Article

Reconstruction of Epidemiological Data in Hungary Using Stochastic Model Predictive Control

Péter Polcz [1,2,*], **Balázs Csutak [1,2]** and **Gábor Szederkényi [1,2,*]**

1 Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/a, H-1083 Budapest, Hungary; csutak.balazs@itk.ppke.hu

2 Systems and Control Laboratory, Institute for Computer Science and Control, Kende u. 13-17, H-1111 Budapest, Hungary

* Correspondence: polcz.peter@itk.ppke.hu (P.P.); szederkenyi@itk.ppke.hu (G.S.)

Abstract: In this paper, we propose a model-based method for the reconstruction of not directly measured epidemiological data. To solve this task, we developed a generic optimization-based approach to compute unknown time-dependent quantities (such as states, inputs, and parameters) of discrete-time stochastic nonlinear models using a sequence of output measurements. The problem was reformulated as a stochastic nonlinear model predictive control computation, where the unknown inputs and parameters were searched as functions of the uncertain states, such that the model output followed the observations. The unknown data were approximated by Gaussian distributions. The predictive control problem was solved over a relatively long time window in three steps. First, we approximated the expected trajectories of the unknown quantities through a nonlinear deterministic problem. In the next step, we fixed the expected trajectories and computed the corresponding variances using closed-form expressions. Finally, the obtained mean and variance values were used as an initial guess to solve the stochastic problem. To reduce the estimated uncertainty of the computed states, a closed-loop input policy was considered during the optimization, where the state-dependent gain values were determined heuristically. The applicability of the approach is illustrated through the estimation of the epidemiological data of the COVID-19 pandemic in Hungary. To describe the epidemic spread, we used a slightly modified version of a previously published and validated compartmental model, in which the vaccination process was taken into account. The mean and the variance of the unknown data (e.g., the number of susceptible, infected, or recovered people) were estimated using only the daily number of hospitalized patients. The problem was reformulated as a finite-horizon predictive control problem, where the unknown time-dependent parameter, the daily transmission rate of the disease, was computed such that the expected value of the computed number of hospitalized patients fit the truly observed data as much as possible.

Keywords: dynamical systems; state estimation; model predictive controller; epidemiological models

1. Introduction

The recent and still ongoing COVID-19 pandemic has brought unprecedented challenges for most countries to protect human lives and operate the economy and society at an acceptable level at the same time [1,2]. In supporting the related difficult decisions, dynamical modeling of the epidemic process has had a key importance in all developed societies [3]. Depending on the modeling goal, a wide range of computational techniques is available to describe, forecast [4], and even control the epidemic process [5]. Here, we can only mention a few selected references from the related extensive literature. The majority of the modeling solutions are based on deterministic compartmental description derived from susceptible–exposed–infected–recovered (SEIR)-type models [6–8]. Due to the rapidly increasing computing power, agent-based models have also become popular for modeling epidemics and studying control possibilities [9,10]. Logistic wavelets were used in [11] to

compute the possible cumulative number of people infected with COVID. The application of artificial intelligence and machine learning has also been successful in COVID-related modeling and prediction [12–14].

The online tracking of informative epidemic parameters and variables proved to be essential for the continuous monitoring and evaluation of the situation, making predictions, or planning interventions. A key parameter in such analyses is the time-varying reproduction number (R_t) which is the population-level transmission potential of the disease at time t, which is known to be non-trivial to estimate [15]. Therefore, several different computational approaches have been proposed to track this quantity for epidemic outbreaks [16,17]. In [18], the authors fit a stochastic compartmental model defined by a random walk to estimate the time-dependent reproduction number of the COVID-19 epidemic in Wuhan from publicly available datasets. A discrete-time Hawkes process was used for the estimation of R_t in [19], which allowed the detection of events such as restrictions and their relaxation. The estimation of other non-measured variables such as the number of people in latent or asymptomatic stages is also a relevant problem in data reconstruction [20]. In [21], a state estimator with proven convergence was proposed to implement model predictive control (MPC) satisfying complex constraints for an eight-compartment model of the COVID-19 pandemic. Essentially the same model was used in [22] for an inversion-based estimation of R_t from Hungarian data. Similar to [21], an MPC approach was presented in [23] to determine optimal social distancing rules (and hence, R_t) to mitigate the epidemic.

The majority of the data analysis approaches use primarily the daily infected data possibly together with the recovery statistics. However, it is well known that only a fraction of the true cases are detected, which also depends on the testing intensity [24–26]. Moreover, the recording of recoveries is also often not immediate and sometimes not precise enough as well. Therefore, similar to [22,27], we used the official data on the daily number of hospitalized people in Hungary, assuming that current testing rules and protocols in hospitals give sufficiently reliable and timely information. The basic system theoretic idea behind our proposed solution is that the transmission parameter β, which is closely related to R_t, can be considered as an input of a nonlinear system describing epidemic spread, and its estimation can be traced back to a trajectory-tracking control problem, where the output to be tracked is the true reported number of hospitalized people. A straightforward choice for the solution is MPC, where we can computationally handle model nonlinearities and parametric uncertainties [28].

The theory of optimization- and prediction-based simultaneous state and parameter estimation is well founded in the literature [29–32], and it is applied in a wide spectrum of sciences, e.g., in geosciences [30], in medicine [33], in agriculture [34,35], in aerospace [36], or in meteorology [37]. The classical approaches [30,37] use variational principles and consider continuous-time models. Due to its simplicity and transparency, the MPC approaches with discrete-time model descriptions are widely used to solve optimal filtering problems, e.g., [32–35] addressed model predictive data assimilation, i.e., optimal state/parameter reconstruction to minimize the deviation between the measurement and model output. In [36], an optimal design parameter was computed through a constrained MPC for a small satellite system, which provides constraint satisfaction during operation time. Furthermore, optimal dosing of cancer chemotherapy was addressed in [33] by solving a predictive control problem with joint state and parameter estimation.

