

# On interval prediction of COVID-19 development in France based on a SEIR epidemic model

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

**Abstract**—This paper aims to identify the parameters of an original modified SEIR model for the COVID-19 epidemic's course in France by using publicly available data and applying such an identified model 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]; it is prevalent 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  $S$ , 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  $R$ , 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 of a small number of parameters to identify (three transition rates in Fig. 1:  $\sigma$ ,  $\gamma$ , and  $b$ ). These model characteristics are an essential issue in a virus attack with a limited amount of data available. In March 2020, that was the case in France under the appearance 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 paper proposes a new modified discrete-time SEIR model for the COVID-19 epidemic's evolution in France. Our model is inspired by the one proposed in [3] used for the epidemic's trend of COVID-19 in China. Other similar SIR/SEIR-type models used recently for modeling SARS-CoV-2 virus are [4]–[7]. The selected model is proposed<sup>1</sup> as follows (we do 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):

<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>1</sup>Compared to the model in [3], we do not consider the inflow/outflow variables for each state in our analysis.

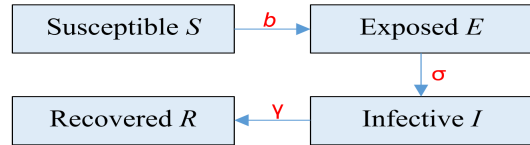


Fig. 1. A schematic representation of SEIR model

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

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

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

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

where  $t \in \mathbb{N}$  (the set of non-negative integers) is the time counted in days ( $t = 0$  corresponds to the beginning of measurements or prediction),  $N \in \mathbb{N}$  denotes the total population, the parameter  $\gamma > 0$  represents the recovery rate (from which the mortality rate can be deduced),  $b > 0$  corresponds to the infection rate of the virus transmission from infectious/exposed to susceptibles during a contact,  $\sigma > 0$  is the incubation rate by which the exposed develop symptoms and  $p_C > 0$  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 reduced). The parameters  $\gamma$ ,  $b$ , and  $\sigma$  are constant. Finally,  $p_C \leq r_t < +\infty$  is the number of contacts per person per day for the exposed population  $E$  (in confinement and depending on its severity,  $r_t$  is time-varying).

Therefore, the SEIR model (1) has only three parameters to be identified:  $\sigma$ ,  $\gamma$ , and  $b$ , representing the rate of change between the states  $E$  to  $I$ ,  $I$  to  $R$  and  $S$  to  $E$  (as 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, many research works devoted to the estimation and identification for SIR/SEIR models were developed by now, and several in the last few years, like [8]–[12], to mention a few.

Since the measured data and parameters contain numerous uncertainties and perturbations, it is not easy to make a

reasonable prediction based on the simulation of such a model with fixed parameter values (also taking into account the model simplicity and generality). However, the interval predictor and observer framework [13]–[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 been applied to different SEIR models (see, *e.g.*, [18]–[20]). In this paper, we apply the interval predictor method for our SEIR model (1) to improve its forecasting quality.

*Remark 1.* It is worth highlighting 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]–[7], stochastic and agent-based 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 a 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 data quickly during the epidemic development. Contrarily to these approaches, the interval predictor method does not use these additional assumptions on distributions. It has also been proposed to estimate a guaranteed interval inclusion of trajectories with a minimal computational effort by the cost of a more complicated mathematical analysis and design [17].

The outline of this paper is as follows. 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 under the variation of parameters and initial states. Final discussions and remarks are provided in Section VI.

## II. AVAILABLE DATA ON COVID-19 IN FRANCE

The current population in France is  $N = 67064000^2$ . The incubation period  $T_i$ , 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  $\frac{1}{2} \leq \sigma \leq \frac{1}{12}$ . The data available from public sources<sup>3</sup>, and which is used in this work, for the time period between March 12th and July 30th can be found in GitHub.

