



HAL
open science

A prediction of COVID-19 development in France based on a modified SEIR epidemic model

Denis Efimov, Rosane Ushirobira

► To cite this version:

Denis Efimov, Rosane Ushirobira. A prediction of COVID-19 development in France based on a modified SEIR epidemic model. [Research Report] Inria Lille Nord Europe - Laboratoire CRISTAL - Université de Lille. 2020. hal-02517866v2

HAL Id: hal-02517866

<https://inria.hal.science/hal-02517866v2>

Submitted on 27 Mar 2020 (v2), 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.

A prediction of COVID-19 development in France based on a modified SEIR epidemic model

Denis Efimov¹ and Rosane Ushirobira¹

Abstract—The goal of this brief report is to identify the parameters of a modified SEIR model for the epidemics course of COVID-19 in France, and to use such an identified model for the prediction of the virus propagation under different conditions of confinement. An interval predictor is also designed, allowing to take into account variations and uncertainties in the model parameters.

I. INTRODUCTION

The SEIR model is one of the most elementary compartmental model of epidemics [1]. It describes the evolution of the relative proportions of four classes of individuals in a population of constant size, 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 susceptibles; and the recovered R , permanently immune (after healing or dying). Such a simple model represents well a generic behavior of epidemics, and a related advantage consists in a small number of parameters to identify. 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 current¹ situation in France under the presence of the SARS-CoV- 2 virus.

There are many types and variations of SEIR models [1]. This note is based on the modified SEIR discrete-time model proposed in [2], where it has been used to model the epidemics trend of COVID-19 in China. The selected model is proposed² 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):

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

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

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

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

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 to susceptibles, $0 < \sigma < +\infty$ is the incubation rate by which the exposed develop symptoms, $0 < pc < +\infty$ 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), $pc \leq r_t < +\infty$ 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).

Roughly speaking, the SEIR model (1) has just three parameters to be identified: σ , γ and b . These parameters σ , γ and b represent, respectively, the rate of change between the states E to I , I to R and S to $E + I$. 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. The numbers of contacts pc 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.

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). However, the interval predictor and observer framework [3], [4] allows a set of trajectories to be obtained corresponding to 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. [5], [6]). In this report we apply the interval predictor method for the considered model (1) to improve its forecasting quality.

Furthermore, in this note we describe the measured data used for parameter identification in Section II, together with some hypotheses used in the sequel. The parameters obtained in [2] 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 variation of parameters.

¹ Denis Efimov and Rosane Ushirobira are with Inria, Villeneuve d'Ascq, France and CRIStal (UMR CNRS 9189), Villeneuve d'Ascq, France Denis.Efimov, Rosane.Ushirobira@inria.fr

²This note was initially written on March 23rd and updated March 27, 2020.

³Compared to the model in [2], the inflow/outflow variables for each state were not considered in our analysis.

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	29567	1698	4955

TABLE I
DATA FOR FRANCE

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 [2], or in a more specialized research between 2 and 12 days [7], so we assume

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

The data available from public sources⁴, and which is used in this note, for the time period between March 12th and 26th is provided in the Table I⁵, 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 α with the estimates from different sources⁶:

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

Formally, such an α 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, 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:

$$\alpha_1 = 5, \alpha_2 = 6;$$

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

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

⁵Sources: Efiens, Johns Hopkins University.

⁶See for example Coronalinks-3-19-20, or a dedicated analysis in CMMID.

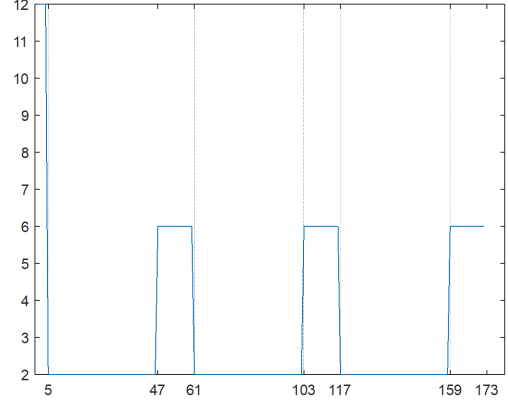


Fig. 1. Variation of the number of contacts r_t in the first scenario

so 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).

III. NUMERICAL EXPERIMENTS WITH PARAMETERS OF [2]

In [2], 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.

Following the ideas presented in [8], for simulation we consider 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.

These scenarios will impact only the variable r_t , the corresponding example of behavior of the number of contacts for asymptomatic infectious is shown in Fig. 1.

