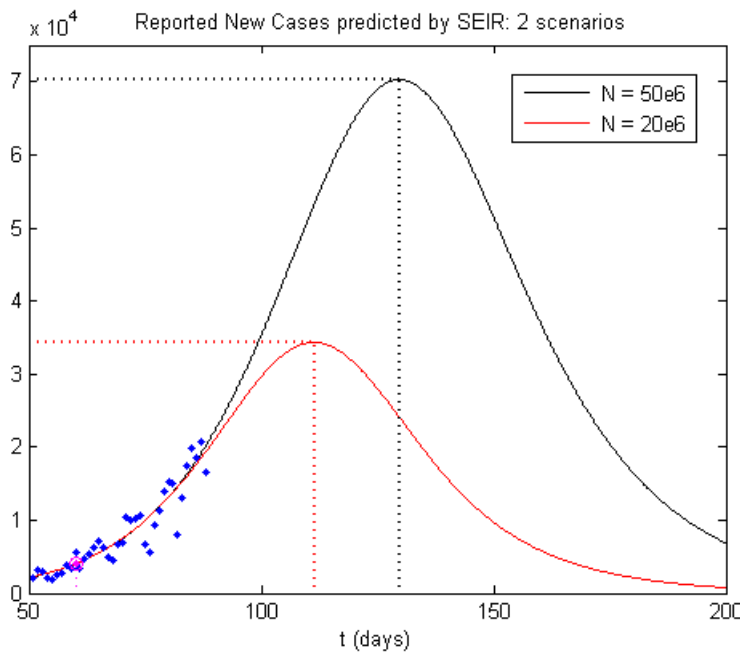
Fig. 2*a*: Prediction by model (1.1) of the daily number of *new cases* of Covid-19 expected to be reported in Brazil between the initial time $t = t_0 = 60$ (April 25th) and $t = 200$ (September 12th), considering susceptible populations of $N = 20$ million (red curve) and $N = 50$ million (black curve). Note the appreciable difference between the predicted peak values (34 and 70 thousand, resp.) and their respective dates, June 6th and July 4th. Actual data points are shown in blue. (Computed from data available at the official site https://covid.saude.gov.br.)
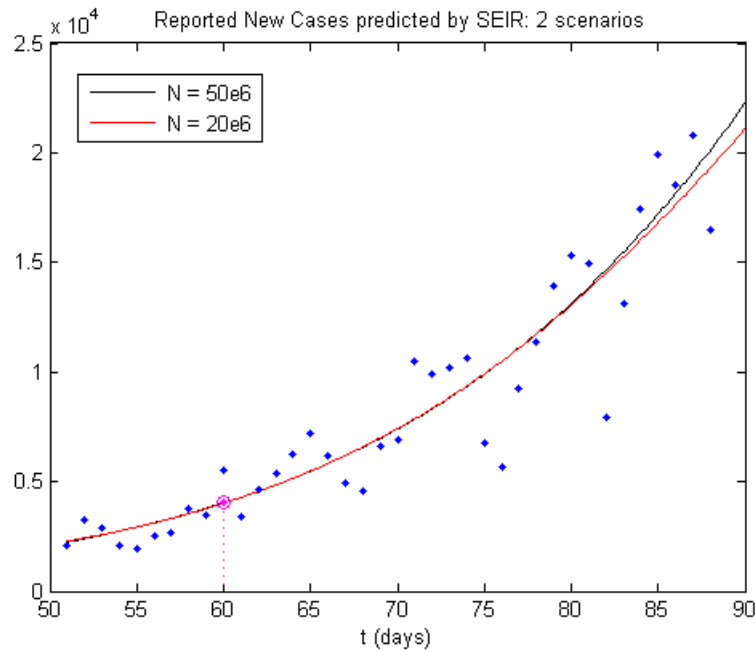
Fig. 2*b***:** Thirty day prediction by model (1.1) of the daily number of *new cases* of Covid-19 to be reported in Brazil between the initial time $t = t_0 = 60$ (04/25) and $t = 90$ (05/25), considering susceptible exposed populations of $N = 20$ million (red curve) and $N = 50$ million (black curve). Note the very close similarity of the two 30D predictions in spite of the appreciable difference in the values of N. Points shown in blue are the official values reported (cf. https://covid.saude.gov.br.)

(*ii*) *generation of initial data* $S(t_0)$, $E(t_0)$, $I(t_0)$, $R(t_0)$, $D(t_0)$

Initial values S_0, E_0, I_0, R_0, D_0 for the five variables are generated from a starting date t_s on, which is taken so as to meet some minimum value chosen of total reported cases (namely, 100). Denoting by $C_r(t)$ the total amount of reported cases up to some time t, and letting EIR(t) be the sum of the populations $E(t)$, $I(t)$ and $R(t)$, we set

$$\text{EIR}(t_s) \; = \; f_c \cdot (C_r(t_s) - D(t_s)), \tag{2.1}$$

where $f_c \geq 1$ denotes a CORRECTION FACTOR to account for likely underreportings on the official numbers given. (In (2.1), we have neglected possible underreportings on the number of deaths, which could of course be similarly accounted for if desired.) Again, this factor f_c will not play an important role in this paper and could be safely ignored, but it should be carefully considered in the case of long time predictions. We have tipically taken $f_c = 5$. Having estimated EIR(t_s), we then set

$$E(t_s) \; = \; E_0(t_s) := \; a \cdot (1 - b) \cdot \text{EIR}(t_s), \tag{2.2a}$$

$$I(t_s) \; = \; I_0(t_s) := \; (1 - a) \cdot (1 - b) \cdot \text{EIR}(t_s), \tag{2.2b}$$

$$R(t_s) = R_0(t_s) := b \cdot \text{EIR}(t_s), \tag{2.2c}$$

$$S(t_s) = S_0(t_s) := N - \big(E(t_s) + I(t_s) + R(t_s) + D(t_s)\big), \tag{2.2d}$$

where $a = T_i/(T_i + T_t)$ and $b = 0.30$. Although (2.2) above might seem reasonable, the expressions (2.2a)-(2.2c) are nevertheless arbitrary and will be probably in error. However, all the errors will eventually fade away (Figure 3) as we compute more values $S_0(t_0), E_0(t_0), I_0(t_0), R_0(t_0), D_0(t_0)$ at later initial times $t_0 = t_s + 1, ..., t_F$, where t_F stands for the final (i.e., most recent) date of reported data available. This can be done as follows. For each t_0, the solution of the equations (1.1) with the previously obtained initial data at $t_0 - 1$ is computed on the interval $J(t_0) = [t_0 - 1, t_1]$, $t_1 = \min\{t_0 - 1 + d_0, t_F\}$, with constant parameters $\beta = \beta_0(t_0 - 1)$, $r = r_0(t_0 - 1)$ determined so that the computed values for $C_r(t)$, $D(t)$ best fit the reported data for these variables on $[t_0, t_1]$ in the sense of LEAST SQUARES [15]. (Here, $d_0 \in [2, 10]$ is chosen according to the data regularity.) Once this solution $(S, E, I, R, D)(t)$ is obtained, we set $S_0(t_0) := S(t_0)$, $E_0(t_0) := E(t_0)$, $I_0(t_0) := I(t_0)$, $R_0(t_0) := R(t_0)$, $D_0(t_0) := D(t_0)$ and move on to the next time level $t_0 + 1$, repeating the procedure until t_F is reached.
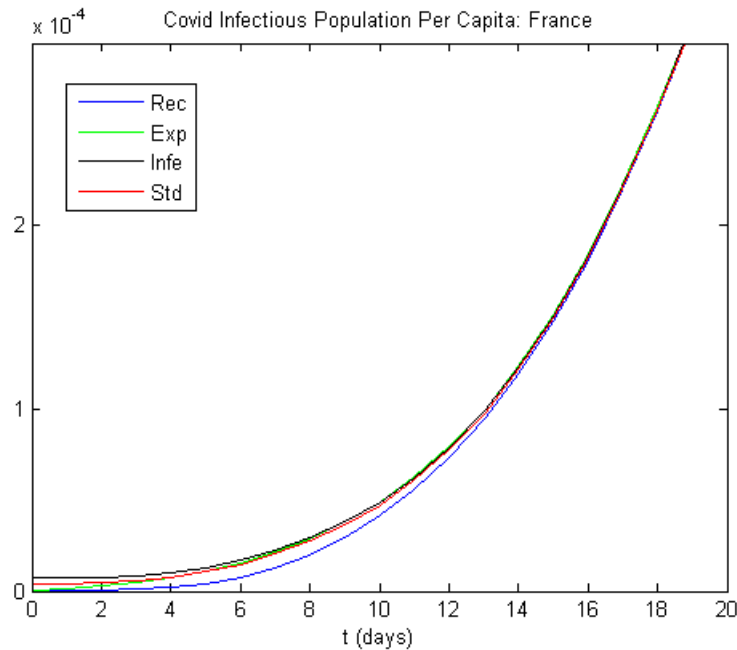
Fig. 3*a*: Self-correction in the initialization procedure (*ii*) to generate $S_0(t_0)$, $E_0(t_0)$, $I_0(t_0)$ and $R_0(t_0)$ for $t_0 = t_s, t_s + 1, ..., t_F$, showing above the case of $I_0(t_0)/N$ in France. At time $t_s = 0$, the day of 100 total cases reported (29/02/2020), four very different sets of values $\{E_0(t_s), I_0(t_s), R_0(t_s)\}$ are considered: our standard choice (2.2a)-(2.2c), shown in red; $E_0(t_s) = I_0(t_s) = 0.05 \times \text{EIR}(t_s)$, $R_0(t_s) = 0.90 \times \text{EIR}(t_s)$, shown in blue; $E_0(t_s) = \text{EIR}(t_s)$, $I_0(t_s) = R_0(t_s) = 0$, shown in green; and $I_0(t_s) = \text{EIR}(t_s)$, $E_0(t_s) = R_0(t_s) = 0$, in black. For $t_0 > t_s + 15$, all four initializations produce roughly the *same* values for $I_0(t_0)$.
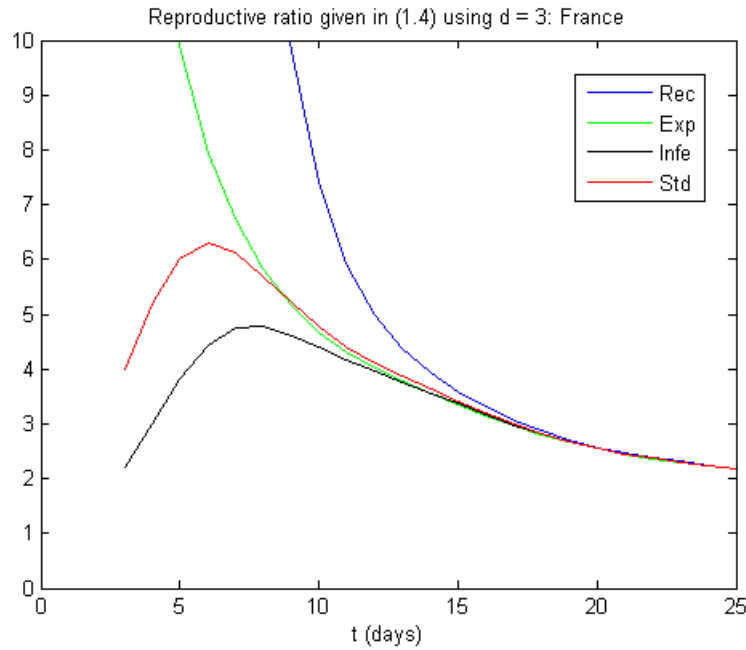
Fig. 3b: Self-correction in the values of $R_t = I_0(t+3)/I_0(t-3)$ induced by the correction of $E_0(t_0)$, $I_0(t_0)$ and $R_0(t_0)$ exhibited in Fig. 3a, considering the same four initializations at $t_s = 0$.

