HAL open science

On an interval prediction of COVID-19 development based on a SEIR epidemic model

Denis Efimov, Rosane Ushirobira

To cite this version:

Denis Efimov, Rosane Ushirobira. On an interval prediction of COVID-19 development based on a SEIR epidemic model. [Research Report] Inria. 2020. hal-02517866v4

HAL Id: hal-02517866

https://inria.hal.science/hal-02517866v4

Submitted on 6 Apr 2020 (v4), last revised 3 Jun 2020 (v6)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On an interval prediction of COVID-19 development based on a SEIR epidemic model

Denis Efimov[1] and Rosane Ushirobira[1]

***Abstract*—In this brief report, a modified version of the well-known mathematical epidemic model SEIR is used to analyze the epidemics course of COVID-19 in six different countries. To this goal, the parameters of the SEIR model are identified by using publicly available data for the corresponding countries: France, Italy, Spain, Germany, Brazil and Russia. The identified model is then applied for the prediction of the SARS-CoV-2 virus propagation under different conditions of confinement. For this purpose, an interval predictor is designed allowing variations and uncertainties in the model parameters to be taken into account. The code and data are available in Github.**

I. Introduction

The SEIR model is one of the most elementary compartmental models of epidemics [1]. It is very popular and widely used in different contexts [2]. This model describes the evolution of the relative proportions of four classes of individuals in a population of constant size, see a general scheme given in Fig. 1. Namely, the susceptible individuals S, capable of contracting the disease and becoming infectious; the asymptomatic E and symptomatic I infectious, capable of transmitting the disease to susceptible; and the recovered R, permanently immune (after healing or dying). Such a simple model represents well a generic behavior of epidemics (plainly as a series of transitions between these populations), and a related advantage consists in a small number of parameters to be identified (three transition rates in Fig. 1: σ, γ and b). This is an important outcome in the case of a virus attack with a limited amount of data available. That is the case of the current worldwide situation[1] under the presence of the SARS-CoV-2 virus.

There are many types and variations of SEIR models [1] (*e.g.*, in the simplest case the classes E and I are modeled at once, leading to a SIR model). This note is based on the modified SEIR discrete-time model proposed in [3], where it has been used to model the epidemics trend of COVID-19 in China (other similar SIR/SEIR-type models used recently for modeling SARS-CoV-2 virus are [4], [5], [6], [7]). The selected model is proposed as follows (we will not consider the influence of the natural birth and mortality, since for the short period of analysis considered here the population may be assumed quasi-constant):

$$
\begin{aligned}
S_{t+1} &= S_t - b\frac{(p_C I_t + r_{t-\tau}E_t)}{N}S_t, && \text{(1a)}\\
E_{t+1} &= (1-\sigma)E_t + b\frac{(p_C I_t + r_{t-\tau}E_t)}{N}S_t, && \text{(1b)}\\
I_{t+1} &= (1-\gamma)I_t + \sigma E_t, && \text{(1c)}\\
R_{t+1} &= R_t + \gamma I_t, && \text{(1d)}
\end{aligned}
$$

where $t \in \mathbb{N}$ (the set of nonnegative integers) is the time counted in days ($t = 0$ corresponds to the beginning of measurements or prediction), $N = S + E + I + R$ denotes the total population, the parameter $0 < \gamma < +\infty$ represents the mortality and recovery rates, the parameter $0 < b < +\infty$ corresponds to the infection rate of the virus transmission from infectious/exposed to susceptible individuals during a contact, $0 < \sigma < +\infty$ is the incubation rate at which the exposed develop symptoms, $0 < p_C < +\infty$ corresponds to the number of contacts for the infectious I (it is supposed that infected people with symptoms are in quarantine, then the number of contacts is decreased), $p_C \leq r_t < +\infty$ is the number of contacts per person per day for the exposed population E (in the presence of confinement and depending on its severity, this number is time-varying), and $\tau > 0$ is the delay in the reaction of the compartments under quarantine conditions (we assume that if $t < \tau$, then $r_{t-\tau} = r_0$). Compared to the model in [3], the inflow/outflow variables from/to other regions for each state are not considered in our analysis, and we introduce the delay τ in the confinement input r_t.

Therefore, the SEIR model (1) has only four parameters to be identified: σ, τ, γ and b. The parameters σ, γ and b represent, respectively, the rate of change between the states E to I, I to R and S to $E+I$ (as in Fig. 1). The parameter σ has a physical meaning: $\sigma = \frac{1}{T_S}$, where T_s is the average duration of incubation period of the virus after contamination, which can be well identified in patients. Similarly for the delay τ, which is also of order T_s, having a natural origin. The numbers of contacts p_C and r_t can be evaluated heuristically based on the population density and social practices. The identification of these parameters must be performed using statistics published by authorities[2]. As a worthy remark, many researches devoted to the estimation and identification for SIR/SEIR models were developed by now, and several in the last few years, like [8], [9], [10], [11], [12], to mention a few.

[1]Denis Efimov and Rosane Ushirobira are with Inria, Univ. Lille, CNRS UMR 9189 - CRIStAL, F-59000 Lille, France, Denis.Efimov, Rosane.Ushirobira@inria.fr

[1]This note was initially written on March 23 and updated April 6, 2020.

[2]As in Imperial College Report, for example.

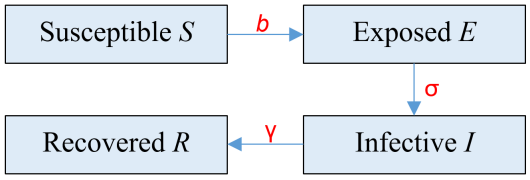


Fig. 1. A schematic representation of SEIR model

Since the measured data and parameters contain numerous uncertainties and perturbations, it is then difficult to make a reasonable prediction based on the simulation of such a model with fixed values of parameters (also taking into account the model simplicity and generality). However, the interval predictor and observer framework [13], [14], [15], [16], [17] allows a set of trajectories to be obtained corresponding to the interval values of parameters and inputs, increasing the model validity without augmenting its complexity. This approach has already been applied to different SEIR models (see, *e.g.*, [18], [19], [20]). In this report, we apply the interval predictor method for the considered SEIR model (1) to improve its forecasting quality.

Remark 1. It is important to highlight that the interval predictor framework used here is not the only method oriented on improving the reliability of prediction using SEIR models. Usually, as in [3], [4], [5], [6], [7], [21], stochastic and multi-agent simulation approaches are used. In those cases, by assuming that the parameters and initial conditions are distributed with some *given* probability, multiple numeric experiments are performed to reconstruct the behavior of possible trajectories of the system. As a first remark, such a methodology needs more computational effort for its realization. Secondly, additional information on the form of probability distribution for all parameters and variables is necessary, demanding either extra hypotheses or more measured data for estimation. As it is currently demonstrated by the SARS-CoV-2 virus attack, it is hard to obtain such data quickly during the epidemic development. Contrarily to these approaches, the interval predictor method does not use these additional assumptions on probability distributions, and it has been also proposed to estimate a *guaranteed* interval inclusion of trajectories with minimal computational effort, by the cost of a more complicated mathematical analysis and design [17].

The outline of this note is as follows. The model validation and identification are first presented considering France as the study case, next the same results are reported for Italy, Spain, Germany, Brazil and Russia. To this end, we describe the measured data applied for parameter identification in Section II, together with some hypotheses used in the sequel. The parameters obtained in [3] for China are tested for France and validated on this data in Section III. The method for parameter identification is presented in Section IV, with validation and some experiments on the influence of confinement parameters on COVID-19 development. An interval predictor is designed in Section V, which allows us to evaluate the situation in France under the variation of parameters and initial states. The accuracy of interval prediction is also evaluated by using a part of data for identification and another part for verification. The results of application of the proposed identification routine and the interval predictor for Italy, Spain, Germany, Brazil and Russia are given in Section VI. Final discussions and remarks are provided in Section VII.

II. Available data on COVID-19 in France

The current population in France is $N = 67064000$[3]. The incubation period T_s, that is widely reported in the literature for COVID-19 studies, is considered to be between 2 and 14 days [3], or in a more specialized research between 2 and 12 days [22], so we assume

$$\frac{1}{2} \leq \sigma \leq \frac{1}{12},$$

it also implies that the delay can be selected in the corresponding limits:

$$2 \leq \tau \leq 12.$$

The data available from public sources[4], and which is used in this note, for the time period between March 12th and 6th April is provided in Github[5], where $\mathscr{I}$, $\mathscr{D}$ and $\mathscr{H}$ represent the number of total detected infected, deceased and recovered individuals, respectively.

Obviously, not all cases can be detected and documented by public health services, so there is a ratio between populations I and $\mathscr{I}$, R and $\mathscr{D}+\mathscr{H}$ as well, which is denoted in this work by α, whose interval of admissible values is estimated from different sources as follows[6]:

$$5 \leq \alpha \leq 20.$$

Formally, such a ratio α has to be time-varying and different for I and R. Due to strict requirements of health services in considered countries, in this report we take the following hypotheses:

$$I_t = \alpha_1(\mathscr{I}_t - \mathscr{D}_t - \mathscr{H}_t),\ R_t = \mathscr{D}_t + \alpha_2 \mathscr{H}_t, \tag{2}$$

i.e., the deaths are reported exactly (see also [6]), but the number of active infected cases and the related recovered individuals can be masked due to the complexity of examination and the actual confirmation of the virus presence. An additional reason is that in many cases the virus symptoms result in a mild reaction of patients (approximately in 80% of cases, see the sources above), hence with no official virus confirmation in such a situation. In this work, we assume modest values for these parameters:

$$\alpha_1 = 5,\ \alpha_2 = \alpha_1,$$

then, roughly speaking, such a choice of α_1 corresponds to registration of all non-mild cases, and α_2 is selected just to be the same. The techniques to identify α_1 from the measurements of $\mathscr{I}$ and $\mathscr{D}$ are described in the last footnote.

[3] www.en.wikipedia.org/wiki/Demographics_of_France.

[4] See www.santepubliquefrance.fr/, www.geodes.santepubliquefrance.fr and www.data.gouv.fr.

[5] Sources: Eficiens, Johns Hopkins University.

[6] See, for example, Coronalinks-3-19-20, or a dedicated analysis in [23], CMMID or University of Melbourne.

So, by fixing α_1 and α_2, the two variables of the model (1), I and R, are available from the beginning of the epidemics via (2).