<sup>2</sup>[www.en.wikipedia.org/wiki/Demographics\\_of\\_France](http://www.en.wikipedia.org/wiki/Demographics_of_France).

<sup>3</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>4</sup> We denote here by  $\mathcal{J}$ ,  $\mathcal{D}$  and  $\mathcal{H}$  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  $\mathcal{J}$ ,  $R$  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<sup>5</sup> as  $5 \leq \alpha \leq 20$ . Formally, such a ratio  $\alpha$  has to be time-varying and different for  $I$  and  $R$ . Due to the strict requirements of France Health Services, in this paper, we take the following hypotheses:

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

*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,  $\alpha_1 = 5$ ,  $\alpha_2 = \alpha_1 + 1$ , so 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  $\mathcal{J}$  and  $\mathcal{D}$  are described in the last footnote (for the data in France, these approaches provide  $\alpha_1 = 2.418$ ). So, by fixing  $\alpha_1$  and  $\alpha_2$ , the two variables of the model (1),  $I$  and  $R$ , are available from the beginning of the epidemics via (2). We also select an average value for the incubation rate,  $\sigma = \frac{1}{7}$ , to simplify further identification (the variation in this value is taken into account later in the interval predictor).

## III. NUMERICAL SIMULATION 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:

$$0.0721 \leq \gamma \leq 0.238, \quad 0.05068 \leq b \leq 0.05429,$$

$$p_C = 3 \text{ (number of contacts in quarantine),}$$

$$p_N = 15 \text{ (number of contacts in normal mode),}$$

$$p_R = 10 \text{ (number of contacts in relaxed quarantine).}$$

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

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

which is related with smaller population density in France, in comparison to the aforementioned Chinese provinces.

In [4], the theory of a cyclic application of quarantine regimes of different severity is evaluated for COVID-19.

<sup>4</sup>Data from France.

<sup>5</sup>See, for example, [Coronainks-3-19-20](https://coronainks-3-19-20.com), or a dedicated analysis in CMMID or University of Melbourne.

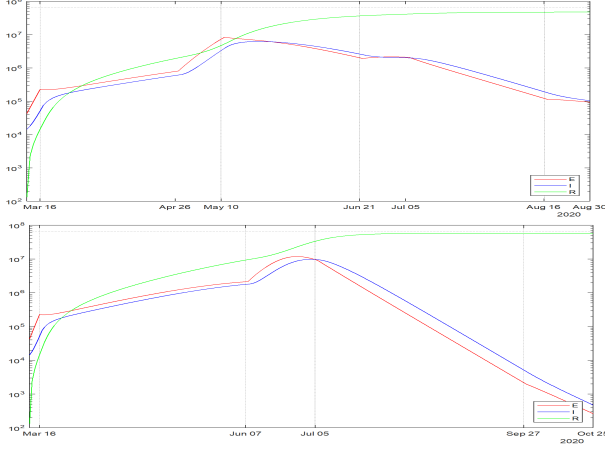


Fig. 2. Simulation results with parameters of [3] for scenarios 1 and 2

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 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.

*Remark 2.* For the chosen model, these scenarios impact only  $r_t$ . 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. 2 (for a better visibility, all populations are plotted in the logarithmic scale in this work). 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. 3 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  $t = 0, 1, \dots, J$ .

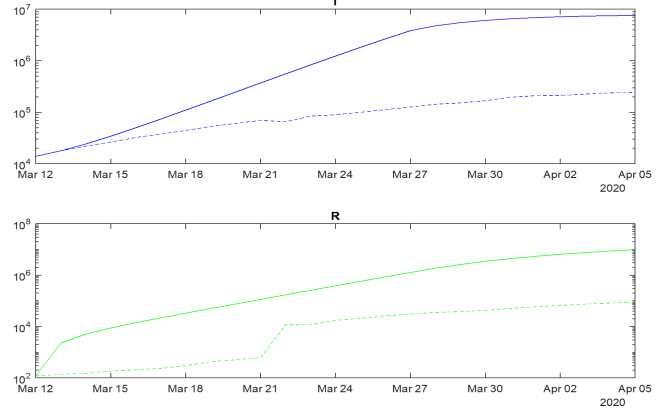


Fig. 3. Results of verification with parameters of [3]

