

Computation of COVID-19 epidemiological data in Hungary using dynamic model inversion *

Balázs Csutak¹, Péter Polcz¹, and Gábor Szederkényi^{†1}

¹Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Práter u. 50/a, H-1083 Budapest, Hungary

Abstract

In this paper, we estimate epidemiological data of the COVID-19 pandemic in Hungary using only the daily number of hospitalized patients, and applying well-known techniques from systems and control theory. We use a previously published and validated compartmental model for the description of epidemic spread. Exploiting the fact that an important subsystem of the model is linear, first we compute the number of latent infected persons in time. Then an estimate can be given for the number of people in other compartments. From these data, it is possible to track the time dependent reproduction numbers via a recursive least squares estimate. The credibility of the obtained results is discussed using available data from the literature.

1 Introduction

Since the spring of 2020, the COVID-19 pandemic has put an enormous burden on the societies, economies and healthcare systems of most countries. It is of fundamental importance to track the evolution of the epidemic process in order to prepare healthcare capacities, or to design efficient measures and vaccination policies against the disease spread. It is well-known that only a fraction of the actually infected persons get registered into the official databases. This ratio varies widely between countries depending on the testing activity and related policies. E.g., as of Apr. 28., 2021, the official data for the cumulative numbers of infected in Czechia and Hungary are 1.626.033 [1] and 774.399 [2], respectively. Considering the similarities between these two countries in terms of geographic location, area, population size, the stringency of interventions, the organization of healthcare, and the number of reported COVID-deaths, it is highly unlikely that there could be such a huge difference in the true numbers. There exist several attempts in the literature for the reconstruction of real epidemic data. In [3] a robust statistical estimate with uncertainty analysis was given for the detection rate in several countries. It turns out from the published results that during the 2020 spring wave of the pandemics, the detection rate could be as small or even lower than 10% for e.g., Sweden, Italy, Spain, or may have reached 40-50% in the case of Australia and South Korea, where huge effort was put into immediate contact tracing and testing. Somewhat tighter upper bounds were computed for the possible total number of cases in several European countries in [4] using a data-driven estimator from the distributions of daily cases and deaths. It is clear from the above that the recorded number of infected people alone is generally not a reliable source for the tracking of the epidemics. Moreover, it is also visible from the Hungarian data that recoveries were not followed and recorded precisely in the second wave of the pandemic until the middle of December, 2020 (see, e.g. at [5]). Therefore, the epidemic curve showing the active cases cannot reflect the real situation even in a qualitative way.

*P. Polcz gratefully acknowledges the support of the New National Excellence Program scholarship (ÚNKP-20-4-I-PPKE-41). B. Csutak was supported by the UNKP-20-3-I-PPKE-66 project of the New National Excellence program of the Ministry for Innovation and Technology. G. Szederkényi acknowledges the partial support by the European Union, co-financed by the European Social Fund through the grant EFOP-3.6.3-VEKOP-16-2017-00002. This work was partially supported by the National Research, Development, and Innovation Office through the grant NKFIH-OTKA-131545.

[†]corresponding author

After the spring wave of COVID-19, a representative serological test was designed and carried out in Hungary which estimated that exposure was approximately 0.68% among the total population by the middle of May, 2020 [6]. However, no similar study has been done since the summer of 2020, although it would certainly have huge significance in exploring the current situation and refining our models.

There is a huge literature on epidemic modeling and recently on the modeling of the COVID-19 pandemic for different purposes such as identification, control, or prediction [7]. E.g., in [8], a hybrid machine learning approach is proposed to predict the number of infected individuals and the mortality rates of the first wave of the COVID-19 outbreak in Hungary. In [9] the authors show that epidemic outbreak prediction can be efficiently supported by integrating machine learning and SEIR models.

In systems and control theory, it is a common practice to estimate unknown quantities (e.g., states or parameters) from measurement data using dynamical models of the observed phenomena [10, 11]. Population-level deterministic epidemic models are most often written in a nonnegative compartmental form with polynomial nonlinearities representing the infection mechanism and linear subsystems describing the transitions between certain compartments [12]. We will use a control oriented ODE model of the pandemics in Hungary which was proposed and used in [13], and was derived from the more detailed model in [14].

Based on the above, the purpose of this paper is to propose a model-based computation approach to study the second and third waves of the COVID-19 pandemic in Hungary. Due to the previously mentioned unreliability of other data sources, we only use the number of hospitalized people available from [15] as input data, assuming that testing is wide-spread and quick enough in hospitals.

2 The applied compartmental model

2.1 Model description

To describe the transmission dynamics of the disease, we use a compartmental model describing the spreading characteristics of COVID-19. We divide the population of N individuals into eight classes, representing the different stages of the illness, discussed in detail in [13]. The compartments used stand for the following: the susceptibles (**S**) are those individuals, who can be infected by the disease (i.e., neither have been infected nor yet obtained immunity by, eg., vaccination). Latents (**L**) are infected people in the very first stage, lacking any symptoms and also incapable to transmit the pathogen. This is followed by the pre-symptomatic (**P**) stage, including individuals already infectious, but still without symptoms. As a significant part of the infected show no characteristic symptoms, and thus they being infected is often not confirmed, we handle the asymptomatic (**A**) and symptomatic infected (**I**) classes differently. Members of **A** always recover (**R**) after a certain time, but for some part of **I** hospitalization is needed. Hospitalized patients (**H**) may recover or decease (**D**). The possible transitions between the compartments are illustrated in Fig. 1 and their dynamics is described formally by the ODEs below.

