



**HAL**  
open science

# On 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 interval prediction of COVID-19 development based on a SEIR epidemic model. [Research Report] Inria Lille Nord Europe - Laboratoire CRISAL - Université de Lille. 2020. hal-02517866v3

**HAL Id: hal-02517866**

**<https://inria.hal.science/hal-02517866v3>**

Submitted on 30 Mar 2020 (v3), 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 interval prediction of COVID-19 development based on a SEIR epidemic model

Denis Efimov<sup>1</sup> and Rosane Ushirobira<sup>1</sup>

**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 five different countries in the world. To this goal, the parameters of the SEIR model are identified by using publicly available data for the corresponding countries: France, Italy, Spain, 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.

## I. INTRODUCTION

The SEIR model is one of the most elementary compartmental models of epidemics [1], which is very popular and widely used in different contexts [2]. It 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  $S$ , 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 series of transitions between these populations), and a related advantage consists in a small number of parameters to identify (three transition rates in Fig. 1:  $\sigma$ ,  $\gamma$  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<sup>1</sup> 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<sup>2</sup> 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):

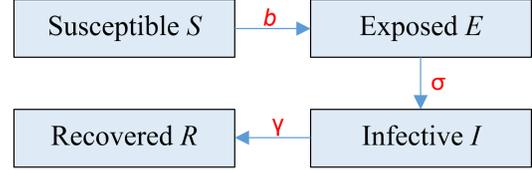


Fig. 1. A schematic representation of SEIR model

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

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

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

$$R_{t+1} = R_t + \gamma I_t, \quad (1d)$$

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),  $N$  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 susceptibles during a contact,  $0 < \sigma < +\infty$  is the incubation rate by 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).

Therefore, the SEIR model (1) has only three parameters to be identified:  $\sigma$ ,  $\gamma$  and  $b$ . These parameters  $\sigma$ ,  $\gamma$  and  $b$  represent, respectively, the rate of change between the states  $E$  to  $I$ ,  $I$  to  $R$  and  $S$  to  $E + I$  (as in in Fig. 1). The parameter  $\sigma$  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. 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. As a worthy remark, very 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.

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

<sup>1</sup>Denis Efimov and Rosane Ushirobira are with Inria, Univ. Lille, CNRS UMR 9189 - CRISTAL, F-59000 Lille, France, Denis.Efimov, Rosane.Ushirobira@inria.fr

<sup>2</sup>This note was initially written on March 23 and updated March 30, 2020.

<sup>3</sup>Compared to the model in [3], the inflow/outflow variables for each state were not considered in our analysis.

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 worth 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], 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 all possible trajectories of the system. As a first remark, such a methodology needs more computational effort for its realization. Second, 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 SARS-CoV-2 virus attack, it is hard to obtain such a data quickly during the epidemic development. Contrarily to these approaches, the interval predictor method does not use these additional assumptions on 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, 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 results of application of the proposed identification routine and the interval predictor for Italy, Spain, 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, is considered to be between 2 and 14 days [3], or in a more specialized research between 2 and 12 days [21], so we assume

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

<sup>3</sup>[www.en.wikipedia.org/wiki/Demographics\\_of\\_France](http://www.en.wikipedia.org/wiki/Demographics_of_France).

Day $t$	$\mathcal{I}_t$	$\mathcal{D}_t$	$\mathcal{H}_t$
1	2876	61	12
2	3661	79	12
3	4499	91	12
4	5423	127	12
5	6633	148	12
6	7730	175	12
7	9134	244	12
8	10995	372	12
9	12612	450	12
10	14459	562	12
11	16018	674	2200
12	19856	860	2200
13	22302	1100	3288
14	25230	1331	3907
15	29155	1696	4948
16	32964	1995	5700
17	37575	2314	6624
18	40174	2606	7132

TABLE I  
DATA FOR FRANCE

The data available from public sources<sup>4</sup>, and which is used in this note, for the time period between March 12th and 30th is provided in Table I<sup>5</sup>, where  $\mathcal{I}$ ,  $\mathcal{D}$  and  $\mathcal{H}$  represent the number of detected infected, deceased and recovered individuals, respectively.

Obviously, not all cases can be detected and documented by the public health services, then there is a ratio between populations  $I$  and  $\mathcal{I}$ ,  $R$  and  $\mathcal{D} + \mathcal{H}$  as well, which is denoted in this work by  $\alpha$ , whose interval of admissible values is estimated from different sources as follows<sup>6</sup>:

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

Formally, such a ratio  $\alpha$  has to be time-varying and different for  $I$  and  $R$ . Due to strict requirements of France Health Services, in this report we take the following hypotheses:

$$I_t = \alpha_1 \mathcal{I}_t, R_t = \mathcal{D}_t + \alpha_2 \mathcal{H}_t, \quad (2)$$

*i.e.*, the deaths are reported exactly (see also [6]), but the actual number of 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:

$$\alpha_1 = 5, \alpha_2 = \alpha_1 + 1,$$

then, roughly speaking, such a choice of  $\alpha_1$  corresponds to registration of all non-mild cases, and  $\alpha_2$  is selected just to be a little bit higher. The techniques to identify  $\alpha_1$  from the measurements of  $\mathcal{I}$  and  $\mathcal{D}$  are described in the last footnote (for the data given in Table I these approaches provide  $\alpha_1 = 2.418$ ). So, by fixing  $\alpha_1$  and  $\alpha_2$ , the two variables of the model

<sup>4</sup>See [www.santepubliquefrance.fr/](http://www.santepubliquefrance.fr/), [www.geodes.santepubliquefrance.fr](http://www.geodes.santepubliquefrance.fr) and [www.data.gouv.fr](http://www.data.gouv.fr).