In general, the nonlinear MPC (NMPC) approaches result in a cumbersome optimization problem, especially when the model equations are stochastic. However, the available sequential convex programing approaches [38,39], the algorithmic differentiation techniques [40,41], and the numerical software tools [42,43] exploit the special structure of a typical MPC problem and provide an efficient toolkit to solve the nonlinear problems precisely in a reasonable time. Among the several approaches to cope with stochastic dynamics [44], we mention two groups of techniques, which are popular in control theory. First, the particle-based approaches [45–49] with scenario trees allow coping with general (not necessarily Gaussian) models. Secondly, the tube-based approaches [50–53] approxi-

mate each predicted state and input by a Gaussian distribution. Therefore, the dynamic equations in these references are recast as a deterministic mean-variance recursion.

In this paper, we propose a generic optimization-based approach to reconstruct epidemiological data through the approximation of unknown time-dependent quantities (such as the state, unknown input, or parameter) of a class of discrete-time nonlinear stochastic dynamical models by a sequence of Gaussian distributions. The problem was formulated as a single stochastic NMPC (SNMPC) computation. However, the solution of an SNMPC over a relatively large prediction horizon is challenging. Therefore, the problem was solved in multiple steps. First, we approximated the expected input and state trajectories through a nonlinear deterministic problem. If the expected values of the unknown quantities are fixed, the variance matrix of the joint distribution is well defined. An optional state feedback gain computation is proposed to reduce the solution's conservatism, i.e., the estimated standard deviation of the unknown quantities. Finally, the computed mean and variance values serve as an initial solution for the SNMPC problem, which ensures a fast convergence for the nonlinear optimization.

The paper is organized as follows. First, we describe the applied compartmental model for the COVID-19 epidemic spread in Section 2. Then, we introduce our computational approach in two steps in Sections 3 and 4. The numerical results with a discussion are presented in Section 5.

Notations and Abbreviations

All random variables are distinguished from the deterministic variables in notation by the accent $\hat{\bullet}$. Namely, $\hat{x}$ is a random variable, whereas, x is deterministic. When $\hat{x}$ is normally distributed with expected value μ and variance Σ, we write that $\hat{x} \sim \mathcal{N}(\mu, \Sigma)$. When $\hat{x} \sim \mathcal{N}(\mu, \sigma^2)$ is a scalar-valued Gaussian variable, the confidence intervals $\mu \pm 1\sigma = [\mu - 1\sigma, \mu + 1\sigma]$ and $\mu \pm 2\sigma = [\mu - 2\sigma, \mu + 2\sigma]$ of probability levels 68.2% and 95.4% are called the 1σ and 2σ confidence intervals, respectively. The value of a time series $x : \{0, 1, \dots\} \to \mathbb{R}^n$ at time instant k is denoted by x_k. Each constant or variable, which denotes a given number of people, is denoted by a boldface letter, e.g., $\mathbf{N}$ constitutes the number or people in a community or the population of a country. When $A \in \mathbb{R}^{n \times m}$ is a matrix, $\mathrm{He}\{A\}$ stands for $A^\top + A$, where $A^\top$ is the transposition of A. Let $x = (x_1, x_2, \dots, x_n)$ denote $x = \left(x_1^\top \ x_2^\top \ \dots \ x_n^\top \right)^\top$. The matrix-valued functions $\frac{\partial f}{\partial x} : \mathbb{R}^{n+m} \to \mathbb{R}^{p \times n}$ and $\frac{\partial f}{\partial (x,y)} : \mathbb{R}^{n+m} \to \mathbb{R}^{p \times (n+m)}$ (with arguments $(x, y) \in \mathbb{R}^{n+m}$) denote the Jacobian of function $f : \mathbb{R}^{n+m} \to \mathbb{R}^p$ with respect to (w.r.t.) $x \in \mathbb{R}^n$ and $(x, y) \in \mathbb{R}^{n+m}$, respectively. Furthermore, the value of $\frac{\partial f}{\partial x}$ at $x_0 \in \mathbb{R}^n$ and $y_0 \in \mathbb{R}^m$ is referred to as $\frac{\partial f}{\partial x}(x_0, y_0) \in \mathbb{R}^{p \times n}$. The Euclidean norm of a vector $x \in \mathbb{R}^n$ is $\|x\|$, whereas the weighted norm of x w.r.t. the symmetric and positive definite matrix Q is referred to as $\|x\|_Q$, namely $\|x\|_Q^2 = x^\top Q x$. Finally, let $\mathbb{I}_a^b = \{a, a+1, \dots, b\}$ denote the set of integers between a and b.

2. Compartmental Model of the Spread of the COVID-19 Epidemic in Hungary

In this section, we present the compartmental model describing our knowledge on the dynamics of disease spreading.

2.1. Transitions between the Phases of the Disease

To capture the spread and the evolution of the COVID-19 epidemic, we considered a modified version of the compartmental model introduced in [21]. This model divides the population of $\mathbf{N}$ individuals into eight classes, representing the different stages of the illness. The compartments of the model correspond to the following subsets/groups of the population: susceptible individuals ($\mathbf{S}$), infected people in the latent ($\mathbf{L}$) and the pre-symptomatic ($\mathbf{P}$) phases, infected people in the main sequence of the disease ($\mathbf{A}$, $\mathbf{I}$), infected people who need hospital treatment ($\mathbf{H}$), and finally, the recovered ($\mathbf{R}$) and deceased ($\mathbf{D}$) people. The main phase of the disease is further divided into those who remain asymptomatic ($\mathbf{A}$) and those who produce symptoms ($\mathbf{I}$).