Remark 1. 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.

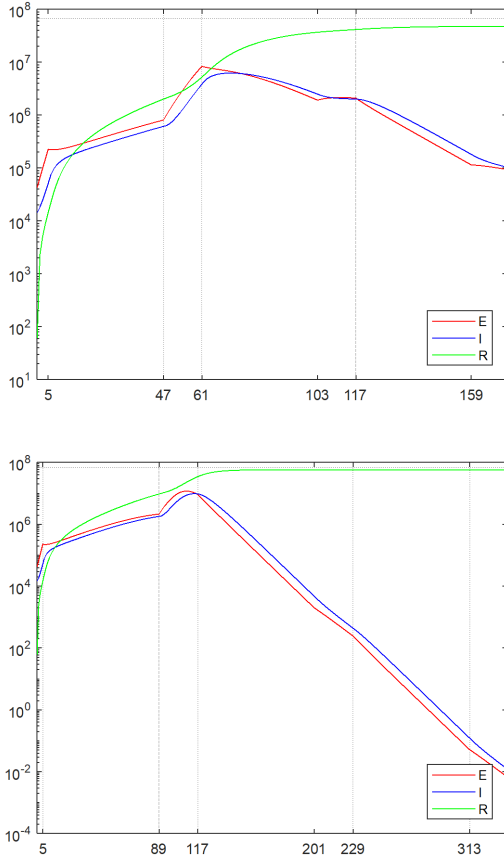


Fig. 2. The results of simulation with parameters of [2] for scenarios 1 and 2

The results of simulation of the model (1) with the described above parameters for both scenarios 1 and 2 are given in Fig. 2. As we can conclude from these results, both scenarios of confinement do not lead to a stabilization of COVID-19 development in France (the 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, a short analysis on how close the obtained curves are to the measured data (see Fig. 3 where dashed lines correspond to the measurements), then we recognize that the model with these parameters from [2] does not correspond well 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 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 $t = 0, 1, \dots, J$.

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

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

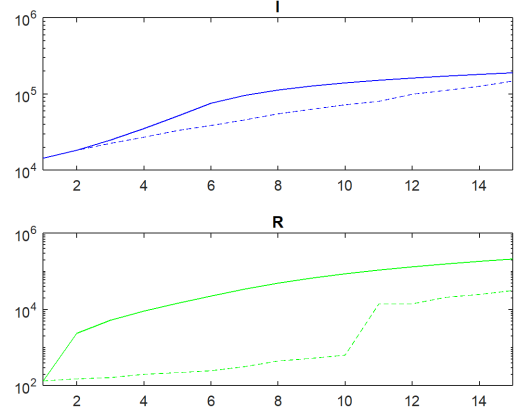


Fig. 3. The results of verification with parameters of [2]

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. 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):

$$\gamma = \frac{1}{K+1} \sum_{k=0}^K \gamma_k = 0.0387.$$

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 and r_i):

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

whose least squares estimation is

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

then the identified value is again the average of these estimates:

$$b = \frac{1}{K+1} \sum_{k=0}^K b_k = 0.0360.$$

The identified values for γ and b lie in the confidence intervals obtained for China in [2].

The simulation results of the model (1) with the identified parameters for both scenarios 1 and 2 are given in Fig. 4, a verification on the measured and reconstructed data is shown in Fig. 5 (in this case the model approximates reasonably well