$$\dot{\mathbf{S}}(t) = -\beta [\mathbf{P}(t) + \mathbf{I}(t) + \delta \mathbf{A}(t)] \mathbf{S}(t)/N, \quad (1a)$$

$$\dot{\mathbf{L}}(t) = \beta [\mathbf{P}(t) + \mathbf{I}(t) + \delta \mathbf{A}(t)] \mathbf{S}(t)/N - \alpha \mathbf{L}(t), \quad (1b)$$

$$\dot{\mathbf{P}}(t) = \alpha \mathbf{L}(t) - p \mathbf{P}(t), \quad (1c)$$

$$\dot{\mathbf{I}}(t) = qp \mathbf{P}(t) - \rho_I \mathbf{I}(t), \quad (1d)$$

$$\dot{\mathbf{A}}(t) = (1 - q)p \mathbf{P}(t) - \rho_A \mathbf{A}(t), \quad (1e)$$

$$\dot{\mathbf{H}}(t) = \rho_I \eta \mathbf{I}(t) - h \mathbf{H}(t), \quad (1f)$$

$$\dot{\mathbf{R}}(t) = \rho_I (1 - \eta) \mathbf{I}(t) + \rho_A \mathbf{A}(t) + (1 - \mu) h \mathbf{H}(t), \quad (1g)$$

$$\dot{\mathbf{D}}(t) = \mu h \mathbf{H}(t). \quad (1h)$$

From now on, we suppress the time arguments in the state variables as it is commonly done in the literature.

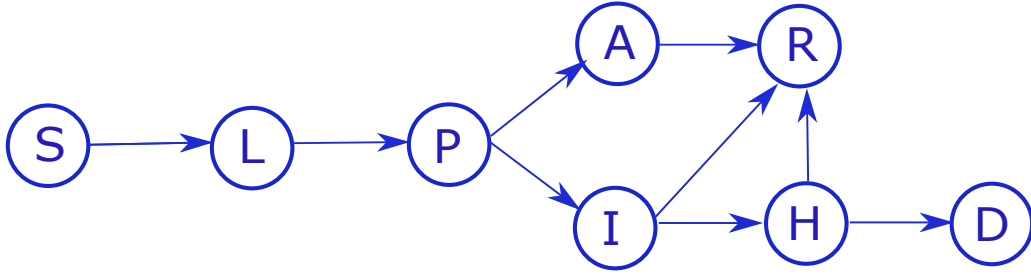


Figure 1: Transition diagram of the disease spreading model. Circles represent compartments and arrows represent transitions between these compartments.

2.2 Model parameters

We set the model parameters according to the estimates published in related literature. We used the following model parameters for (1) which had been determined by updating the parameter set published in [13] with newer Hungary-specific data. The population of Hungary is $N = 9.8 \cdot 10^6$ people. The latent period (α^{-1}) is 2.5 days on average, whereas, the pre-symptomatic period (p^{-1}) is generally 3 days. The infectious period of both the symptomatic (ρ_I) and asymptomatic (ρ_A) classes is approximately 4 days. The average length of hospitalization (h^{-1}) before recovery or death is 10 days. The hospitalization probability of the symptomatic (η) is 0.076. The probability of developing symptoms ($q = 0.6$) and the relative infectiousness of the asymptomatic ($\delta = 0.75$) are based on CDC estimates. The death ratio among hospitalized is $\mu = 0.185$.

The basic reproduction number, expressing the average number of new infections generated by a single infected individual in a fully susceptible population, is given as

$$R_0 = \beta \left(\frac{1}{p} + \frac{q}{\rho_I} + \frac{\delta(1-q)}{\rho_A} \right). \quad (2)$$

Using $R_0 = 2.2$ estimated from early Hungarian data, and the already determined parameters, a nominal value of $\beta = 1/3$ can be derived, but note that this parameter largely depends on the actual restrictions and conditions, therefore, it will be estimated as a time-varying parameter. The time-dependent reproduction number for our model taking into account the change of β and the decrease of the susceptible population is given by

$$R_c(t) = \beta(t) \left(\frac{1}{p} + \frac{q}{\rho_I} + \frac{\delta(1-q)}{\rho_A} \right) \frac{\mathbf{S}(t)}{N}. \quad (3)$$

3 Computation methodology

3.1 Data processing

Dynamic inversion is known to be sensitive to the abrupt changes in the derivatives of measured signals caused by fluctuations and noise. Therefore, we applied the following preprocessing to smoothen the input data.

Let \check{y}_k denote the number of hospitalized patients on day k collected from [15], $k = 1, \dots, T$. As the starting date ($k = 1$), we considered the 20th of August, 2020. The available data was collected up till the 28th of April, 2021.