(iii) computing the solution on some final interval $[t_0, T]$ (PREDICTION PHASE)

Having completed the previous steps, we can address the possibility of *prediction*. Although this is not important for our present goals, it is included for completeness. Choosing an initial time $t_0 \in (t_s, t_F]$, we then take the initial values

$$S(t_0) = S_0(t_0), \ E(t_0) = E_0(t_0), \ I(t_0) = I_0(t_0), \ R(t_0) = R_0(t_0), \ D(t_0) = D_0(t_0).$$

In order to predict the values of the variables $S(t), E(t), I(t), R(t), D(t)$ for $t > t_0$, it is important to have good estimates for the evolution of the key parameters $\beta(t)$ and $r(t)$ beyond t_0. This is the most computationally intensive part of the algorithm and is better executed in large computers. Such estimates can be given in the form

$$\beta(t) = \beta_0 + a_\beta \, e^{-\lambda_\beta(t - t_0)} \qquad (2.3a)$$

$$r(t) = r_0 + a_r \, e^{-\lambda_r(t - t_0)} \qquad (2.3b)$$

where $\beta_0, a_\beta, \lambda_\beta, r_0, a_r, \lambda_r \in \mathbb{R}$ are determined so as to minimize the maximum size of weighted RELATIVE ERRORS in the computed values for $C_r(t), D(t)$ as compared to the official data reported for these variables on some previous interval $[t_0 - \tau_0, t_0]$ (weighted CHEBYCHEFF PROBLEM) for some chosen $\tau_0 > 0$ (usually, $20 \leq \tau_0 \leq 30$). This problem is solved iteratively starting with an initial guess obtained from the analysis of the previous values $\beta_0(t), r_0(t)$ computed in the step (ii) above. The result is illustrated in Figure 4 for the case of $\beta(t)$, with similar considerations for $r(t)$.
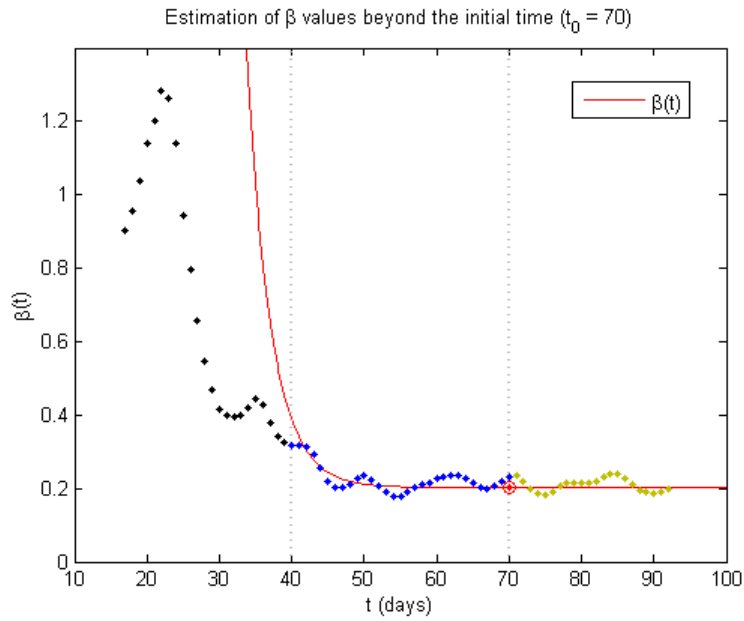
Fig. 4: Estimation of future values of the transmission parameter $\beta(t)$ beyond the initial time t_0 = 70 (05/05/2020) for the outbreak of Covid-19 in Brazil, assuming the basic form (2.3a), after solving the Chebycheff problem (red curve). The data points in the interval $[40, 70]$, shown here in blue, are values of the function $\beta_0(t)$ computed in step (ii), which are used to obtain the first approximation to $\beta(t)$. Values of $\beta_0(t)$ previous to $t = 40$ (04/05/2020), shown in black, are disregarded. The golden points beyond $t_0 = 70$ are future values of $\beta_0(t)$, not available on 05/05/2020, displayed to allow comparison with the predicted values $\beta(t)$.

Once $\beta(t)$, $r(t)$ have been obtained, the equations (1.1) are finally solved (Figure 5).
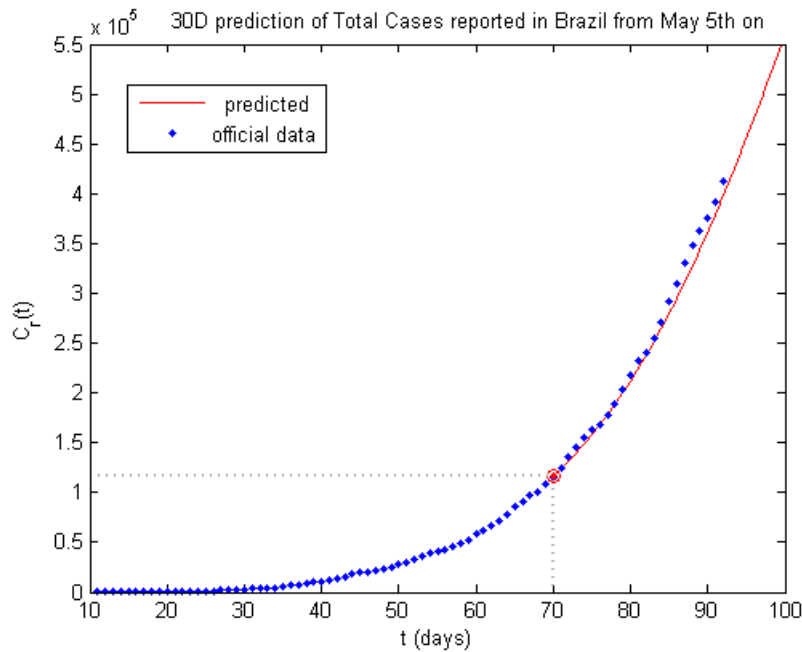
Fig. 5: Computation of $C_r(t) = \big(E(t) + I(t) + R(t)\big)/f_c + D(t)$ for $t > t_0 = 70$ (05/05/2020), with initial data $C_r(t_0) = \big(E_0(t_0) + I_0(t_0) + R_0(t_0)\big)/f_c + D_0(t_0)$, after obtaining $\beta(t)$, $r(t)$ – see Fig. 3 for $\beta(t)$. The numerical solution of equations (1.1) is easily obtained by any method.

3. Reproduction numbers

Reproduction numbers or ratios measure the intensity of transmission of a contagious disease as it unfolds in some population, and, in doing so, are useful indicators to monitor the situation and the effects of intervention procedures that might have been taken. Using the generic symbol R_t to denote such quantities,[1] they signal a rise in the number of infections in the case $R_t > 1$, their decrease when $R_t < 1$, and temporary steadiness if $R_t = 1$. There are basically two ways of generating them.

(i) *reproduction ratios obtained from the analysis of the model equations*

This tends be complicated for complex models [10, 18], but is a common method otherwise. For the SEIR equations (1.1), for instance, we may proceed as follows. With an eye on the critical infectious population, we rewrite the equation (1.1c) as

$$\frac{dI}{dt} = \alpha(t)\,I(t), \qquad \alpha(t) := \delta\cdot E(t)/I(t) - r(t) - \gamma, \qquad (3.1a)$$

showing that $I(t)$ will increase if $\alpha(t) > 0$, decrease when $\alpha(t) < 0$ and stay about the same if $\alpha(t) = 0$ — or, in terms of the ratio

$$R_t := \frac{\delta\cdot E(t)/I(t)}{r(t) + \gamma}, \qquad (3.1b)$$

whether we have $R_t > 1$, $R_t < 1$ or $R_t = 1$, respectively. Likewise, selecting the total infected population to be tracked, we get, adding the equations (1.1b) and (1.1c),

$$\frac{d\mathbb{I}}{dt} = \tilde{\alpha}(t)\,I(t), \qquad \tilde{\alpha}(t) := \beta(t)\cdot S(t)/N - r(t) - \gamma, \qquad (3.2a)$$

where $\mathbb{I}(t) = E(t) + I(t)$. This leads to the (numerically similar) reproductive ratio

$$R_t := \frac{\beta(t)\cdot S(t)/N}{r(t) + \gamma}. \qquad (3.2b)$$

Expressions like (3.2b) are useful to investigate herd immunity levels (see Section 5).

(ii) *reproduction ratios obtained directly from numerical simulations of the model*

This method, which is always easy to apply, leads in the case of complex models to distinct reproduction numbers yielding different information. Given some quantity of interest $W(t)$ computed by/from the simulations, the ratios $W(t+d)/W(t)$

---

[1]The notation R_t is natural in stochastic models, and is adopted here as we have already used $R(t)$, $R_0(t)$ with other meanings (size of the recovered population and their initial values, resp.).