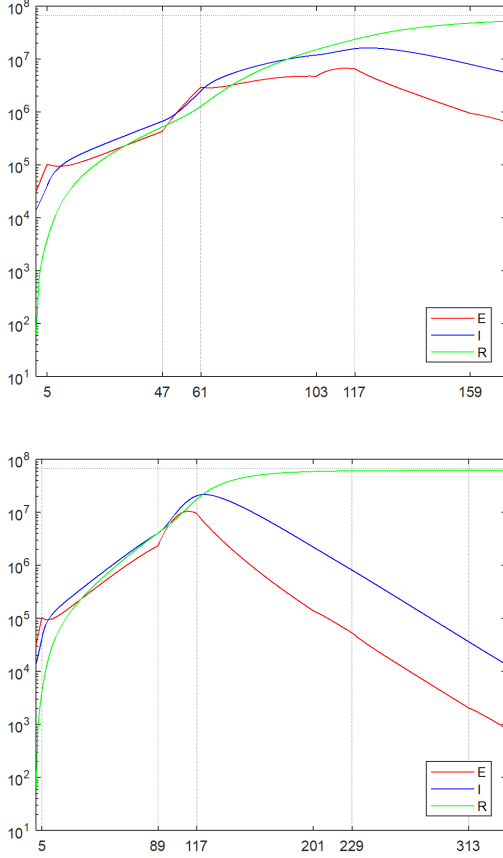


Fig. 4. The results of simulation with identified parameters

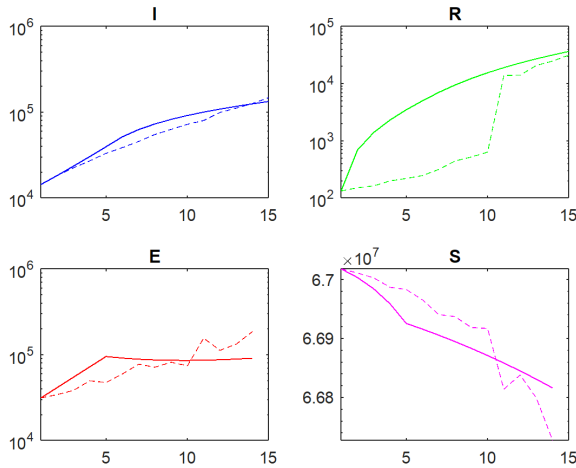


Fig. 5. The results of verification with identified parameters

the virus propagation). These results are more consistent with the available statistics for France, and also demonstrate the lack of efficiency of the confinement (even of a long length).

Let us enlarge the validity of the model by considering intervals of admissible values for parameters.

V. INTERVAL PREDICTION

The idea of 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 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 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 [4] answers this question, and a possible solution (that utilizes the non-negativity 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. 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:

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

together with the initial conditions in (1):

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

where all 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 values 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:

$$\begin{aligned} \underline{S}_{t+1} &= \max \left\{ 0, 1 - \bar{b} \frac{(p_c \bar{I}_t + \bar{r}_t \bar{E}_t)}{N} \right\} \underline{S}_t, \\ \underline{E}_{t+1} &= \left(1 - \bar{\sigma} + \underline{b} \frac{\underline{r}_t}{N} \underline{S}_t \right) \underline{E}_t + p_c \underline{b} \frac{\underline{I}_t \underline{S}_t}{N}, \\ \underline{I}_{t+1} &= (1 - \bar{\gamma}) \underline{I}_t + \underline{\sigma} \underline{E}_t, \\ \underline{R}_{t+1} &= \underline{R}_t + \underline{\gamma} \underline{I}_t, \\ \bar{S}_{t+1} &= \min \left\{ N, \max \left\{ 0, 1 - \underline{b} \frac{(p_c \underline{I}_t + \underline{r}_t \underline{E}_t)}{N} \right\} \bar{S}_t \right\}, \\ \bar{E}_{t+1} &= \min \left\{ N, \left(1 - \underline{\sigma} + \bar{b} \frac{\bar{r}_t}{N} \bar{S}_t \right) \bar{E}_t + p_c \bar{b} \frac{\bar{I}_t \bar{S}_t}{N} \right\}, \\ \bar{I}_{t+1} &= \min \left\{ N, (1 - \underline{\gamma}) \bar{I}_t + \bar{\sigma} \bar{E}_t \right\}, \\ \bar{R}_{t+1} &= \min \left\{ N, \bar{R}_t + \bar{\gamma} \bar{I}_t \right\}, \end{aligned}$$