We also select another average value for the incubation rate:

$$\sigma = \frac{1}{7}$$

to simplify further identification (the variation in this value will be taken into account later in the interval predictor). The nominal value of the delay is chosen as

$$\tau = 10,$$

the algorithm of its identification is discussed below.

III. Numerical experiments with parameters of [3]

In [3], using the data available for the Chinese provinces of Zhejiang, Guangdong and Hubei, which have been impacted differently by the virus, the following parameter bounds have been evaluated:

$$0.0721 \leq \gamma \leq 0.238,$$
$$0.05068 \leq b \leq 0.05429,$$
$$p_C = 3 \text{ (number of contacts in quarantine)},$$
$$p_N = 15 \text{ (number of contacts in normal mode)},$$
$$p_R = 10 \text{ (number of contacts in relaxed quarantine)}.$$

Selecting the average values $\gamma = 0.155$ and $b = 0.0525$, we also choose to decrease the number of contacts for France as follows:

$$p_C = 2,\ p_N = 12,\ p_R = 6,$$

which is related with smaller population density in France.

In [4], the theory of a cyclic application of quarantine regimes of different severity is evaluated for COVID-19. By iterating the periods of complete isolation for everybody (*suppression*), which decelerates the virus advancement, with a time of mild regulation (*mitigation*), which allows the economy balance to be maintained on an arguable level, and when only fragile parts of the population are isolated, it is possible to attenuate the material consequences of epidemics while decreasing the load on health services. Following this idea, for simulation we consider here two scenarios of confinement:

1) Six weeks of strict quarantine and two weeks of a relaxed one, which is further periodically repeated.
2) Twelve weeks of strict quarantine and four weeks of a relaxed one, which is further periodically repeated.

For the chosen model, these scenarios will impact only the variable r_t. The example of behavior in the first scenario of the number of contacts for asymptomatic infectious is shown in Fig. 2.

Remark 2. In other words, r_t can be considered as a kind of control for the virus propagation, by imposing different periods and strictness levels for the confinement.

The results of simulation of the model (1) with the parameters described above for both scenarios, 1 and 2, are given in Fig. 3 (for a better visibility, all populations are plotted in the logarithmic scale in this report). As we can conclude from these results, these scenarios of confinement do not lead to a stabilization of COVID-19 development in France (the black dotted line in the top of the plots correspond to N (the total population), then according to these graphics the epidemics is going to stop after a total infection of the country). However, after a short analysis on how close are the obtained curves to the measured data (see Fig. 4 where dashed lines correspond to the measurements), we recognize that the model with the parameters from [3] does not correspond well (it overestimates) to the situation in France. Therefore, an identification of parameters is needed.

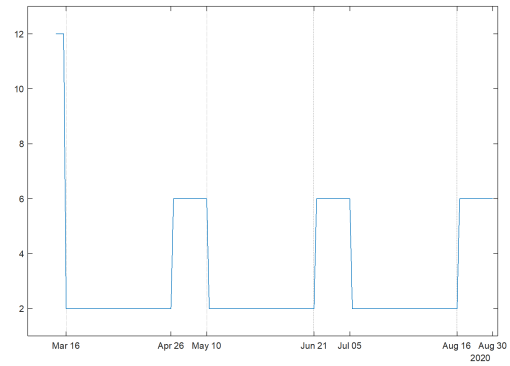


Fig. 2. Variation of the number of contacts r_t in the first scenario

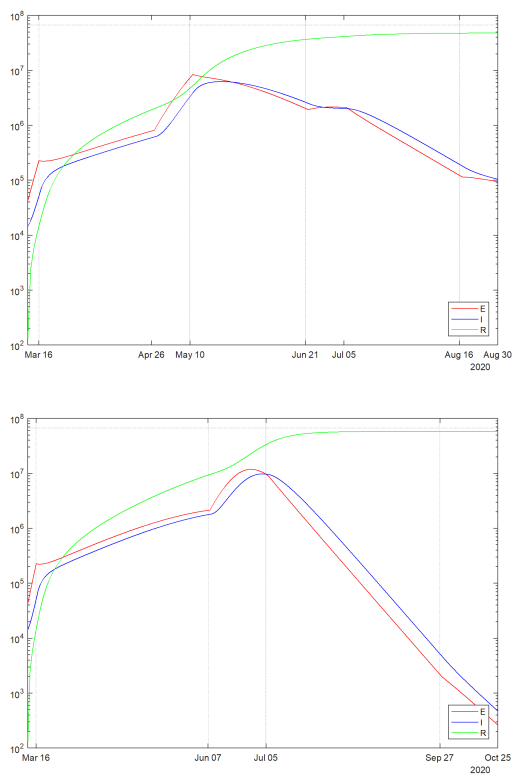


Fig. 3. The results of simulation with parameters of [3] for scenarios 1 and 2

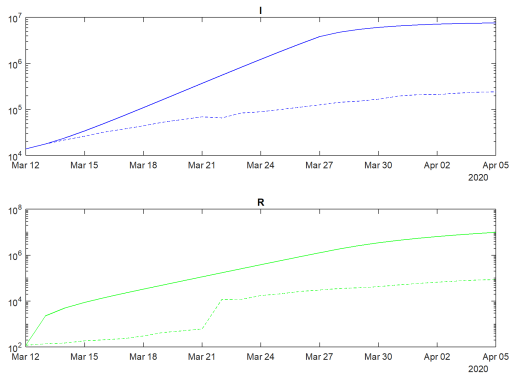


Fig. 4. The results of verification with parameters of [3]

IV. Parameter identification

For the parameter identification, we assume that the incubation rate σ is fixed as above, and that the symptomatic infectious I_t and the recovered people R_t are measured for the first J days of the virus attack as in (2) for $t = 0, 1, \ldots, J$.

In this section, we begin by discussing an approach to the identification of the delay τ. Then, the method for the identification of the recovery rate γ and the infection rate b is presented. Finally, the model (1d) with the obtained values for the parameters is validated in simulations.

A. Delay dentification

From the dynamics of (1d), the increment of E_t (*i.e.*, $E_{t+1} - E_t$) is directly proportional to $r_{t-\tau}$. The number of contacts r_t instantaneously changes its value after the imposition of the quarantine (it jumps from p_N to p_C), and the same occurs after the suppression of the confinement (from p_C to p_N or p_R), see Fig. 2. It implies that the increment of E_t shows discontinuities in these time instants. The variable E_t is not available for measurements, but the same (filtered) behavior will also be observed in the increment of the variable I_t. Since both variables, I_t and E_t, in (1d) have an exponential rate of changes, then the signal

$$dI_t = \ln I_t - \ln I_{t-1}$$

for $t = 2, \ldots, J$ should have a step-like form (the logarithm of the increment of an exponentially growing or decaying signal is a constant) with the change of value in the time instant $t_c \geq 2$, when a modification of the confinement rules starts to influence the variable I_t. Therefore, the delay can be estimated as (with a mild ambiguity in this work we use the same symbol to denote a parameter and its estimate)

$$\tau = t_c - t',$$

where $t' \geq 0$ is the instant of application of the new confinement rule. Hence, to estimate the value t_c the following algorithm is proposed:

$$t_c = \underset{t=3,\ldots,J}{\arg\min} \sqrt{\sum_{\ell=2}^{J} |dI_\ell - dI_\ell^t|^2},$$

where

$$dI_\ell^t = \begin{cases} \frac{1}{t-2}\sum_{s=2}^{t-1} dI_s & ,\text{if } \ell < t \\ \frac{1}{J-t+1}\sum_{s=t}^{J} dI_s & ,\text{if } \ell \geq t \end{cases}$$

is a step-like varying signal, which jumps at the instant t. The main drawback of this approach is the noise in the measurements (as for any approach that indirectly use a derivative estimation). Applying it for all countries, taking into account its sensitivity with respect to the used data, we finally justified the choice of the nominal value

$$\tau = 10$$

for further use.

Remark 3. Note that if the value of γ is known (see below how we can estimate it), then the variable $E_t = \frac{1}{\sigma}(I_{t+1} - (1-\gamma)I_t)$ can be reconstructed from the measurements, and the same approach can be applied to the increment $dE_t = \ln E_t - \ln E_{t-1}$, which depends on $r_{t-\tau}$. Unfortunately, we have a very noise data for COVID-19, so the calculated variables E_t contain a lot of noise, and the above approach does not provide a reliable estimation using dE_t.

B. Rates identification

From the equation (1d) we can identify the value of the parameter γ:

$$\gamma = \frac{R_{t+1} - R_t}{I_t},$$

whose least squares estimation is

$$\gamma_k = \frac{\sum_{t=0}^{J-k-1} I_t (R_{t+1} - R_t)}{\sum_{t=0}^{J-k-1} I_t^2}$$

for $k = 0, 1, \ldots, K$, where $0 < K < J - 1$ is the number of the last days used for identification (in this work we selected $K = J - 10$). Another possible approach is the moving window estimation:

$$\gamma_k = \frac{\sum_{t=k}^{k+K_w} I_t (R_{t+1} - R_t)}{\sum_{t=k}^{k+K_w} I_t^2}$$

for $k = 0, 1, \ldots, K$ with $K = J - K_w$, where $K_w > 1$ is the window length. Then the average value is used for further analysis and design:

$$\gamma = \frac{1}{K+1}\sum_{k=0}^{K} \gamma_k = 0.0340.$$

Next, the equation (1c) allows us to calculate the related number of asymptomatic infectious (σ is chosen and γ is estimated):

$$E_t = \frac{1}{\sigma}(I_{t+1} - (1-\gamma)I_t),$$

while the number of susceptible individuals can be evaluated using the total population:

$$S_t = N - I_t - R_t - E_t.$$

Finally, from the equation (1b) we can derive the infection rate (for the selected p_C, τ and r_t):