In this work, the model of [21] was complemented with a new compartment (**U**) comprising all individuals who became immune through vaccination. The members of this compartment are governed by the daily number of vaccinated people (**V**). A discrete-time (DT) version of the augmented continuous-time model was obtained through the explicit Euler method with a 1 d-long sampling period. The dynamic equations are given as follows:

$$
\begin{cases}
\mathbf{S}_{k+1} = \mathbf{S}_k - \beta_k(\mathbf{P}_k+\mathbf{I}_k+\delta\mathbf{A}_k)\mathbf{S}_k/\mathbf{N} - \nu\,\frac{\mathbf{S}_k}{\mathbf{S}_k+\mathbf{R}_k}\mathbf{V}_k, & \text{(1a)} \\[2mm]
\mathbf{L}_{k+1} = \mathbf{L}_k + \beta_k(\mathbf{P}_k+\mathbf{I}_k+\delta\mathbf{A}_k)\mathbf{S}_k/\mathbf{N} - \alpha\,\mathbf{L}_k, & \text{(1b)} \\[2mm]
\mathbf{P}_{k+1} = \mathbf{P}_k + \alpha\,\mathbf{L}_k - \zeta\,\mathbf{P}_k, & \text{(1c)} \\[2mm]
\mathbf{I}_{k+1} = \mathbf{I}_k + \gamma\,\zeta\,\mathbf{P}_k - \rho_I\mathbf{I}_k, & \text{(1d)} \\[2mm]
\mathbf{A}_{k+1} = \mathbf{A}_k + (1-\gamma)\,\zeta\,\mathbf{P}_k - \rho_A\,\mathbf{A}_k, & \text{(1e)} \\[2mm]
\mathbf{H}_{k+1} = \mathbf{H}_k + \rho_I\,\eta\,\mathbf{I}_k - \lambda\,\mathbf{H}_k, & \text{(1f)} \\[2mm]
\mathbf{R}_{k+1} = \mathbf{R}_k + \rho_I(1-\eta)\,\mathbf{I}_k + \rho_A\,\mathbf{A}_k + (1-\mu)\,\lambda\,\mathbf{H}_k - \nu\,\frac{\mathbf{R}_k}{\mathbf{S}_k+\mathbf{R}_k}\mathbf{V}_k, & \text{(1g)} \\[2mm]
\mathbf{D}_{k+1} = \mathbf{D}_k + \mu\,\lambda\,\mathbf{H}_k, & \text{(1h)} \\[2mm]
\mathbf{U}_{k+1} = \mathbf{U}_k + \nu\,\mathbf{V}_k. & \text{(1i)}
\end{cases}
$$

The transitions between the compartments are illustrated in Figure 1. It is worth mentioning that the recovered's compartment **R** contains the recovered people who are not yet vaccinated. To represent all the recovered people including the vaccinated, we can consider the following additional equation:

$$
\mathbf{R}_{k+1}^{(\mathrm{all})} = \mathbf{R}_k^{(\mathrm{all})} + \rho_I(1-\eta)\,\mathbf{I}_k + \rho_A\,\mathbf{A}_k + (1-\mu)\,\lambda\,\mathbf{H}_k, \tag{2}
$$

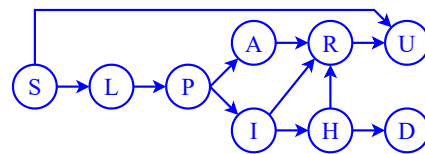which clearly does not affect the dynamics of other states.

Figure 1. Transition graph of the epidemic model. Compartments and transitions are represented by the nodes and edges, respectively.

The detailed worldwide [54] and local serological [55] and dynamical [7,20,56,57] analysis results provide estimates for the average lengths of the phases of the illness and the probabilities of transitions between the compartments. These parameters are aligned with the Hungary-specific data and were presented in detail in [21]. After infection, the latent period of the disease (**L**) usually lasts approximately $\alpha^{-1} = 2.5$ d. This period is followed by a pre-symptomatic phase (**P**) of $\zeta^{-1} = 3$ d. A person in the main sequence of the disease (**I** or **A**) remains infectious for about $\rho_I^{-1} = \rho_A^{-1} = 4$ d. The empirical probability of producing symptoms in the main sequence is $\gamma = 0.6$; furthermore, an $\eta = 0.076$ portion of the symptomatic cases require hospitalization. The average length of a hospital treatment is $\lambda^{-1} = 10$ d. A hospitalized patient dies with a probability of $\mu = 0.205$, or recovers. The recovered people are assumed to be immune to reinfection. We assumed that the disease is transmitted by the members of compartments **P**, **A**, and **I**, such that the relative infectiousness of the asymptomatic individuals (**A**) is $\delta = 0.75$, compared to those who produce symptoms (**I**). The transmission rate β of the disease is the most prominent parameter of the epidemic spread, which is typically time-dependent. The nominal values of the above-mentioned model parameters and their assumed uncertainty are collected in Table 1. For simplicity, each uncertain parameter was assumed to have a normal distribution, such that its nominal value (μ) is the expectation and its uncertainty ($\pm a\%$) gives the 2σ confidence interval with the standard deviations $\sigma = a\mu/200$.

Table 1. Presumed model parameters with their short description and their estimated uncertainty.

Description	Nominal Value	Uncertainty
The population of Hungary	$\mathbf{N} = 9.8 \times 10^6$	
Inverse of … (1/day)		
latent period	$\alpha = 1/2.5$	$\pm 20\%$
presymptomatic infectious period	$\zeta = 1/3$	$\pm 30\%$
infectious period of symptomatic individuals	$\rho_I = 1/4$	$\pm 25\%$
infectious period of asymptomatic individuals	$\rho_A = 1/4$	$\pm 25\%$
average length of hospitalization	$\lambda = 1/10$	$\pm 10\%$
Relative infectiousness of asymptomatic	$\delta = 0.75$	$\pm 10\%$
Probability of developing symptoms	$\gamma = 0.6$	$\pm 10\%$
Hospitalization probability of symptomatic cases	$\eta = 0.076$	$\pm 10\%$
Probability of fatal outcome (if already hospitalized)	$\mu = 0.205$	$\pm 10\%$
Effectiveness of vaccination	$\nu = 0.75$	$\pm 10\%$

Here, we examined the evolution of the epidemic in a fixed time window between 1 March 2020 ($k = 0$) and 30 June 2021 ($k = T$). This interval contains the first three waves of the epidemic in Hungary. Subscript $k \in \{0, 1, \ldots, T\}$ denotes the number of days elapsed in the given time window of length $T + 1$.