<sup>5</sup>Sources: Eficiens, Johns Hopkins University.

<sup>6</sup>See, for example, [Coronalinks-3-19-20](https://coronalinks-3-19-20.com), or a dedicated analysis in CMMID or University of Melbourne.

(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).

### 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 by the virus differently, 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 economics to be maintained on 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 described above parameters 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

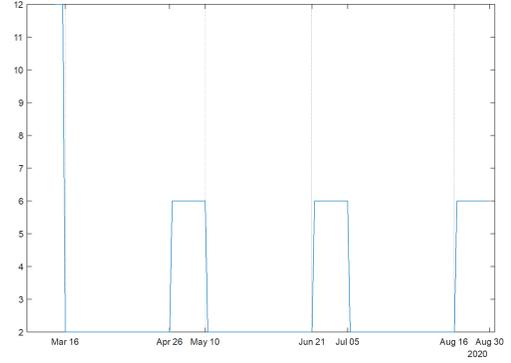


Fig. 2. Variation of the number of contacts  $r_t$  in the first scenario

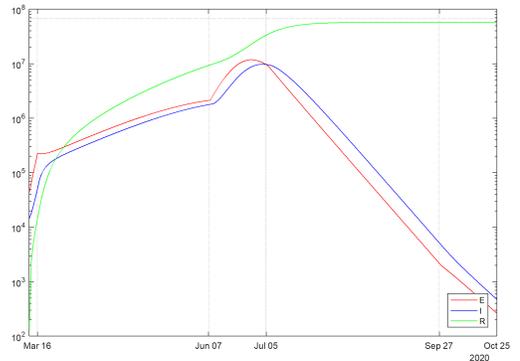
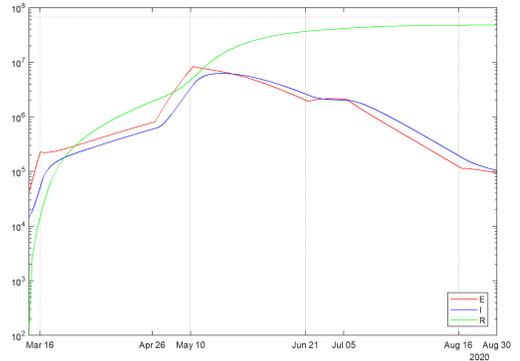


Fig. 3. The results of simulation with parameters of [3] for scenarios 1 and 2

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 (overestimates) to the situation in France. Therefore, an identification of parameters is needed.

### IV. PARAMETER IDENTIFICATION

For the parameter identification, we assume that the incubation rate  $\sigma$  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  $t = 0, 1, \dots, J$ .

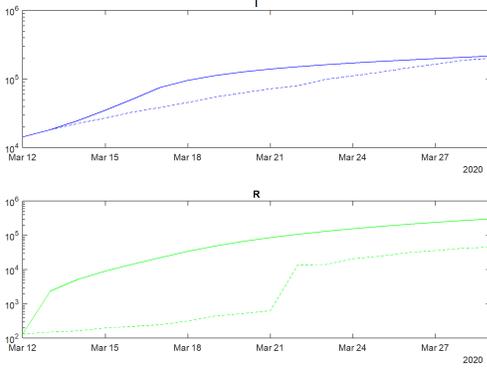


Fig. 4. The results of verification with parameters of [3]

From the equation (1d) we can identify the value of the parameter  $\gamma$ :

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

whose least squares estimation is

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

for  $k = 0, 1, \dots, K$ , where  $0 < K < J - 1$  is the number of the last days used for identification (in this work we selected  $K = J - 10$ ). Then the average value is used for further analysis and design (with a mild ambiguity we are using the same symbol to denote the parameter and its estimate):

$$\gamma = \frac{1}{K+1} \sum_{k=0}^K \gamma_k = 0.0391.$$

Next, the equation (1c) allows us to calculate the related number of asymptomatic infectious ( $\sigma$  is chosen and  $\gamma$  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  $p_C$  and  $r_t$ ):

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

whose least squares estimation is

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

for  $k = 0, 1, \dots, K$ , then the identified value is again the average of these estimates:

$$b = \frac{1}{K+1} \sum_{k=0}^K b_k = 0.0387.$$

The obtained values  $\gamma_k, b_k$  (solid lines) together with the selected estimates  $\gamma, b$  (dot lines) are shown in Fig. (5). The

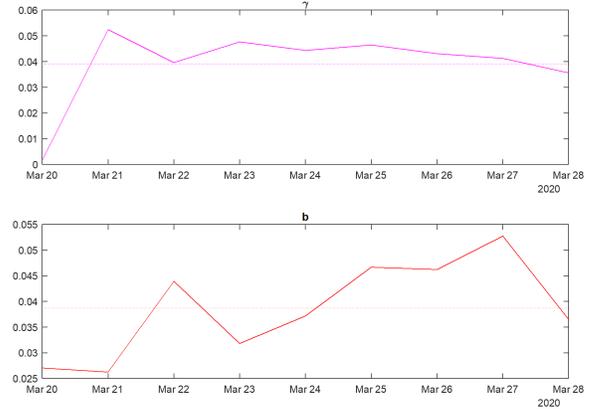


Fig. 5. The identified parameters for France

identified values for  $\gamma$  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.

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 these results are more 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 epidemics will be reached in the middle of July 2020 (the model with parameters from [3] predicts the peaks even faster).

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

## V. INTERVAL PREDICTION

In the previous section, to make a prediction the model (1) with identified values of the parameters  $b, \gamma$  and selected choices for  $\alpha_1, \alpha_2, \sigma$  was used. The initial values for the state of the model  $S_0, I_0, E_0$  and  $R_0$  were selected 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 the intervals of admissible values for all variables and parameters used for simulation, so 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 framework [17].