$$b = N\frac{E_{t+1} - (1-\sigma)E_t}{(p_C I_t + r_t E_t) S_t},$$

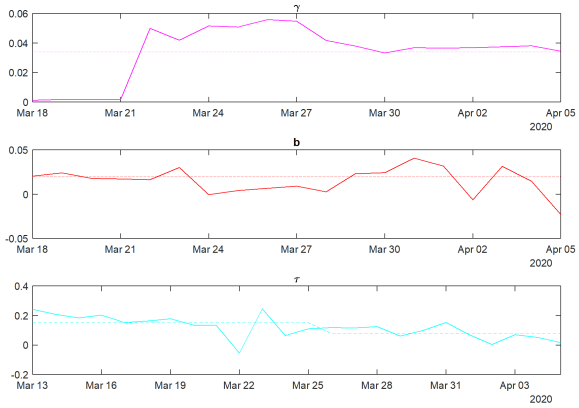


Fig. 5. The identified parameters for France

whose least squares estimation is

$$b_k = N\frac{\sum_{t=0}^{J-k-1}(p_C I_t + r_{t-\tau}E_t)(E_{t+1}-(1-\sigma)E_t)S_t}{\sum_{t=0}^{J-k-1}(p_C I_t + r_{t-\tau}E_t)^2 S_t^2}$$

for $k = 0,1,\ldots,K$, or the moving window estimation version:

$$b_k = N\frac{\sum_{t=k}^{k+K_w-1}(p_C I_t + r_{t-\tau}E_t)(E_{t+1}-(1-\sigma)E_t)S_t}{\sum_{t=k}^{k+K_w-1}(p_C I_t + r_{t-\tau}E_t)^2 S_t^2}$$

for $k = 0,1,\ldots,K$ with $K = J - K_w$, then the identified value is again the average of these estimates:

$$b = \frac{1}{K+1}\sum_{k=0}^{K} b_k = 0.0197.$$

Remark 4. Due to measurement noise, the derived values of E_t and b_k can be negative (that is physically impossible), then a previous positive estimate can be taken into account, *i.e.*, $E_t = E_{t-1}$, or only positive quantities for average calculation can be used: $b = \frac{1}{K+1}\sum_{k=0}^{K}\rho_k b_k$ with $\rho_k = 0.5(\mathrm{sign}(b_k)+1)$ (it is 0 for negative b_k and 1 otherwise).

The obtained values γ_k, b_k (solid lines) together with the selected estimates γ, b (dot lines) and the signals dI_t (solid line) with dI_t^τ (dash line) are shown in Fig. (5). The identified values for γ and b are not included in the confidence intervals obtained for China in [3] (they are reported at the beginning of Section III), which explains the probable bad forecast of the model (1) with the parameters validated for China in that work.

C. Simulation and validation

The simulation results of the model (1) with the identified values of parameters for both scenarios 1 and 2 are given in Fig. 6, a verification on the measured and reconstructed data is shown in Fig. 7. In this case the model can approximate reasonably well the virus propagation, since the identified parameters are consistent with the available statistics for France.

Unfortunately, the obtained curves also demonstrate the lack of efficiency of the confinement (even of a long length). In both scenarios the peak of the epidemics will be reached in

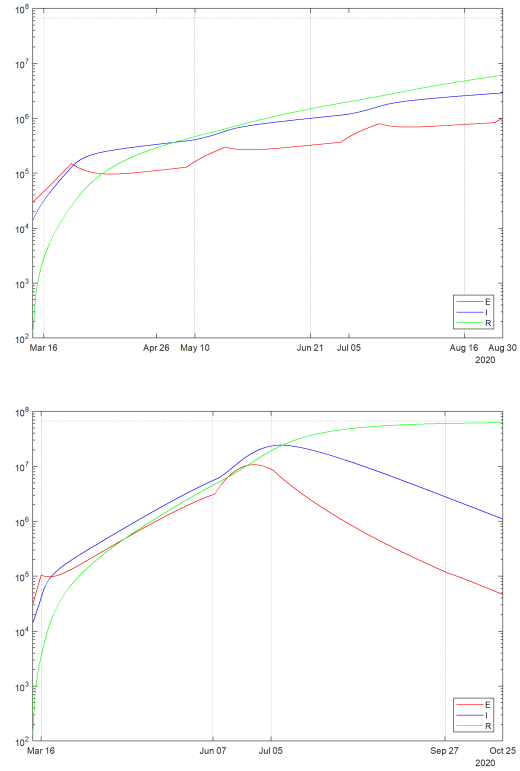


Fig. 6. The results of simulation with identified parameters

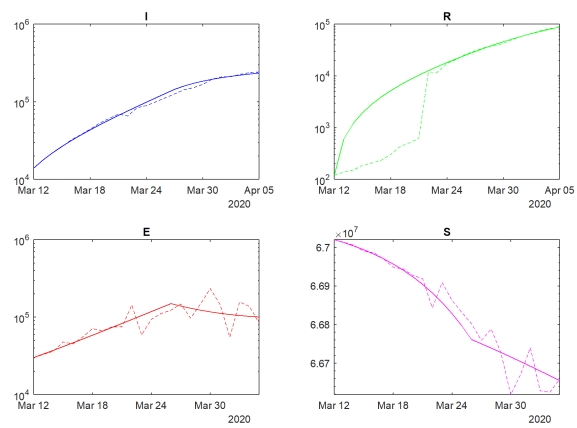


Fig. 7. The results of verification with identified parameters

the middle of July 2020 (the model with parameters from [3] predicts the peaks even earlier).

Let us enlarge the validity of the prediction based on (1) by considering intervals of admissible values for the parameters and initial conditions.

V. Interval prediction

In the previous section, we achieved a prediction for the model (1) with identified values of the parameters b, γ, τ and selected values for $\alpha_1, \alpha_2, \sigma$. The initial values for the state of the model S_0, I_0, E_0 and R_0 were chosen from measured/reconstructed sets. However, due to the generic

structure of the model, uncertainties in the values of the auxiliary parameters, and noises in the measured information, it is obvious that the reliability of the obtained prognosis is limited. A way to overcome such a weakness consists in taking into account intervals of admissible values for all variables and parameters used for simulation, hence enlarging the validity of the model. In such a case, we calculate/evaluate the sets of the resulted trajectories. In this work, we use for this purpose the interval prediction framework [17].

A. Explanation of idea

The idea of the interval prediction can be illustrated on a simple scalar system:

$$x_{t+1} = a_t x_t + d_t,$$

where $x_t \in \mathbb{R}_+$ is a nonnegative system state, whose initial conditions belong to a given interval:

$$x_0 \in [\underline{x}_0, \overline{x}_0];$$

$a_t \in \mathbb{R}_+$ and $d_t \in \mathbb{R}$ are uncertain inputs, which also take values in known intervals:

$$a_t \in [\underline{a}_t, \overline{a}_t],\ d_t \in [\underline{d}_t, \overline{d}_t]$$

for all $t \in \mathbb{N}$. So, we assume that $\underline{x}_0 \le \overline{x}_0$, $0 \le \underline{a}_t \le \overline{a}_t$ and $\underline{d}_t \le \overline{d}_t$ are known for all $t \in \mathbb{N}$. The imposed nonnegativity constraints on x_t and a_t correspond to the case of the model (1). We would like to calculate the lower $\underline{x}_t$ and upper $\overline{x}_t$ predictions of the state x_t of this system under the introduced hypotheses on all uncertain variables, requiring the relations:

$$\underline{x}_t \le x_t \le \overline{x}_t \quad \forall t \in \mathbb{N}.$$

The theory of interval observers and predictors [17], [24] answers this question, and a possible solution (that utilizes the nonnegativity of x_t and a_t) is as follows:

$$\begin{aligned} \underline{x}_{t+1} &= \underline{a}_t \underline{x}_t + \underline{d}_t, \\ \overline{x}_{t+1} &= \overline{a}_t \overline{x}_t + \overline{d}_t, \end{aligned}$$

which is rather straightforward. To substantiate the desired interval inclusion for x_t by $\underline{x}_t, \overline{x}_t$, we can consider the lower $\underline{e}_t = x_t - \underline{x}_t$ and the upper $\overline{e}_t = \overline{x}_t - x_t$ prediction errors, whose dynamics take the form:

$$\begin{aligned} \underline{e}_{t+1} &= (a_t x_t - \underline{a}_t \underline{x}_t) + (d_t - \underline{d}_t), \\ \overline{e}_{t+1} &= (\overline{a}_t \overline{x}_t - a_t x_t) + (\overline{d}_t - d_t), \end{aligned}$$

then it is easy to verify that the terms $d_t - \underline{d}_t, \overline{d}_t - d_t$ are nonnegative by the definition of $\underline{d}_t, \overline{d}_t$, and the terms $a_t x_t - \underline{a}_t \underline{x}_t, \overline{a}_t \overline{x}_t - a_t x_t$ have the same property for $t = 0$ by the definition of $\underline{a}_t, \overline{a}_t$ and $\underline{x}_0, \overline{x}_0$, therefore, $\underline{e}_1 \ge 0$, $\overline{e}_1 \ge 0$ (that implies $x_1 \in [\underline{x}_1, \overline{x}_1]$) and the analysis can be iteratively repeated for all $t \in \mathbb{N}$. Let us apply this method to the model (1) (clearly, each equation there has the form as above).

B. Equations of interval predictor and its properties

To this end, we assume that all parameters belong to the known intervals (for simplicity we will not deviate the value of τ):

$$\begin{gathered} \sigma \in [\underline{\sigma}, \overline{\sigma}],\ \gamma \in [\underline{\gamma}, \overline{\gamma}],\ b \in [\underline{b}, \overline{b}], \\ r_t \in [\underline{r}_t, \overline{r}_t] \quad \forall t \in \mathbb{N}, \end{gathered} \tag{3}$$

together with the initial conditions in (1):

$$S_0 \in [\underline{S}_0, \overline{S}_0],\ I_0 \in [\underline{I}_0, \overline{I}_0],\ E_0 \in [\underline{E}_0, \overline{E}_0],\ R_0 \in [\underline{R}_0, \overline{R}_0], \tag{4}$$

where nonnegative values $\underline{\sigma}, \overline{\sigma}$, $\underline{\gamma}, \overline{\gamma}$, $\underline{b}, \overline{b}$, $\underline{r}_t, \overline{r}_t$, $\underline{S}_0, \overline{S}_0$, $\underline{I}_0, \overline{I}_0$, $\underline{E}_0, \overline{E}_0$ and $\underline{R}_0, \overline{R}_0$ are obtained from the ones used in the previous section by applying $\pm\delta\%$ deviation from those nominal quantities. Then applying the approach explained just above, we derive the equations of the interval predictor:

$$\begin{aligned} \underline{S}_{t+1} &= \left(1 - \overline{b}\frac{\left(p_C \overline{I}_t + \overline{r}_{t-\tau}\overline{E}_t\right)}{N}\right)\underline{S}_t, \\ \underline{E}_{t+1} &= \left(1 - \overline{\sigma} + \underline{b}\frac{\underline{r}_{t-\tau}}{N}\underline{S}_t\right)\underline{E}_t + p_C \underline{b}\frac{\underline{I}_t \underline{S}_t}{N}, \\ \underline{I}_{t+1} &= (1 - \overline{\gamma})\underline{I}_t + \underline{\sigma}\underline{E}_t, \\ \underline{R}_{t+1} &= \underline{R}_t + \underline{\gamma}\underline{I}_t, \\ \overline{S}_{t+1} &= \min\left\{N, \left(1 - \underline{b}\frac{\left(p_C \underline{I}_t + \underline{r}_{t-\tau}\underline{E}_t\right)}{N}\right)\overline{S}_t\right\}, \\ \overline{E}_{t+1} &= \min\left\{N, \left(1 - \underline{\sigma} + \overline{b}\frac{\overline{r}_{t-\tau}}{N}\overline{S}_t\right)\overline{E}_t + p_C \overline{b}\frac{\overline{I}_t \overline{S}_t}{N}\right\}, \\ \overline{I}_{t+1} &= \min\left\{N, (1 - \underline{\gamma})\overline{I}_t + \overline{\sigma}\overline{E}_t\right\}, \\ \overline{R}_{t+1} &= \min\left\{N, \overline{R}_t + \overline{\gamma}\overline{I}_t\right\}, \end{aligned}$$

where $\underline{S}_t, \overline{S}_t$, $\underline{I}_t, \overline{I}_t$, $\underline{E}_t, \overline{E}_t$ and $\underline{R}_t, \overline{R}_t$ are the lower and upper interval predictions for S_t, I_t, E_t and R_t, respectively.

Theorem 5. *For the model* (1) *satisfying the relations* (3) *and* (4) *with*

$$2\overline{b}\sup_{t\in\mathbb{N}} \overline{r}_t \le 1,\ \overline{\sigma} \le 1,\ \overline{\gamma} \le 1, \tag{6}$$

the interval predictor (5) *guarantees the interval inclusions for the state of* (1) *for all* $t \in \mathbb{N}$*:*

$$S_t \in [\underline{S}_t, \overline{S}_t],\ I_t \in [\underline{I}_t, \overline{I}_t],\ E_t \in [\underline{E}_t, \overline{E}_t],\ R_t \in [\underline{R}_t, \overline{R}_t]$$

with boundedness of all predictions for all $t \in \mathbb{N}$*:*

$$\underline{S}_t, \overline{S}_t, \underline{I}_t, \overline{I}_t, \underline{E}_t, \overline{E}_t, \underline{R}_t, \overline{R}_t \in [0, N].$$

Proof: By direct calculation, we can check that

$$\begin{gathered} \underline{b}\frac{(p_C \underline{I}_t + \underline{r}_{t-\tau}\underline{E}_t)}{N} \le b\frac{(p_C I_t + r_{t-\tau}E_t)}{N} \le \overline{b}\frac{\left(p_C \overline{I}_t + \overline{r}_{t-\tau}\overline{E}_t\right)}{N}, \\ \underline{b}\frac{\underline{r}_{t-\tau}}{N}\underline{S}_t - \overline{\sigma} \le b\frac{r_{t-\tau}}{N}S_t - \sigma \le \overline{b}\frac{\overline{r}_{t-\tau}}{N}\overline{S}_t - \underline{\sigma}, \\ p_C \underline{b}\frac{\underline{I}_t \underline{S}_t}{N} \le p_C b\frac{I_t S_t}{N} \le p_C \overline{b}\frac{\overline{I}_t \overline{S}_t}{N}, \\ \underline{\sigma}\underline{E}_t \le \sigma E_t \le \overline{\sigma}\overline{E}_t, \\ \underline{\gamma}\underline{I}_t \le \gamma I_t \le \overline{\gamma}\overline{I}_t \end{gathered}$$

due to (3) and (4) for $t=0$. Since (recall that $r_t \geq p_C$, $\overline{I}_t+\overline{E}_t \leq 2N$, thus $\underline{S}_t \geq 0$)

$$\overline{b}\frac{\left(p_C\overline{I}_t+\overline{r}_{t-\tau}\overline{E}_t\right)}{N} \leq \overline{b}\overline{r}_{t-\tau}\frac{\overline{I}_t+\overline{E}_t}{N} \leq 2\overline{b}\overline{r}_{t-\tau} \leq 2\overline{b}\sup_{t\in\mathbb{N}}\overline{r}_t,$$

$$1-\overline{\sigma}+\underline{b}\frac{\underline{r}_{t-\tau}}{N}\underline{S}_t \geq 1-\overline{\sigma},$$

we obtain that

$$1 \geq \overline{b}\frac{\left(p_C\overline{I}_t+\overline{r}_{t-\tau}\overline{E}_t\right)}{N},\ 1+\underline{b}\frac{\underline{r}_{t-\tau}}{N}\underline{S}_t \geq \overline{\sigma}$$

due to (6), then as we demonstrated above

$$S_1 \in [\underline{S}_1,\overline{S}_1],\ I_1 \in [\underline{I}_1,\overline{I}_1],\ E_1 \in [\underline{E}_1,\overline{E}_1],\ R_1 \in [\underline{R}_1,\overline{R}_1],$$

and such a verification can be repeated for all $t \in \mathbb{N}$. In the same way we can show that if the relations

$$\underline{S}_t \leq \overline{S}_t,\ \underline{I}_t \leq \overline{I}_t,\ \underline{E}_t \leq \overline{E}_t,\ \underline{R}_t \leq \overline{R}_t$$

are satisfied for some $t \in \mathbb{N}$, then they also hold for $t+1$ in (5), and to substantiate boundedness of the state of the interval predictor, it is enough to guarantee that $\overline{S}_t$, $\overline{I}_t$, $\overline{E}_t$ and $\overline{R}_t$ doe not exceed N (as it is done by construction in (5)), while nonnegativity of $\underline{S}_t$, $\underline{I}_t$, $\underline{E}_t$ and $\underline{R}_t$ is ensured by (6). ■

Remark 6. The boundedness of the state of (5) established in Theorem (5) does not imply the stability of the internal dynamics of the interval predictor (it is also a reason to impose the explicit saturation in (5)), which is a frequent and challenging problem for these predictors [16], [24].

Remark 7. The dynamics of lower and upper interval predictions are interrelated through the update equations of $\underline{S}_t,\overline{S}_t$. Thus, the dimension of the predictor (5) is twice higher than in the system (1). The values of the variables $\underline{S}_t,\overline{S}_t$ can be evaluated using the population equation $S_t+E_t+I_t+R_t=N$:

$$\underline{S}_t = N-\overline{I}_t-\overline{E}_t-\overline{R}_t,$$
$$\overline{S}_t = N-\underline{I}_t-\underline{E}_t-\underline{R}_t,$$

which however does not isolate the dynamics of lower and upper interval predictions. In addition, preliminary simulations show that such a modification leads to more conservative results, so we keep (5) for all further utilization.

C. The results of numeric experiments

The simulation results of the interval predictor (5) with $\delta = 10\%$ are presented in Fig. 8 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the results of simulation obtained in the previous section, the circles depict measured and reconstructed data points used for identification and validation). The same results for the case when only σ is deviated with $[\underline{\sigma},\overline{\sigma}] = \left[\frac{1}{12},\frac{1}{2}\right]$ are shown in Fig. 9 (the parameters γ, b and the signal r_t take the nominal values, initial conditions are the same as in the previous experiment), where we can recognize how strong the influence of the parameter σ and how significant its deviation for the COVID-19 trend. The width of the predicted interval of admissible values for the state of (1) is growing, which is related with a high level of uncertainty reflected by δ and chosen for these simulations.

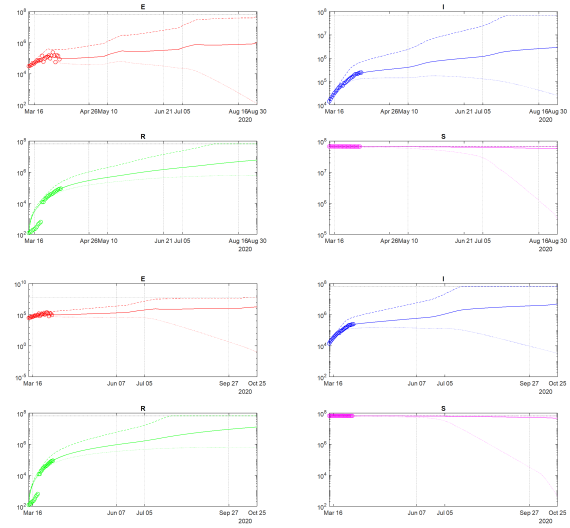


Fig. 8. The results of simulation of (5) for scenarios 1 and 2 in France under $\pm 10\%$ variation of all parameters

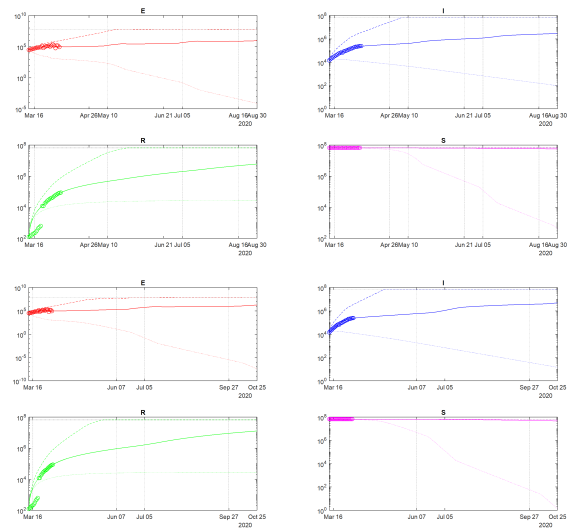


Fig. 9. The results of simulation of (5) for scenarios 1 and 2 under variation of $\sigma \in [\frac{1}{12},\frac{1}{2}]$

As we can conclude from these curves, under sufficiently big deviations of the parameters (which correspond to the small amount of data publicly available now), the confinement may slow down the epidemics. The measurements are nearly included in the obtained intervals validating the prediction (the value of δ was selected to ensure this property). There are two variants of epidemic development demonstrated in these results: optimistic, which corresponds to the lower bounds of I and E, and pessimistic presented by the respective upper bounds. Under the current level of uncertainty, in the optimistic setting, the maximum of infected people can be reached in approximately two months.