From (1d) we can identify the value of the parameter  $\gamma$  (for convenience, we write  $\gamma$  for its estimate as well):

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

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, (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  $S_t = N - I_t - R_t - E_t$ . From (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. (4). The identified values for  $\gamma$  and  $b$  are not included in the confidence intervals obtained for China in [3] (they are

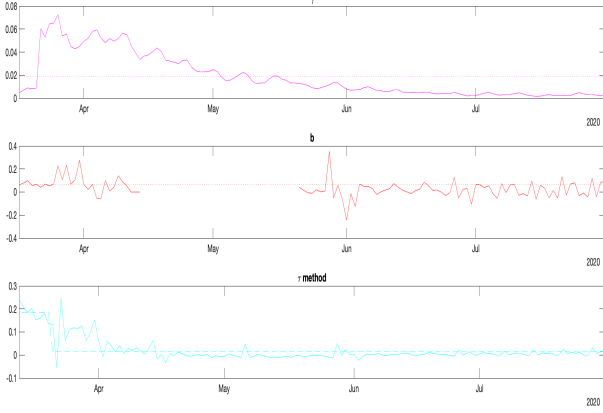


Fig. 4. Identified parameters

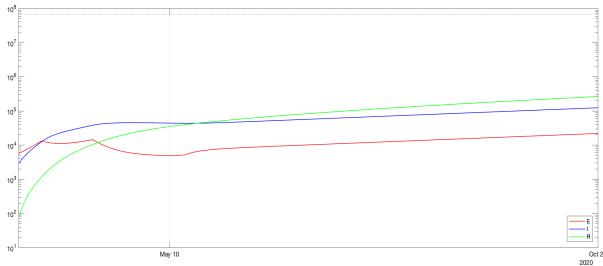


Fig. 5. Simulation results with identified parameters

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.

By the time the final version of this paper was completed, the strict lockdown has ended in France (May 11th). We chose to show in Fig. 5, the simulation results of the model (1) with the identified values of parameters for the actual scenario. Notice that the missing points in the identification of  $b$  is caused by an additional rule of the identification algorithm to stop the calculation of  $b$  if the values of  $p_C$  and  $r_t$  are too small. Verification on the measured and reconstructed data is shown in Fig. 6. The model can approximate the virus propagation reasonably well since these results are more consistent with France's COVID-19 statistics.

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 states 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 evident that the reliability of the obtained prognosis is limited. A way to overcome such weakness is to consider the

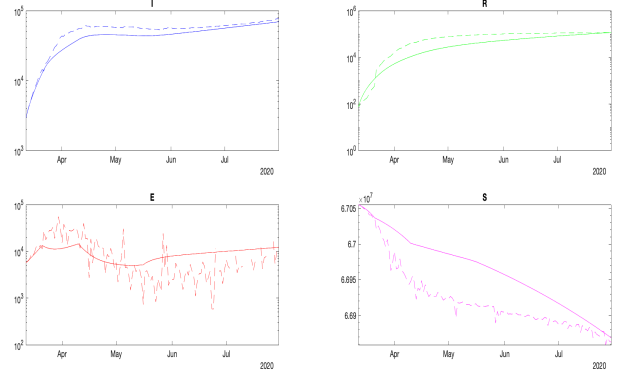


Fig. 6. Results of verification with identified parameters

intervals of admissible values for all variables and parameters used for simulation, so enlarging the model's validity. 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 in a simple scalar system:

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

where  $x_t \in \mathbb{R}_+$  is a non-negative 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  $\underline{x}_0 \leq \bar{x}_0$ ,  $0 \leq \underline{a}_t \leq \bar{a}_t$  and  $\underline{d}_t \leq \bar{d}_t$  are known for all  $t \in \mathbb{N}$ . The imposed non-negativity constraints on  $x_t$  and  $a_t$  correspond to the case of the model (1). We wish 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  $\underline{x}_t \leq x_t \leq \bar{x}_t, \forall t \in \mathbb{N}$ . The theory of interval observers and predictors [17], [22] answers this question, and a possible solution (that utilizes the non-negativity of  $x_t$  and  $a_t$ ) is as follows:

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

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:

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

then it is easy to verify that the terms  $d_t - \underline{d}_t, \bar{d}_t - d_t$  are non-negative 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:

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

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 non-negative 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:

$$\underline{S}_{t+1} = \left( 1 - \underline{b} \frac{(pC\bar{I}_t + \bar{r}_t\bar{E}_t)}{N} \right) \underline{S}_t, \quad (5)$$

$$\underline{E}_{t+1} = \left( 1 - \bar{\sigma} + \underline{b} \frac{r_t}{N} \underline{S}_t \right) \underline{E}_t + pC\underline{b} \frac{I_t \underline{S}_t}{N},$$

$$\underline{I}_{t+1} = (1 - \bar{\gamma}) \underline{I}_t + \underline{\sigma} \underline{E}_t,$$

$$\underline{R}_{t+1} = \underline{R}_t + \underline{\gamma} \underline{I}_t,$$

$$\bar{S}_{t+1} = \min \left\{ N, \left( 1 - \underline{b} \frac{(pC\underline{I}_t + r_t \underline{E}_t)}{N} \right) \bar{S}_t \right\},$$

$$\bar{E}_{t+1} = \min \left\{ N, \left( 1 - \underline{\sigma} + \bar{b} \frac{\bar{r}_t}{N} \bar{S}_t \right) \bar{E}_t + pC\bar{b} \frac{\bar{I}_t \bar{S}_t}{N} \right\},$$

$$\bar{I}_{t+1} = \min \left\{ N, (1 - \underline{\gamma}) \bar{I}_t + \bar{\sigma} \bar{E}_t \right\},$$

$$\bar{R}_{t+1} = \min \left\{ N, \bar{R}_t + \bar{\gamma} \bar{I}_t \right\},$$

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

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

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

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

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

$$\underline{b} \frac{(pC\underline{I}_t + r_t \underline{E}_t)}{N} \leq \underline{b} \frac{(pC\underline{I}_t + r_t \underline{E}_t)}{N} \leq \bar{b} \frac{(pC\bar{I}_t + \bar{r}_t \bar{E}_t)}{N},$$

$$\underline{b} \frac{r_t}{N} \underline{S}_t - \bar{\sigma} \leq \underline{b} \frac{r_t}{N} \underline{S}_t - \bar{\sigma} \leq \bar{b} \frac{\bar{r}_t}{N} \bar{S}_t - \underline{\sigma},$$

$$pC\underline{b} \frac{I_t \underline{S}_t}{N} \leq pC\underline{b} \frac{I_t \underline{S}_t}{N} \leq pC\bar{b} \frac{\bar{I}_t \bar{S}_t}{N},$$

$$\underline{\sigma} \underline{E}_t \leq \underline{\sigma} \underline{E}_t \leq \bar{\sigma} \bar{E}_t, \underline{\gamma} \underline{I}_t \leq \underline{\gamma} \underline{I}_t \leq \bar{\gamma} \bar{I}_t$$