(or, if preferred, $W(t+d)/W(t-d)$, and so on) are examined, for some $d > 0$ chosen. For the equations (1.1), natural examples are, say,

$$R_t := \frac{I(t+d)}{I(t-d)}, \qquad R_t := \frac{E(t+d)+I(t+d)}{E(t-d)+I(t-d)}, \qquad (3.3)$$

which are again very similar (Figure 6). Note that the choice $d = T_t/2$ corresponds to the standard *basic/effective reproduction number*, i.e., the mean number of secondary infections caused by a typical infected individual during his entire transmission period [13, 15]. In our code, the corresponding expressions for (3.1b) and (3.3) with $d = 3$ would be, recalling the variables computed in the step (ii) of the algorithm,

$$R_t^{(1)} := \frac{\delta \cdot E_0(t)/I_0(t)}{r_0(t)+\gamma} \qquad (3.4)$$

and

$$R_t^{(2)} := \frac{I_0(t+3)}{I_0(t-3)}, \qquad R_t^{(3)} := \frac{E_0(t+3)+I_0(t+3)}{E_0(t-3)+I_0(t-3)}, \qquad (3.5)$$

and similarly for (3.2b). All these indicators registrate similar scenarios (Figure 6a). We have found $R_t^{(2)}$ particularly convenient for Covid-19, producing results that are very similar to those obtained in other studies (see e.g. [16]). For the time scales involved, the choice $d = 3$ is good to zoom in the details and facilitate the reading (Figure 6b), while not compromising robustness (Figure 7).
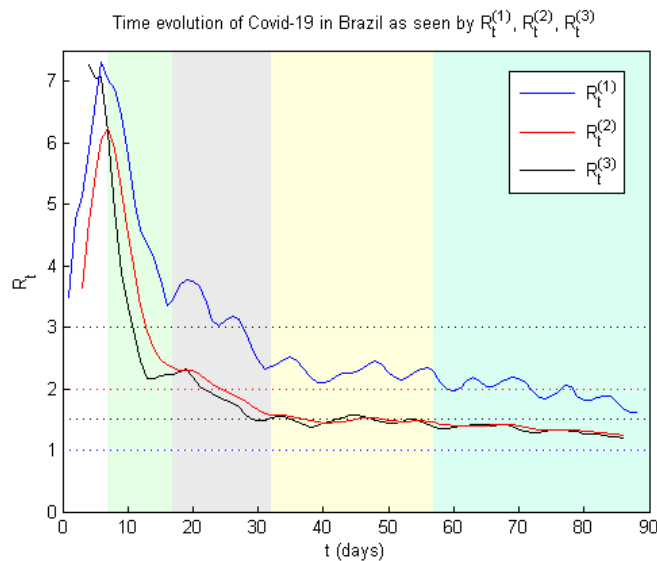
Fig. 6a: Comparison of the time evolution of Covid-19 in Brazil (since 100 cases reported) as seen by the indicators defined in (3.4) and (3.5), all pointing to similar scenarios. In the three cases it is clear that Brazil has never reached a state of control of the epidemic ($R_t < 1$) during the period considered. (From official data reported at https://covid.saude.gov.br.)

11
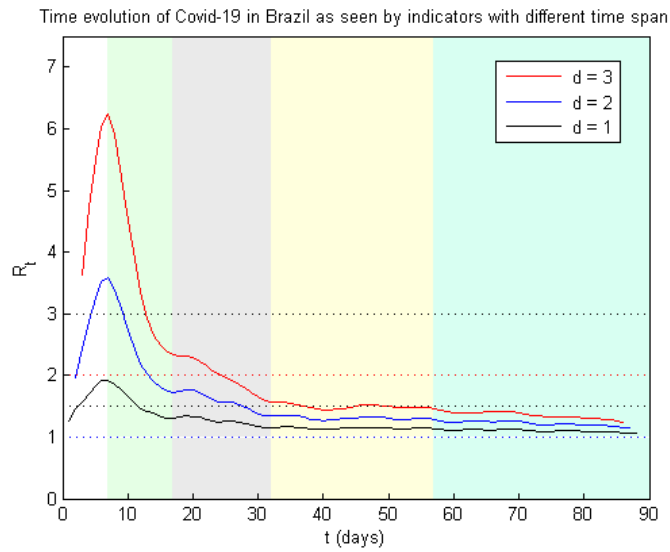
Fig. 6b: Comparison of the time evolution of Covid-19 in Brazil (since 100 cases reported) as seen by $R_t = I(t+d)/I(t-d)$ for different values of d, showing similar scenarios. In the three cases it is clear that Brazil has never reached a state of control of the epidemic ($R_t < 1$) during the period considered. (From official data reported at https://covid.saude.gov.br.)
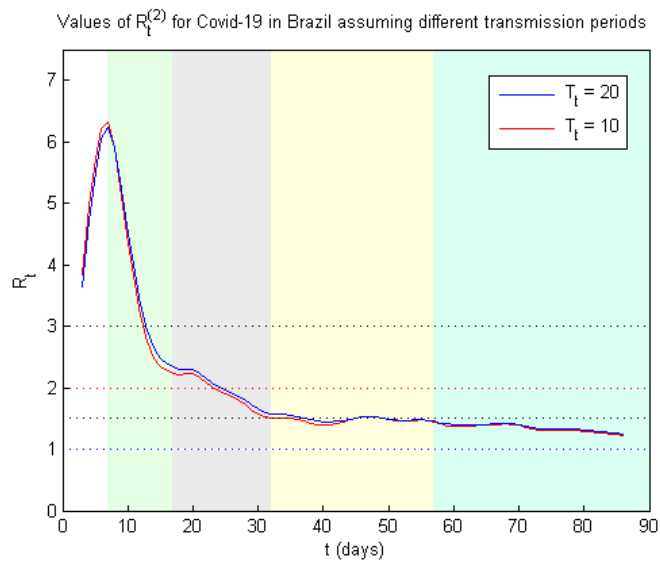
Fig. 7: Robustness of $R_t^{(2)}$ with respect to large uncertainties on the value of transmission time. Date zero refers to 100 cases reported, that is: 03/13/2020. (As in Fig. 5 and Fig. 6 above, calculations were based upon official data reported at https://covid.saude.gov.br.)

Next, we will apply the indicator $R_t^{(2)}$ above to examine the evolution of the Covid-19 epidemic in several countries around the world during the year 2020, including our original comments made in late June of that year.

## 4. Applications to the Covid-19 epidemic in 2020

In this section we will illustrate the use of reproduction numbers by examining the evolution of Covid-19 in eight different countries under the optics of the ratio $R_t^{(2)}$ associated with SEIR model (1.1), cf. (3.5) above. Thus, we set

$$R_t = \frac{I_0(t+3)}{I_0(t-3)} \qquad (4.1)$$

where $I_0(\tau)$ is the size of the ACTIVE INFECTED POPULATION at time τ as computed in the step (ii) of the SEIR algorithm (see Section 2). The computation of the curves below was based on the data for each country available at worldometers/coronavirus. For each case, we reproduce our previous analysis in late June/2020 (shown in blue), followed by an update (in black) covering the rest of the year. In spite of the efforts, the overall picture is a general failure in achieving a sustained control of the epidemic through the use of non-pharmacological intervention.

Taking right decisions about contention or relaxation measures is a very difficult and complex process that involves a careful consideration of several mathematical indicators and other factors including many health, economic and social issues. In the following examples we consider only the single viewpoint of reproductive ratios.

Example 1: Time evolution of Co-vid-19 in Argentina since 03/18/2020 ($t = 0$), the date of 97 total cases reported. Strong containment measures had begun 3 days earlier ($t = -3$) and managed to keep the number of cases and deaths down low, with R_t decreasing continually until 05/04/2020 ($t = 47$), when it reached a minimum value of 1.08. Following that, the situation deteriorated with R_t increasing to 1.54 on 05/24/2020 ($t = 67$), despite the reinforcement of most intervention

Time evolution of 7D reproduction number of Covid-19 in Argentina

procedures. Partial relaxation of some of these measures was introduced on 06/01/2020 ($t = 75$) and, in this new period, R_t has remained relatively stable at 1.30 (yellow band). Bringing the epidemic to a state of nationwide control ($R_t < 1$) still seems far away. This example illustrates the unfortunate fact that having low numbers of infections does not necessarily mean having the epidemic under control.

13
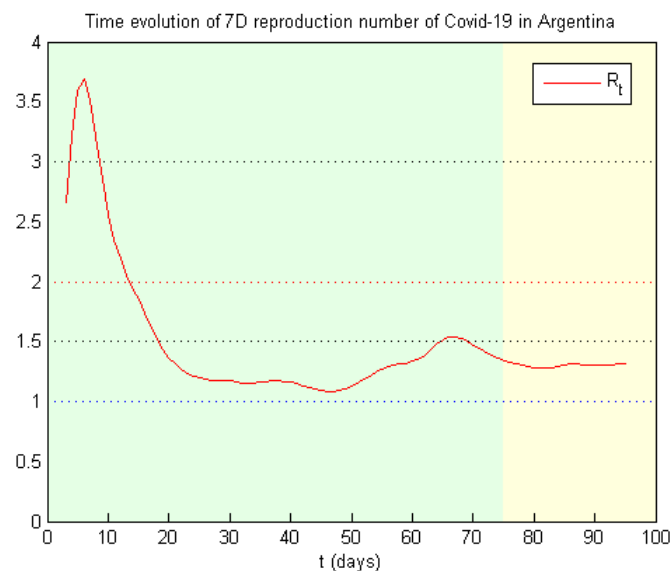
The low numbers of covid cases reported in the first two months after the outbreak (reaching a total of 7.5 thousand cases by mid-May, less than 0.02% of the population) led many to believe in a successfull containment of the disease up to that point — which was elusive, as the R_t values clearly show. The number of infected people kept steadily increasing until reaching almost 3% of the total population in late October, and is presently on the rise again (Figure 8). This quantity is conveniently measured by the sum of the percentages

$$e(t) \; = \; 100 \, \frac{E_0(t)}{N}, \qquad i(t) \; = \; 100 \, \frac{I_0(t)}{N} \tag{4.2}$$