In order to check the accuracy of the prediction we can select a part of the data for identification, and another part for verification of prediction reliability. The results of such a validation are shown in Fig. 10, where the interval prediction

for the infectious population I is presented in the first scenario with a deviation of all parameters. As previous, blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the case given in Fig. 8, the bold lines are calculated using previous day initial conditions), the blue circles and squares are the measured information used for identification and validation, and the red line is the average behavior. In the top plot, all data (*i.e.*, for $t = 0, 1, \ldots, J$) is used for identification as described above, then the interval predictor is launched from the last day initial conditions (bold dashed and dotted blue lines and the red one). In the bottom plot only the data points for $t = 0, 1, \ldots, J - 10$ are utilized, they are shown by circles, and the interval predictor is initiated with the data for $t = J - 10$, then square data points (which were not taken into account during identification for $t = J - 9, \ldots, J$) can be compared with the predictor trajectories (bold dashed and dotted blue lines and the red one). As we can see, the points marked by squares are well included in the predicted interval, which confirms the reliability of (5) at least for 10 days, then returning to the top plot we can similarly use it for prediction of future development of COVID-19 for $10 - 14$ days.

In general, a further precision of the model and the parameters is needed, but as a recommendation after these preliminary simulations: an *augmentation* of the severity of the quarantine is desirable. This suggestion is illustrated by Fig. 11, where the interval prediction for the infectious population I is presented in the first scenario with deviation of all parameters. As in Fig. 10, blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using previous day initial conditions), and the magenta circles are the measured information, the red line is the average behavior. The difference with respect to the top graphic of Fig. 10 is that it is computed for $p_C = 1$ (*i.e.* for a more strict quarantine). In the latter case, in the optimistic scenario, the confinement stops the virus progression, while at the peak the quantity of infected people can be decreased (which is an important achievement representing a significant load decay on public health services).

VI. Other countries

In this section we do not repeat all the considerations and analysis previously presented for France. For the other countries, we introduce the used data together with the selected parameters, if they differ from the ones accepted above, identify the parameters (as in Fig. 5) and simulate the interval predictor (as in Fig. 8 together with the plots of figures 10 and 11). The common parameters assigned to all countries (to simplify the analysis) are:

$$\alpha_1 = 5,\ \alpha_2 = \alpha_1,\ \sigma = \frac{1}{7},\ \tau = 10,$$
$$p_C = 2,\ p_N = 12,\ p_R = 6.$$

Of course, adjusting these values for each country will definitely improve the forecast precision, but our goal here is mainly illustrate a wide applicability of the proposed method for virus propagation interval prediction.

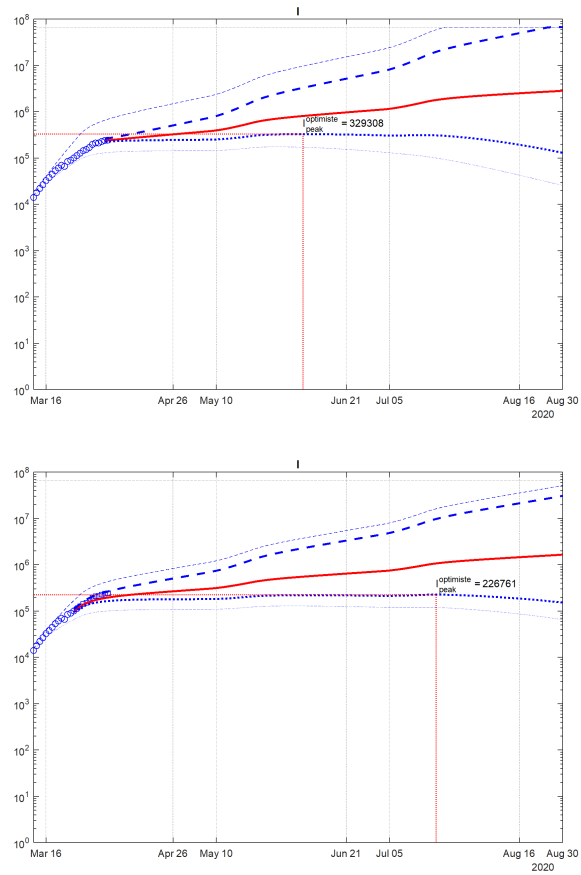


Fig. 10. Validation of prediction of I for scenario 1 in France with J and $J - 10$ points of data under deviations of values of all parameters

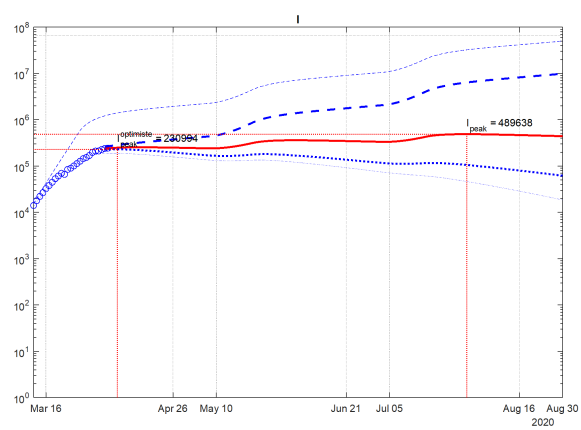


Fig. 11. Prediction of the growth of I for scenario 1 in France with $p_C = 1$ under deviations of values of all parameters

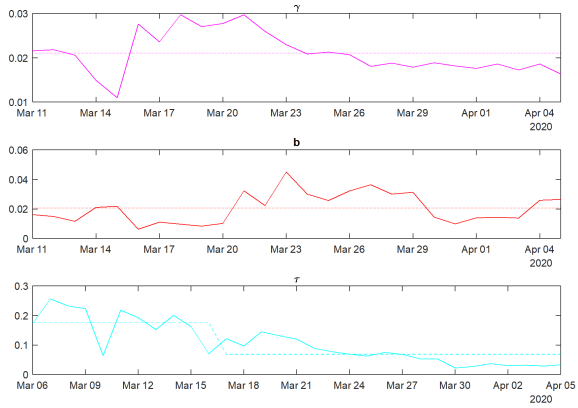


Fig. 12. The identified parameters for Italy

A. Italy

The current population in Italy is $N = 60359546$[7]. The data available from public sources for the time period between March 5th and April 6th is provided in Github[8].

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0211,\ b = 0.0206$$

with the values γ_k, b_k, dI_t and dI_t^τ given in Fig. 12. France and Italy have the virus infection rate b approximately close, but in Italy the recovery rate γ is lower compared to France.

The simulation results of the interval predictor (5) with $\delta = 10\%$ are presented in Fig. 13 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the average behavior, the circles depict measured and reconstructed data points used for identification). As we can conclude from these plots, the obtained model follows well the measured statistics for Italy (probably the value of the delay τ should be increased).

A good fit of the model to the data is demonstrated in Fig. 14 (blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using the last day included in the identification data), the red line is the average, the magenta circles and squares are the measured information used for identification and validation), where the square points belong to the middle of the predicted interval in the bottom plot, while the top plot can be used for prediction of the number of infectives in $10-14$ days. Finally, the influence of the strictness of confinement is illustrated in Fig. 15. A more strict quarantine decreases and makes it sooner the peak of epidemics in the optimistic scenario.

B. Spain

The current population in Spain is $N = 46600396$[9]. The data available from public sources for the time period between March 12th and April 6th is provided in Github[10].

[7]https://en.wikipedia.org/wiki/Demographics_of_Italy.
[8]Sources: Eficiens, Johns Hopkins University.
[9]https://en.wikipedia.org/wiki/Demographics_of_Spain.
[10]Sources: Statista, Johns Hopkins University.

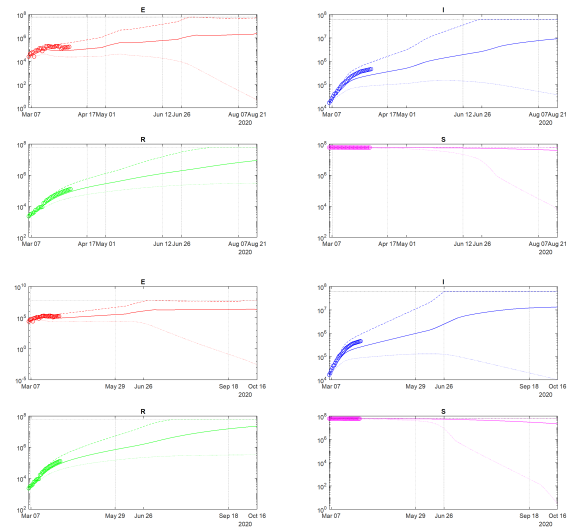


Fig. 13. The results of simulation of (5) for scenarios 1 and 2 in Italy under $\pm 10\%$ variation of all parameters

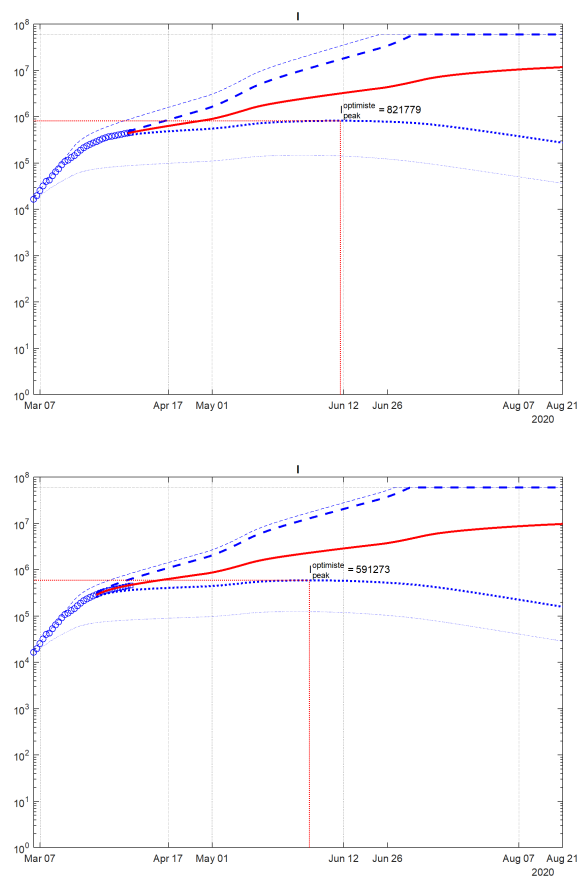


Fig. 14. Validation of prediction of I for scenario 1 in Italy with J and $J-10$ points of data under deviations of values of all parameters

