

RESEARCH ARTICLE

 OPEN ACCESS

Assessment of COVID-19 Hospitalization Forecasts from a Simplified SIR Model

P.-A. Absil^a, Ousmane Diao^a, Mouhamadou Diallo^b

^aICTEAM Institute, UCLouvain, B-1348 Louvain-la-Neuve, Belgium; ^bMolecular Biology Unit/Bacteriology-Virology Lab, CNHU A. Le Dantec / Université Cheikh Anta Diop, Dakar, Sénégal

ABSTRACT

We propose the SH model, a simplified version of the well-known SIR compartmental model of infectious diseases. With optimized parameters and initial conditions, this time-invariant two-parameter two-dimensional model is able to fit COVID-19 hospitalization data over several months with high accuracy (e.g., the root relative squared error is below 10% for Belgium over the period from 2020-03-15 to 2020-07-15). Moreover, we observed that, when the model is trained on a suitable three-week period around the first hospitalization peak for Belgium, it forecasts the subsequent two months with mean absolute percentage error (MAPE) under 4%. We repeated the experiment for each French department and found 14 of them where the MAPE was below 20%. However, when the model is trained in the increase phase, it is less successful at forecasting the subsequent evolution.

ARTICLE HISTORY

Received August 24, 2020
Accepted September 17, 2021

KEYWORDS

COVID-19 prediction,
COVID-19 forecast,
SARS-CoV-2,
coronavirus,
SIR model,
hospitalization prediction

1 Introduction

Compartmental models, and in particular the well-known SIR model (Kermack et al., 1927), have been widely used to model infectious diseases since the early 20th century (see Hethcote, 2000, and references therein). Letting $S(t)$, $I(t)$, and $R(t)$ denote the number of susceptible, infectious and removed (or recovered) individuals at time t , and letting $\dot{S}(t)$, $\dot{I}(t)$, and $\dot{R}(t)$ denote their time derivatives, the SIR model consists in the following three-dimensional continuous-time autonomous dynamical system

$$\dot{S}(t) = -\frac{\beta}{N}S(t)I(t) \quad (1a)$$

$$\dot{I}(t) = \frac{\beta}{N}S(t)I(t) - \gamma I(t) \quad (1b)$$

$$\dot{R}(t) = \gamma I(t), \quad (1c)$$

where $N = S(t) + I(t) + R(t)$ is the constant total population and β and γ are parameters. The SIR model, and several (sometimes deep) variations thereof, have been applied in several works to model the COVID-19 dynamics (see, e.g., Liu et al., 2020; Atkeson, 2020; Calafiore et al., 2020; Fanelli and Piazza, 2020; Carletti et al., 2020; Kozyreff, 2021; Nesterov, 2020b,a; Djidjou-Demasse et al., 2020; Abrams et al., 2021) with known limitations (Roda et al., 2020; Baek et al., 2020; Wang and Flessa, 2020; Comunian et al., 2020). In (Bhanot and DeLisi, 2020), an SIR-like model is used to make long-term forecasts. However, at the time of writing this paper, it appears that studies are still rare (see Singh and Mohapatra, 2020; Ghosh and Chakraborty, 2020; Ghanam et al., 2020) where the SIR-like model parameters and initial conditions are learned on a “train” part of the available data in order to predict a subsequent “test” part, making it possible to assess the forecast accuracy of the model.

In this paper, we adapt the SIR model to the situation where (i) S , I and R are hidden variables but $I(t)$ is observed through a “proxy” $H(t) = \alpha I(t)$, where α is unknown but constant, and (ii) not only β and γ but also the total concerned population N are unknown. In the context of the COVID-19 application, the proxy H will be the total number of lab-confirmed hospitalized patients; the equation $H(t) = \alpha I(t)$ thus posits that a constant fraction of the infected people is hospitalized. The proposed adapted SIR model, which we term *SH model*, is given in (8). It has two state variables (\tilde{S} —a scaled version S —and H) and two parameters ($\tilde{\beta}$ —which lumps together the parameters β , N , and α —and γ).

We leverage the proposed SH model as follows in order to make forecasts regarding hospitalization. Given observed values $(H_o(t))_{t=t_s, \dots, t_e}$, we estimate the parameters $\tilde{\beta}$, γ , and the initial conditions $\tilde{S}(t_i)$ and $H(t_i)$ of the SH model. Then we simulate the SH model in order to predict $(H(t))_{t=t_e+1, \dots, t_f}$ for a specified final prediction time t_f . This approach thus relates to the areas of parameter estimation (for obvious reasons), data assimilation (for the generation of the initial conditions) and machine learning (for the train-test approach).

The paper is organized as follows. The available data is reviewed in Section 2. The SH model is derived in Section 3. Several methods to estimate the parameters and initial conditions are presented in Sections 4 and 5. Experimental results are reported in Section 6 and conclusions are drawn in Section 7.

2 Data and Notation

In the experiments (Section 6), we will use a COVID-19 dataset for Belgium¹ that provides us with the following data for $t = t_s, \dots, t_e$, where t_s is 2020-03-15 and t_e is 2021-05-15:

- $H_o(t)$: total number of lab-confirmed hospitalized (i.e., at the hospital) COVID-19 patients on day t , including intensive care unit (TOTAL_IN column);
- $E_o(t)$: number of COVID-19 patients entering the hospital (i.e., number of lab-confirmed hospital intakes) on day t (NEW_IN column);
- $L_o(t)$: number of COVID-19 patients discharged from the hospital on day t (NEW_OUT column).

The subscript $_o$ stands for “observed”.

We will also mention results obtained with a dataset for France² where t_s is 2020-03-18 and t_e is 2021-05-16.

2.1 Discussion

In the data for Belgium, there is a mismatch between $H_o(t)$ and $H_o(t-1) + E_o(t) - L_o(t)$ for most t , and $H_o(t_s) + \sum_{t=t_s+1}^{t_e} E_o(t) - L_o(t)$ is significantly larger than $H_o(t_e)$. This can be due to the patients who get infected at the hospital (they would be counted in H_o without appearing in E_o) and to the patients who die at the hospital (they would be removed from H_o without appearing in L_o). In order to remedy this mismatch, we redefine $L_o(t)$ by $L_o(t) := -H_o(t) + H_o(t-1) + E_o(t)$.

For the French data, we sum the “rad” (daily number of new home returns) and “dc” (daily number of newly deceased persons) columns to get $L_o(t)$. Since there is no column for E_o , we define $E_o(t) = H_o(t) - H_o(t-1) + L_o(t)$.

Several other COVID-19-related data are available for Belgium and France. In particular, the daily number of infected individuals, $I_o(t)$, is also reported by health authorities. However, the decision process to admit individuals at the hospital is believed to be more stable over time than the decision process to test individuals, hence the hospitalization data is a priori more prone to be modeled by a dynamical system with time-invariant parameters. Moreover, for the authorities, predicting H is more crucial than predicting I . Therefore, as in (Kozyreff, 2021), we focus on H .

3 Models

3.1 Case Hospitalization Ratio

We assume that, for all t ,

$$H(t) = \alpha I(t) \quad (2)$$

where α is unknown but constant over time. As already mentioned, (2) posits that a constant fraction of the infected people is hospitalized.