2.2. Vaccination Model

For simplicity, our vaccination model assumed that only susceptible and recovered people are eligible for vaccination, and we neglected those rare cases when the shot is given during an unidentified infection. Based on the serological test data [58], our model assumed that a subject becomes immune $T_V = 21$ d after the first dose with an average probability of $\nu = 0.75$. Correspondingly, variable $\mathbf{V}_k$ in (1) denotes the number of individuals who received the first dose of vaccine on day $k - T_V$. In our model,

> an individual is said to be *immune* *to the disease if he/she will not be infected within the modeled time horizon.*

With this simplification, the people in the **R** and **U** compartments do not transmit the disease any more, as well as those who are still in the hospital. It is worth remarking that a positive IgG test does not necessarily ensure immunity in this sense. Serological tests suggest that even a relatively high IgG level does not exclude the possibility of reinfection [59].

On the other hand, we assumed that the willingness of the susceptible and recovered patients to vaccinate is roughly the same. Namely, the proportion of susceptible and recovered people vaccinated coincides with the proportion of all susceptibles and recovereds on each day. Therefore, the model counts with $\nu\, \mathbf{S}_k \mathbf{V}_k / (\mathbf{S}_k + \mathbf{R}_k)$ susceptible and $\nu\, \mathbf{R}_k \mathbf{V}_k / (\mathbf{S}_k + \mathbf{R}_k)$ recovered individuals who achieved immunity at time k. It is worth remarking that the value of $\mathbf{V}_k$ is known, but cannot be manipulated as the computations were performed on *past* data of the epidemic spread. Correspondingly, $\mathbf{V} : k \mapsto \mathbf{V}_k$ can be considered as a preliminarily known input, or a *scheduling variable* [60], or a *measured disturbance* [36,61]. The official European vaccination data including Hungary are available at [62]. Official Hungarian COVID-related data with additional analyses are also available at [63].

2.3. Computing the Reproduction Number

To give meaningful estimates from an epidemiological perspective, we computed the basic and the time-dependent effective reproduction numbers. The basic reproduction number, namely the average number of new infections generated by a single infected individual in a fully susceptible population, can be given by the following closed-form expression [21]:

$$R_0 = \beta \left(\frac{1}{\zeta} + \frac{\gamma}{\rho_I} + \frac{\delta(1-\gamma)}{\rho_A} \right), \tag{3}$$

It must be noted that using $R_0 = 2.2$, an early estimate for Hungary commonly used in the literature, and expressing the transmission rate β from the equation, a nominal value of $\beta = 1/3$ can be given. However, as this parameter is highly influenced by the circumstances varying in time (e.g., the stringency of restrictions, new variants of the virus, etc.), it is considered to be a time-dependent parameter and estimated as such (the corresponding R_0 being calculated afterwards).

The time-dependent effective reproduction number R_t shows the average number of infections caused by a single individual, given the state of the model at time t (thus, taking into account the time-varying nature of beta and the decrease of the susceptible population). This is calculated as follows,

$$R_t = \beta_t \left(\frac{1}{\zeta} + \frac{\gamma}{\rho_I} + \frac{\delta(1-\gamma)}{\rho_A} \right) \frac{\mathbf{S}_t}{\mathbf{N}} \quad \text{with} \quad t \in \{0, 1, \dots, T\} \tag{4}$$

It can be a base for comparison between different epidemic-handling strategies, incorporating the strictness of the restrictions as well. In accordance with the traditional notation "R_t", we used $t \in \mathbb{I}_0^T$ (instead of k) as the time parameter of the time-dependent reproduction number.

Inspired by [21], we considered the daily transmission rate β_k of the disease an unknown time-dependent parameter. The past values of β_k were computed such that the evolution of the epidemic spread matched the available observations.

2.4. Available Measurements

It is commonly stated in the literature [24,25] that the daily number of infected people is not well observable, as the measurement relies on aggressive and exhaustive contact tracing and testing strategies [64,65]. Though it is reasonable to assume that testing is wide-spread and quick enough in the hospitals, the registered numbers of hospitalized patients [66] are still influenced by practical considerations. The limited healthcare capacity on weekends and holidays usually results in a lower number of performed and documented tests, as well as in a delayed hospital discharge. Therefore, following the common engineering practice, we applied a 7 d-long moving average filter to the published data ($\mathbf{H}_k^{\text{Off,raw}}$) [66] to avoid biased estimates caused by these administrative inaccuracies. The smoothed time series $\mathbf{H}_k^{\text{Off}}$ is formally calculated as:

$$\mathbf{H}_k^{\text{Off}} = \frac{1}{\min(3, T-k) + \min(3, k) + 1} \sum_{i=-\min(3,k)}^{\min(3, T-k)} \mathbf{H}_{k+i}^{\text{Off,raw}}. \tag{5}$$

Obviously, the 7 d-long sliding window must be truncated at both ends of the series. Finally, the filtered hospitalization data were considered as the single available processed measurement, which reveal relevant information about the time-evolution of the process.

3. Optimization-Based Reconstruction of Past Epidemiological Data

State-Space Model Representation and Problem Statement

In Section 2.1, we presented the dynamic Equation (1) of the epidemic spread. We introduced a possible vaccination model in Section 2.2, where $\mathbf{V}_k$ acts as a *measured disturbance* input of the dynamical model. In Section 2.3, we explained why parameter β_k can be considered as an *unknown input* of the system. Finally, in Section 2.4, we proposed to consider the hospitalization data ($\mathbf{H}$) as the model output. These "ingredients" allowed us to embed the epidemic spread model into the following discrete-time state-space representation:

$$x_{k+1} = f(x_k, u_k, \theta, v_k), \quad y_k = C x_k, \tag{6}$$

where $x_k = (\mathbf{S}_k, \mathbf{L}_k, \mathbf{P}_k, \mathbf{I}_k, \mathbf{A}_k, \mathbf{H}_k, \mathbf{R}_k, \mathbf{D}_k, \mathbf{U}_k) \in \mathbb{R}^n$ is the state, $u_k = \beta_k \in \mathbb{R}^m$ is the unknown input, $\theta = (\alpha, \zeta, \rho_I, \rho_A, \lambda, \delta, \gamma, \eta, \mu, \nu) \in \mathbb{R}^p$ is the vector of model parame-

ters, $v_k = \mathbf{V}_k \in \mathbb{R}^q$ is a measured disturbance, and $y_k = \mathbf{H}_k \in \mathbb{R}^s$ is the output with $C = (0\,0\,0\,0\,0\,1\,0\,0\,0)$.