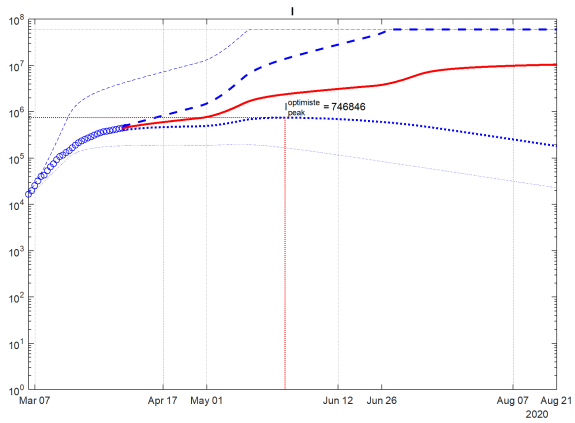


Fig. 15. Prediction of the growth of I for scenario 1 in Italy with $p_C = 1$ under deviations of values of all parameters

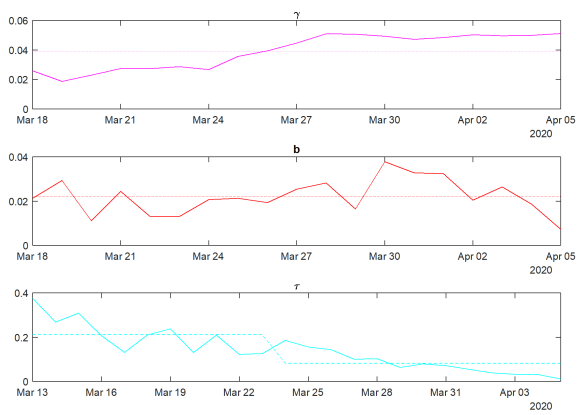


Fig. 16. The identified parameters for Spain

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0393,\ b = 0.0221$$

with the values γ_k, b_k, dI_t and dI_t^τ given in Fig. 16. Thus, the recovery rate γ and the infection rate b in Spain belong to the same range of values as in France (having slightly higher values in Spain, which implies a faster virus propagation in the country).

The simulation results of the interval predictor (5) with $\delta = 10\%$ are presented in Fig. 17 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the average behavior, the circles depict measured and reconstructed data points used for identification). As we can conclude from these plots, the obtained model follows well the measured statistics for Spain. In addition, due to a higher virus transition rate established previously the country infection peak seems to be attained in May.

A good fit of the model to the data is demonstrated in Fig. 18 (blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using the last day included in the identification data), the red line

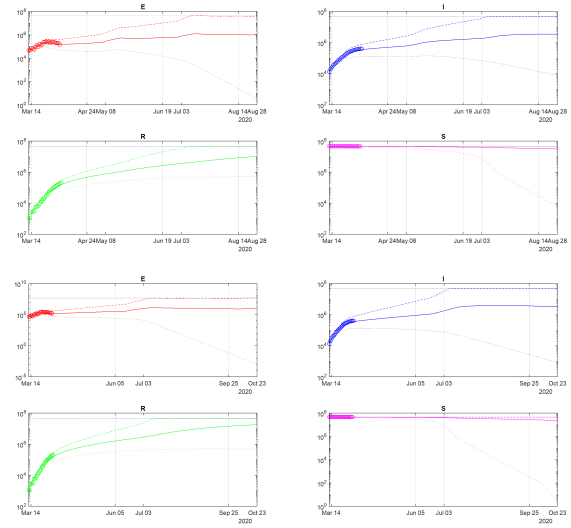


Fig. 17. The results of simulation of (5) for scenarios 1 and 2 in Spain under $\pm 10\%$ variation of all parameters

is the average, the blue circles and squares are the measured information used for identification and validation), where the square points lie close to the middle of the predicted interval in the bottom plot, while the top plot can be used for prediction of the number of infectives in $10-14$ days (both models indicate that the peak of infection in the optimistic case will be passed at the end of May). Finally, the influence of the strictness of confinement is illustrated in Fig. 19. A more strict quarantine decreases the peak amplitude in the optimistic setting.

C. Germany

The current population in Germany is $N = 46600396$[11]. The data available from public sources for the time period between March 12th and April 6th is provided in Github[12].

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0305,\ b = 0.0202$$

with the values γ_k, b_k, dI_t and dI_t^τ given in Fig. 20. Thus, the recovery rate γ and the infection rate b in Germany belong to the same range of values as in France.

The simulation results of the interval predictor (5) with $\delta = 15\%$ are presented in Fig. 21 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the average behavior, the circles depict measured and reconstructed data points used for identification). As we can conclude from these plots, the obtained model follows well the measured statistics for Germany. The optimistic country infection peak seems to be attained in May.

The predictor validation is demonstrated in Fig. 22 (blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using the last day included in the identification data), the red line is the average,

[11] https://en.wikipedia.org/wiki/Demographics_of_Germany.
[12] Sources: www.rki.de, Johns Hopkins University.

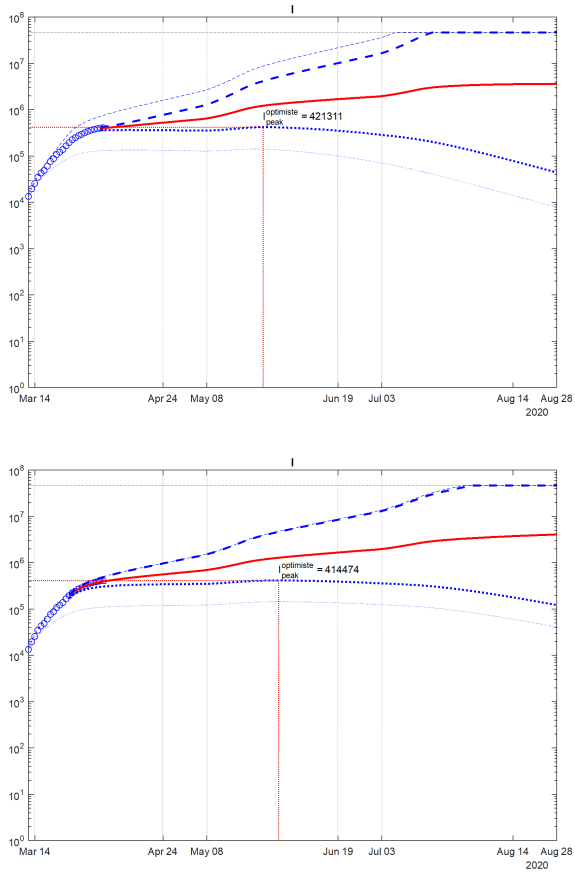


Fig. 18. Validation of prediction of I for scenario 1 in Spain with J and $J-10$ points of data under deviations of values of all parameters

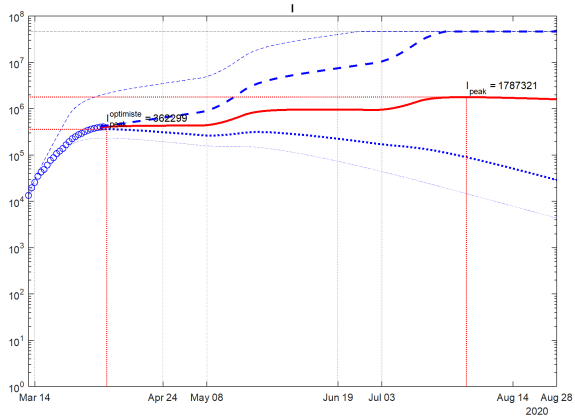


Fig. 19. Prediction of the growth of I for scenario 1 in Spain with $p_C = 1$ under deviations of values of all parameters

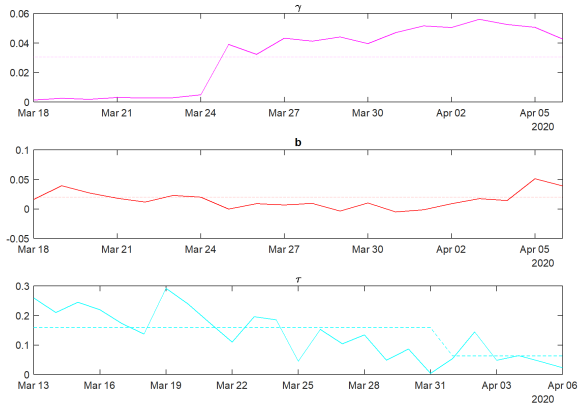


Fig. 20. The identified parameters for Germany

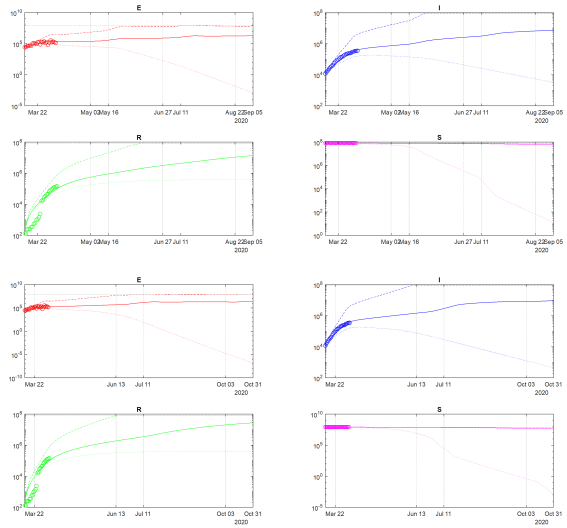


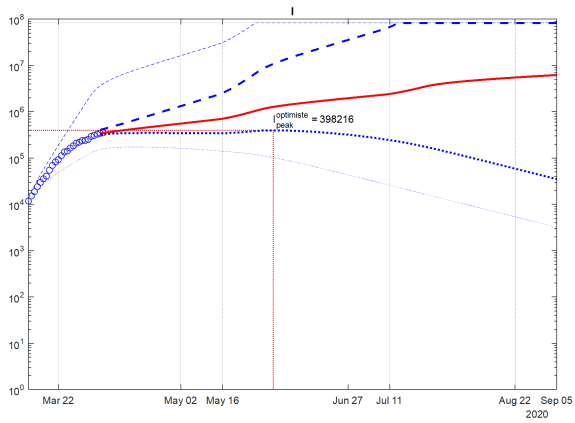
Fig. 21. The results of simulation of (5) for scenarios 1 and 2 in Germany under $\pm 15\%$ variation of all parameters

the blue circles and squares are the measured information used for identification and validation), where the square points are included again in the predicted interval in the bottom plot, while the top plot can be used for prediction of the number of infectives in $10-14$ days (both models indicate that the peak of infection in the optimistic case will be passed at the end of May or middle of June). Finally, the illustration of influence of the strictness of confinement is shown in Fig. 23. A more strict quarantine decreases the peak amplitude in the optimistic setting.