Equation (2) is reminiscent of Carletti et al. (2020, eq. (3)), where the number of dead individuals plays the role of H and α is time dependent.

¹https://epistat.sciensano.be/Data/COVID19BE_HOSP.csv obtained from <https://epistat.wiv-isp.be/covid/>

²Data file `donnees-hospitalieres-covid19-2021-05-16-19h08.csv` was obtained from the website <https://www.data.gouv.fr/en/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/>. Department 46 was removed from the data due to abnormal values in the “rad” column.

3.2 Observation Models

We assume the following observation models with additive noise:

$$H_o(t) = H(t) + \varepsilon_H(t) \quad (3a)$$

$$E_o(t) = E(t) + \varepsilon_E(t) \quad (3b)$$

$$L_o(t) = L(t) + \varepsilon_L(t). \quad (3c)$$

The variables with the subscript $_o$ stand for the values provided in the datasets; see Section 2. The variables without the subscript stand for the values given by the forthcoming model. The ε variables account for the discrepancies between the former and the latter.

We do not make any explicit statistical assumption on the ε variables in this work. However, we point out that if we (questionably) assume them to be independent Gaussian centered random variables, then some subsequent estimators admit a maximum likelihood interpretation.

3.3 Proposed SH Model

Multiplying (1a) and (1b) by α , and further multiplying the numerators and denominators by α , we obtain the equivalent system

$$\alpha \dot{S}(t) = -\frac{\beta}{N\alpha} \alpha S(t) \alpha I(t) \quad (4)$$

$$\alpha \dot{I}(t) = \frac{\beta}{N\alpha} \alpha S(t) \alpha I(t) - \gamma \alpha I(t). \quad (5)$$

Letting

$$\bar{S} := \alpha S \quad (6)$$

$$\bar{\beta} := \frac{\beta}{N\alpha} \quad (7)$$

and using (2), we obtain the simplified SIR model

$$\dot{\bar{S}}(t) = -\bar{\beta} \bar{S}(t) H(t) \quad (8a)$$

$$\dot{H}(t) = \bar{\beta} \bar{S}(t) H(t) - \gamma H(t) \quad (8b)$$

which we term the *SH model*. (The “S” in this SH model can be interpreted as the number of individuals susceptible of being hospitalized.) The SH model has only two parameters ($\bar{\beta}$ and γ), one hidden state variable (\bar{S}) and one observed state variable (H) with observation model (3a).

Note that, in the SH model (8), the number of patients entering the hospital by unit of time is

$$E(t) := \bar{\beta} \bar{S}(t) H(t) \quad (9)$$

and the number of patients leaving the hospital by unit of time is

$$L(t) := \gamma H(t). \quad (10)$$

4 Estimation and Prediction Method

The goal is now to leverage the SH model (8) in order to predict future values of H based on its past and current observations $(H_o(t))_{t=t_i, \dots, t_c}$. To this end, we have to estimate (or “learn”) four variables, which we term *estimands*: the two parameters $\bar{\beta}$ and γ and the two initial values $\bar{S}(t_i)$ and $H(t_i)$, where t_i is the chosen initial time for the SH model (8). One possible approach is to minimize some error measure between the simulated values $(H(t))_{t=t_i, \dots, t_c}$ and the observed values $(H_o(t))_{t=t_i, \dots, t_c}$ as a function of the four estimands. However, the error measure is not available as a closed-form expression of the four estimands, and this makes this four-variable optimization problem challenging. We show in this section that it is possible to estimate $H(t_i)$ and γ separately. This leaves us with an optimization problem in the two remaining estimands $\bar{\beta}$ and $\bar{S}(t_i)$, making it possible to visualize the objective function by means of a contour plot.

4.1 Train and Test Sets

To recap, we have $t_s \leq t_i < t_c < t_e$. The provided dataset goes from t_s to t_e . The *test set*, $(H_o(t), E_o(t), L_o(t))_{t \in [t_{c+1}, t_e]}$, is not available to the forecasting algorithm; it is revealed only to compute the forecasting error (see Section 6.1). The SH model is initialized at t_i , and we refer to the data $(H_o(t), E_o(t), L_o(t))_{t \in [t_i, t_c]}$ as the *train set*, though it is legitimate to widen it to $t \in [t_s, t_c]$.

4.2 Estimation of $H(t_i)$

We simply take

$$H(t_i) := H_o(t_i).$$

In view of (3a), this amounts to $\varepsilon_H(t_i) = 0$.

We could have chosen instead $H(t_i)$ as a filtered version of $H_o(t_i)$ in order to reduce the impact of the weekly variations in $H_o(t)$. Those weekly variations, visible in Figure 1 in the form of equally spaced ripples, are due to the fact that fewer patients are discharged during the weekend. However, such a filter did not yield better forecasts in our experiments, and we thus opted for the above-mentioned $H(t_i) := H_o(t_i)$.

4.3 Estimation of γ

We have $L(t) = \gamma H(t)$, see (10). In view of the observation model (3), we can estimate γ by a ratio of means:

$$\hat{\gamma}^{\text{RM}} = \frac{\sum_{t=t_i}^{t_c} L_o(t)}{\sum_{t=t_i}^{t_c} H_o(t)}. \quad (11)$$

A theoretical justification of the choice of the ratio-of-means estimator can be found in (Schoukens and Pintelon, 1991); or see (Schoukens, 2018, §1.2).

Note that t_i in the expression of $\hat{\gamma}$ can legitimately be replaced by any time between t_s and t_c . Only data in the test set, i.e., occurring after t_c , are unavailable in the variable estimation phase.

4.4 Joint Estimation of $\tilde{\beta}$ and $\tilde{S}(t_i)$

Now we have to estimate the two remaining estimands, namely $\tilde{\beta}$ and $\tilde{S}(t_i)$. We choose the following sum-of-squared-errors objective function

$$\phi(\tilde{\beta}, \tilde{S}(t_i)) = c_H \sum_{t=t_i}^{t_c} (H(t) - H_o(t))^2 + c_E \sum_{t=t_i}^{t_c} (E(t) - E_o(t))^2 + c_L \sum_{t=t_i}^{t_c} (L(t) - L_o(t))^2, \quad (12)$$

where the c coefficients are parameters, all set to 1 in our experiments unless otherwise stated. In (12), $H(t)$, $E(t)$ (see (9)), and $L(t)$ (see (10)) are given by the (approximate) solution of the SH model (8) in which (i) $H(t_i)$ and γ take the values estimated as above, and (ii) $\tilde{\beta}$ and $\tilde{S}(t_i)$ take the values specified in the argument of ϕ .

Several methods exist to compute the required (approximate) solution of the SH model (8); see (Barlow and Weinstein, 2020). Since the method will have to be repeatedly called by the optimization solver, we opt for the simplicity of the explicit Euler method with a time step of one day (as in, e.g., Vinitsky et al., 2021), yielding, for $t = t_i, \dots, t_c - 1$,

$$\tilde{S}(t+1) = \tilde{S}(t) - \tilde{\beta} \tilde{S}(t) H(t) \quad (13a)$$

$$H(t+1) = H(t) + \tilde{\beta} \tilde{S}(t) H(t) - \gamma H(t). \quad (13b)$$

As an alternative numerical integration method for (8), we tested SciPy's explicit Runge–Kutta method of order 5(4) with default parameters. In comparison with Euler (13), the forecast error was most often in favor of the Runge–Kutta method, but always by a narrow margin, and the execution time with Runge–Kutta was more than four times the one with Euler. For these reasons, we decided to retain (13). This choice also leads to simpler developments in Section 5.1.