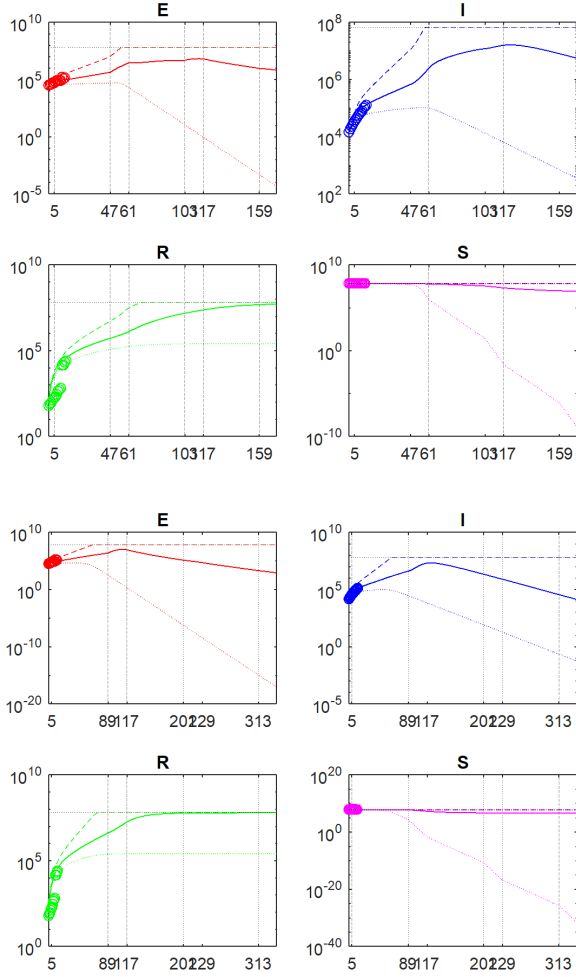


Fig. 6. The results of simulation of (3) for scenarios 1 and 2 under $\pm 15\%$ variation of all parameters

where for all $t \in \mathbb{N}$ we obtain the interval inclusions for the state of (1):

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

and $\underline{S}_t, \bar{S}_t, \underline{I}_t, \bar{I}_t, \underline{E}_t, \bar{E}_t$ and $\underline{R}_t, \bar{R}_t$ are the interval lower and upper bound predictions.

Remark 2. The boundedness of these variables are not discussed in this note.

The results of simulation of this interval predictor (3) with $\delta = 15\%$ are presented in Fig. 6 (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 $[\underline{\sigma}, \bar{\sigma}] = [\frac{1}{12}, \frac{1}{2}]$ are shown in Fig. 7 (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 this parameter and how significant its deviation for the COVID-19 trend.

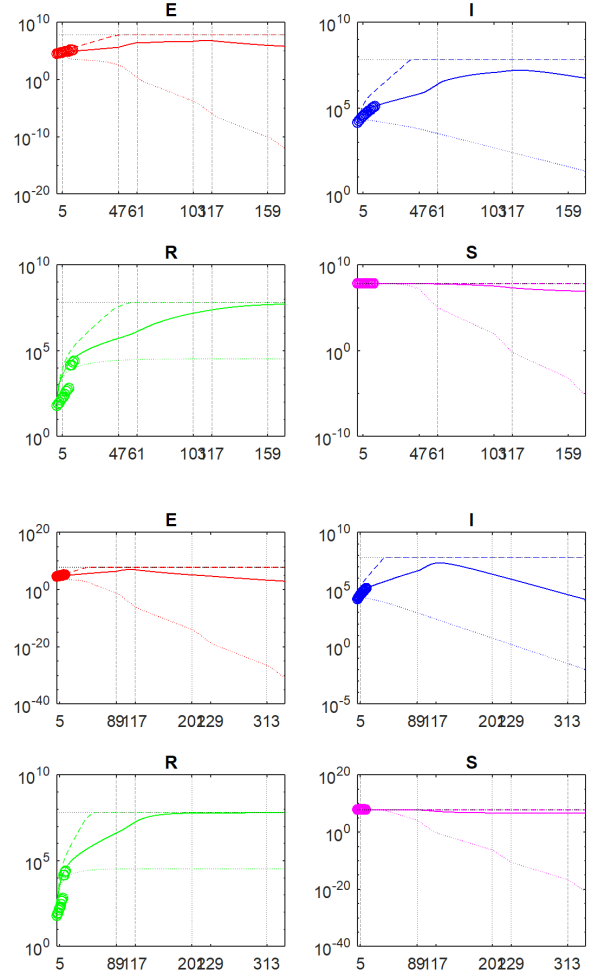


Fig. 7. The results of simulation of (3) for scenarios 1 and 2 under variation of $\sigma \in [\frac{1}{12}, \frac{1}{2}]$

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 be efficient. The measurements are nearly included in the obtained interval predictions. 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 is needed, but as a recommendation after these preliminary simulations: an augmentation of severity of the quarantine is desirable.

VI. 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. To enlarge the model prediction performance an interval predictor was also used to analyze the COVID-19 course. The prediction showed

that a longer confinement may be a little bit more efficient, but under the current level of uncertainty, a more strict as possible confinement seems to be advisable.

Machine learning tools can be further used to identify and optimize time profile for the confinement.

REFERENCES

- [1] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008. I
- [2] 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, I, II, III, 2, III, 3, IV
- [3] J. Gouzé, A. Rapaport, and M. Hadj-Sadok, "Interval observers for uncertain biological systems," *Ecological Modelling*, vol. 133, pp. 46–56, 2000. I
- [4] 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, V
- [5] 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
- [6] 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
- [7] 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
- [8] 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. III