Generally, at weekends the healthcare system has a decreased capacity to perform tests and document the confirmed cases. Therefore, as it is commonly done in the literature and engineering practice, we apply a 7-day long moving average filter to the officially published hospitalization data record. Formally, we consider the following smoothed time series of the numbers of hospitalized patients:

$$\bar{y}_k = \frac{1}{7} \sum_{i=-3}^3 \check{y}_{k+i}, \text{ if } k = 4, \dots, T-3. \quad (4)$$

Obviously, at the two ends of the time series, the sliding window has to be truncated as follows:

$$\bar{y}_k = \frac{1}{\min(3, T-k) + \min(3, k) + 1} \sum_{i=-\min(3, k)}^{\min(3, T-k)} \tilde{y}_{k+i}. \quad (5)$$

Next, we used a cubic spline interpolation technique, ensuring the result being twice continuously differentiable. We divided the data series into $n = 15$ equally long segments, the endpoints being at $t_i = i \cdot (T/n)$, $t = 0, 1, \dots, n$ in time. As these time points (expressed in days) are not necessarily integers, we used a linear interpolation to calculate \bar{y}_{t_i} from $\bar{y}_{\lfloor i \cdot (T/n) \rfloor}$ and $\bar{y}_{\lfloor i \cdot (T/n) \rfloor + 1}$. The spline was fitted on data points (t_i, \bar{y}_{t_i}) , $i = 0, 1, \dots, n$ using not-a-knot conditions for the slope endpoints. Let the spline interpolated hospitalization data be denoted by y_t , $t \in [1, T]$. The original measurement (\tilde{y}) and the results (\bar{y} and y) of the smoothing process can be seen in Fig. 2. Additionally, Fig. 3 illustrates that the data preprocessing steps make it possible to compute the first three derivatives of y relatively noise-free. Henceforth, in all computations, we use y as the single available information about the time evolution of the epidemic spread.

3.2 Input computation by inversion

Consider the $(\mathbf{P}, \mathbf{I}, \mathbf{H})$ subsystem of (1) in the following continuous-time linear time-invariant (LTI) state-space form:

$$\Sigma_{\mathbf{PIH}} : \begin{cases} \dot{x} = Ax + Bu, \\ y = Cx, \end{cases} \quad \text{where } x = \begin{pmatrix} \mathbf{P} \\ \mathbf{I} \\ \mathbf{H} \end{pmatrix}, \quad u = \mathbf{L}, \quad (6)$$

$$A = \begin{pmatrix} -p & 0 & 0 \\ pq & -\rho_I & 0 \\ 0 & \eta\rho_I & -h \end{pmatrix}, \quad B = \begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix}, \quad C = (0 \ 0 \ 1). \quad (7)$$

The input of $\Sigma_{\mathbf{PIH}}$ is the daily number of people in the latent phase ($u = \mathbf{L}$), whereas, the output is (*the filtered time series of*) the daily number of hospitalized patients ($y = \mathbf{H}$). The continuous-time transfer function of $\Sigma_{\mathbf{PIH}}$ is the following:

$$G(s) = \frac{0.00152}{s^3 + 0.6833s^2 + 0.1417s + 0.008333}, \quad (8)$$

from which we can see that the analytical inverse of the system is non-causal.

To compute the unknown input of $\Sigma_{\mathbf{PIH}}$, we consider the discrete-time (DT) model of $\Sigma_{\mathbf{PIH}}$ with $T_s = 1$ (day) sampling period. Using zero order hold, the transfer function of the discretized model is the following:

$$W(z) = \frac{b_2 z^2 + b_1 z + b_0}{z^3 + a_2 z^2 + a_1 z + a_0}, \quad (9)$$

where $a_0 = -0.5049$, $a_1 = 1.911$, $a_2 = -2.4$, $b_0 = 1.52 \cdot 10^{-4}$, $b_1 = 7.225 \cdot 10^{-4}$, $b_2 = 2.139 \cdot 10^{-4}$. The corresponding DT input-output model can be written as follows:

$$y_{k+3} + a_2 y_{k+2} + a_1 y_{k+1} + a_0 y_k = b_2 u_{k+2} + b_1 u_{k+1} + b_0 u_k. \quad (10)$$

First, we compute the input from the output using a standard least squares deconvolution. Using the available filtered time series of $y_{k=1, \dots, T}$, the input-output difference equation (10) can be written in the following matrix-vector format:

$$M_a \mathbf{y} = M_b \mathbf{u}, \quad \text{where} \quad (11)$$

$$M_a = \begin{pmatrix} a_0 & a_1 & a_2 & 1 & 0 & \dots & \dots \\ 0 & a_0 & a_1 & a_2 & 1 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 0 & a_0 & a_1 & a_2 & 1 \end{pmatrix}, \quad M_b = \begin{pmatrix} b_0 & b_1 & b_2 & 0 & 0 & \dots \\ 0 & b_0 & b_1 & b_2 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 0 & b_0 & b_1 & b_2 \end{pmatrix}, \quad (12)$$

$$\mathbf{y} = (y_1 \ \dots \ y_T)^\top, \quad \mathbf{u} = (u_1 \ \dots \ u_{T-1})^\top. \quad (13)$$

The system of linear equations in (11) is under-determined in \mathbf{u} , since $M_a \in \mathbb{R}^{(T-3) \times T}$ and $M_b \in \mathbb{R}^{(T-3) \times (T-1)}$. The least-squares solution for \mathbf{u} can be computed by considering the Moore–Penrose right pseudo-inverse $M_b^\ddagger = M_b^\top (M_b M_b^\top)^{-1}$ of the row-shaped matrix M_b , namely:

$$\check{\mathbf{u}}_{\text{ls}} = M_b^\ddagger M_a \mathbf{y}. \quad (14)$$

After computation, we filtered the time series $\check{\mathbf{u}}_{\text{ls}}$ with a 7-day moving average filter, which resulted in $u_{\text{ls},k}$, $k = 1, \dots, T-1$. Having an estimate of the input of the system $\Sigma_{\mathbf{PIH}}$, we can apply a standard linear state observer to estimate the population in the non-measured compartments ($\mathbf{P}, \mathbf{I}, \mathbf{A}, \mathbf{R}$). An estimate for \mathbf{D} can be computed from (1h), where the official governmental data can be used as well. However, we do not use the published data on COVID deaths for inversion, and the dynamics of the other compartments do not depend on \mathbf{D} .