In [22], we presented two possible linear time-invariant (LTI) methods to reconstruct the past states x_{k+1} and the unknown inputs u_k using the measured output y_k, $k = \mathbb{I}_0^{T-1}$. Both techniques in [22] rely on the dynamic inversion of the LTI subsystem of Model (1). In this paper, we revisited the unknown input filtering problem and reformulated it as an optimal predictive tracking control problem. Namely, we computed an optimal input sequence u_k, $k = \mathbb{I}_0^{T-1}$ such that the output y_k of the system follows the reference signal $r_k = \mathbf{H}_k^{\mathrm{Off}}$, which contains the past output measurements, i.e., the daily number of hospitalized patients ($\mathbf{H}_k^{\mathrm{Off}}$). The simultaneous unknown input and state reconstruction can be formulated as the following optimization task.

Problem 1 (NMPC for epidemiological data reconstruction with fixed model parameters)**.** *Given the dynamical model* (6) *with initial condition x_0, a vector of constants θ, a measured disturbance v_k, and a reference output trajectory r_{k+1} to track ($k \in \mathbb{I}_0^{T-1}$), we looked for a sequence of inputs u_k and states x_{k+1} that solve the state recursion* (6), *satisfy the constraint $u_k \in \mathcal{U}$, and minimize the following weighted cost function:*

$$J(X, U) = \sum_{k=0}^{T-1} \|Cx_{k+1} - r_{k+1}\|_Q^2 + \sum_{k=0}^{T-2} \|u_{k+1} - u_k\|_R^2, \tag{7}$$

where $X = \begin{pmatrix} x_1 & \dots & x_T \end{pmatrix}$ and $U = \begin{pmatrix} u_0 & \dots & u_{T-1} \end{pmatrix}$ collect the free decision variables of the optimization, Q, R are positive definite weight matrices, and $\mathcal{U}$ is a closed subset of the input space $\mathbb{R}^m$.

In Problem 1, we formulated a *data assimilation* problem in the form of a *nonlinear model predictive controller* (NMPC) computation. The available numerical optimization tools [40–43] make it possible to solve Problem 1 precisely in a reasonable time. From an epidemiological point of view, the first term of the cost function (7) minimizes the deviation of the computed number of hospitalized patients from the official data, whereas the second term minimizes the slope of the transmission rate of the pathogen. In this way, the NMPC design provides an optimal *smooth* solution for the unknown transmission rate function $\beta : \{0, \dots, T-1\} \rightarrow \mathcal{U} = [0.06, 1]$, which does not have sudden changes.

Remark 1. *The daily transmission rate of the disease β_k is an unknown time-dependent (but, supposedly not abruptly varying) parameter. During an outbreak of the epidemic, the number of infected people is not negligible, namely the sum $\mathbf{P}_k + \mathbf{I}_k + \delta\mathbf{A}_k$ is significant. In this case, the transmission rate function $\beta : k \mapsto \beta_k$ influences the overall dynamics significantly and determines the shape of the epidemic wave. Therefore, parameter β_k is generally well identifiable from the measurements during an outbreak of the epidemic, and it becomes uncertain when the number of infectious people is small.*

The unknown input filtering task becomes complicated when the model parameters and the initial state (from which the state reconstruction was performed) are probabilistic variables. In the next section, we address the stochastic extension of Problem 1.

4. Statistical Analysis for Normally Distributed Model Parameters

4.1. Gaussian Assumptions

In this section, we allowed the model parameters to vary in time ($\hat{\theta} : k \mapsto \hat{\theta}_k$), but we assumed that the parameter process $\hat{\theta}$ is a collection of *independent identically distributed* (i.i.d.) Gaussian random variables:

$$\hat{\theta}_k \sim \mathcal{N}(\mu^\theta, \Sigma^\theta) \text{ for all } k = 0, 1, \dots, T-1, \tag{8}$$

where μ^θ is the expected value of $\hat{\theta}_k$ corresponding to the values presented in Section 2.1 and the diagonal Σ^θ is its variance. The expected value of the parameter vector contains

the nominal values from Table 1, whereas the variances are determined such that the uncertainty intervals from Table 1 resemble the 2σ confidence intervals. Moreover, we assumed that the initial state, from which the prediction was performed, is itself a random variable, namely:

$$\hat{x}_0 \sim \mathcal{N}(\mu_0^x, \Sigma_0^x). \tag{9}$$

Consequently, every further state and output are random variables, which obey the following stochastic recursion and output equation:

$$\hat{x}_{k+1} = f(\hat{x}_k, \hat{u}_k, \hat{\theta}_k, v_k), \quad \hat{y}_k = C\hat{x}_k. \tag{10}$$

Due to the nonlinear terms in the state transition function f, the distribution of the predicted states $\hat{x}_k$ becomes more and more complicated as we look forward in time ($k = 1, 2, \dots$). Therefore, it is very inefficient to compute or at least approximate the non-Gaussian probability density functions of the predicted states for the nonlinear stochastic model (10). As is commonly done in the literature (see, e.g., [50–53]), we performed a tube-like trajectory estimation. With this technique, each predicted state $\hat{x}_k$ is described by the first two moments, the expected value μ_k^x, and the variance Σ_k^x, namely the states are approximated by normal distributions:

$$\hat{x}_k \sim \mathcal{N}(\mu_k^x, \Sigma_k^x) \text{ for all } k = 0, 1, 2, \dots \tag{11}$$