Now that the objective function ϕ (also termed “cost function” or “loss function”) is defined, we let the estimated $(\tilde{\beta}, \tilde{S}(t_i))$ be the (approximate) minimizer of ϕ returned by some optimization solver. The initial guess that we give to the solver is the outcome of the successive estimation method that we will present in Section 5.1. (If this proposed initial guess is negative, a situation that we observed on rare occasions in the late part of the decrease phase, then we take its opposite.)

4.5 Prediction of H

Recall that the time range between t_i and t_c is the train period and the time range between $t_c + 1$ and t_e is the test period. In order to predict (i.e., forecast) the values of H over the test period, we apply the above procedure to estimate the four estimand variables $\tilde{\beta}$, γ , $\tilde{S}(t_i)$, and $H(t_i)$, and we compute the solution $H(t)$ of (13) for t from t_i to t_e . The prediction is then $(H(t))_{t=t_c+1, \dots, t_e}$. The discrepancy between $(H(t))_{t=t_c+1, \dots, t_e}$ and $(H_o(t))_{t=t_c+1, \dots, t_e}$ reveals the accuracy of the prediction.

5 Alternative Estimation and Prediction Methods

5.1 Successive Estimation of $\tilde{\beta}$ and $\tilde{S}(t_i)$

As an alternative to Section 4.4, we now present a method to estimate $\tilde{\beta}$ independently. We do not recommend this alternative, as we have observed that it usually yields less accurate forecasts. However, it provides a convenient initial guess for Section 4.4, and it also sheds light on the various forecast accuracies observed in Section 6.

From (13a) and (9), we obtain

$$\frac{E(t+1)}{H(t+1)} - \frac{E(t)}{H(t)} = -\tilde{\beta}E(t).$$

Summing both sides yields the following estimator of $\tilde{\beta}$:

$$\widehat{\tilde{\beta}} = -\frac{\frac{E_o(t_c)}{H_o(t_c)} - \frac{E_o(t_i)}{H_o(t_i)}}{\sum_{t=t_i}^{t_c-1} E_o(t)}. \tag{14}$$

Finally, in view of (9), we choose the following estimator for $\tilde{S}(t_i)$:

$$\widehat{\tilde{S}}(t_i) = \frac{E_o(t_i)}{\widehat{\tilde{\beta}}H_o(t_i)}.$$

For the observation model (3), under the assumption that $|\varepsilon_H(t)| \leq \varepsilon$ and $|\varepsilon_E(t)| \leq \varepsilon$ for all t , we have the following error bound, where higher order terms in ε are neglected:

$$\left| \frac{\frac{E_o(t_c)}{H_o(t_c)}}{\sum_{t=t_i}^{t_c-1} E_o(t)} - \frac{\frac{E(t_c)}{H(t_c)}}{\sum_{t=t_i}^{t_c-1} E(t)} \right| \lesssim \frac{\varepsilon}{H(t_c) \sum_{t=t_i}^{t_c-1} E(t)} \left(1 + \frac{E(t_c)}{H(t_c)} + \frac{E(t_c)(t_c - t_i)}{\sum_{t=t_i}^{t_c-1} E(t)} \right),$$

and likewise for the second term in (14).

Consequently, the estimation of $\tilde{\beta}$ should be the most accurate when $E_o(t)H_o(t)$ is the largest. This occurs slightly before the peak of $H_o(t)$. This means that the estimation of $\tilde{\beta}$ should be the most accurate for a train period slightly before the peak. However, this does not mean that this position of the train period gives the most accurate forecasts, as we will see below.

Let us consider the situation where the train period is located *before* the peak. Then the estimation of $\tilde{\beta}$ is less accurate, and this impacts $\widehat{\tilde{S}}(t_i)$. At the initial time t_i , this does not impact the right-hand term of (8b) in view of the definition of $\widehat{\tilde{S}}(t_i)$. However, an overestimation of $\tilde{\beta}$ will induce an underestimation of $\tilde{S}(t_i)$ and, in view of (8a), a subsequent even stronger underestimation of $\tilde{S}(t)$. Hence the first term of (8b) will be underestimated. As a consequence, the peak in H will appear sooner and lower. The case of an underestimation of $\tilde{\beta}$ leads to the opposite conclusion, namely a peak in H that appears later and higher. In summary, the further before the peak the train period is located, the more inaccurate the position and height of the peak is expected to be.

Finally, let us consider the situation where the train period is located *after* the peak. Then we can make the same observations as in the previous paragraph, except that predicting the peak is now irrelevant. Moreover, we are in the decrease phase, where the first term of (8b) (which involves $\tilde{\beta}$ and $\tilde{S}(t)$) is smaller than the second term (which does not involve these quantities). Consequently, the possibly large estimation errors on $\tilde{\beta}$ and $\tilde{S}(t)$ will only slightly affect the forecast of $H(t)$.

5.2 Joint Estimation of the Four Estimands

An alternative to Sections 4.2–4.4 is to reconsider (12) as a function of all four estimands:

$$\tilde{\phi}(\tilde{\beta}, \tilde{S}(t_i), \gamma, H(t_i)) = c_H \sum_{t=t_i}^{t_c} (H(t) - H_o(t))^2 + c_E \sum_{t=t_i}^{t_c} (E(t) - E_o(t))^2 + c_L \sum_{t=t_i}^{t_c} (L(t) - L_o(t))^2. \tag{15}$$

In (15), $H(t)$, $E(t)$ (see (9)), and $L(t)$ (see (10)) are given by the solution of the discrete-time SH model (13) where the parameters $\hat{\beta}$ and γ and the initial conditions $\tilde{S}(t_i)$ and $H(t_i)$ take the values specified in the argument of $\tilde{\phi}$. Minimizing $\tilde{\phi}$ is a more challenging problem than minimizing ϕ (12) in view of the larger number of optimization variables. It may be essential to give a good initial guess to the optimization solver, and a natural candidate for this is the values obtained by the procedure described in Sections 4.2–4.4.

In our experiments, we have found that this alternative does not present a clear advantage in terms of the prediction mean absolute percentage error (MAPE). The results reported in Section 6 are obtained with the prediction approach of Section 4, unless otherwise specified.

6 Results

We now apply the method of Section 4 (by default) or Section 5.2 (when specified) to the data of Section 2.

The methods are implemented in Python 3 and run with Anaconda 2019.10. The code to reproduce the results is available from <https://sites.uclouvain.be/absil/2020.05>.

6.1 Accuracy Measures

We will conduct experiments for more than 100 districts with various train and test periods. Since plotting all the results is out of the question, we have to return measures that give an account of the accuracy of the predictions. Among the many possibilities, we will favor the measures defined as follows for an arbitrary period from t_a to t_b :

- Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{t=t_a}^{t_b} (H_o(t) - H(t))^2}{t_b - t_a + 1}}. \quad (16)$$

The numerator corresponds to the first term of the objective function (12). From an optimization viewpoint, it has the advantage of being a smooth function.