3.3 Unknown-input state estimation

As another possible solution, in this subsection we present an unknown input observer approach for system inversion and state estimation. It can be seen from the transfer function model (8) that the LTI subsystem $\Sigma_{\mathbf{PIH}}$ of (1) has relative degree 3, i.e., the input signal does appear only in the third (and higher order) derivatives of the output equation:

$$\dot{y} = CAx + CBu = CAx, \quad (CB = 0) \quad (15)$$

$$\ddot{y} = CA^2x + CABu = CA^2x, \quad (CAB = 0) \quad (16)$$

$$\dddot{y} = CA^3x + CA^2Bu. \quad (17)$$

We can check that $\Sigma_{\mathbf{PIH}}$ in itself does not satisfy the necessary criteria of [16, Lemma 3.1] or [17, Criterion 2] for the existence of an *unknown-input observer* which may seem contradictory knowing the result of the previous subsection.

Observe that, if we consider (15) and (16) as two additional outputs of $\Sigma_{\mathbf{PIH}}$, then the augmented state-space model

$$\begin{cases} \dot{x} = Ax + Bu, \\ \mathcal{Y} = \mathcal{O}_3 x, \end{cases} \quad \text{where } \mathcal{Y} = \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \end{pmatrix}, \quad \mathcal{O}_3 = \begin{pmatrix} C \\ CA \\ CA^2 \end{pmatrix} \quad (18)$$

is unknown-input observable. Furthermore, (17) allows us to algebraically express the unknown input from the *third* derivative of signal y .

Following the ideas of [16, Section 3.2.1], we construct the following unknown-input observer for (18):

$$\begin{cases} \dot{z} = Fz + K\mathcal{Y}, \\ x_{\text{uio}} = z + H\mathcal{Y}, \\ u_{\text{uio}} = (CA^2B)^{-1}(\ddot{y} - CA^3x_{\text{uio}}). \end{cases} \quad (19)$$

The coefficient matrices of (19) are:

$$\begin{aligned} K &= K_1 + K_2, \quad H = B(\mathcal{O}_3B)^\dagger, \\ F &= A - H\mathcal{O}_3A - K_1\mathcal{O}_3, \quad K_2 = FH, \end{aligned} \quad (20)$$

where $(\mathcal{O}_3B)^\dagger = ((\mathcal{O}_3B)^\top(\mathcal{O}_3B))^{-1}(\mathcal{O}_3B)^\top$ constitutes the Moore–Penrose left pseudo-inverse of matrix \mathcal{O}_3B , and K_1 is selected such that F is a stability matrix.

To summarize, the non-measured state variables (\mathbf{P}, \mathbf{I}) of $\Sigma_{\mathbf{PIH}}$ can be reconstructed using the first two derivatives of y , whereas, the input reconstruction requires also the third derivative of y . The remaining state variables ($\mathbf{A}, \mathbf{R}, \mathbf{D}$) of (1) can be reconstructed by considering the estimated trajectories of (\mathbf{P}, \mathbf{I}) and the measured trajectory of y . Finally, using the fact that the whole population is conserved in our compartmental model, the number of susceptibles can be estimated simply as

$$\mathbf{S} = N - (\mathbf{L} + \mathbf{P} + \mathbf{I} + \mathbf{A} + \mathbf{H} + \mathbf{R} + \mathbf{D} + \mathbf{V}), \quad (21)$$

where \mathbf{V} denotes the cumulative number of people how have become immune due to vaccination.

Remark 1 *In the computation, we considered a simple vaccination model taking into consideration the documented efficacy of vaccines applied in Hungary. Based on this, we assumed that 85% of vaccinated people do not take part in the infection process any more from 21 days after administering the first dose. The time series of the number of vaccinated people were collected from the official website [18] of the Government of Hungary.*

3.4 Estimation of the reproduction numbers

It is visible from Eqs. (2) and (3) that the parameter β is needed to compute estimates for the reproduction numbers. The right hand side of Eq. (1a) depends linearly on β , therefore, a recursive least squares (RLS) algorithm with exponential forgetting can be applied [11]. For this, we discretize (1b) using a simple forward-Euler scheme with sampling time T_s which can be written in linear regression form as

$$\underbrace{\mathbf{L}_{k+1} + (\alpha T_s - 1)\mathbf{L}_k}_{\pi_k} = \beta_k \underbrace{T_s(\mathbf{P}_k + \mathbf{I}_k + \delta \mathbf{A}_k)\mathbf{S}_k/N}_{\varphi_k}. \quad (22)$$

Then, the RLS algorithm for estimating β can be written as

$$\hat{\beta}_{k+1} = \hat{\beta}_k + P_k \varphi_k (\pi_k - \varphi_k \hat{\beta}_k) \quad (23)$$

$$P_{k+1} = (\lambda \cdot P_k^{-1} + \varphi_k^2)^{-1} \quad (24)$$

where $\hat{\beta}_k$ is the estimate for β , P_k is the auxiliary variable, and $\lambda = 0.9$ is the forgetting factor. The estimates for the reproduction numbers R_0 and R_c are given by substituting $\hat{\beta}$ into Eqs. (2) and (3). We note that the computation of R_c uses the simple vaccination model presented in Remark 1.