4.2. Closed-Loop Control Policy

In the literature [44], the values of the optimal control input u are often searched as functions of the states as follows:

$$\hat{u}_k = \mu_k^u - K(\mu_k^x)(\hat{x}_k - \mu_k^u), \tag{12}$$

where μ_k^u are free decision variables and K is (not necessarily a closed-form) function of the expected state. Thus, the control input is inherently a random variable and is normally distributed as the state (11) itself is approximated by a Gaussian. If $\Sigma_k^{x\theta} = (\Sigma_k^{\theta x})^\top$ denotes the covariance between $\hat{x}_k$ and $\hat{\theta}_k$, the joint distribution of $\hat{x}_k$, $\hat{u}_k$, and $\hat{\theta}_k$ is:

$$\begin{pmatrix} \hat{x}_k \\ \hat{u}_k \\ \hat{\theta}_k \end{pmatrix} \sim \mathcal{N}(\mu_k, \Sigma_k), \text{ where } \mu_k = \begin{pmatrix} \mu_k^x \\ \mu_k^u \\ \mu^\theta \end{pmatrix}, \; \Sigma_k = (\star)^\top \begin{pmatrix} \Sigma_k^x & \Sigma_k^{x\theta} \\ \Sigma_k^{\theta x} & \Sigma^\theta \end{pmatrix} \begin{pmatrix} I & -K^\top(\mu_k^x) & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{13}$$

Remark 2 (Nonlinear state-dependent input policy). *When $K = 0$, the optimal tracking problem is said to be an open-loop MPC problem [67], whereas $K \neq 0$ results in a so-called closed-loop MPC problem [51], where the optimal input policy is parameterized by the state. A stabilizing state feedback (12) is typically useful when the predicted states are random variables, and their actual realizations may deviate from the predicted expectations. When the prediction model is stochastic, a sequence of deterministic input values ($K = 0$) may result in a diverging sequence of state variances, and hence in a conservative (overly cautious) prediction. When the input is parameterized by the state ($K \neq 0$), the adaptability of the input may reduce the uncertainty of the predicted states significantly if the feedback function (12) is determined appropriately. In this sense, the gain function quantifies the trade-off between the uncertainty of the state and the input. Unfortunately, it is not straightforward to compute a stabilizing gain function K for the nominal model (6). Later, in Section 4.5, we demonstrate that a reference state trajectory (if available) makes it possible to compute the values of K separately in each operating reference state through a classical LTI state feedback approach, e.g., a pole placement or a linear quadratic regulator (LQR) design ([68] Section 6.4.2).*

4.3. Probabilistic Cost and Input Constraint

Problem 1 with the stochastic state Equation (10) and the joint distribution (13) results in a stochastic optimal control problem, where both the cost function (7) and the input

constraint are probabilistic. Therefore, the inputs and the states are meant to be found such that they minimize the *expected cost*, namely:

$$J(M, S, V) = \sum_{k=0}^{T-1} \left\| C\mu_{k+1}^x - r_{k+1} \right\|_Q^2 + \sum_{k=0}^{T-2} \left\| \mu_{k+1}^u - \mu_k^u \right\|_R^2 \tag{14}$$

$$+ \sum_{k=0}^{T-1} \mathrm{Tr}\left(Q\, C\, \Sigma_{k+1}^x\, C^\top \right) + \sum_{k=0}^{T-2} \mathrm{Tr}\left(R\, K(\mu_{k+1}^x)\, \Sigma_{k+1}^x\, K^\top(\mu_{k+1}^x) \right)$$

$$+ \sum_{k=0}^{T-2} \mathrm{Tr}\left(R\, K(\mu_k^x)\, \Sigma_k^x\, K^\top(\mu_k^x) - R\, \mathrm{He}\Big\{ K(\mu_{k+1}^x)\, \mathrm{Cov}(\hat{x}_{k+1}, \hat{x}_k)\, K^\top(\mu_k^x) \Big\} \right).$$

$$\text{where } M = \begin{pmatrix} \mu_1^x & \dots & \mu_T^x \end{pmatrix}, S = \begin{pmatrix} \Sigma_1^x & \dots & \Sigma_T^x \\ \Sigma_1^{\theta x} & \dots & \Sigma_T^{\theta x} \end{pmatrix}, \text{ and } V = \begin{pmatrix} \mu_0^u & \dots & \mu_{T-1}^u \end{pmatrix}. \tag{15}$$

The expanded Formula (14) of the expectation of cost (7) is derived in Appendix A.

Remark 3. *The term* $\mathrm{Cov}(\hat{x}_{k+1}, \hat{x}_k) = \mathrm{Cov}(f(\hat{x}_k, \hat{u}_k, v_k, \hat{\theta}_k), \hat{x}_k)$ *in* (14) *is typically a non-quadratic function of the mean and the variance of the joint distribution* (13). *This term introduces a potential difficulty to the optimization, which is addressed later in Section 4.4.*

The conditions on the input can be formulated as chance constraints of the form $\mathrm{Pr}(\hat{u}_k \in \mathcal{U}) \geq p_u$, where p_u denotes the *probability level* of the *confidence set* $\mathcal{U}$. When $\hat{u}_k$ comprise a single input and the input domain $\mathcal{U}$ is an interval, the chance constraint $\mathrm{Pr}(\hat{u}_k \in [\underline{u}, \overline{u}]) \geq p_u$ is equivalent to the following deterministic interval constraint [51]:

$$\mu_k^u \in [\underline{u} + c, \overline{u} - c], \text{ with } c = \Phi^{-1}\left(\frac{p_u + 1}{2}\right)\sqrt{K(\mu_k^x)\, \Sigma_k^x\, K^\top(\mu_k^x)}, \tag{16}$$

where $\Phi : \mathbb{R} \to (0, 1)$ denotes the (cumulative) distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. This technique for the reformulation of a probabilistic condition is referred to as *constraint tightening* [69].