D. Brazil

The current population in Brazil is $N = 212559417$[13]. The data available from public sources for the time period between March 12th and April 6th is provided in Github[14].

[13] https://en.wikipedia.org/wiki/Demographics_of_Brazil.
[14] Sources: Johns Hopkins University.



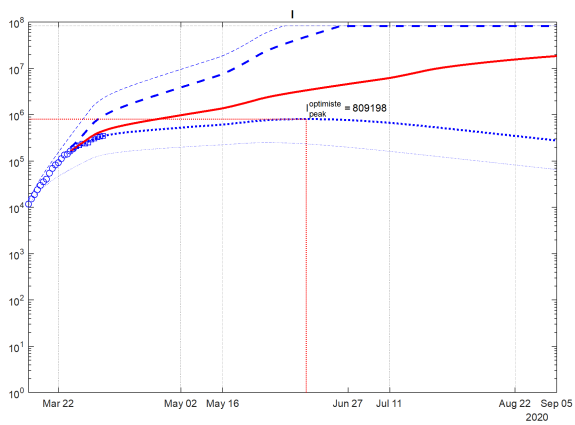


Fig. 22. Validation of prediction of I for scenario 1 in Germany with J and $J-10$ points of data under deviations of values of all parameters

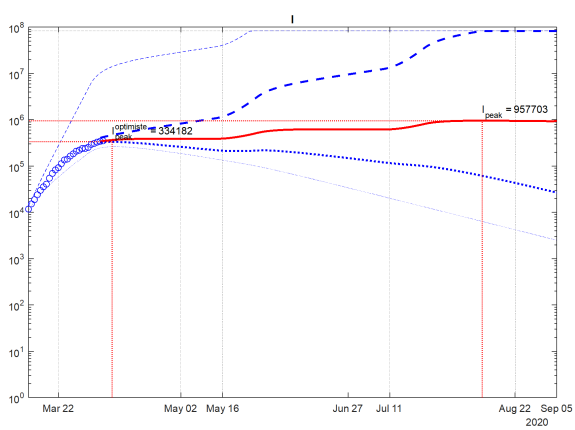


Fig. 23. Prediction of the growth of I for scenario 1 in Germany with $p_C = 1$ under deviations of values of all parameters

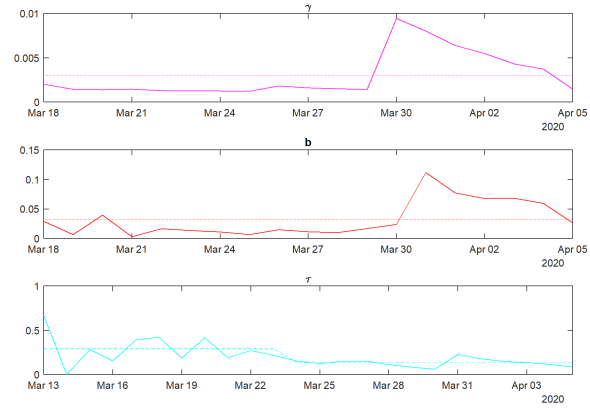


Fig. 24. The identified parameters for Brazil

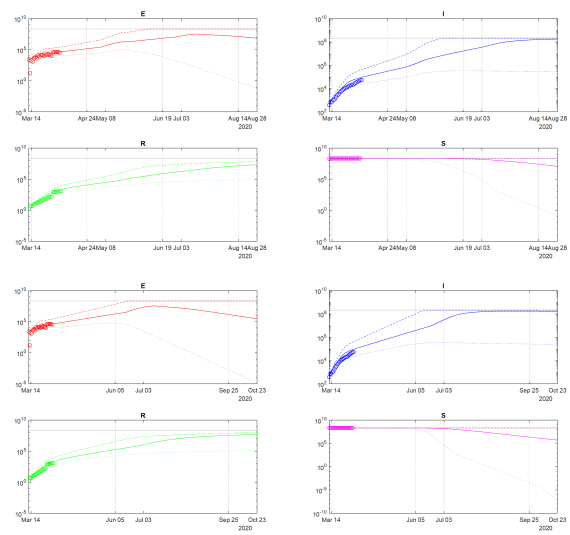


Fig. 25. The results of simulation of (5) for scenarios 1 and 2 in Brazil under $\pm 10\%$ variation of all parameters

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0030,\ b = 0.0323$$

with the values γ_k, b_k, dI_t and dI_t^{τ} given in Fig. 24. The infection rate b is within the limits of the same rate for France, Italy, Spain and Germany, but the value of the recovery rate γ is much smaller, which can be related with the fact that the epidemics in Brazil is currently on an initial phase, or that different protocols are used for registration.

The simulation results of the interval predictor (5) with $\delta = 10\%$ are presented in Fig. 25 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the average behavior, the circles depict measured and reconstructed data points used for identification). As we can conclude from these plots, the obtained model follows the measured statistics for Brazil, but with higher variations.

Verification of accuracy of the predictor using a part of data is demonstrated in Fig. 26 (blue dashed and dotted

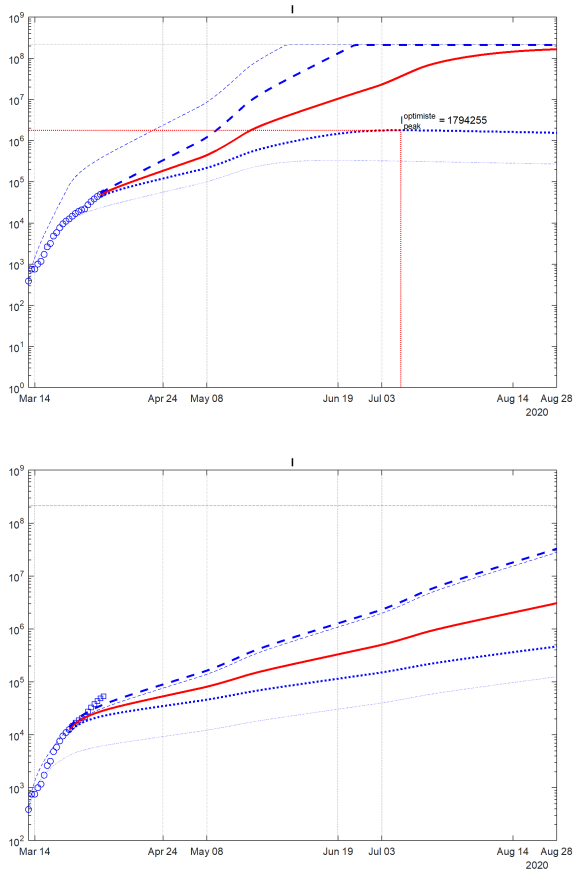


Fig. 26. Validation of prediction of I for scenario 1 in Brazil with J and $J-10$ points of data under deviations of values of all parameters

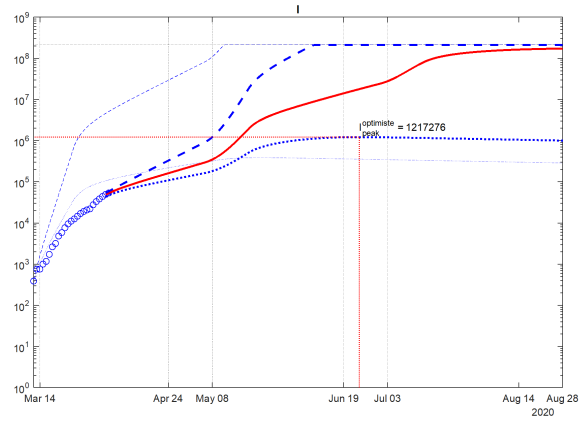


Fig. 27. Prediction of the growth of I for scenario 1 in Brazil with $p_C = 1$ under deviations of values of all parameters

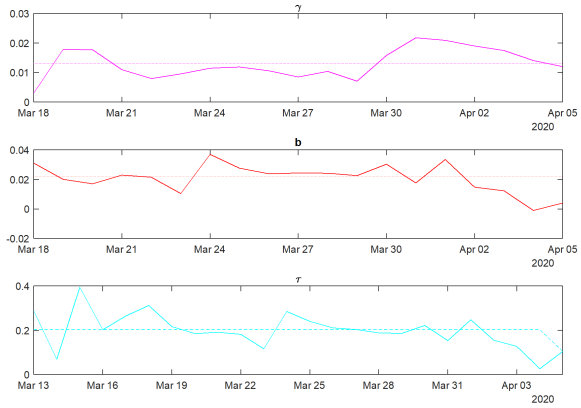


Fig. 28. The identified parameters for Russia

lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using the last day included in the identification data), the red line is the average, the blue circles and squares are the measured information used for identification and validation), where the square points exit the predicted interval in the bottom plot, which confirm again a less alignment of the selected model (1) and the corresponding predictor (5) with the available data. Finally, the illustration of influence of the strictness of confinement is shown in Fig. 27. A more strict quarantine decreases the epidemic peak amplitude in two times for the optimistic scenario.

E. Russia

The current population in Russia is $N = 146745098$[15]. The data available from public sources for the time period between March 12th and April 6th is provided in Github[16].

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0131,\ b = 0.0220$$

with the values γ_k, b_k, dI_t and dI_t^τ given in Fig. 28. In this case the values of the parameter γ is smaller than those for European countries, which can be related with a preliminary phase of the epidemics in Russia or different rules for registration.

The results of simulation of the interval predictor (5) with $\delta = 10\%$ are presented in Fig. 29 (the dashed and dotted lines represent, respectively, upper and lower interval bounds, the solid lines correspond to the average behavior, the circles depict measured and reconstructed data points used for identification). As we can conclude from these plots, the obtained model follows the measured statistics for Russia for an increased value of uncertainty.