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

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

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

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

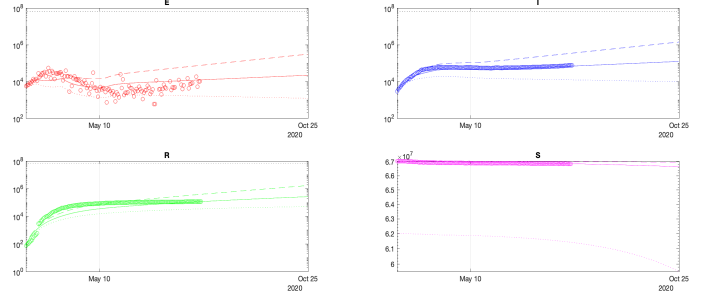


Fig. 7. Simulation results of (5) under  $\pm 10\%$  variation of all parameters

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

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

are satisfied for some  $t \in \mathbb{N}$ , then they also hold for  $t = t + 1$  in (5). 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 non-negativity 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 twice higher than in the system (1). The values of the variables  $\underline{S}_t, \bar{S}_t$  can be evaluated using the population equation  $S_t + E_t + I_t + R_t = N$ :

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

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.

The simulation results of the interval predictor (5) with  $\delta = 10\%$  are presented in Fig. 7 (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). As we can conclude from this curve, under sufficiently significant deviations of the parameters (which correspond to the amount of data publicly available now), the lockdown slows 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, corresponding to the lower bounds of  $I$  and  $E$ , and pessimistic presented by the respective upper bounds.

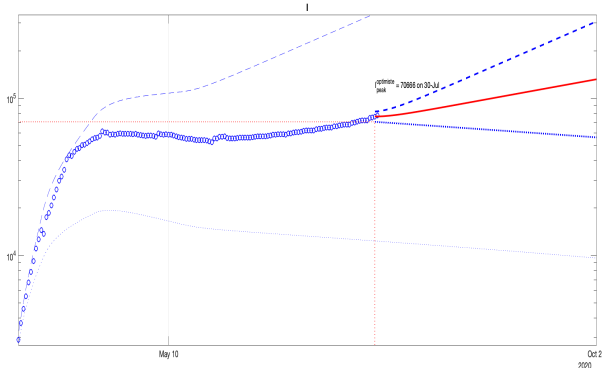


Fig. 8. Prediction of the growth of  $I$  for scenario 1 with  $p_C = 2$  or  $p_C = 1$  under deviations of values of all parameters

A further precision of the model and the parameters is needed, but as a direction after these preliminary simulations is that in March, an *augmentation* of the quarantine's severity was desirable. This suggestion is illustrated by Fig. 8 that presents the interval prediction for the infectious population  $I$  in the actual scenario with a 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. This graphic corresponds to the case given previously in Fig. 7 for  $p_C = 0.5$ . In the optimistic scenario, the confinement constraints the virus, which is an important achievement representing a significant decay of the load to the public health services.

## VI. CONCLUSION

A novel simple discrete-time SEIR epidemic model was identified and used to predict the quarantine's influence on the SARS-CoV-2 virus propagation in France. 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 uncertainty level, a more strict as possible confinement seems to be advisable.

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

## REFERENCES

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

- [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," covid-19 reports, 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, 2020.
- [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.
- [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.
- [7] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," *arXiv e-prints*, 2020.
- [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.
- [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.
- [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.
- [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.
- [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)*, pp. 1170–1175, June 2019.
- [13] J. Gouzé, A. Rapaport, and M. Hadj-Sadok, "Interval observers for uncertain biological systems," *Ecological Modelling*, vol. 133, pp. 46–56, 2000.
- [14] F. Mazenc and O. Bernard, "Interval observers for linear time-invariant systems with disturbances," *Automatica*, vol. 47, no. 1, pp. 140–147, 2011.
- [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.
- [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.
- [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.
- [18] K. H. Degue, D. Efimov, and A. Iggidr, "Interval estimation of sequestered infected erythrocytes in malaria patients," in *2016 European Control Conference (ECC)*, pp. 1141–1145, June 2016.
- [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.
- [20] K. H. Degue and J. Le Ny, "An interval observer for discrete-time SEIR epidemic models," in *2018 Annual American Control Conference (ACC)*, pp. 5934–5939, June 2018.
- [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.
- [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.