4.4. Linear Approximation of the State Dynamics around the Expectation

In the literature, there exist different stochastic sample-based optimization approaches for a predictive optimal controller design; see, e.g., [45–50]. However, these approaches are computationally tractable only for a shorter prediction horizon. Alternatively, we have the possibility to formulate deterministic recursions for the first two moments of the state vector, e.g., [51] proposed the state transition function f to be approximated by its first-order Taylor polynomial around the expected values $\mu_k = (\mu_k^x, \mu_k^u, \mu^\theta)$ of the probabilistic variables $(\hat{x}_k, \hat{u}_k, \hat{\theta}_k)$, namely:

$$\hat{x}_{k+1} \approx \frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k) \begin{pmatrix} \hat{x}_k \\ \hat{u}_k \\ \hat{\theta}_k \end{pmatrix} + f(\mu_k, v_k) - \frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k)\, \mu_k. \tag{17}$$

This approach leads to a deterministic mean-variance ("$\mu\Sigma$") dynamics, which is typically nonlinear in the free variables μ_k^x and μ_k^u.

$$\begin{cases} \mu_{k+1}^x = f(\mu_k, v_k), & \text{(18a)} \\[2mm] \Sigma_{k+1}^{x\theta} = \frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k) \begin{pmatrix} I & -K^\top(\mu_k^x) & 0 \\ 0 & 0 & I \end{pmatrix}^\top \begin{pmatrix} \Sigma_k^{x\theta} \\ \Sigma^\theta \end{pmatrix}, & \text{(18b)} \\[3mm] \Sigma_{k+1}^x = \frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k)\, \Sigma_k \left(\frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k) \right)^\top. & \text{(18c)} \end{cases}$$

Note that the linear Taylor approximation of $\hat{x}_{k+1}$ allowed us to express the non-quadratic term in the cost function (14) as follows:

$$\mathrm{Cov}(\hat{x}_{k+1}, \hat{x}_k) = \frac{\partial f}{\partial(x,u,\theta)}(\mu_k, v_k) \begin{pmatrix} I & -K^\top(\mu_k^x) & 0 \\ 0 & 0 & I \end{pmatrix}^\top \begin{pmatrix} \Sigma_k^x \\ \Sigma_k^{\theta x} \end{pmatrix} \tag{19}$$

The dynamic equations in (18) constitute a possible deterministic *prediction model* for System (10) and result in the following nonlinear optimal predictive control problem.

Problem 2 ($\mu\Sigma$-NMPC for unknown-input state reconstruction). *Given the dynamical model* (18) *with an initial state distribution* (9)*, an i.i.d. parameter process* (8)*, a measured disturbance* v_k*, an input policy* (12) *with a fixed gain function K, and a reference output trajectory* r_{k+1} *to track* ($k \in \mathbb{I}_0^{T-1}$)*, we looked for a sequence of deterministic values* μ_k^u*, state moments* μ_{k+1}^x*,* Σ_{k+1}^x*, and covariance matrices* $\Sigma_{k+1}^{x\theta}$ *with* $\Sigma_0^{x\theta} = 0$*, which solve* (18)*, satisfy the input constraint* (16)*, and minimize the cost* (14)*. The free variables of the optimization are collected in* (15)*.*

Problem 2 is a *stochastic data assimilation* problem, reformulated as an optimal predictive tracking problem with a deterministic nonlinear $\mu\Sigma$-prediction model (10). Henceforth, we refer to Problem 2 as a *Gaussian* or *mean-variance* NMPC problem ($\mu\Sigma$-NMPC). In general, the variance dynamics (18b,c) significantly increase the complexity of the control problem. If n and p denote the dimension of the state $\hat{x}_k$ and the parameter $\hat{\theta}_k$, respectively, the equations in (18) comprise $np + n(n+1)/2$ separate scalar equations, whereas the deterministic model (6) constitutes a system of n scalar equations. Therefore, the $\mu\Sigma$-NMPC in Problem 2 is typically (at least) an order of magnitude more demanding than the ordinary NMPC in Problem 1. However, an appropriate initial guess for the solution of $\mu\Sigma$-NMPC may reduce the computational complexity of the optimization substantially by providing a fast convergence of the solution.

4.5. Initial Solution for the $\mu\Sigma$-NMPC Problem

In this section, we compute a pseudo-optimal (i.e., feasible, but not necessarily optimal) solution of Problem 2, which satisfies the dynamic equations (10) and the input constraint (16), but it does not necessarily minimize the cost (14). The computed solution can be considered an initial value for the $\mu\Sigma$-NMPC problem. The solution relies on three observations.

First, observe that the mean equation in (18a) resembles the deterministic state recursion in (6) as the mean dynamics is not affected by the variances nor the state-dependent feedback gain $K(\mu_k^x)$. Therefore, Problem 2 simplifies to Problem 1 if we neglect the variances ($S = 0$) and their dynamics (18b,c) from the optimization. Accordingly, a possible guess for the expectation (M, V), which solves the mean Equation (18a), can be given by the optimal solution (X^*, U^*) of Problem 1 with initial condition $x_0 \leftarrow \mu_0^x$ and parameter vector $\theta \leftarrow \mu^\theta$.

Secondly, we note that the gain function K depends (by design) on the expected states only. This allows computing an appropriate gain K_k at each operating point $x_0^* = \mu_0^x$, x_k^*, $k \in \mathbb{I}_1^{T-1}$ along the computed mean solution. We determined K_k through the DT version of the LQR design applied to the controllable modes of the pair (A_k, B_k), where:

$$A_k = \tfrac{\partial f}{\partial x}(\mu_k, v_k), \ B_k = \tfrac{\partial f}{\partial u}(\mu_k, v_k). \tag{20}$$

Through a sequence of DT-LQR computations, we selected a static feedback gain matrix K_k at each time instant k, which minimizes the quadratic cost:

$$\sum_{t=k}^{\infty} \left(x_t^\top \mathbf{Q}_k^{\mathrm{LQR}} x_t + u_t^\top \mathbf{R}_k^{\mathrm{LQR}} u_t \right), \ \text{with } u_t = -K_k x_t. \tag{21}$$