corresponding to the exposed and infectious populations, respectively, where $E_0(t)$, $I_0(t)$ are computed in the step (ii) of the algorithm (Section 2) and N is the total population.
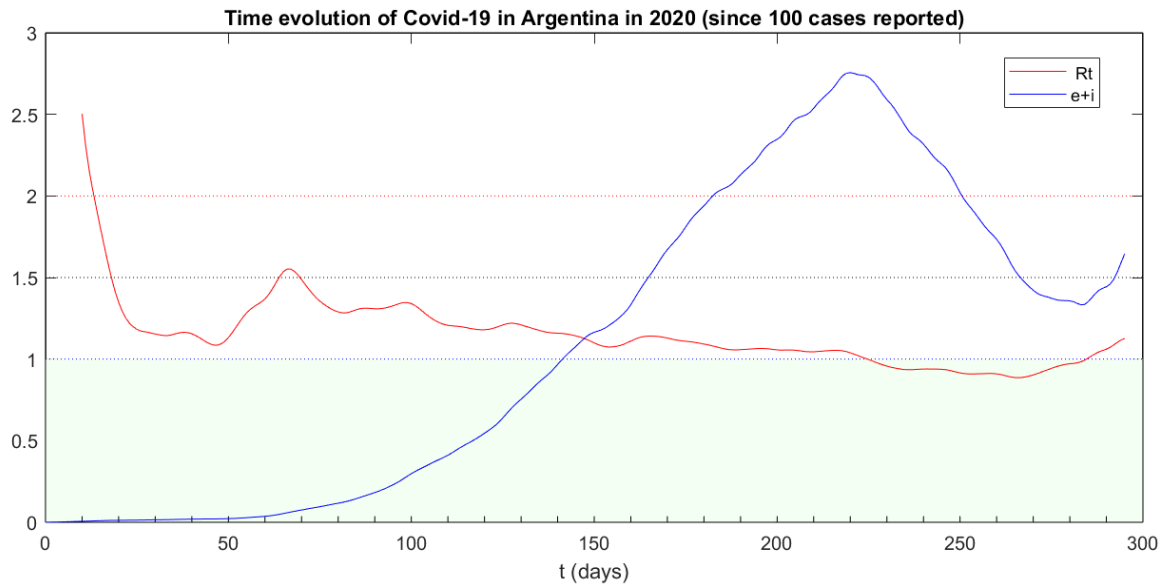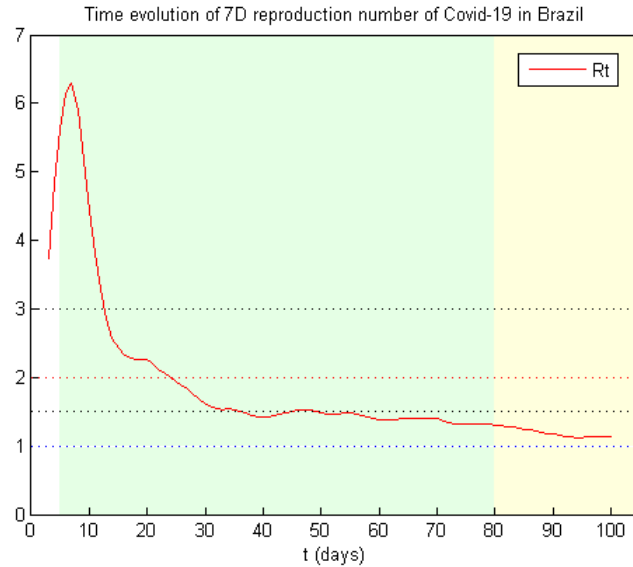

Fig. 8: History of Covid-19 in Argentina in 2020 since the date of 100 cases reported (03/18). For the computation of the percentage of the infected population at day t (blue curve), see (4.2), it was assumed a typical underreporting level with factor $f_c = 5$ (Section 2).

Argentina, Brazil, Colombia and Peru are countries in South America that barely got any control of the epidemic in 2020. In contrast, the disease management in Uruguay was much more successful, keeping the disease under control until the beginning of October. After peaking at 1.42 in mid-December, the values of R_t in Uruguay have been subsiding, reaching 1.17 on 12/31/2020.
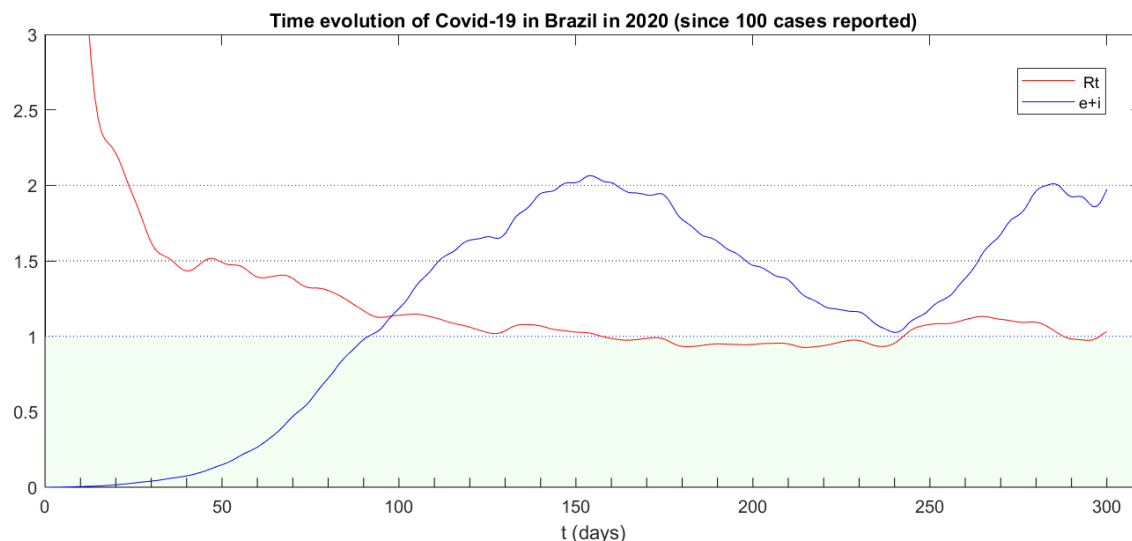
country	estimated population never infected until 12/31/2020
New Zealand	99.8%
Uruguay	97.3%
Argentina	82.5%
UK	82.1%
Brazil	80.7%
France	80.4%
Spain	79.7%

Example 2: Time evolution of Covid-19 in Brazil since 03/13/2020, the date of 98 total cases reported ($t = 0$). With very poor coordination between the central and local authorities and with different levels of intervention in the various states of the country, the decreasing of R_t after reaching 1.5 by mid-April proceeded very slowly (green band) due to the spread of the epidemic and the emergence of new infection foci. Relaxation measures began to be implemented on different dates according to the individual regions, but can be traced on average back to 06/01/2020 ($t = 80$). Despite the encouraging behavior of R_t shown in the last fortnight (yellow band), the indicator is likely to resume increasing due to further disease development in less affected areas of the country, particularly the southern and central western states. This has indeed been the case, with R_t presently increasing. Another negative factor is that flexibilization of control measures has been introduced before the various regions had attained a state of epidemic control ($R_t < 1$), which is *not* ideal.

In fact, R_t continued increasing very slightly for still another week before beginning a slow general decrease that lasted about 10 weeks, followed by a relatively stable two-month period around the value 0.95. Control of the epidemic was then lost from mid-November on, initiating what was then called the second covid wave in the country. On 12/31/2020 an estimated 1.3% of the total population was infectious (i.e., transmitting the disease).

Time evolution of 7D reproduction number of Covid-19 in Brazil

Time evolution of Covid-19 in Brazil in 2020 (since 100 cases reported)

Example 3: Time evolution of Covid-19 in France since 02/29/2020 ($t = 0$), the date of 100 total cases reported. Containment measures began relatively late on 03/16/2020 ($t = 16$), with a strict eight-week lockdown that reduced the value of R_t down to 0.81 (green band). Restrictions were afterwards relaxed (yellow band), with R_t stable for a couple of weeks, when it began increasing. A peak value of 0.99 was reached on 05/30/2020, followed by a reduction to 0.89 on 06/08/2020

Time evolution of 7D reproduction number of Covid-19 in France

($t = 100$), staying on a slow ascent ever since (latest value being 0.95). The situation requires careful monitoring, with the possibility of having to impose some containment restrictions back to keep the epidemic under nationwide control ($R_t < 1$).
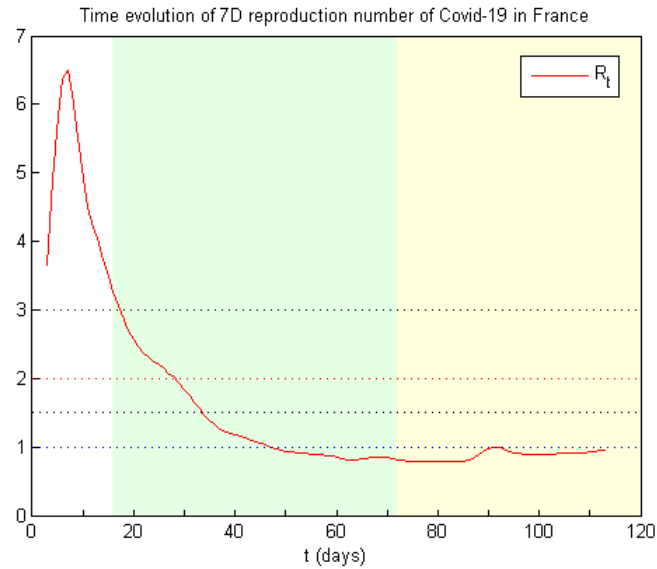
Looking at the final values of R_t in the graph above, we see that the control of the epidemic begun in mid-April ($t \approx 50$) after four weeks of lockdown was being lost by late May ($t \approx 90$), with the situation becoming critical in mid-June ($t \approx 110$). With estimated tens of thousands of infected people still remaining in France and the arrival of summer, the excessive removal of contention rules that followed led to a slow but steady build-up of a second wave that would hit the nation hard (graph below) and claim tens of thousands of lives. Control of the epidemic was regained in the last six weeks of 2020 but deteriorated towards the end, with 1.5% of the population estimated to be infectious on 12/31/2020.
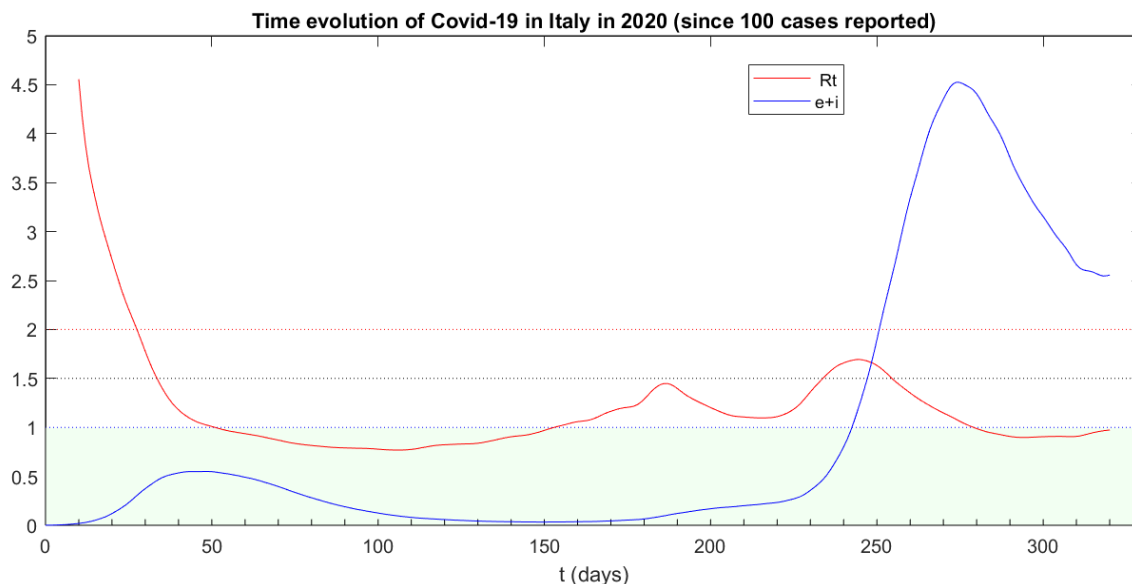