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

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

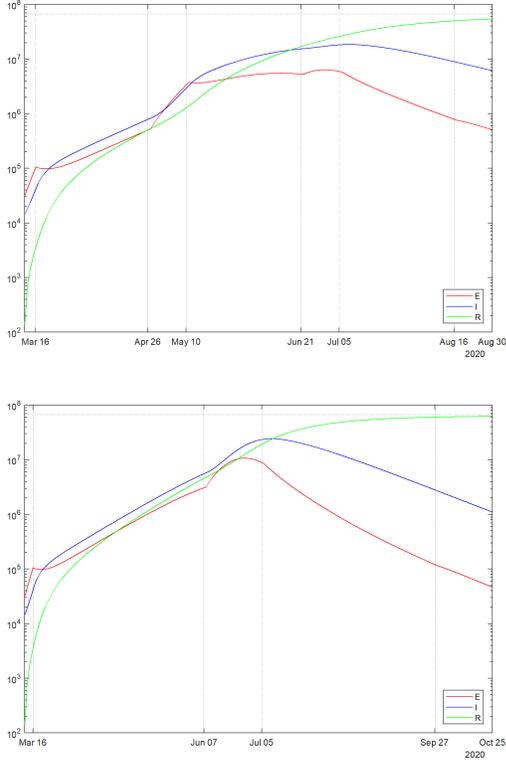


Fig. 6. The results of simulation with identified parameters

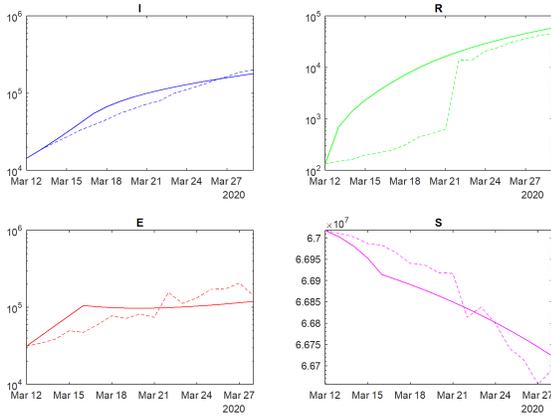


Fig. 7. The results of verification with identified parameters

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

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

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

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

for all  $t \in \mathbb{N}$ . So, we assume that  $x_0 \leq \bar{x}_0$ ,  $0 \leq a_t \leq \bar{a}_t$  and  $\underline{d}_t \leq \bar{d}_t$  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  $\bar{x}_t$

predictions on the state  $x_t$  of this system under the introduced hypotheses on all uncertain variables, which have to satisfy the relations:

$$\underline{x}_t \leq x_t \leq \bar{x}_t \quad \forall t \in \mathbb{N}.$$

The theory of interval observers and predictors [17], [22] answers this question, and a possible solution (that utilizes the nonnegativity of  $x_t$  and  $a_t$ ) is as follows:

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

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

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

then it is easy to verify that the terms  $d_t - \underline{d}_t, \bar{d}_t - d_t$  are nonnegative by the definition of  $\underline{d}_t, \bar{d}_t$ , and the terms  $a_t x_t - \underline{a}_t \underline{x}_t, \bar{a}_t \bar{x}_t - a_t x_t$  have the same property for  $t = 0$  by the definition of  $\underline{a}_t, \bar{a}_t$  and  $\underline{x}_0, \bar{x}_0$ , hence,  $\underline{e}_1 \geq 0, \bar{e}_1 \geq 0$  (that implies  $x_1 \in [\underline{x}_1, \bar{x}_1]$ ) 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).

To this end, we assume that all parameters belong to the known intervals:

$$\begin{aligned} \sigma &\in [\underline{\sigma}, \bar{\sigma}], \gamma \in [\underline{\gamma}, \bar{\gamma}], b \in [\underline{b}, \bar{b}], \\ r_t &\in [\underline{r}_t, \bar{r}_t] \quad \forall t \in \mathbb{N}, \end{aligned} \quad (3)$$

together with the initial conditions in (1):

$$S_0 \in [\underline{S}_0, \bar{S}_0], I_0 \in [\underline{I}_0, \bar{I}_0], E_0 \in [\underline{E}_0, \bar{E}_0], R_0 \in [\underline{R}_0, \bar{R}_0], \quad (4)$$

where nonnegative values  $\underline{\sigma}, \bar{\sigma}, \underline{\gamma}, \bar{\gamma}, \underline{b}, \bar{b}, \underline{r}_t, \bar{r}_t, \underline{S}_0, \bar{S}_0, \underline{I}_0, \bar{I}_0, \underline{E}_0, \bar{E}_0$  and  $\underline{R}_0, \bar{R}_0$  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:

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

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

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

$$2\bar{b} \sup_{t \in \mathbb{N}} \bar{r}_t \leq 1, \quad \bar{\sigma} \leq 1, \quad \bar{\gamma} \leq 1, \quad (6)$$