- Root relative squared error (RRSE):

$$\text{RRSE} = \sqrt{\frac{\sum_{t=t_a}^{t_b} (H_o(t) - H(t))^2}{\sum_{t=t_a}^{t_b} (H_o(t) - \text{mean}(H_o))^2}} \quad (17)$$

where $\text{mean}(H_o) = \frac{1}{t_b - t_a + 1} \sum_{t=t_a}^{t_b} H_o(t)$. This measure can be interpreted as the ratio between the RMSE of H and the RMSE of $\text{mean}(H_o)$.

- Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{t_b - t_a + 1} \sum_{t=t_a}^{t_b} |H_o(t) - H(t)|. \quad (18)$$

The MAE is considered to have a better interpretability than the RMSE. However, for comparisons between datasets, scale-invariant versions, given next, should be favored.

- Mean absolute scaled error (MASE):

$$\text{MASE} = \frac{\text{MAE}}{\frac{1}{t_b - t_a} \sum_{t=t_a}^{t_b-1} |H_o(t+1) - H_o(t)|}. \quad (19)$$

This measure is inspired from Hyndman and Koehler (2006). It consists of the ratio between the MAE and the mean daily variation of the observed values. An MASE value around 1 or below indicates an excellent accuracy.

- Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{t_b - t_a + 1} \sum_{t=t_a}^{t_b} \frac{|H_o(t) - H(t)|}{|H_o(t)|}. \quad (20)$$

While easy to interpret, this measure may be considered to give an undue importance to errors on small values. Moreover, in the case, encountered in practice, where $H_o(t) = 0$ for some t , it is undefined.

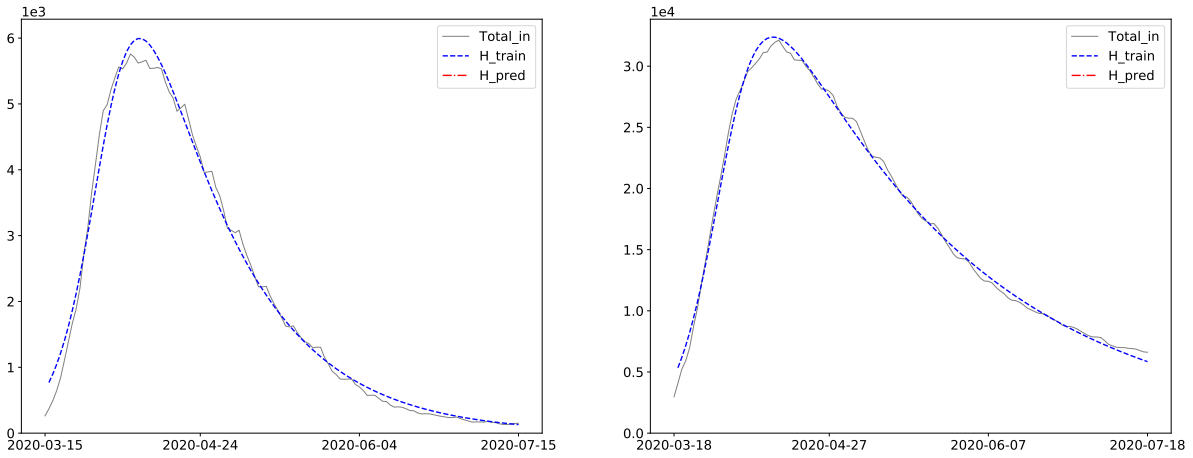


Figure 1: Fitting the SH model to the H_o (total hospitalized) curve. In this experiment, the train set is the whole dataset, hence there is no test (i.e., forecast) curve. Left: Belgium. Right: France.

- Symmetric mean absolute percentage error (sMAPE):

$$\text{sMAPE} = \frac{1}{t_b - t_a + 1} \sum_{t=t_a}^{t_b} \frac{|H_o(t) - H(t)|}{(|H_o(t)| + |H(t)|)/2}. \quad (21)$$

This measure overcomes the undefinedness issue of the MAPE since $H(t)$ is always (strictly) positive in the SH model.

6.2 Fitting Experiment

We first check how well the SH model (13) can fit the available data. For this experiment, we use the method of Section 5.2 with $c_E = c_L = 0$ in order to get the best possible fit (in the least squares sense) to the H_o time series. The results are shown in Figure 1 for Belgium³ and France.⁴ In this figure, the data is restricted to the first four months, where only one wave is present. Note that a tight fit over a period containing more than one wave is unachievable since the solutions of model (13) have only one peak.

For Belgium, the RRSE is 9% and the MAPE 14%. For France, the RRSE is 6% and the MAPE 3%. The comparatively large value of the MAPE for Belgium illustrates the comment made about MAPE in Section 6.1.

The fitting error is remarkably small in view of the fact that there are some 120 data points for only 4 estimands. Recall indeed that the parameters of the SH model are constant with respect to time in our experiments. This contrasts with Kozyreff (2021) where there are two phases, and with Nesterov (2020b) where the infection rate is piecewise constant with several pieces.

We stress that Figure 1 in itself does not imply that the model leads to accurate forecasts. If the optimal fit over some period is bad, then forecasts over that period can only be bad. But if the fit is good (as it is the case here), forecasts can still be bad due to their sensitivity with respect to the data preceding the to-be-forecasted period. For example, a better fit (in the RMSE sense) than in Figure 1 (left) can be obtained with a polynomial of degree 8; however, its forecast accuracy is abysmal.

In order to assess the forecast accuracy of the model, we have to learn the estimand variables over a *train period* that we make available to the algorithm, then use the learned estimands in order to predict H over a subsequent *test period* whose data is not available to the algorithm, and finally compare the prediction with the data on the test period. This is what we proceed to do in the rest of this Section 6.

6.3 Forecasts from Various Train Periods

Since the solutions of the SH model (13) have only one peak, it is already known that its forecast accuracy will be poor if the train and test periods overlap with two successive waves. In this section, we investigate if there exist train periods such that the subsequent few weeks are forecasted with high accuracy by the SH model.

For the Belgian data, Figure 2 shows,⁵ as a function of t_c , the MASE over the 60-day period $[t_c + 1, t_c + 60]$ obtained with the two-week train period $[t_c - 13, t_c]$. As anticipated, for some train periods, the forecast error is high, with an MASE around 100 or above. For most train periods, the MASE is around 10. However, when the train period is located around the first peak,

³Reproduce with SHR_22PA_py_BELsum_1sttraintstart1_1sttraintend123_1sttesttend123_c111.zip.

⁴Reproduce with SHR_22PA_py_FRAsum_1sttraintstart1_1sttraintend123_1sttesttend123_c111.zip.

⁵Reproduce with SHR_22PA_py_BELsum_1sttraintstart1_1sttraintend15_1sttesttend75_c111.zip.