4 Results and Discussion

The studied period is from August 20, 2020 to April 28, 2021. We considered this time independently of the drastically and successfully suppressed first wave of the epidemic in the spring of 2020. During the first half of August, the main infection indicators (daily new infections, test positivity, number of hospitalized people) were so low that the zero initial condition assumption for the transfer function representation is acceptable. (We remark that the correct convergence of the state estimates would be ensured even in the case of non-zero initial conditions after a transient period, since the subsystem (6) is asymptotically stable and observable.)

4.1 Evaluation of the inversion

In Fig. 4, we illustrate the least squares solution \tilde{u}_{1s} for the unknown input, its 14-day long moving average filtered data u_{1s} , and the dynamically reconstructed input u_{uio} . In Fig. 5, the solid lines illustrate the simulated state variables of the $(\mathbf{P}, \mathbf{I}, \mathbf{H}, \mathbf{A})$ subsystem of (1) driven by the filtered least squares solution \tilde{u}_{1s} . Whereas, the dashed lines of Fig. 5, constitute the observed state variables of $\Sigma_{\mathbf{PIH}}$ and (1e), and the reconstructed unknown input u_{uio} in the approximative knowledge of the daily number of hospitalized patients (y) and its first three derivatives.

To evaluate the reliability of the state observations, we simulate $\Sigma_{\mathbf{PIH}}$ with the computed (and linearly interpolated) unknown input functions \tilde{u}_{1s} , u_{1s} , and u_{uio} . Then, the reconstructed hospitalization data, denoted by \tilde{y}_{1s} , y_{1s} , and y_{uio} , respectively, are compared to the actual and the filtered measurements \tilde{y} , \bar{y} , and y . Fig. 6 illustrates that the shape of y , y_{1s} , and y_{uio} are qualitatively the same.

To quantify the difference between the measured and the reconstructed signals, (instead of the standard squared error model) we use the following relative squared error formula:

$$d_r(y, \hat{y}) = \frac{\|y - \hat{y}\|_2}{\|y\|_2} = \left(\sum_{k=1}^{T-3} |y_k - \hat{y}_k|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{T-3} |y_k|^2 \right)^{-\frac{1}{2}}. \quad (25)$$

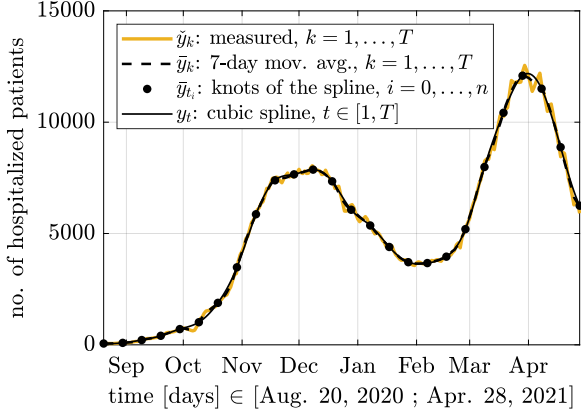


Figure 2: Smoothed time series (solid black) compared to the officially published hospitalization data (thick yellow) and the 7-day long moving average filtered data (dashed black).

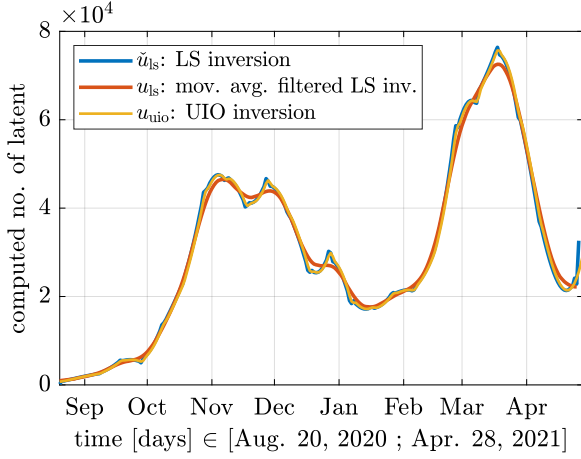


Figure 4: Computed estimate for the number of people in the latent phase using a least squares solution and a dynamical input reconstruction.

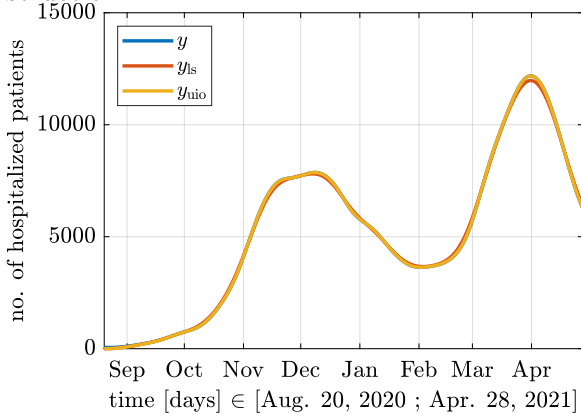


Figure 6: Measured and filtered hospitalization data and its reconstruction using both the least squares solution u_{ls} and the dynamically computed u_{uio} .