Time evolution of Covid-19 in France in 2020 (since 100 cases reported)

16
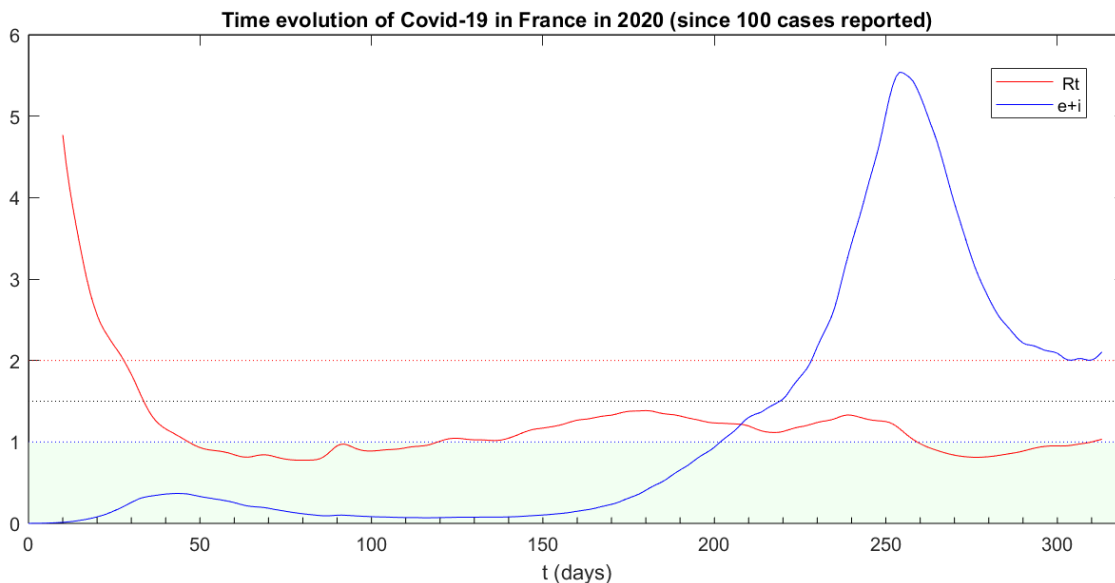
Example 4: Time evolution of Covid-19 in Italy since 02/22/2020, the date of 79 total cases reported ($t = 0$). Containment measures began fifteen days later, with a strict eight-week national lockdown imposed on 03/10/2020 ($t = 17$). The strong intervention, embraced by the population and maintained for the whole period, succeeded in reducing R_t continually down to a safe value of 0.80 on 05/18/2020 ($t = 86$), when some of the contention rules began being relaxed

Time evolution of 7D reproduction number of Covid-19 in Italy
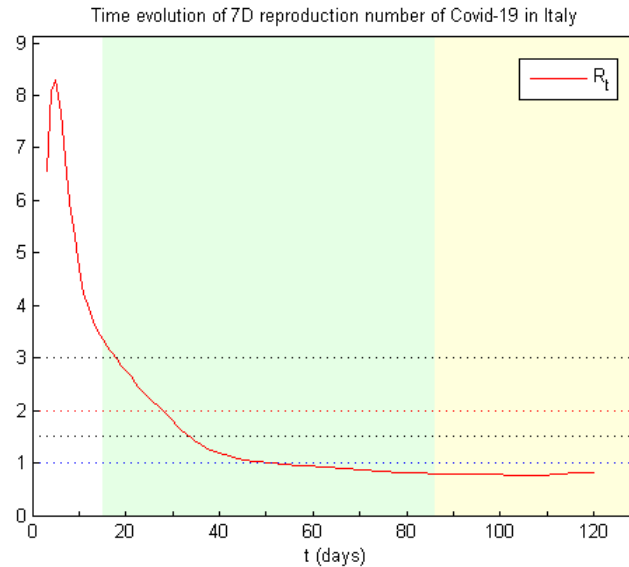
(yellow band). The descent continued for nineteen days, reaching a bottom value of 0.77 on 06/06/2020 ($t = 105$). After this, a steady and very slow increase set in leading to the final value of 0.82 ($t = 120$).

The rising trend of R_t near the end of the period above continued to hold during the summer, increasing R_t to a peak value of 1.45 in late August (graph below). A second and higher peak of 1.7 was reached after mid-October, when the effects of a second wave to hit Italy were clearly visible (blue curve below). Reintroduction of intervention measures managed to bring the epidemic under some control in December, with $R_t = 0.93$ on 12/31/2020 but again on the rise. Assuming a standard underreporting factor $f_c = 5$, the year closed with an estimated infectious population of 1.2 million people in the country. With such critical situation, very careful attention continues to be required.
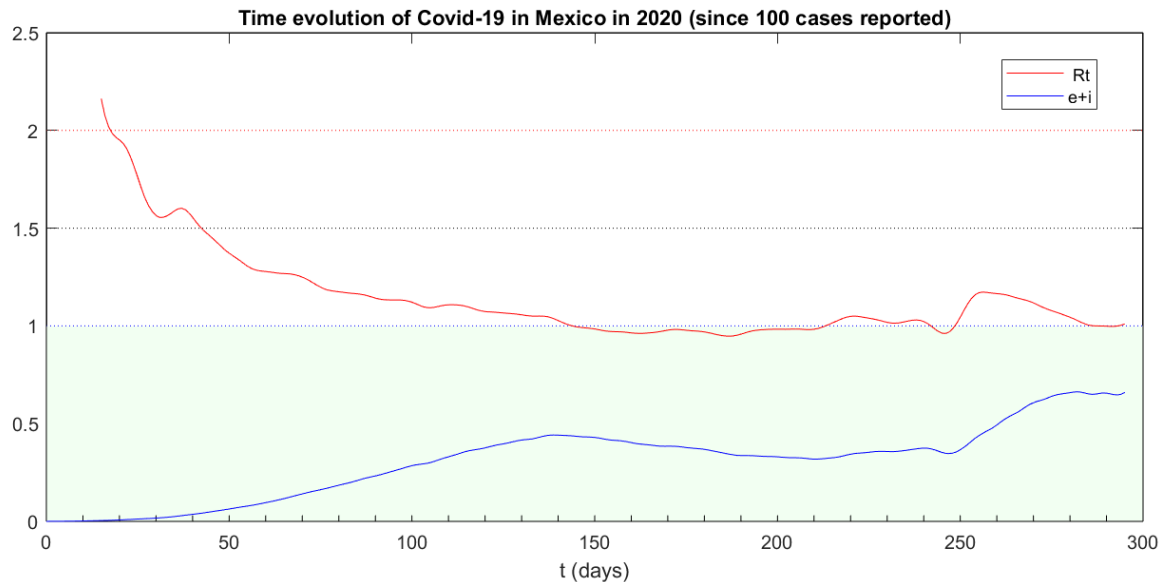

Time evolution of Covid-19 in Italy in 2020 (since 100 cases reported)

17

Example 5: Time evolution of Co-vid-19 in Mexico since 03/18/2020, the date of 93 total cases reported ($t = 0$). After containment measures began on 03/22/2020 ($t = 4$), the value of R_t continually decreased to 1.20 (green band), when restrictions began to be relaxed on 06/01/2020 (yellow band). Relaxation measures have apparently not changed the be-havior of R_t afterwards, but reach-ing a state of control still looks a few weeks away. Similarly to Argentina and Brazil, the flexibilization started

before the country had properly entered the safe zone $R_t < 1$.
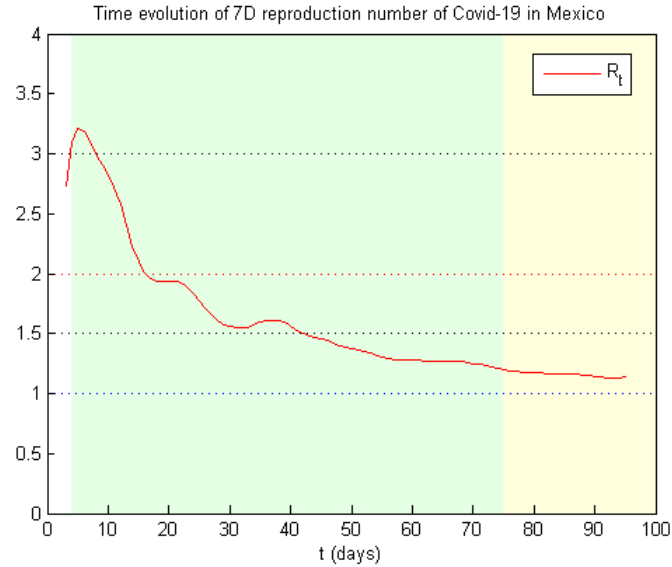
Following this, the R_t values continued their descent, finally reaching slightly under 1 after 7 weeks, with little change in the next 10 weeks or so. This accounted for the long plateau observed for the number of infected people (blue curve below) and other similar quantities, locked from significant variation in this relatively steady period where $R_t \approx 1$. This delicate equilibrium lasted until late November ($t = 250$), when the country was hit by a new infection surge. This led to new intervention measures imposed in mid-December.