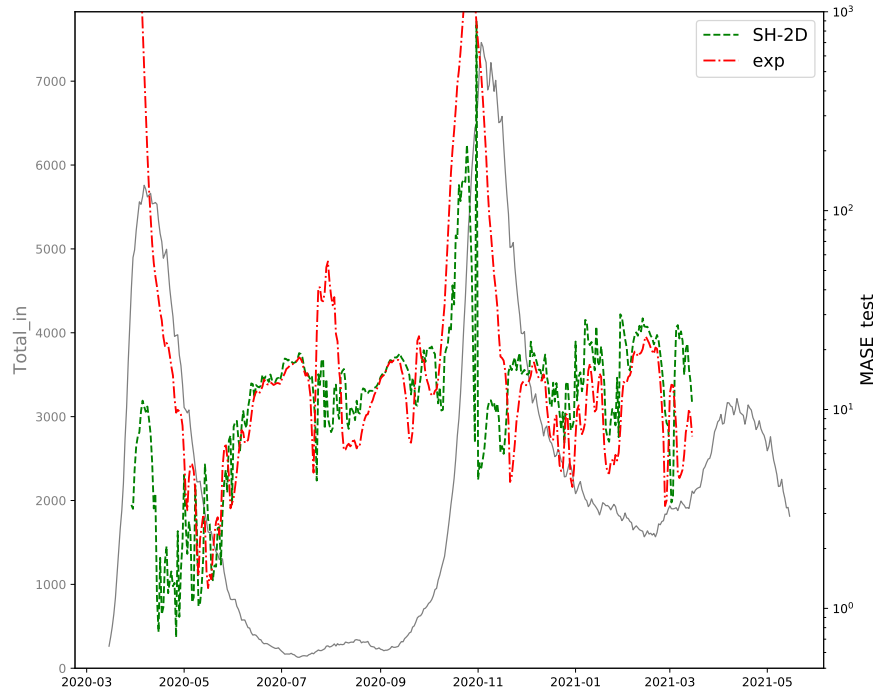


Figure 2: Forecast accuracy assessment for Belgian data. As a function of t (horizontal axis), the figure shows $(H_o(t))_{t=t_s, \dots, t_e}$ (observed hospitalizations, gray solid line, left-hand vertical axis), and the MASE over the test period $[t + 1, t + 60]$ for the train period $[t - 13, t]$ obtained respectively with the SH model (13) using the estimation method of Section 4 (green dashed line, right-hand vertical axis) and with the exponential model presented at the end of Section 6.3 (red dash-dotted line, right-hand vertical axis).

the MASE goes down to around 1. This means that the forecast error over the 60 days is so low that it is comparable to the daily variation of the data.

Figure 3 compares⁶ the observed hospitalization curve with the hospitalizations forecasted by the model for a few train periods chosen around the first peak. The forecast errors are indeed remarkably small. This finding is compatible with the considerations in Section 5.1, where we concluded that the estimation of $\hat{\beta}$, under a simple noise model, should be most accurate slightly before the peak of $H_o(t)$.

In Figure 4, we superpose the results⁷ obtained with various train periods of 14 days for Belgium. The figure further corroborates the comments of Section 5.1. In particular, if the train period is fully located before the peak, then the forecasts are rather inaccurate. Placing the train period around the peak gives excellent forecast results. When the train period is fully located in the decrease phase, the estimation of $\hat{\beta}$ and $\hat{S}(t_i)$ is seen to be very sensitive, but this does not affect much the quality of the forecast of $H(t)$.

Figure 5 is the equivalent of Figure 4 for France⁸. Again, the experiments are compatible with the comments of Section 5.1. In experiments not reported here, we also considered some departments separately, with fairly similar results.

A disconcerting aspect is the evolution of the estimated γ as a function of the location of the train period. In Figure 4 (Belgium), the estimation of γ is grouped around 0.08 for several train periods. However, in Figure 5 (France), the estimation of γ keeps decreasing, indicating that the daily number of patients leaving the hospital is an increasingly small fraction of the number of patients at the hospital. In view of (11), this casts doubt on the L_o values (see Section 2).

Since, for some train periods, highly accurate two-months-ahead forecasts are obtained with the SH model (13), whose order is as low as two, we also considered the forecasts obtained by an even simpler model, namely an exponential decay (or growth if $\gamma < 0$) of the form $H(t) = H(t_i)e^{-\gamma(t-t_i)}$. The two parameters $H(t_i)$ and γ are estimated in the least-square sense over the train period. Figure 2 shows that, for most train periods, the forecast errors obtained with the exponential model and the SH model are fairly similar. However, the SH model produces much more accurate forecasts than the exponential model when the train period is located around the peak.

⁶Reproduce with SHR_19PA_py_BELsum_1sttraintstart16_1sttraintend32_1sttesttend123_c111.zip.

⁷Reproduce with SHR_19PA_py_BELsum_1sttraintstart1_1sttraintend15_1sttesttend123_c111.zip.

⁸Reproduce with SHR_19PA_py_FRAsum_1sttraintstart1_1sttraintend15_1sttesttend122_c111.zip.

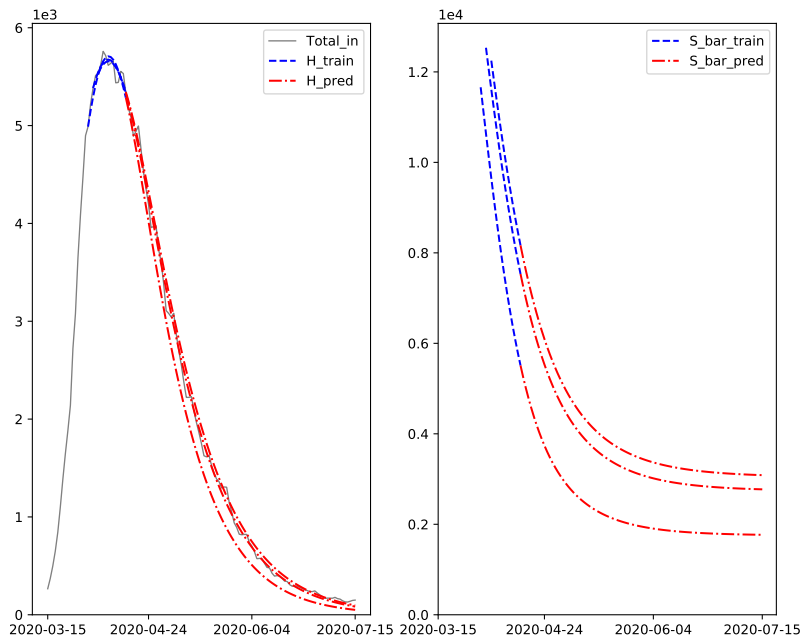


Figure 3: Belgium, train periods around the first peak. The left-hand plot shows $(H_0(t))_{t=t_1, \dots, t_c}$ (observed hospitalizations, gray solid line), $(H(t))_{t=t_1, \dots, t_c}$ (hospitalizations given by the model over the train period, blue dashed line), and $(H(t))_{t=t_c+1, \dots, t_e}$ (hospitalizations forecasted over the test period, in red dash-dot line). In order to give a sense of the sensitivity of the results, we superpose the curves obtained for three slightly different train periods. The test MASE values for the three curves are 3.16, 0.77, and 1.15. These values are compatible with the remark in Section 6.1 that MASE values around 1 or lower indicate an excellent accuracy. The test MAPE values for the three curves are 27%, 7%, and 8%. The right-hand plot shows the evolution of $\bar{S}(t)$.