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

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

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

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

*Proof:* By direct calculations we can check that

$$\begin{aligned} \underline{b} \frac{(p_c \underline{I}_t + r_t \underline{E}_t)}{N} &\leq \underline{b} \frac{(p_c \bar{I}_t + r_t \bar{E}_t)}{N} \leq \bar{b} \frac{(p_c \bar{I}_t + \bar{r}_t \bar{E}_t)}{N}, \\ \underline{b} \frac{r_t}{N} \underline{S}_t - \bar{\sigma} &\leq \underline{b} \frac{r_t}{N} \underline{S}_t - \bar{\sigma} \leq \bar{b} \frac{\bar{r}_t}{N} \bar{S}_t - \bar{\sigma}, \\ p_c \underline{b} \frac{\underline{I}_t \underline{S}_t}{N} &\leq p_c \bar{b} \frac{\bar{I}_t \bar{S}_t}{N} \leq p_c \bar{b} \frac{\bar{I}_t \bar{S}_t}{N}, \\ \underline{\sigma} \underline{E}_t &\leq \sigma \underline{E}_t \leq \bar{\sigma} \bar{E}_t, \\ \underline{\gamma} \underline{I}_t &\leq \gamma \underline{I}_t \leq \bar{\gamma} \bar{I}_t \end{aligned}$$

due to (3) and (4) for  $t = 0$ , and

$$1 \geq \bar{b} \frac{(p_c \bar{I}_t + \bar{r}_t \bar{E}_t)}{N}, \quad 1 + \underline{b} \frac{r_t}{N} \underline{S}_t \geq \bar{\sigma}$$

due to (6) (recall that  $r_t \geq p_c$ ,  $\bar{I}_t + \bar{E}_t \leq 2N$ , thus  $\underline{S}_t \geq 0$ ), then as we demonstrated above

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

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

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

are satisfied for some  $t \in \mathbb{N}$ , then they also hold for  $t = t + 1$  in (5), and to substantiate boundedness of the state of the interval predictor, it is enough to guarantee that  $\bar{S}_t$ ,  $\bar{I}_t$ ,  $\bar{E}_t$  and  $\bar{R}_t$  do 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 4.* The boundedness of the state of (5) established in Theorem (3) 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 the predictors [16], [22].

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

$$\begin{aligned} \underline{S}_t &= N - \bar{I}_t - \bar{E}_t - \bar{R}_t, \\ \bar{S}_t &= N - \underline{I}_t - \underline{E}_t - \underline{R}_t, \end{aligned}$$

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.

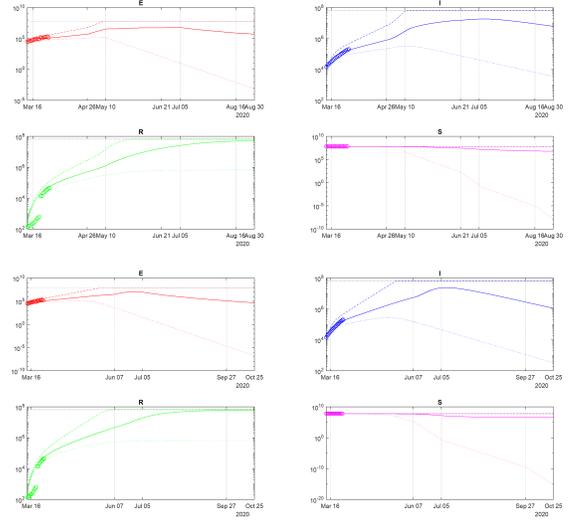


Fig. 8. The results of simulation of (5) for scenarios 1 and 2 in France under  $\pm 10\%$  variation of all parameters

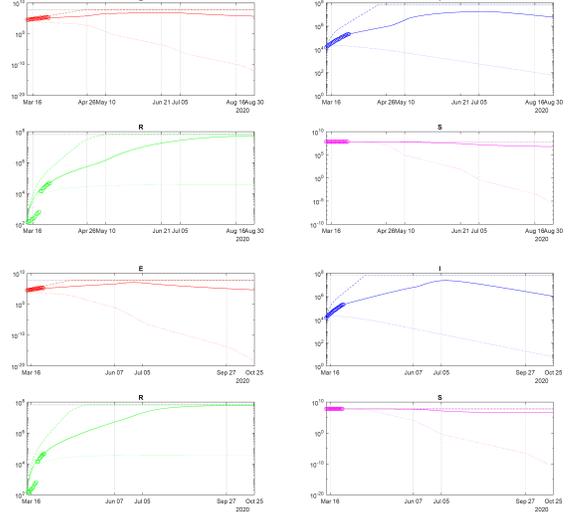


Fig. 9. The results of simulation of (5) for scenarios 1 and 2 under variation of  $\sigma \in [\frac{1}{12}, \frac{1}{2}]$

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  $\sigma$  is deviated with  $[\underline{\sigma}, \bar{\sigma}] = [\frac{1}{12}, \frac{1}{2}]$  are shown in Fig. 9 (the parameters  $\gamma, b$  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 this parameter 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  $\delta$  and chosen for these simulations.

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  $\delta$  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 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. 10, where the interval prediction for the infectious population  $I$  is presented in the first scenario with deviation of all parameters. As previous, blue dashed and dotted lines correspond to the upper  $\bar{I}$  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 top graphic corresponds to the case given previously in Fig. 8 for  $p_C = 2$ , while the bottom plot 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, while at the peak the quantity of infected people can be decreased in almost three times (which is an important achievement representing a significant decay of the load to the 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 Fig. 10). The common parameters assigned to all countries (to simplify the analysis) are:

$$\alpha_1 = 5, \alpha_2 = \alpha_1 + 1, \sigma = \frac{1}{7},$$

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

### A. Italy

The current population in Italy is  $N = 60359546^7$ . The data available from public sources for the time period between March 13th and 30th is provided in Table II<sup>8</sup>.

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0207, b = 0.0447$$

with the values  $\gamma_k$  and  $b_k$  given in Fig. 11. France and Italy have the virus infection rate  $b$  approximately close, but in Italy the recovery rate  $\gamma$  is lower compared to France.