Assuming an underreporting factor $f_c = 5$, as in the other examples discussed here, our SEIR code obtained the curve $e + i$ (total infected population at time t) shown below, having significantly lower values in the last few months of 2020 than those observed for the other countries. With little evidence of superior covid control strategies applied there, this is suggestive of higher underreporting levels under practice in Mexico.


18

Example 6: Time evolution of Covid-19 in Spain since 03/01/2020, the date of 84 total cases reported ($t = 0$). After containment measures began on 03/13/2020 ($t = 12$), the value of R_t continually decreased to 0.89 on 05/11/2020 ($t = 71$), when restrictions began to be relaxed (yellow band). A minimum value of 0.74 was finally reached on 06/07/2020 ($t = 98$), after which a slow, steady increase set in towards the present value of 0.81 ($t = 112$), in much the same way as Italy.

Time evolution of 7D reproduction number of Covid-19 in Spain
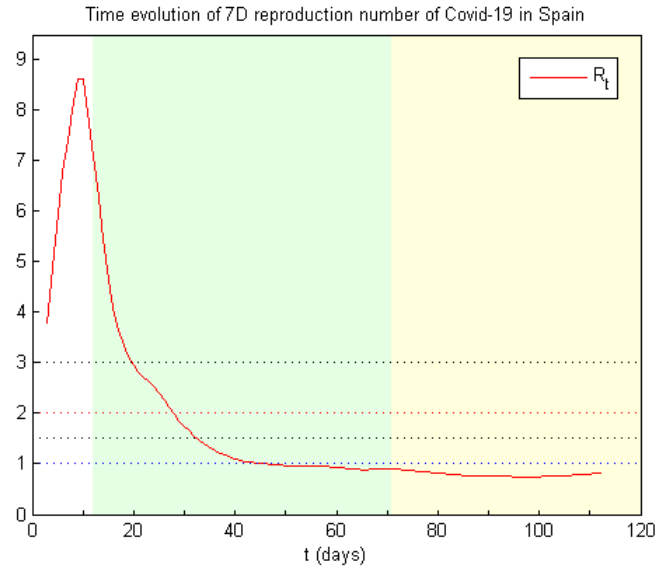
This trend continued for six weeks more until a maximum value $R_t = 1.55$ was reached in mid-July. With an average summer value of about 1.25, superior to the other countries in the continent like France (1.19), Italy (1.09) and UK (1.04), Spain was the first European nation to experience a second infection wave, already visible in mid-July (graph below). This wave extended for the rest of the year and claimed more than 20 thousand lives. With poor coordination among central and local authorities and plenty of political tension in the country, new infections remained on the rise for 4 months, with 2.5% of the population estimated as actively infected in mid-November. New intervention measures were imposed in October, including mobility restrictions and partial lockdown in some areas, which faced considerable resistance. As the year closed, disease control was being lost ($R_t \approx 1.05$ and rising), with 1.25% of the population estimated to be infectious by our code.

Time evolution of Covid-19 in Spain in 2020 (since 100 cases reported)

19
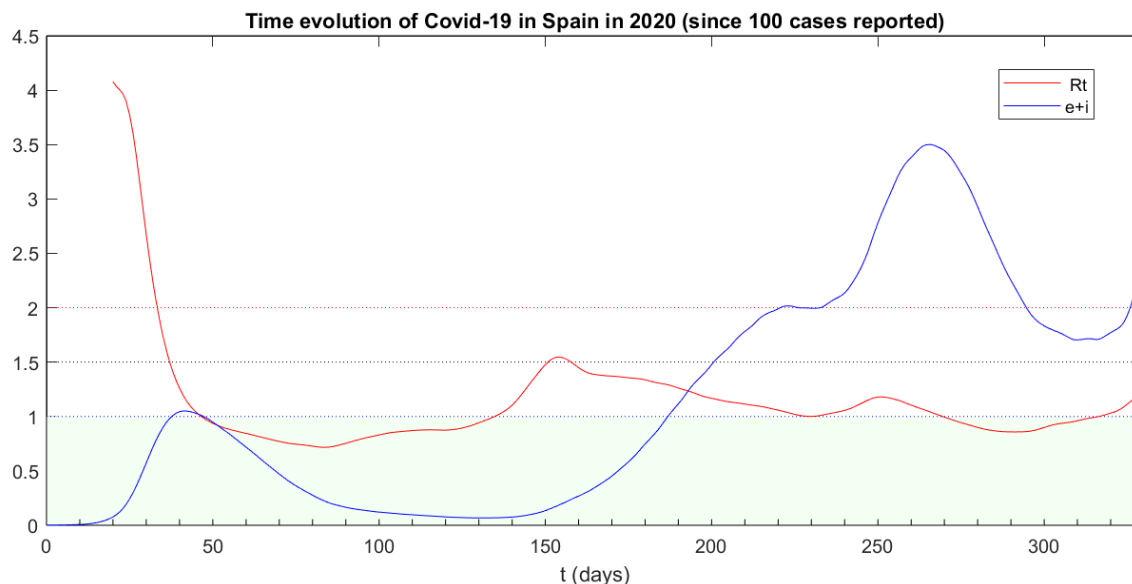
Example 7: Time evolution of Covid-19 in the UK since 03/04/2020, the date of 87 total cases reported $(t = 0)$. After containment measures began relatively late on 03/20/2020 $(t = 16)$, including strict national lockdown and other rules three days later, the value of R_t continually decreased to 0.98 on 05/13/2020 $(t = 70)$, when restrictions began to be relaxed, and then further down to 0.86 nineteen days later, when the lockdown was removed (yellow band). Despite successfully bringing the epidemic under control, the number of reported cases and deaths was very high due to the initial delay in taking intervention action.
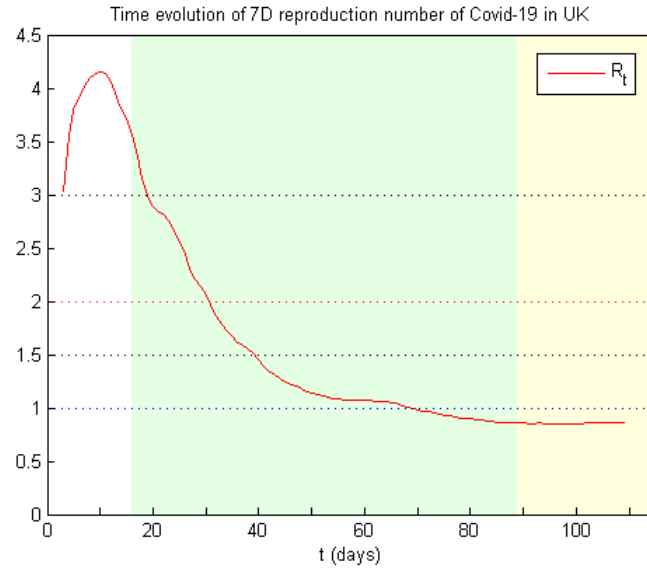
Two weeks later, however, the excessive relaxation of contention measures took its toll and let reproductive ratios start increasing again on July 7. From August to the end of 2020 and further on into the new year, covid control was never really regained $(R_t > 1)$, with a second wave becoming visible in mid-September already — in the last three weeks the infectious population triplicated and continued to rise ever faster. The reality of the new wave was officially acknowledged on 07/18 and intervention rules were reintroduced, including a 4-week partial lockdown beginning on 11/05 $(t = 245)$. On its release $(t = 273)$, the weekly ratio R_t had reached below 1 $(R_t = 0.92)$, but immediately resumed increasing and left the control zone $(R_t < 1)$ ten days later. On 12/31/2020 $(t = 301)$, an estimated 2.8% of the population was infectious, with $R_t = 1.25$ and still rising.

20
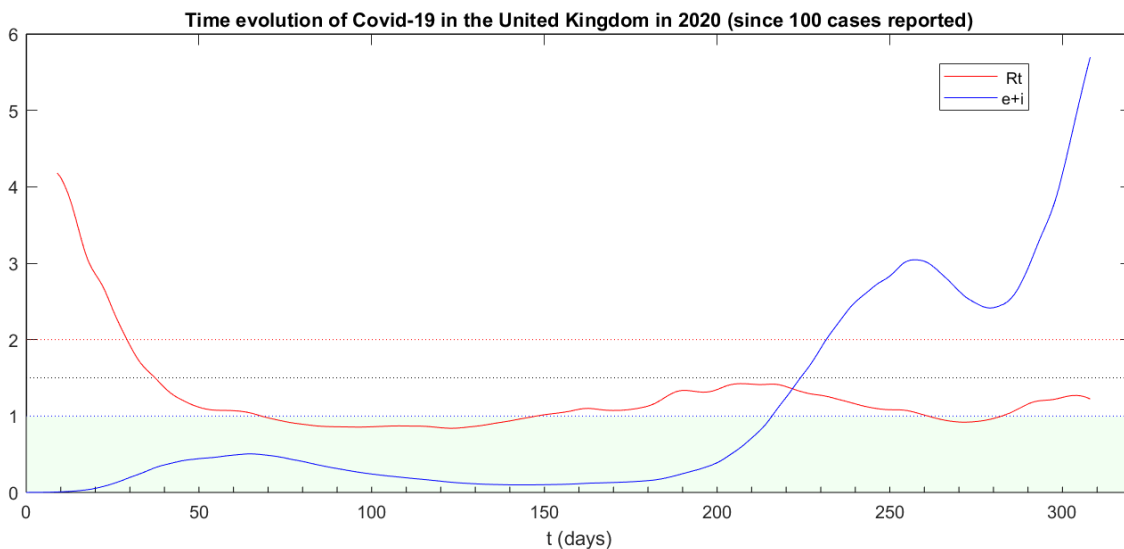
Example 8: Time evolution of Covid-19 in the US since 03/02/2020, the date of 100 total cases reported ($t = 0$). After containment measures began on 03/15/2020 ($t = 13$), R_t successfully decreased continually to 0.97 on 05/15/2020 ($t = 74$), when restrictions began to be relaxed, and then slightly down to 0.96 on 05/27/2020 ($t = 86$), followed by a slow and steady increase to the present value of 1.08 (yellow band). With a poor coordination between central and local authorities in the

beginning of the epidemic, the country suffered a high mortality rate (0.0376 %) and number of infections (almost 2.5 million cases reported). As of 06/24/2020, the United States have not succeeded in bringing the epidemic under nationwide control. A second peak ("second wave") in the size of the active infected population is now clear to happen sometime in the future, as indicated by the curve of $I_0(t)$.