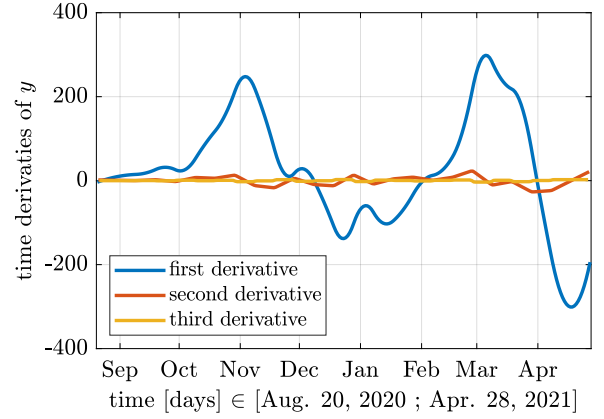


Figure 3: The first three derivatives of the cubic spline interpolated measurement data.

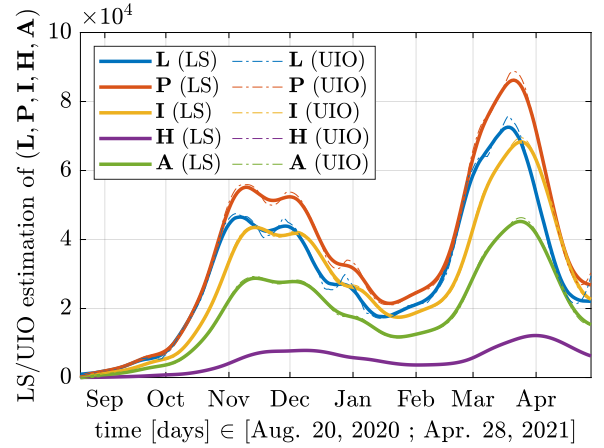


Figure 5: Estimated state variables obtained through the least squares solution u_{ls} (solid) and by the unknown-input observer (dashed).

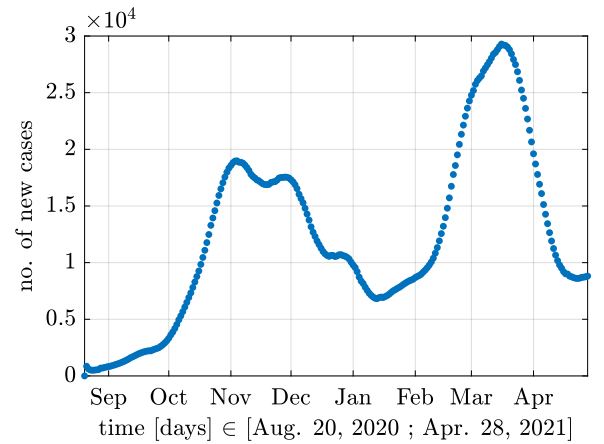


Figure 7: Computed number of new infections per day.

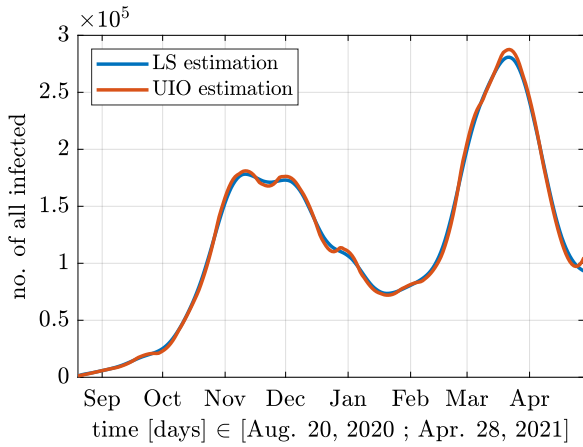


Figure 8: Computed number of all infected people (sum of \mathbf{L} , \mathbf{P} , \mathbf{A} , \mathbf{I} , and \mathbf{H}).

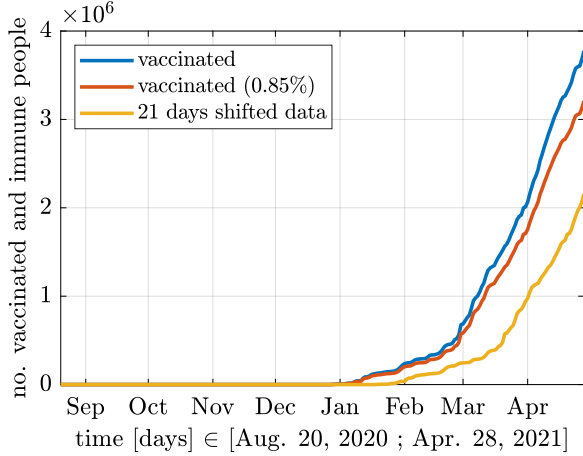


Figure 10: Number of vaccinated and estimated number of truly immune people.