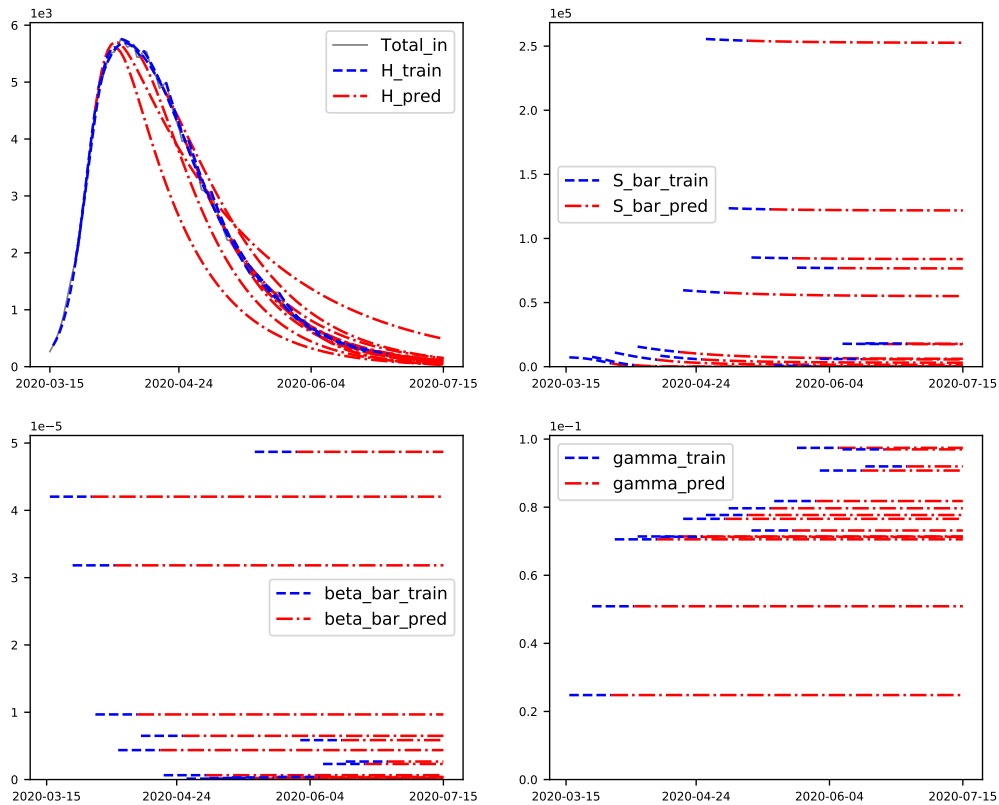


Figure 4: Belgium, various train periods.

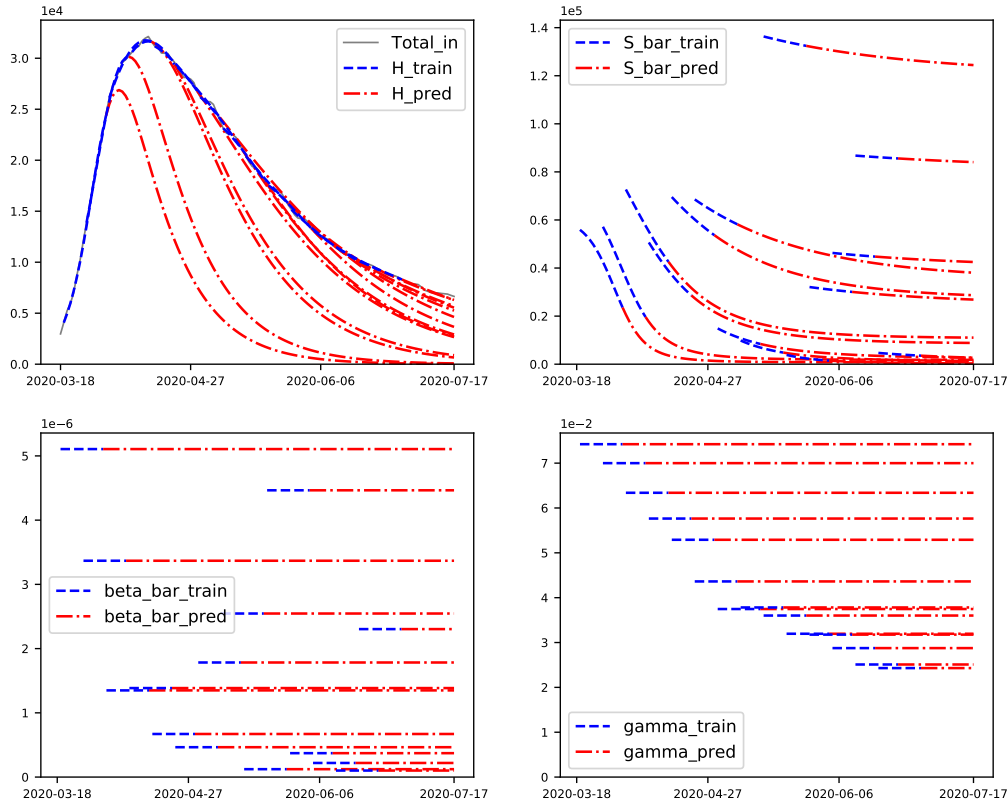


Figure 5: France, various train periods.

6.4 Forecasts from an Automatically Chosen Region Around the Peak

In this section, we investigate if the accurate forecasts obtained when the train period is around the first peak are also observed in other datasets.

In order to conduct experiments on a large number of datasets, we have to automatize the selection of the train period around the peak. Moreover, we want the procedure to be free of data leakage, i.e., it has to select the end time t_c of the train period without ever reading the data for $t > t_c$.

The proposed train-period selection procedure goes as follows. We first select a time span N ; in our experiments, we choose $N = 7$. In order to robustify the procedure, we smooth out the H_o time series by computing its moving average (MA):

$$H_o^{\text{MA}}(t) = \text{mean}(H_o(t - N), \dots, H_o(t + N)). \quad (22)$$

Then we compute $\hat{t} = \min\{t : \arg \max_{\tau \leq t} H_o^{\text{MA}}(\tau) = t - N\}$; in other words, the highest value of $(H_o^{\text{MA}}(t))_{t \leq i}$ is at $t = \hat{t} - N$, which we declare to be the location of the peak. This \hat{t} is computed by letting t increase until the argmax condition is satisfied. In view of (22), the data for $t > \hat{t} + N$ remains unseen, hence we choose $t_c := \hat{t} + N$ as the end time of the train period, thereby avoiding data leakage. The start of the train period is chosen as $t_i = \hat{t} - 2N$.

The test accuracies are measured over the 60-day period $[t_c + 1, \min(t_c + 60, t_c)]$.

The results obtained for each district (province in Belgium and department in France) are summarized in Table 1 (Belgium)⁹ and Table 2 (France)¹⁰. Moreover, for the whole of Belgium,¹¹ we obtain the excellent $\text{MASE}_{\text{test}} = 0.70$, and for the whole of France,¹² we get a considerably poorer $\text{MASE}_{\text{test}} = 14.07$.