This second peak in the infectious population happened six weeks later on 08/07/2020 ($t = 158$), with an estimated 1.2% of the American people then transmitting the disease. R_t values soared to 1.22 in the beginning of July and stayed above 1 for half of the Summer. Intervention measures regained some control of the epidemic for a while, managing to bring the wave down and reducing R_t to values slightly below 1. After 09/28/2020 ($t = 210$), the epidemic was back to expanding ($R_t > 1$). This continued through the rest of the year and into 2021, with 3.7% of the population estimated to be infectious on 12/31/2020.
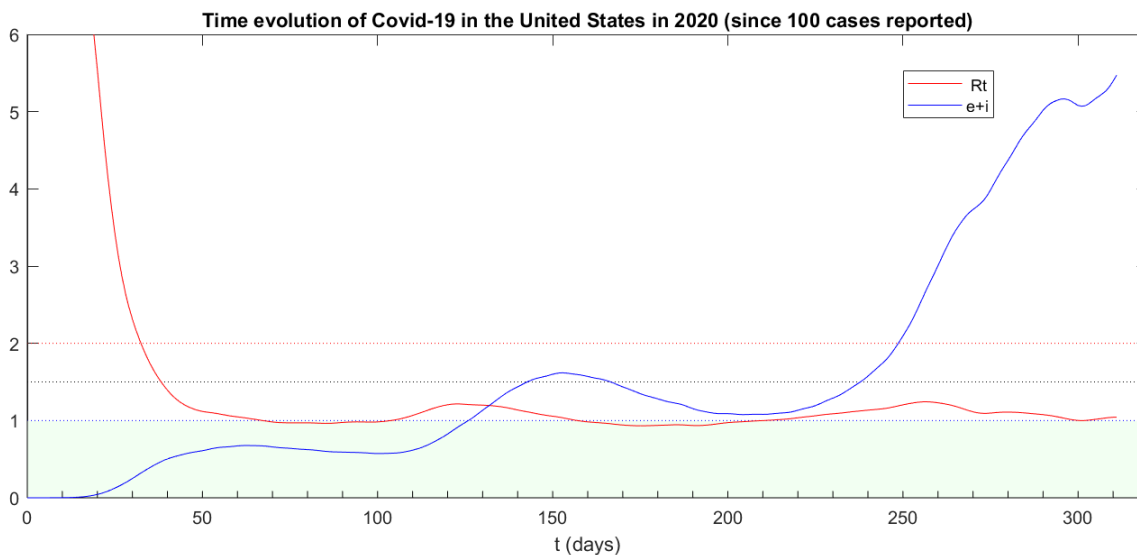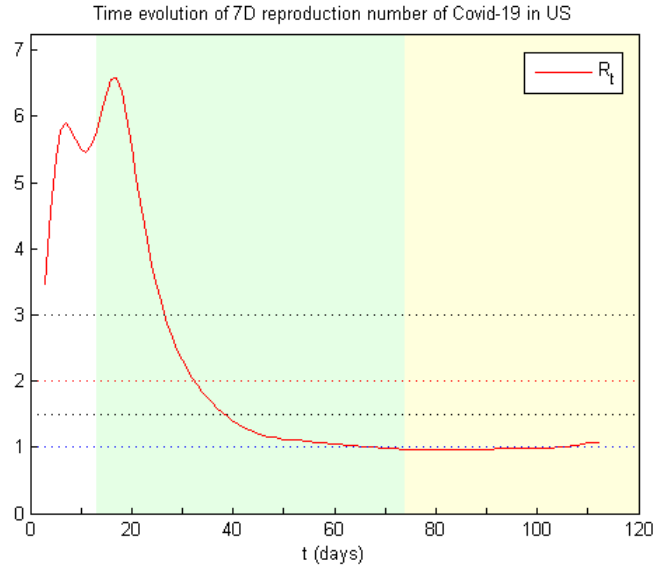
5. Herd Immunity for Covid-19

In this section we use our SEIR model and code to estimate immunization levels in a population that would likely prevent a Covid-19 outbreak to happen. Since the answer is strongly dependent on the transmission rates that would be observed, two distinct scenarios are considered: the case of an unsuspecting population caught off guard against the disease, and a wary population that remains somewhat on guard. Whether or not an outbreak will happen can be answered by examining the values of reproduction numbers associated with the infected population. A convenient choice is given by (3.2b), that is, the infected reproductive ratio

$$\mathcal{R}_t = \frac{\beta \cdot S/N}{r + \gamma} \tag{5.1}$$

(obtained by adding the equations (1.1b) and (1.1c) above), where γ is given in (1.2) and β, r are computed as explained in Section 2. Assuming that a fraction p of the population has been immunized, so that $S/N = 1 - p$, the condition $\mathcal{R}_t < 1$ becomes

$$p > \hat{p} := 1 - \frac{1}{\mathbb{R}_t}, \qquad \mathbb{R}_t = \frac{\beta}{r + \gamma} \tag{5.2}$$

(see e.g. [15], p. 217) and it only remains to decide on the appropriate values of $\mathbb{R}_t$.

(i) *the case of an unsuspecting population caught off guard*: $85\% < \hat{p} < 90\%$

That is, the population is not aware of the presence of any infected individuals, and contention or hygienic measures are not being observed. This was the case, for example, of the countries above in the first weeks of their 2020 covid-19 outbreak. We can then obtain the $\mathbb{R}_t$ values needed to compute $\hat{p}$ by looking at what happened there in the days before the application of intervention measures (Figure 9 below). This gives the numerical ranges of $\mathbb{R}_t$ and $\hat{p} = 1 - 1/\mathbb{R}_t$ shown in the next table, which indicate that $85\% < \hat{p} < 90\%$.

| country	$\mathbb{R}_t$ **range**	$\hat{p}$ **range**
France	$6.67 < \mathbb{R}_t < 9.63$	$85.0\% < \hat{p} < 89.7\%$
Italy	$6.46 < \mathbb{R}_t < 8.38$	$84.5\% < \hat{p} < 88.1\%$
Spain	$6.07 < \mathbb{R}_t < 9.02$	$83.5\% < \hat{p} < 89.0\%$
United Kingdom	$6.64 < \mathbb{R}_t < 9.00$	$84.9\% < \hat{p} < 88.9\%$
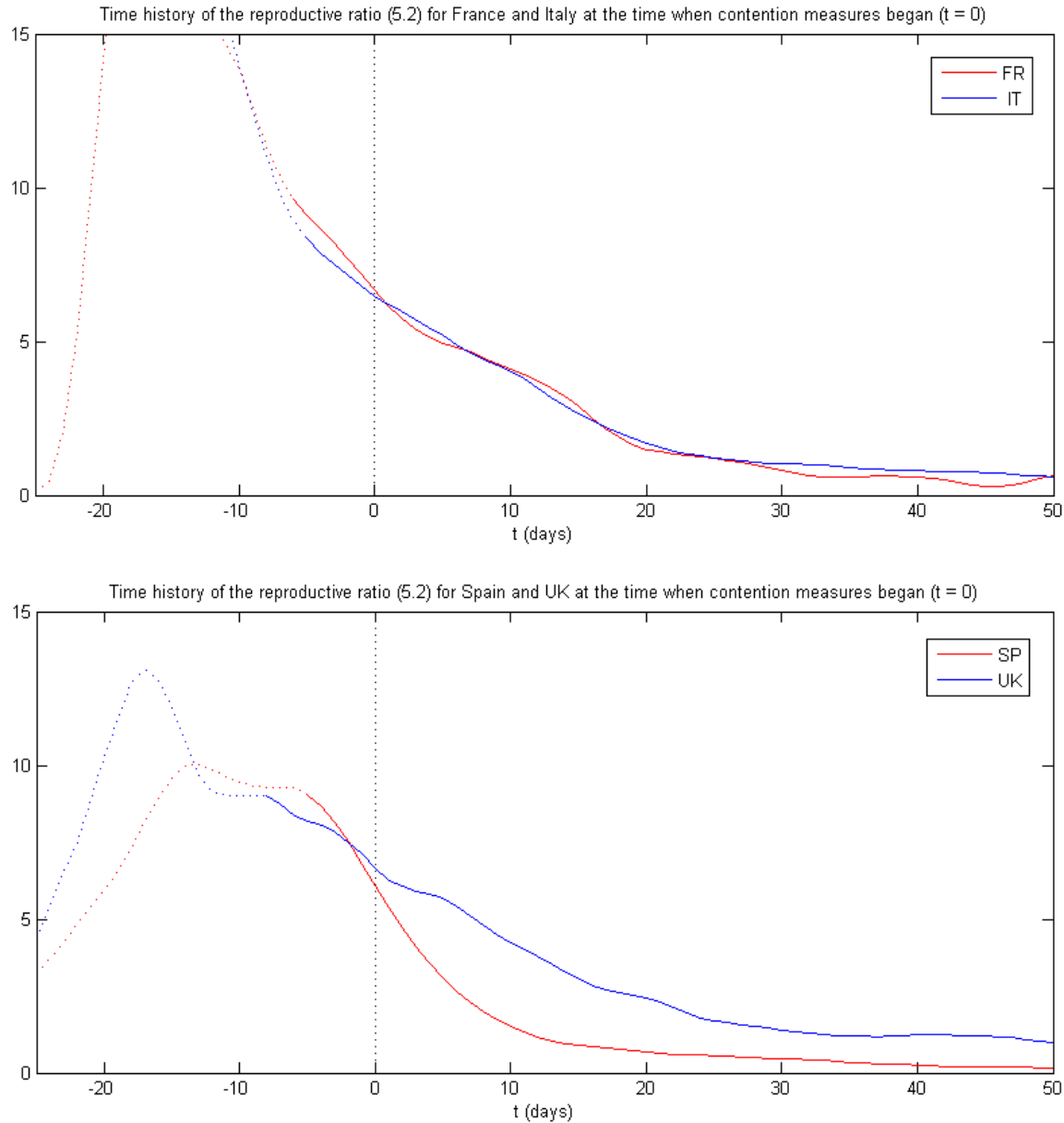
Fig. 9: Time history of the reproductive ratios $\mathbb{R}_t$ defined by (5.2) in France and Italy (top) and in Spain and UK (bottom) in the first weeks about the day when intervention measures were applied ($t = 0$). The dashed parts came from the initialization phase of the algorithm before self-correction set in (Section 2) and were disregarded.