The simulation results of the interval predictor (5) with  $\delta = 10\%$  are presented in Fig. 12 (the dashed and dotted

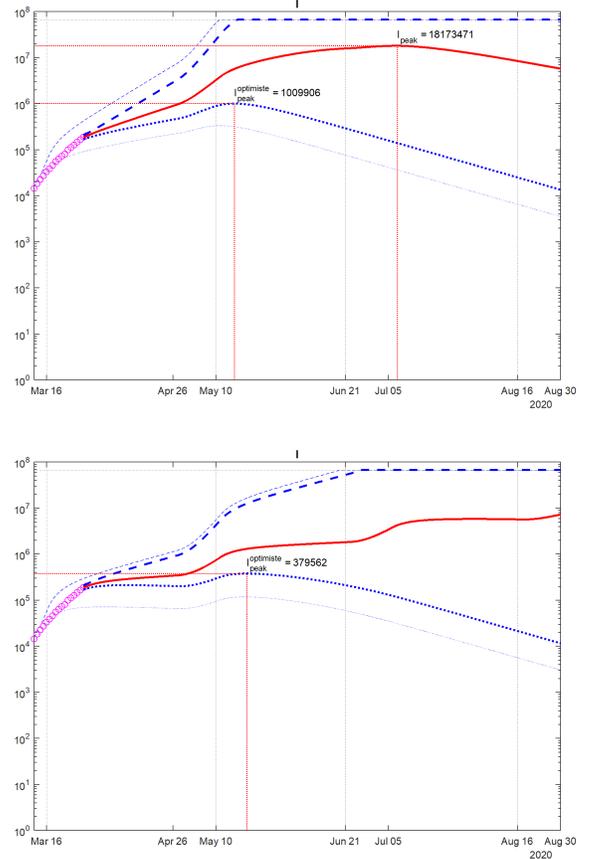


Fig. 10. Prediction of the growth of  $I$  for scenario 1 in France with  $p_C = 2$  or  $p_C = 1$  under deviations of values of all parameters

Day $t$	$\mathcal{I}_t$	$\mathcal{D}_t$	$\mathcal{H}_t$
1	17660	1266	1439
2	21157	1441	1439
3	24747	1809	1439
4	27980	2158	2749
5	31506	2503	2941
6	35713	2978	4025
7	41035	3405	4440
8	47021	4032	5129
9	53578	4825	6072
10	59158	5476	7024
11	63927	6077	7432
12	69176	6820	8326
13	74386	7503	9362
14	80589	8215	10361
15	86498	9134	10950
16	92472	10023	12384
17	97689	10779	13030

TABLE II  
DATA FOR ITALY

<sup>7</sup>[https://en.wikipedia.org/wiki/Demographics\\_of\\_Italy](https://en.wikipedia.org/wiki/Demographics_of_Italy).

<sup>8</sup>Sources: Eficiens, Johns Hopkins University.

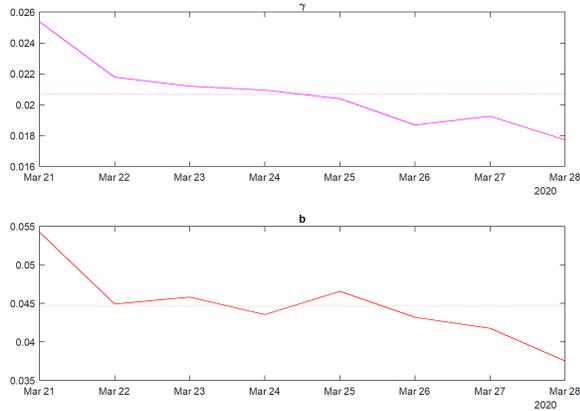


Fig. 11. The identified parameters for Italy

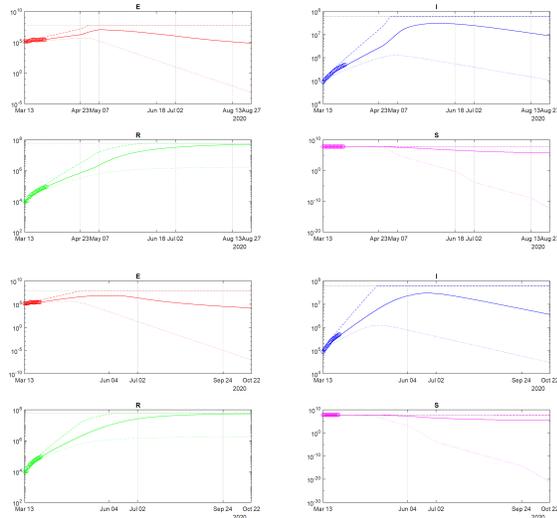


Fig. 12. The results of simulation of (5) for scenarios 1 and 2 in Italy under  $\pm 10\%$  variation of all parameters

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.

Finally, the illustration of influence of the strictness of confinement is shown in Fig. 13. A more strict quarantine decreases and makes it sooner the peak of epidemics in 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 30th is provided in Table III<sup>10</sup>.

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0353, b = 0.0761$$

<sup>9</sup>[https://en.wikipedia.org/wiki/Demographics\\_of\\_Spain](https://en.wikipedia.org/wiki/Demographics_of_Spain).

<sup>10</sup>Sources: Statista, Johns Hopkins University.