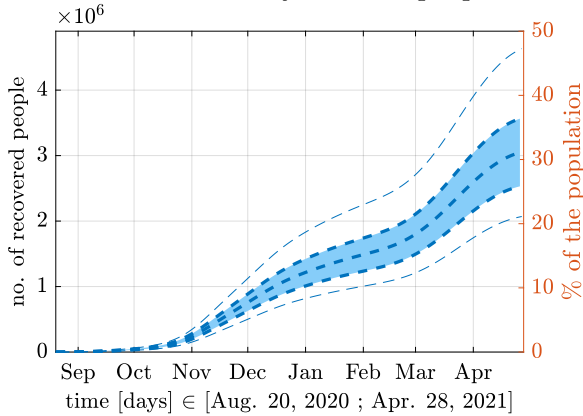


Figure 12: Estimated number of recovered people. The shaded blue area illustrates the standard deviation of the simulated trajectories. The light blue dashed lines show the simulated minimum/maximum number of recovered people.

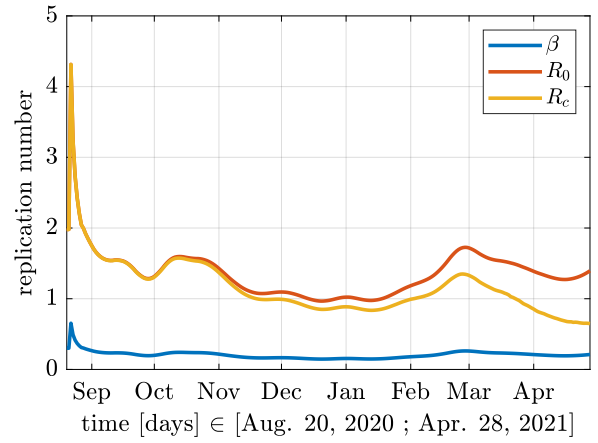


Figure 9: Time evolution of the estimated reproduction number.

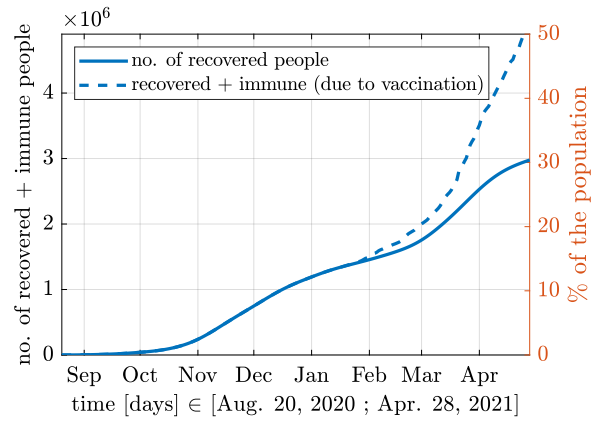


Figure 11: Estimated number of recovered or immune people.

Table 1: Relative distance of the reconstructed hospitalization data to the actual and filtered data.

$d_r(\downarrow, \rightarrow)$	\check{y}_{ls}	y_{ls}	y_{uio}
\check{y}	0.02764	0.02941	0.02541
\bar{y}	0.01476	0.01579	0.01064
y	0.01111	0.0128	0.00167

Practically, (25) is the relative distance between the two series obtained from their absolute ℓ_2 distance $\|y - \hat{y}\|_2$ and divided by the ℓ_2 norm $\|y\|_2$ of the *reference* sequence y . It is worth mentioning that the relative distance between y and the identically zero function is 1. Furthermore, note that the last *three* values of the least squares solution \check{u}_{ls} cannot be considered, since (9) is a *third-order* DT-LTI model and thus the summation in (25) is performed only up to $T - 3$.

In Table 1, we present the (pairwise) relative ℓ_2 distance between the measurement \check{y} , \bar{y} , y and the reconstructed hospitalization data \check{y}_{ls} , y_{ls} , y_{uio} . The results show a slightly smaller relative error when an unknown-input observer is applied. However, it is worth remarking that the preliminary filtering steps of Section 3.1 are essential when the derivative-based observer filter (19) is applied.

4.2 Discussion of the obtained results

Fig. 7 shows the computed number of daily new infections, where two distinct waves (called the 2nd and 3rd waves, respectively) can be observed. The peaks of this curve during November-December 2020 and March 2021 match the available data, and the computation suggests a roughly 4-fold ratio between the model computed and officially detected new cases. The effect of the restrictions (closure of restaurants and gyms, online education in secondary schools, banning of most gatherings, curfew from 8pm to 5am, etc.) introduced in the first half of November is clearly visible. In Fig. 8, the sum of all the infected compartments (**L**, **P**, **I**, **A**, **H**) is shown. The obtained numbers for October-November are comparable to the results of a nationwide testing in Slovakia at the end of October 2020, where about 1.06 percent of the population proved to be COVID-positive [19]. The estimated reproduction numbers are visible in Fig. 9. In February 2021, R_c increased above 1 again which indicates the spread of a more contagious variant (B.1.1.7) causing the large 3rd wave although the November restrictions were not at all relaxed at that time. The ratio of the peaks of the estimated R_0 in December 2020 and late February 2021 is 1.618, which is completely in agreement with the 1.4 - 1.8 interval reported in the UK [20]. On March 8, 2021, further measures were introduced including the closing of all schools which clearly contributed to the decrease of the number of daily infections from the second half of March. Fig. 11 illustrates the estimated number of recovered people. The computations suggest that approximately 30 percent of the Hungarian population might have already gone through the disease. This result is extended with a straightforward uncertainty analysis (illustrated in Fig. 12). For each constant model parameter in (1a) - (1h) we assumed an uncertainty interval of $\pm 20\%$ around the nominal value. 5000 simulation runs were performed with uniformly distributed random parameters taken from the multidimensional uncertainty interval. The mean value of the sum of the estimated number of recovered people is 4.0 times higher on April 28, 2021 than the official cumulative number of COVID-positive cases mentioned in the Introduction. This multiplier is comparable to [4], where the calculated ratio of probable total and detected cases is between 3.93 and 7.94 for 10 European countries not including Hungary, with a mean of 4.85. For this comparison, we have to take into consideration that testing intensity was significantly increased during the 2nd and 3rd waves.