The tables reveal that none of the districts admits an $\text{MASE}_{\text{test}}$ as low as the 0.70 obtained for the whole of Belgium. However, Section 6.3 has shown that an $\text{MASE}_{\text{test}}$ around 3 is still appreciably low. In view of Table 1, more than half of the Belgian provinces have an $\text{MASE}_{\text{test}}$ below 3.15. Table 2 indicates that an $\text{MASE}_{\text{test}}$ below 3 occurs for at least 10% of the French departments.

⁹Reproduce with SHR_22PA_py_BELeach_1sttraintstart19_1sttraintend41_1sttestend101_c111.zip.

¹⁰Reproduce with SHR_22PA_py_FRAeach_1sttraintstart50_1sttraintend72_1sttestend132_c111.zip.

¹¹Reproduce with SHR_22PA_py_BELsum_1sttraintstart17_1sttraintend39_1sttestend99_c111.zip.

¹²Reproduce with SHR_22PA_py_FRAsum_1sttraintstart20_1sttraintend42_1sttestend102_c111.zip.

Table 1: Statistics of forecast errors over all Belgian provinces for a train period automatically selected around the first peak. The left column specifies the various accuracy measures, as defined in Section 6.1. The notation P_n stands for the n th percentile.

Percentiles	min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	max
RRSE_train	0.30	0.32	0.33	0.33	0.45	0.55	0.65
RRSE_test	0.13	0.16	0.20	0.32	0.36	0.40	0.99
RRSE_test/RRSE_train	0.24	0.53	0.63	0.79	0.92	1.15	1.52
MASE	0.61	0.79	1.18	1.43	2.47	2.60	8.89
MASE_train	0.65	0.71	0.76	0.88	0.94	1.07	1.28
MASE_test	1.51	1.81	2.23	3.15	3.87	4.73	14.90
MASE_test/MASE_train	1.54	2.21	2.57	3.69	4.78	6.08	16.62
sMAPE_train	0.02	0.02	0.02	0.03	0.03	0.04	0.06
sMAPE_test	0.09	0.11	0.20	0.22	0.47	0.49	0.83
sMAPE_test/sMAPE_train	4.31	5.20	6.44	8.00	14.60	20.01	35.69

Table 2: Same as Table 1, now for the French departments.

Percentiles	min	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}	max
RRSE_train	0.17	0.35	0.40	0.52	0.67	0.86	1.26
RRSE_test	0.13	0.31	0.45	0.77	1.34	2.12	5.79
RRSE_test/RRSE_train	0.33	0.50	0.77	1.38	2.95	4.43	11.85
MASE	0.37	1.26	2.12	4.16	7.77	11.43	23.46
MASE_train	0.50	0.75	0.87	0.99	1.20	1.39	1.89
MASE_test	1.88	2.95	4.37	8.50	14.08	18.83	90.10
MASE_test/MASE_train	2.05	2.83	4.23	7.66	16.15	21.14	68.08
sMAPE_train	0.01	0.01	0.02	0.04	0.05	0.08	0.36
sMAPE_test	0.06	0.18	0.26	0.45	0.79	1.09	1.80
sMAPE_test/sMAPE_train	1.33	4.17	6.94	12.88	23.25	33.34	56.45

These experiments thus indicate that the SH model (13), trained by the procedure described in Sections 4.2–4.4, can be an asset for planning health care resources and the easing of lockdown restrictions over the two–three forthcoming months as soon as the peak has been reached.

6.5 Challenges

The MASE curves in Figure 2 have a rather jagged shape. For the exponential model, this can be attributed to the weekly variations of the hospitalization curve due to the fact that fewer patients are discharged during the weekend. For the SH model, the phenomenon is more pronounced because, more than the exponential model, it is able to fit the technical weekly variations in the data, resulting in poorer forecast accuracies for some train periods. A preliminary filtering aiming at reducing these technical weekly variations might lead to improved forecasts.

Another cause, specific to the SH model, is that the objective function (12), which has to be minimized to find the estimands $\tilde{\beta}$ and $\tilde{S}(t_i)$, has a narrow valley¹³ as illustrated in Figure 6. This makes it challenging to accurately compute the minimizer. The situation is worst for periods where the data is accurately fitted by the exponential model, because this means that an accurate fit is obtained with $\tilde{S}(t)$ almost constant in (13b), which makes it difficult to separately estimate $\tilde{\beta}$ and $\tilde{S}(t_i)$. For some train periods, the optimization solver `scipy.optimize.fmin` terminates far away from the global minimum of the objective function, resulting in a significantly suboptimal fit. Developing optimization methods tailored to the specific landscape of the objective function (12) is an important topic for further research.

¹³Reproduce with `SHR_22PA_py_BELsum_1sttraintstart17_1sttraintend39_1sttesttend99_c111.zip`.

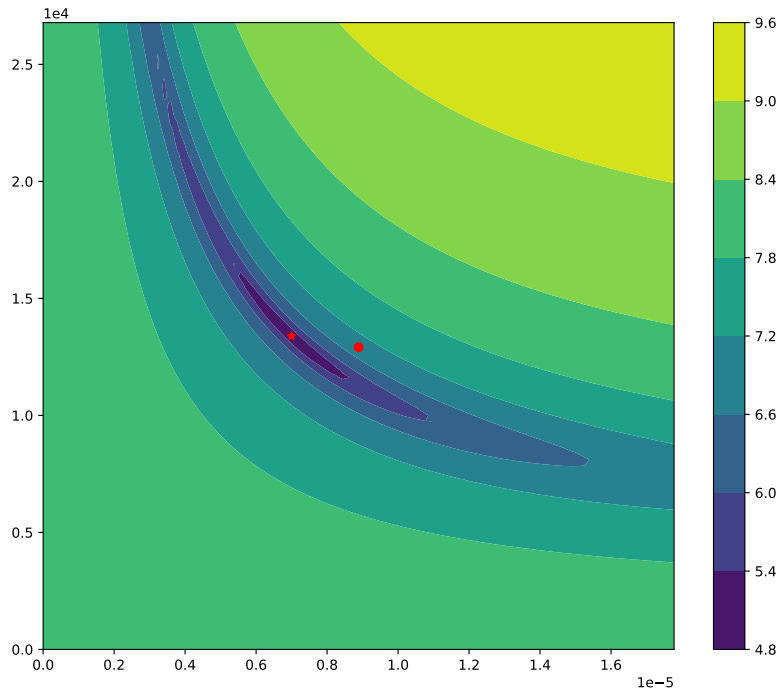


Figure 6: Belgium, contour plot of the objective function (12) for a train period around the peak. Horizontal axis: $\tilde{\beta}$; vertical axis: $\tilde{S}(t_i)$. In order to make the minimizer easier to visualize, the plot shows equispaced-level curves of $\log(\phi - 0.90\phi_*)$, where ϕ_* is the approximation of the minimal value of ϕ provided by the optimization solver. The red circle shows the initial guess obtained from Section 5.1 and the red star shows the approximate minimizer returned by the optimization solver. In our experiments, the optimization solver is `scipy.optimize.fmin` with its default parameters.