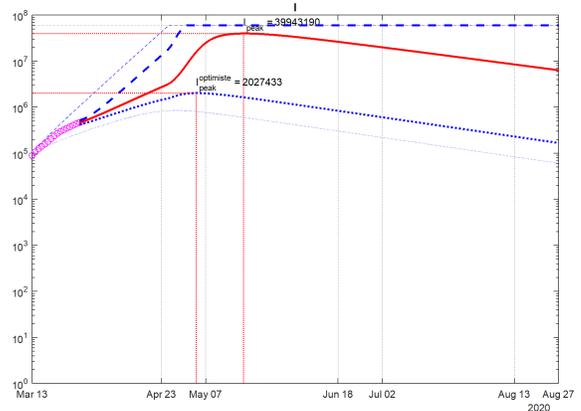
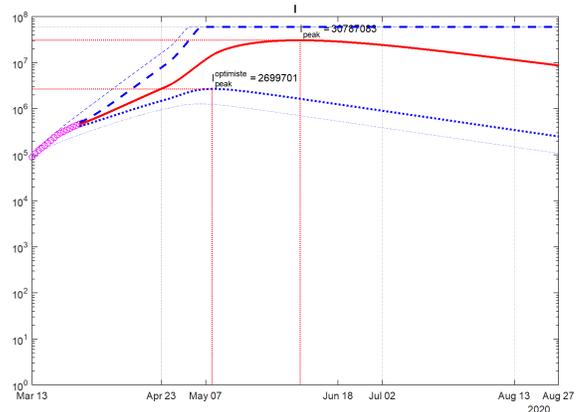


Fig. 13. Prediction of the growth of  $I$  for scenario 1 in Italy with  $p_C = 2$  or  $p_C = 1$  under deviations of values of all parameters

Day $t$	$\mathcal{I}_t$	$\mathcal{D}_t$	$\mathcal{H}_t$
1	2950	84	189
2	4209	118	193
3	5753	136	517
4	7753	288	517
5	9383	309	530
6	11273	497	1028
7	13716	598	1081
8	17147	767	1107
9	19980	982	1588
10	24926	1326	2125
11	28572	1720	2575
12	33089	2182	3355
13	39675	2696	3794
14	47600	3434	5367
15	56188	4365	7015
16	64059	5138	9357
17	73235	5982	12285
18	80110	6803	14709

TABLE III  
DATA FOR SPAIN

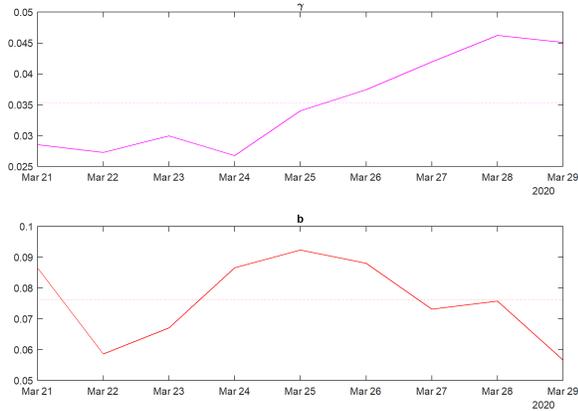


Fig. 14. The identified parameters for Spain

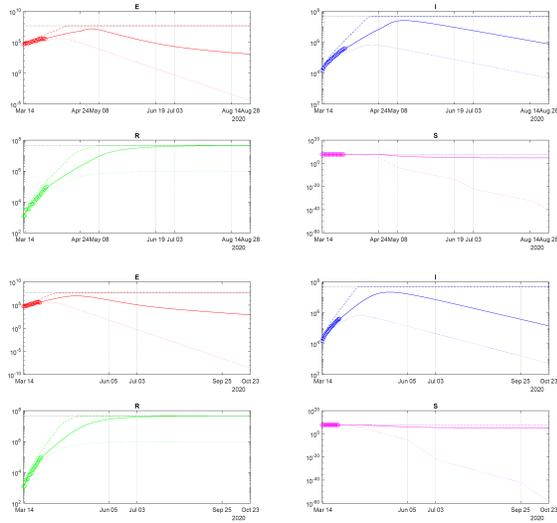


Fig. 15. The results of simulation of (5) for scenarios 1 and 2 in Spain under  $\pm 10\%$  variation of all parameters

with the values  $\gamma_k$  and  $b_k$  given in Fig. 14. Thus, the recovery rate  $\gamma$  in Spain belongs to the same range of values as in France and Italy, but the infection rate  $b$  takes much higher values, which implies a faster virus propagation in the country.

The simulation results of the interval predictor (5) with  $\delta = 10\%$  are presented in Fig. 15 (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 the beginning of May.

Finally, the illustration of influence of the strictness of confinement is shown in Fig. 16. A more strict quarantine decreases the peak amplitude in the optimistic setting.

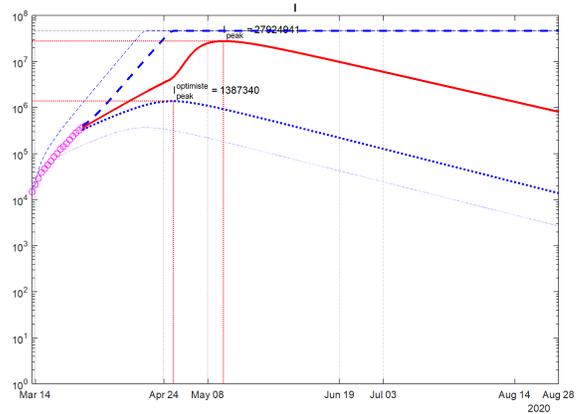
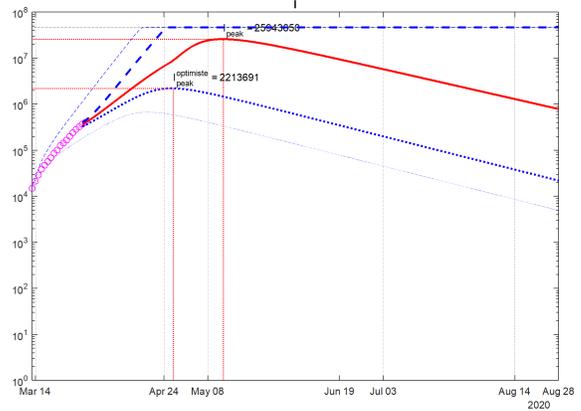


Fig. 16. Prediction of the growth of  $I$  for scenario 1 in Spain with  $p_C = 2$  or  $p_C = 1$  under deviations of values of all parameters

Day $t$	$\mathcal{I}_t$	$\mathcal{D}_t$	$\mathcal{H}_t$
1	77	0	0
2	151	0	1
3	151	0	1
4	200	0	1
5	234	0	2
6	346	1	2
7	529	4	2
8	640	7	2
9	970	11	2
10	1178	18	2
11	1546	25	2
12	1924	34	2
13	2247	46	2
14	2554	59	2
15	2985	77	6
16	3477	93	6
17	3904	114	6
18	4256	136	6

TABLE IV  
DATA FOR BRAZIL

### C. Brazil

The current population in Brazil is  $N = 212559417^{11}$ . The data available from public sources for the time period between March 12th and 30th is provided in Table IV<sup>12</sup>.