5 Conclusions

The data of the COVID-19 pandemic in Hungary was studied in this paper using a systems theoretic approach. The approach is based on the dynamic inversion of a linear subsystem of the used compartmental model. We have shown two possible approaches leading to similar results to track the non-measured state variables, which then allows the estimation of the time dependent reproduction numbers. We remark that the modeling of vaccination is not necessary to give an estimation for the

compartments containing infected people, although it is definitely required to estimate the number of susceptible people and the time-dependent reproduction number during the 3rd wave. The described approach not only gives an estimate for the cumulative number of infected and recovered people but it is also able to unravel information on the whole time-course of the epidemic process. The obtained results fall into the ranges that have been published in the literature for different European countries. Future work will be focused on the extension of the method to take model uncertainty into consideration in a systematic way.

References

- [1] Ministry of Health of the Czech Republic, “COVID-19: Overview of the current situation in the Czech Republic,” <https://onemocneni-aktualne.mzcr.cz/covid-19>, (Accessed on 04/28/2021).
- [2] Government of Hungary, “3 million 774 thousand vaccinated, 1,692 new infected, 188 patients deceased,” <https://koronavirus.gov.hu/cikkek/3-millio-774-ezer-beoltott-1692-az-uj-fertozott-elhunyt-188-beteg>, (Accessed on 04/28/2021).
- [3] S. J. Phipps, R. Q. Grafton, and T. Kompas, “Robust estimates of the true (population) infection rate for COVID-19: A backcasting approach,” *Royal Society Open Science*, vol. 7, no. 11, p. 200909, 2020.
- [4] I. Rocchetti, D. Böhning, H. Holling, and A. Maruotti, “Estimating the size of undetected cases of the COVID-19 outbreak in Europe: An upper bound estimator,” *Epidemiologic Methods*, vol. 9, no. s1, 2020.
- [5] “Hungary coronavirus (live): 391 170 cases — The coronavirus app,” <https://coronavirus.app/tracking/hungary>, (Accessed on 04/26/2021).
- [6] B. Merkely, A. J. Szabó, A. Kosztin, E. Berényi, A. Sebestyén, C. Lengyel, G. Merkely, J. Karády, I. Várkonyi, C. Papp *et al.*, “Novel coronavirus epidemic in the Hungarian population, a cross-sectional nationwide survey to support the exit policy in Hungary,” *GeroScience*, vol. 42, no. 4, pp. 1063–1074, 2020.
- [7] G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, “Forecasting models for coronavirus disease (covid-19): a survey of the state-of-the-art,” *SN Computer Science*, vol. 1, no. 4, pp. 1–15, 2020.
- [8] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, “Covid-19 pandemic prediction for hungary; a hybrid machine learning approach,” *Mathematics*, vol. 8, no. 6, p. 890, 2020.
- [9] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, and P. M. Atkinson, “Covid-19 outbreak prediction with machine learning,” *Algorithms*, vol. 13, no. 10, p. 249, 2020.
- [10] E. D. Sontag, *Mathematical control theory: Deterministic finite dimensional systems*. Springer Science & Business Media, 2013, vol. 6.
- [11] L. Lennart, “System identification: Theory for the user,” *PTR Prentice Hall, Upper Saddle River, NJ*, vol. 28, 1999.
- [12] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*. Springer, 2012, vol. 2.
- [13] T. Péni, B. Csutak, G. Szederkényi, and G. Röst, “Nonlinear model predictive control with logic constraints for COVID-19 management,” *Nonlinear Dynamics*, vol. 102, no. 4, pp. 1965–1986, Dec. 2020.

- [14] G. Röst, F. A. Bartha, N. Bogyá, P. Boldog, A. Dénes, T. Ferenci, K. J. Horváth, A. Juhász, C. Nagy, T. Tekeli *et al.*, “Early phase of the COVID-19 outbreak in Hungary and post-lockdown scenarios,” *Viruses*, vol. 12, no. 7, p. 708, 2020.
- [15] “Data on hospital and ICU admission rates and current occupancy for COVID-19,” <https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19>, (Accessed on 04/28/2021).
- [16] J. Chen and R. Patton, *Robust model-based fault diagnosis for dynamic systems*. Boston, MA: Kluwer Academic Publisher, 1999.
- [17] J. Moreno, “Existence of unknown input observers and feedback passivity for linear systems,” in *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*, vol. 4, Orlando, FL, USA, Dec. 2001, pp. 3366–3371 vol.4.
- [18] Government of Hungary, “Information page on the coronavirus,” <https://koronavirus.gov.hu>, (Accessed on 04/26/2021).
- [19] E. Holt, “COVID-19 testing in Slovakia,” *The Lancet. Infectious Diseases*, vol. 21, no. 1, p. 32, 2021.
- [20] E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O’Toole *et al.*, “Transmission of SARS-CoV-2 lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data,” *medRxiv*, pp. 2020–12, 2021.