An accurate fit of the model to the data is demonstrated in Fig. 30 (blue dashed and dotted lines correspond do the upper $\overline{I}$ and the lower bounds $\underline{I}$ (the bold lines are calculated using the last day included in the identification data), the red line is the average, the blue circles and squares are the measured information used for identification and validation), where the square points belong to the middle of the predicted interval in the bottom plot, while the top plot can be used for prediction of the number of infectives in $10-14$ days (in both cases, in the optimistic scenario the model predicts that the peak will be passed in July). Finally, the illustration of influence of the strictness of confinement is shown in Fig. 31. A more strict

[15] www.en.wikipedia.org/wiki/Demographics_of_Russia.
[16] Sources: Wikipedia, /www.rbc.ru.

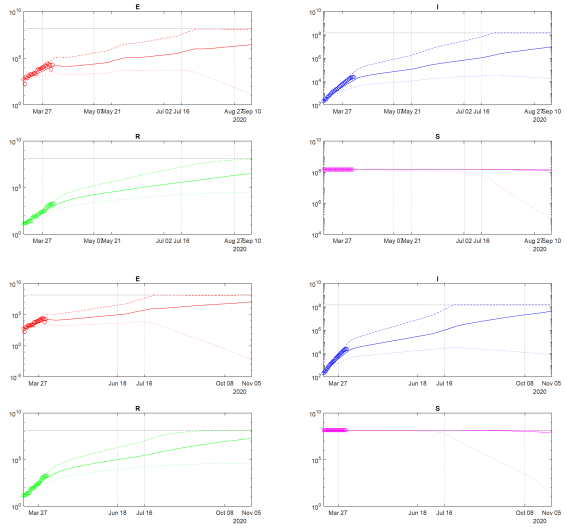


Fig. 29. The results of simulation of (5) for scenarios 1 and 2 in Russia under $\pm 10\%$ variation of all parameters

quarantine almost stabilizes the virus propagation giving the time for public health services to attenuate the consequences.

VII. Conclusion

A simple discrete-time SEIR epidemic model was identified and used to predict the influence of the quarantine on the SARS-CoV-2 virus propagation in France, Italy, Spain, Germany, Brazil and Russia. To enlarge the model prediction performance, an interval predictor method was also used to analyze the COVID-19 course. It was demonstrated that the reliability of the interval prediction for $10-14$ days is rather good, even by such a simple model. The prediction showed that a longer confinement may be a bit more efficient, but under the current level of uncertainty, a more strict as possible quarantine seems to be advisable.

The six considered countries can be divided on two groups: four European states (France, Italy, Spain and Germany), where the virus presence is already well developed with several weeks of quarantine, and two BRICS countries (Brazil and Russia), where the epidemics just started and general confinement has been imposed only recently. The identified models for these groups of countries have common patterns (*e.g.*, a low recovery rate γ for Brazil and Russia). Our prediction showed that in European countries the peak of infections will occur in May in the optimistic scenario. An increased severity of the confinement could significantly decrease the amplitude of the peak decharging the load on health services.

Machine learning tools can be further used to identify and optimize time profile for the confinement. Another possible direction of improvement of the proposed approach is to consider a SEIR model with population separation either by age or by region (or by both), but this implies an increasing number of parameters to be identified (that can be impossible) and also needs a special structured data to be available.

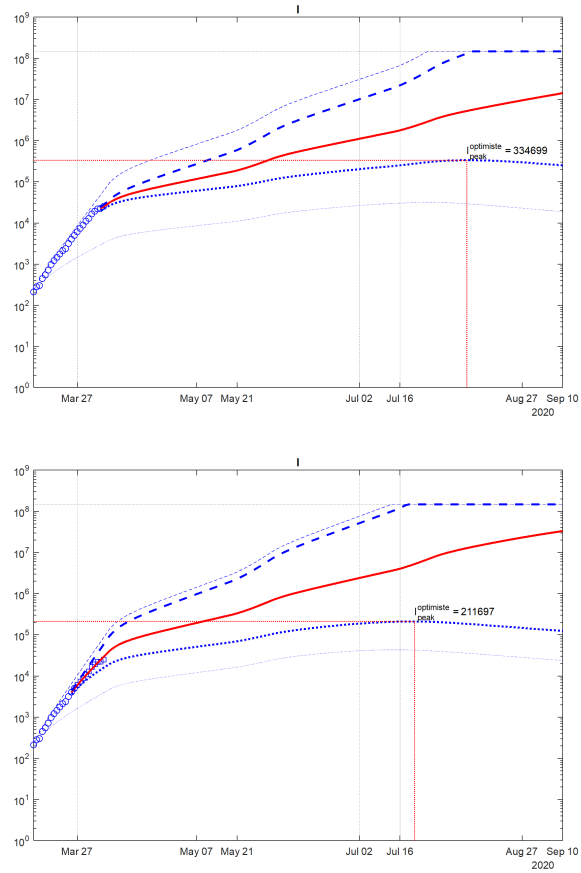


Fig. 30. Validation of prediction of I for scenario 1 in Russia with J and $J-10$ points of data under deviations of values of all parameters

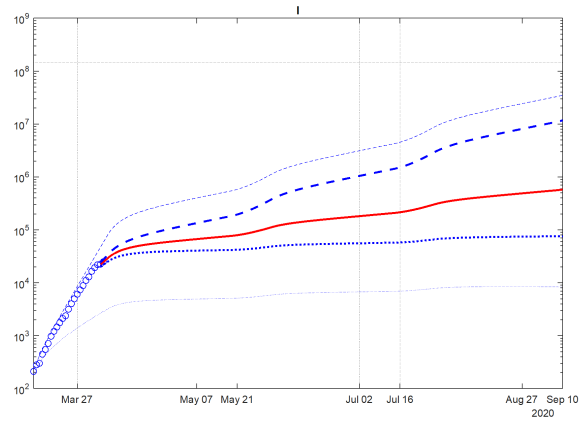


Fig. 31. Prediction of the growth of I for scenario 1 in Russia with $p_C = 1$ under deviations of values of all parameters

References

[1] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008. I

[2] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, "Statistical physics of vaccination," *Physics Reports*, vol. 664, pp. 1–113, 2016. I

[3] Z. Yang, Z. Zeng, K. Wang, S.-S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, 2020. [Online]. Available: http://jtd.amegroups.com/article/view/36385 I, I, 1, I, II, III, 3, III, 4, IV-B, IV-C

[4] 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. G. 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," WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease Analysis, Abdul Latif Jameel Institute for Disease and Emergency Analytics Imperial College London, COVID-19 reports, 2020. I, 1, III

[5] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China," *medRxiv*, 2020. I, 1

[6] J. Lourenco, R. Paton, M. Ghafari, M. Kraemer, C. Thompson, P. Simmonds, P. Klenerman, and S. Gupta, "Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic," *medRxiv*, 2020. I, 1, II

[7] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," 2020. I, 1

[8] A. d'Onofrio, P. Manfredi, and P. Poletti, "The interplay of public intervention and private choices in determining the outcome of vaccination programmes," *PLOS ONE*, vol. 7, no. 10, p. e45653, 2012. I

[9] B. Cantó, C. Coll, and E. Sánchez, "Estimation of parameters in a structured SIR model," *Advances in Difference Equations*, vol. 2017, no. 1, p. 33, 2017. [Online]. Available: https://doi.org/10.1186/s13662-017-1078-5 I

[10] P.-A. Bliman, D. Efimov, and R. Ushirobira, "A class of nonlinear adaptive observers for SIR epidemic model," in *Proceedings of ECC'18, the 16th annual European Control Conference*, June 2018. I

[11] P. Magal and G. Webb, "The parameter identification problem for SIR epidemic models: identifying unreported cases," *J. Math. Biol.*, vol. 77, pp. 1629–1648, 2018. I

[12] R. Ushirobira, D. Efimov, and P. Blirnan, "Estimating the infection rate of a SIR epidemic model via differential elimination," in *2019 18th European Control Conference (ECC)*, June 2019, pp. 1170–1175. I

[13] J. Gouzé, A. Rapaport, and M. Hadj-Sadok, "Interval observers for uncertain biological systems," *Ecological Modelling*, vol. 133, pp. 46–56, 2000. I

[14] F. Mazenc and O. Bernard, "Interval observers for linear time-invariant systems with disturbances," *Automatica*, vol. 47, no. 1, pp. 140–147, 2011. I

[15] T. Raïssi, D. Efimov, and A. Zolghadri, "Interval state estimation for a class of nonlinear systems," *IEEE Trans. Automatic Control*, vol. 57, no. 1, pp. 260–265, 2012. I

[16] F. Mazenc, T. N. Dinh, and S. I. Niculescu, "Interval observers for discrete-time systems," *International Journal of Robust and Nonlinear Control*, vol. 24, pp. 2867–2890, 2014. I, 6

[17] D. Efimov and T. Raïssi, "Design of interval observers for uncertain dynamical systems," *Autom. Remote Control*, vol. 77, no. 2, pp. 191–225, 2016. I, 1, V, V-A

[18] K. H. Degue, D. Efimov, and A. Iggidr, "Interval estimation of sequestered infected erythrocytes in malaria patients," in *2016 European Control Conference (ECC)*, June 2016, pp. 1141–1145. I

[19] M. S. Aronna and P.-A. Bliman, "Interval observer for uncertain time-varying SIR-SI epidemiological model of vector-borne disease," in *2018 16th European Control Conference (ECC)*, Limassol, 2018. I

[20] K. H. Degue and J. Le Ny, "An interval observer for discrete-time SEIR epidemic models," in *2018 Annual American Control Conference (ACC)*, June 2018, pp. 5934–5939. I

[21] Z. Hu, Q. Ge, S. Li, E. Boerwincle, L. Jin, and M. Xiong, "Forecasting and evaluating intervention of Covid-19 in the World," *arXiv e-prints*, p. arXiv:2003.09800, Mar. 2020. 1

[22] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application," *Annals of Internal Medicine*, 2020. II

[23] P. Magal and G. Webb, "Predicting the number of reported and unreported cases for the covid-19 epidemic in south korea, italy, france and germany," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/03/24/2020.03.21.20040154 6

[24] E. Leurent, D. Efimov, T. Raïssi, and W. Perruquetti, "Interval prediction for continuous-time systems with parametric uncertainties," in *Proc. IEEE Conference on Decision and Control (CDC)*, Nice, 2019. V-A, 6