<sup>11</sup>[https://en.wikipedia.org/wiki/Demographics\\_of\\_Brazil](https://en.wikipedia.org/wiki/Demographics_of_Brazil).

<sup>12</sup>Sources: Johns Hopkins University.

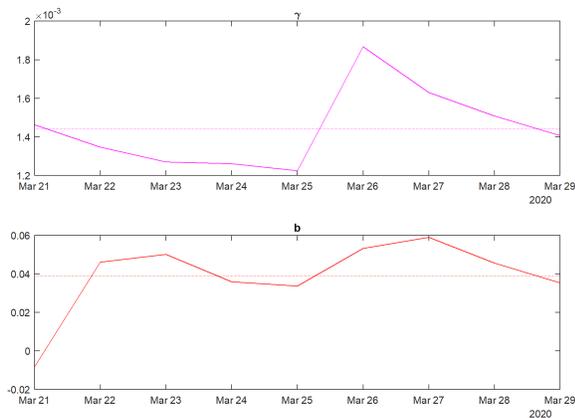


Fig. 17. The identified parameters for Brazil

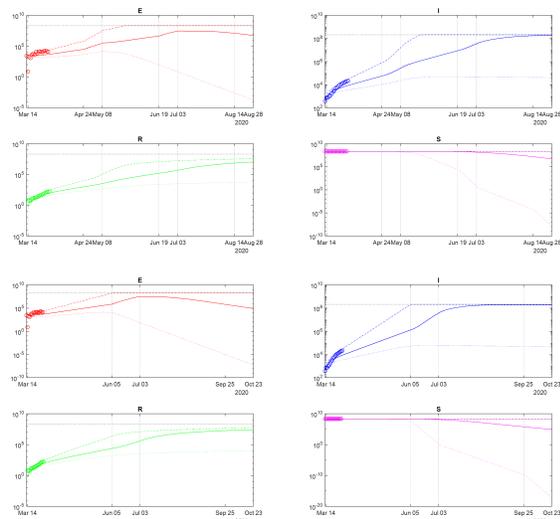


Fig. 18. The results of simulation of (5) for scenarios 1 and 2 in Brazil under  $\pm 20\%$  variation of all parameters

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0014, b = 0.0390$$

with the values  $\gamma_k$  and  $b_k$  given in Fig. 17. The infection rate  $b$  is within the limits of the same rate for France, Italy and Spain, but the value of the recovery rate  $\gamma$  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  $\delta = 20\%$  are presented in Fig. 18 (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 Brazil, but at the price of an increased level of uncertainty of two times with respect to European countries.

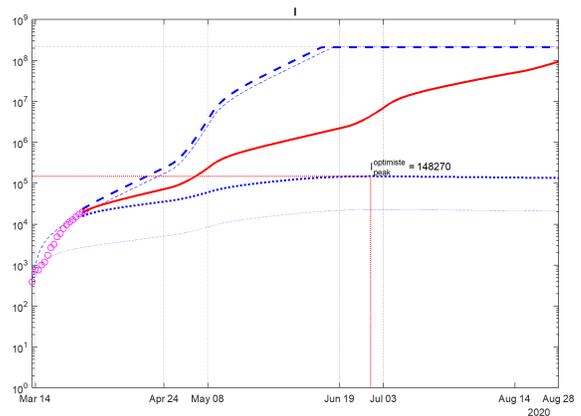
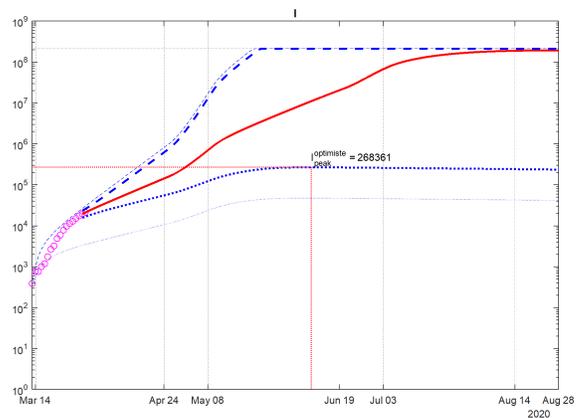


Fig. 19. Prediction of the growth of  $I$  for scenario 1 in Brazil with  $p_C = 2$  or  $p_C = 1$  under deviations of values of all parameters

Finally, the illustration of influence of the strictness of confinement is shown in Fig. 19. A more strict quarantine decreases the epidemic peak amplitude in two times for the optimistic scenario.

#### D. Russia

The current population in Russia is  $N = 146745098$ <sup>13</sup>. The data available from public sources for the time period between March 12th and 30th is provided in Table V<sup>14</sup>.

Applying the proposed procedure to parameter identification leads to

$$\gamma = 0.0112, b = 0.0234$$

with the values  $\gamma_k$  and  $b_k$  given in Fig. 20. In this case the values of these parameters are smaller those for European countries, which again 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  $\delta = 15\%$  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

<sup>13</sup>[www.en.wikipedia.org/wiki/Demographics\\_of\\_Russia](http://www.en.wikipedia.org/wiki/Demographics_of_Russia).