7 Conclusion

We have assessed the accuracy of COVID-19 hospitalization forecasts obtained with the SH model (13), a simple discrete-time dynamical system with only two state variables and two (time-independent) parameters. The experiments in Section 6 have shown that the proposed method has a remarkably good fitting accuracy over the whole first wave. It also produces remarkably accurate forecasts over certain time ranges for some areas (Belgium, some Belgian provinces, and a few French departments).

However, there are also time ranges and areas where the forecasts are very inaccurate. In particular, when it is trained before the peak, the model produces rather poor forecasts for the position and height of the peak and for the subsequent decrease. The model is also unable to produce multiple peaks in order to fit or forecast rebounds. The forecasts returned by the model should thus be taken with much caution.

Another source of caution is that, even though it requires to estimate only four (time-invariant) estimands, fitting the SH model (13) to the data is already not a trivial task. We have not ruled out the situation where the considered objective function would be multimodal. The optimization solver might thus get stuck in a local nonglobal minimum, yielding a suboptimal fit of the train data and possibly poorer forecasts than what the global minimum would achieve. Moreover, even if the objective function is unimodal, the stopping criterion of the solver may trigger before a suitably accurate approximation of the minimum is reached, as discussed in Section 6.5.

If the proposed model is used to guide prevention policies, then further caveats are in order. We have seen that the estimation of $\tilde{\beta}$ is very sensitive. Hence the proposed model can hardly help assess the impact of prevention measures on $\tilde{\beta}$. Without knowing sufficiently accurately the impact of prevention measures on $\tilde{\beta}$, we may not aptly use the model to predict their impact on the evolution of the hospitalizations.

Yet another caveat is that it may be tempting to deduce from the excellent fit with a constant-parameter model (Figure 1) that the evolution of the prevention measures over the dataset period has had no impact on $\tilde{\beta}$. But the deduction is flawed. Indeed, in view of the comments made in Section 5.1, the available data could also be very well explained with fairly large jumps in $\tilde{\beta}$ during the decrease phase.

In spite of all these caveats, the hospitalization forecasts returned by the method, and also the evolution of $\tilde{S}(t)$, might be of practical use in the context of various disease outbreaks, e.g., for resource planning. To this end, it will be important

to understand which specific features of the COVID-19 outbreak in Belgium made it possible to forecast so accurately the hospitalization decrease several months ahead.

Acknowledgment

The second author is supported by a fellowship awarded by UCLouvain's Conseil de l'action internationale.

This work benefited from discussions with several colleagues, in particular Frédéric Crevecoeur, Pierre Dupont, Alexey Medvedev, Pierre Schaus, and Loïc Van Hoorebeeck.

References

- Abrams, S., J. Baetens, J. Vergeynst, T. Alleman, I. Nopens, K. Barbé, F. Vermolen, N. Franco, S. Clesse, L. Willem, C. Faes, G. Molenberghs, P. Beutels, and N. Hens (2021, 04). Long-term scenarios for the number of new hospitalizations during the Belgian COVID-19 epidemic. <http://doi.org/10.13140/RG.2.2.31704.83208>. 215
- Atkeson, A. (2020, March). What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. Working Paper 26867, National Bureau of Economic Research. <http://doi.org/10.3386/w26867>. 215
- Baek, J., V. F. Farias, A. Georgescu, R. Levi, T. Peng, D. Sinha, J. Wilde, and A. Zheng (2020). The limits to learning an SIR process: Granular forecasting for COVID-19. <https://arxiv.org/abs/2006.06373>. 215
- Barlow, N. S. and S. J. Weinstein (2020). Accurate closed-form solution of the SIR epidemic model. *Physica D: Nonlinear Phenomena* 408, 132540. 218
- Bhanot, G. and C. DeLisi (2020). Predictions for Europe for the Covid-19 pandemic from a SIR model. *medRxiv*. <https://www.medrxiv.org/content/early/2020/05/28/2020.05.26.20114058.1>. 215
- Calafiore, G. C., C. Novara, and C. Possieri (2020). A modified SIR model for the COVID-19 contagion in Italy. <https://arxiv.org/abs/2003.14391>. 215
- Carletti, T., D. Fanelli, and F. Piazza (2020). COVID-19: The unreasonable effectiveness of simple models. *Chaos, Solitons & Fractals: X* 5, 100034. 215, 216
- Comunian, A., R. Gaburro, and M. Giudici (2020). Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic. *Physica D: Nonlinear Phenomena*, 132674. 215
- Djidjou-Demasse, R., Y. Michalakis, M. Choisy, M. T. Sofonea, and S. Alizon (2020). Optimal COVID-19 epidemic control until vaccine deployment. *medRxiv*. <https://www.medrxiv.org/content/early/2020/05/15/2020.04.02.20049189>. 215
- Fanelli, D. and F. Piazza (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals* 134, 109761. 215
- Ghanam, R., E. L. Boone, and A.-S. G. Abdel-Salam (2020). COVID-19: SEIRD model for Qatar COVID-19 outbreak. *Letters in Biomathematics*. 215
- Ghosh, I. and T. Chakraborty (2020). An integrated deterministic-stochastic approach for predicting the long-term trajectories of COVID-19. *medRxiv*. <https://www.medrxiv.org/content/early/2020/07/04/2020.05.13.20101303>. 215
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review* 42(4), 599–653. 215
- Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4), 679 – 688. 220
- Kermack, W. O., A. G. McKendrick, and G. T. Walker (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115(772), 700–721. 215
- Kozyreff, G. (2021). Hospitalization dynamics during the first COVID-19 pandemic wave: SIR modelling compared to Belgium, France, Italy, Switzerland and New York City data. *Infectious Disease Modelling* 6, 398–404. 215, 216, 221

- Liu, Y., A. A. Gayle, A. Wilder-Smith, and J. Rocklöv (2020, 02). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine* 27(2). 215
- Nesterov, Y. (2020a). Online analysis of epidemics with variable infection rate. <https://arxiv.org/abs/2007.11429>. 215
- Nesterov, Y. (2020b). Online prediction of COVID19 dynamics. Belgian case study. CORE Discussion Paper 2020/22, UCLouvain. <https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers.html>. 215, 221
- Roda, W. C., M. B. Varughese, D. Han, and M. Y. Li (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* 5, 271 – 281. 215
- Schoukens, J. (2018). System identification. http://homepages.vub.ac.be/~pdreesen/edu/course_introduction_to_system_identification.pdf. 218
- Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems*. Oxford: Pergamon. 218
- Singh, S. S. and D. Mohapatra (2020). Predictive analysis for COVID-19 spread in India by adaptive compartmental model. *medRxiv*. <https://www.medrxiv.org/content/early/2020/07/09/2020.07.08.20148619>. 215
- Vinitzky, S. I., A. A. Gusev, V. L. Derbov, P. M. Krassovitskiy, F. M. Pen'kove, and G. Chuluunbaatar (2021). Reduced SIR model of COVID-19 pandemic. *Computational Mathematics and Mathematical Physics* 61, 376–387. 218
- Wang, M. and S. Flessa (2020). Modelling Covid-19 under uncertainty: what can we expect? *The European Journal of Health Economics*. <http://doi.org/10.1007/s10198-020-01202-y>. 215