(*ii*) *the case of a wary population observing basic measures*: $40\% < \hat{p} < 60\%$

By basic measures we mean simple hygienic procedures like washing hands and wearing masks, and perhaps some light social distancing if possible, observed by the majority of the population. Estimates for $\mathbb{R}_t$ in this situation can be obtained by looking at what happened to this indicator in various countries after the initial

23

contention measures have been relaxed or even removed and searching for the maximum values in this period (Figure 10). In the cases below, for example, this gives $\mathbb{R}_t < 2.19$ ($\hat{p} < 54.4\%$) for Argentina, $\mathbb{R}_t < 1.74$ ($\hat{p} < 42.6\%$) for Brazil, $\mathbb{R}_t < 2.39$ ($\hat{p} < 58.2\%$) for Mexico and France, $\mathbb{R}_t < 2.51$ ($\hat{p} < 60.2\%$) for UK, and $\mathbb{R}_t < 2.20$ ($\hat{p} < 54.6\%$) for US, suggesting $40\% < \hat{p} < 60\%$ in general. This interval is consistent with the values obtained for other regions, but, as $\hat{p}$ reflects the particular attitude — which may vary significantly — of a population with regard to uncertain threat, larger values are possible: for example, we found $\mathbb{R}_t < 2.98$ ($\hat{p} < 66.5\%$) for Spain and $\mathbb{R}_t < 3.42$ ($\hat{p} < 70.8\%$) for Italy, so that a case-by-case analysis is recommended.
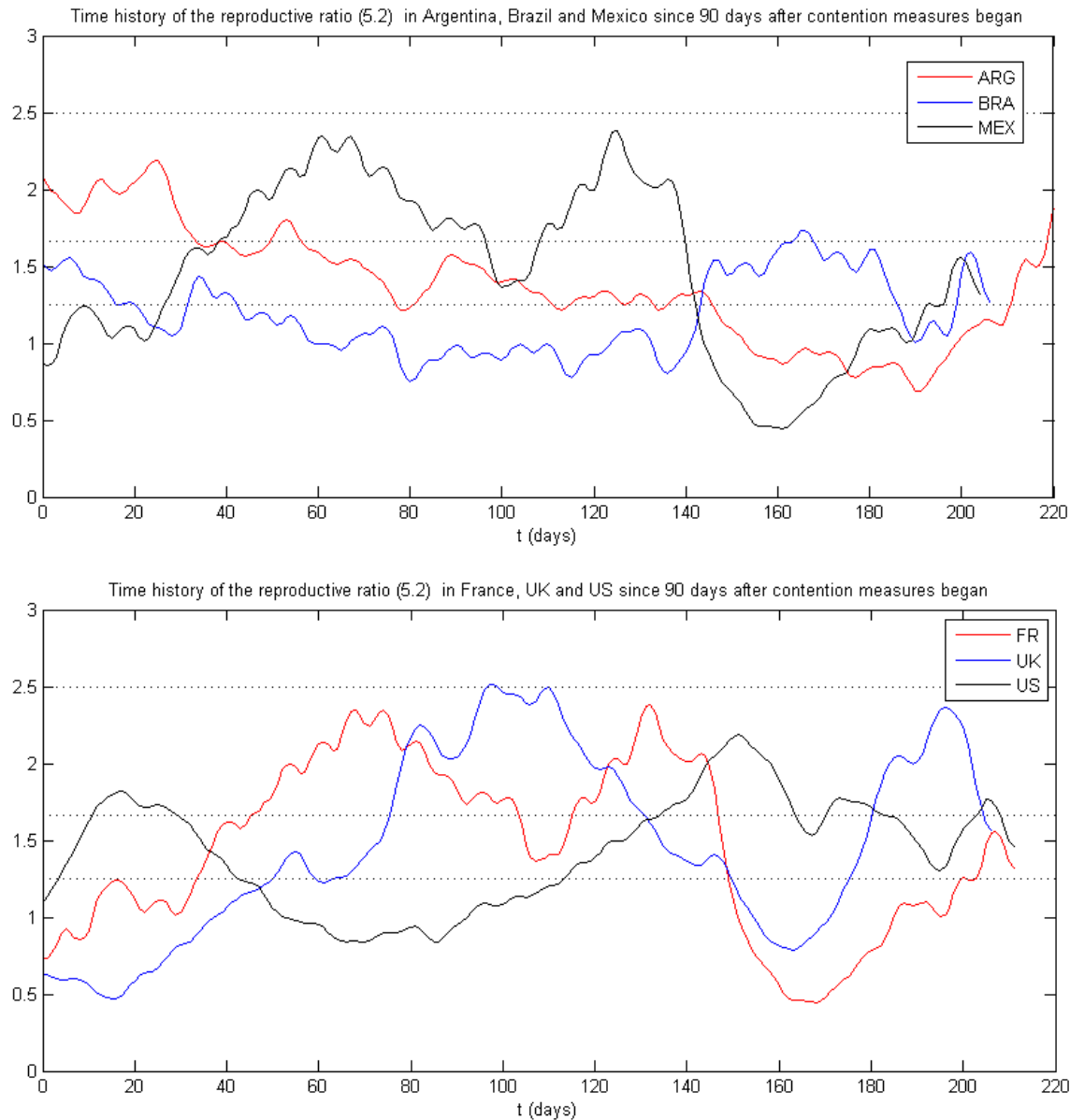
Fig. 10: Time evolution of $\mathbb{R}_t$ in 2020 for six countries since 90 days after the first intervention. The values 1.25, 1.67 and 2.5 (dashed lines) correspond to $\hat{p} = 20\%$, 40% and 60%.

References

[1] F. Brauer, P. van den Driessche and J. Wu (Eds), *Mathematical Epidemiology*, Lecture Notes in Mathematics, vol. 1945, Springer, New York, 2008.

[2] A. Cori, N. M. Ferguson, C. Fraser and S. Cauchemez, *A new framework and software to estimate time-varying reproduction numbers during epidemics*, Amer. J. Epidemiology 178 (2013), 1505-1512.

[3] O. Diekmann, J. P. Heesterbeek and J. J. Metz, *On the definition and the computations of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol. 28 (1990), 365-382.

[4] O. Diekmann and J. A. Heesterbeek, *Mathematical epidemiology of infectious diseases*: model building, analysis and interpretation, Wiley, New York, 2000.

[5] K. Dietz, *The estimation of the basic reproduction number for infectious diseases*, Stat. Meth. Med. Res. 2 (1993), 23-41.

[6] P. van den Driessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), 29-48.

[7] P. van den Driessche and J. Watmough, *Further notes on the basic reproduction number*, in: F. Brauer et al (Eds), *Mathematical Epidemiology*, Lecture Notes in Mathematics, vol. 1945, Springer, New York, 2008, pp. 159-178.

[8] K. M. Gostic, L.McGough, E. Baskerville, S. Abott et al., *Practical considerations for measuring the effective reproductive number R_t*, MedRxiv (2020), 121, DOI: 10.1101/2020.06.18.20134858.

[9] J. A. P. Heesterbeek, *A brief history of R_0 and a recipe for its calculation*, Acta Biotheret. 50 (2002), 189-204.

[10] J. M. Heffernan, R. J. Smith and L. M. Wahl, *Perspectives on the basic reproductive ratio*, J. R. Soc. Interface, 2 (2005), 281-293.

[11] H. W. Hethcote, *The mathematics of infectious diseases*, SIAM Rev. 42 (2000), 599-653.

[12] S. Kim, Y. B. Seo and E. Jung, *Prediction of Covid-19 transmission dynamics using a mathematical model considering behavior changes in Korea*, Epidemiology and Health, 42 (2020), DOI: 10.4178/epih.e2020026.

[13] 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*, Lancet Infectious Diseases 2020, 20:553-558, DOI: 10.1016/S1473-3099(20)30144-4.

[14] S. A. Lauer, K. H. Grantz, Q. Bi, F. L. Jones, Q. Zheng et al, *The incubation period of coronavirus disease 2019* (*Covid-19*) *from publicly reported confirmed cases*: *estimation and application*, Ann. Intern. Med. 172 (2020), 577-582.

[15] M. Martcheva, *An Introduction to Mathematical Epidemiology*, Springer, New York, 2015.

[16] A. Mellan, H. H. Hoeltgebaum, S. Mishra, C. Whittaker et al., *Estimating COVID-19 cases and reproduction number in Brazil*, Report #21, Imperial College London, May/2020, DOI: 10.25561/78872.

[17] T. Obadia, R. Haneef and P. Y. Boelle, *The R0 package*: *a toolbox to estimate reproduction numbers for epidemic outbreaks*, BMC Med. Inform. Decis. Mak. **12**, 147 (2012), DOI: 10.1186/1472-6947-12-147.

[18] B. Ridenhour, J. M. Kowalik and D. K. Shay, *Unravelling R_0*: *Considerations for Publich Health Applications*, Amer. J. Public Health 108 (2018), S445-S454.

[19] H. Salje, C. Tran Kiem, N. Lefrancq, N. Courtejoie, P. Bosetti et al, *Estimating the burden of SARS-CoV-2 in France*, Science 369 (July 2020), 208-211, DOI: 10.1126/science.abc3517.

[20] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen et al., *Improved inference of time-varying reproduction numbers during infectious disease outbreaks*, Epidemics 29 (2019), 100356.

[21] S. Vaid, C. Cakan and M. Bhandari, *Using machine learning to estimate unobserved COVID-19 infections in North America*, J. Bone Joint Surg. Am. 2020, 00:1-5 (DOI: 10.2106/JBJS.20.00715).

[22] J. Wallinga and P. Teunis, *Different epidemic curves for Severe Acute Respiratory Syndrome reveal similar impacts of control measures*, Am. J. Epidemiol. **160** (2004), 509-516.