<sup>14</sup>Sources: Wikipedia, Johns Hopkins University.

Day $t$	$\mathcal{I}_1$	$\mathcal{D}_1$	$\mathcal{H}_1$
1	45	0	3
2	59	0	3
3	63	0	4
4	93	0	5
5	114	0	5
6	147	0	5
7	199	0	5
8	253	0	12
9	306	0	16
10	367	0	16
11	438	0	17
12	495	0	22
13	658	0	29
14	840	2	38
15	1036	3	45
16	1264	4	49
17	1534	8	64
18	1836	9	66

TABLE V  
DATA FOR RUSSIA

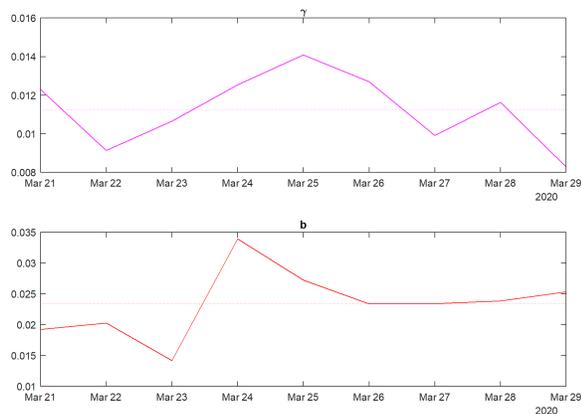


Fig. 20. The identified parameters for Russia

for identification). As we can conclude from these plots, the obtained model follows the measured statistics for Russia for an increased value of uncertainty.

Finally, the illustration of influence of the strictness of confinement is shown in Fig. 22. A more strict 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, Brazil and Russia. To enlarge the model prediction performance, an interval predictor method was also used to analyze the COVID-19 course. 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 confinement seems to be advisable.

The five considered countries can be divided on two groups: three European states (France, Italy and Spain), where the virus presence is already well developed with several weeks of quarantine, and two BRICS countries (Brazil and Russia), where the epidemics is just started and the confinement has

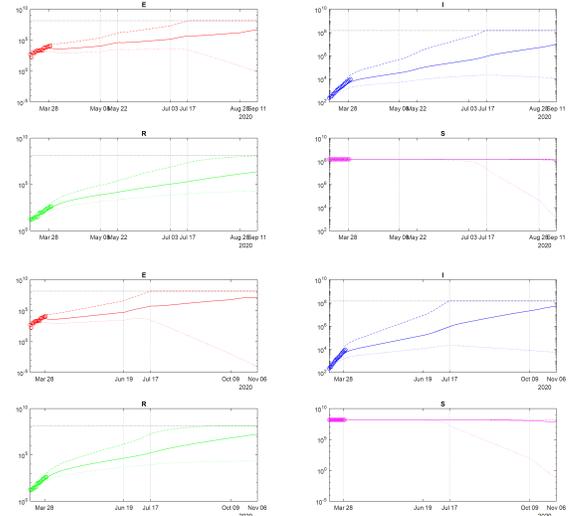


Fig. 21. The results of simulation of (5) for scenarios 1 and 2 in Russia under  $\pm 15\%$  variation of all parameters

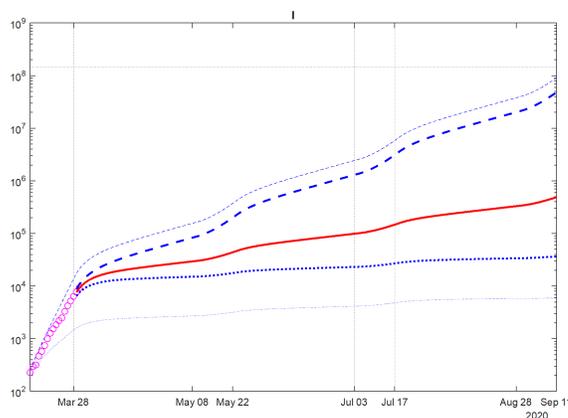
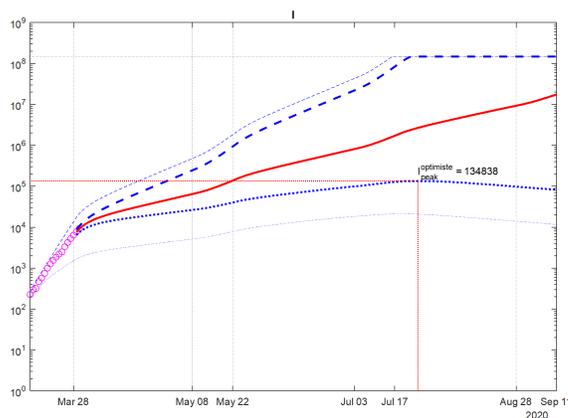


Fig. 22. Prediction of the growth of  $I$  for scenario 1 in Russia with  $pc = 2$  or  $pc = 1$  under deviations of values of all parameters

been imposed just recently. The identified models for these groups of countries have common patterns (e.g., low recovery rate  $\gamma$  for Brazil and Russia, and also lower precision of the model required a higher value of  $\delta$ ). The prediction shows that in European countries the peak of infections will be passed at the beginning of May in the optimistic scenario. An increased severity of the confinement could significantly decrease the amplitude of the peak discharging 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 increasing of the number of parameters to identify and also needs a special structured data to be available.

## 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, 2, 1, I, II, III, 3, III, 4, IV, IV
- [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 sub-exponential 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. Bliman, “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, 4
- [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
- [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] 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
- [22] 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, 4