For a DT-LTI state-space model $x_{t+1} = A_k x_t + B_k u_t$ (with $t = 0, 1, 2, \ldots$, but a fixed k), the constant gain matrix K_k can be computed through simple linear algebra operations ([68] Section 6.4.2), which were implemented in function `dlqr` of the Control System Toolbox [70] for MATLAB. When selecting the weight matrices $\mathbf{Q}_k^{\mathrm{LQR}}$ and $\mathbf{R}_k^{\mathrm{LQR}}$ of the LQR problem at time $k \in \mathbb{I}_0^{T-1}$, we needed to take into consideration that the value of K_k quantifies the trade-off between the uncertainty of $\hat{x}_{k+1}$ and $\hat{u}_k$. If the locally stabilizing gain has a higher value, the input $\hat{u}_k$ is more adaptive (hence, more uncertain), but the uncertainty of $\hat{x}_{k+1}$ is smaller. However, the chance constraint (16) does not allow the uncertainty of $\hat{u}_k$ to

increase beyond any bounds. Therefore, the gain should be selected carefully, such that it generates an input distribution satisfying (16). If the first computed value for K_k does not result in an admissible input distribution, we are allowed to compute K_k multiple times with a gradually decreasing value for $\mathbf{R}_k^{\text{LQR}}$.

Finally, with the knowledge of μ_k, v_k, $K(\mu_k^x) = K_k$, $k \in \mathbb{I}_0^{T-1}$, Σ_0^x, and $\Sigma_0^{x\theta} = 0$, we computed the variances according to (18b,c), which give the value of S in (15). If the expected values and K are fixed, the variances are well-defined by the variance Equation (18b,c). By construction, the tuple (M, S, V) is a feasible solution for Problem 2 as it satisfies both the $\mu\Sigma$-Equation (18) and the input constraint (16). The computed solution is a good initial guess for the optimal solution of Problem 2. In Algorithm 1, we summarize the proposed operations with a single input $u_k \in \mathbb{R}$ and a simple LQR weight selection.

Algorithm 1 Computing a pseudo-optimal solution for Problem 2.

1: Fix $\hat{x}_0 \sim \mathcal{N}(\mu_0^x, \Sigma_0^x)$, $\Sigma_0^{x\theta} \leftarrow 0$, μ^θ, and Σ^θ. (Optionally, fix μ_0^u.)
2: Collect data v_k and r_{k+1}, then solve Problem 1 to obtain μ_k^x and μ_k^u, where $k \in \mathbb{I}_0^{T-1}$.
3: **for** $k \in \mathbb{I}_0^{T-1}$ **do**
4:　　 $i \leftarrow 1$.
5:　　 **repeat**
6:　　　　 Compute K_k for the pair (A_k, B_k) given in (20) through a DT-LQR design
　　　　　 with weight matrices $\mathbf{Q}^{\text{LQR}} \leftarrow I_n$ and $\mathbf{R}_k^{\text{LQR}} \leftarrow 2^{i-1}$;　 $i \leftarrow i + 1$.
7:　　 **until** condition (16) is met. (If no such K_k is found, let $K_k \leftarrow 0$.)
8:　　 Compute $\Sigma_{k+1}^{x\theta}$, Σ_{k+1}^x, as given in (18b) and (18c), respectively, using $K(\mu_k^x) = K_k$.
9: **end for**

Remark 4. *From the authors' experience, the computationally demanding $\mu\Sigma$-NMPC optimization for Model (1) will generally not result in a significantly lower expected cost (14) compared to the computed pseudo-optimal solution (M, S, V).*

5. Results and Discussion

In this section, we present the numerical results we obtained through the MPC-based reconstruction of the unknown epidemiological data. The results were computed in the MATLAB environment with the Control System Toolbox [70]. For algorithmic differentiation, we used CasADi [40,41]. To solve nonlinear MPC problems, we used IPOPT [42], an interior point line search algorithm, with the MUltifrontal Massively Parallel sparse direct Solver (MUMPS) [71,72]. The MPC implementations are available online in the public repository [73].

To compute the unknown epidemiological data, we followed the operations of Algorithm 1 to find a pseudo-optimal solution for the $\mu\Sigma$-NMPC in Problem 2. On 1 March 2020 ($k = 0$), we assumed a susceptible and almost healthy population, with a small uncertainty as follows:

$$\mu_0^x = (\mathbf{N} - 40 \;\; 10 \;\; 10 \;\; 10 \;\; 10 \;\; 0 \;\; 0 \;\; 0 \;\; 0)^\top, \;\; \Sigma_0^x = \text{diag}(7, 1, 1, 1, 1, 1, 1, 1, 0). \quad (22)$$

We note that the effect of the initial number of infected people on the reconstructed state vanishes after a transient period due to the stability properties of the compartmental model (1). Furthermore, we considered pairwise independent random parameters with μ^θ and a diagonal Σ^θ as presented in Table 1. We fixed the initial value for β to $\mu_0^u = 1/3$ [21]. We solved the ordinary NMPC in Problem 1 to find the candidate mean functions for $\hat{x}$ and $\hat{u}$. Then, we computed the gain matrices and the variances of the joint distribution (13). Using the obtained *feasible* solution as an initial guess, we solved the $\mu\Sigma$-NMPC in Problem 2. We learned that the local optimum (M^*, S^*, V^*) found for Problem 2 is qualitatively the same as the initial guess (M, S, V). The relative difference in the cost obtained by the optimal and the pseudo-optimal solution is negligible, namely:

$$\frac{J(M,S,V) - J(M^*,S^*,V^*)}{J(M^*,S^*,V^*)} \approx 7.1 \times 10^{-7}, \text{ where } J(M^*, S^*, V^*) \approx 17{,}075.129. \quad (23)$$

If $\mathfrak{X}, \mathfrak{X}^* \in \mathbb{R}^{(n+m+np+n(n+1)/2)\cdot T-1}$ denote the vectors of independent variables of (M, S, V) and (M^*, S^*, V^*), respectively, the relative difference between $\mathfrak{X}$ and $\mathfrak{X}^*$ is:

$$\|\mathfrak{X} - \mathfrak{X}^*\| \, / \, \|\mathfrak{X}^*\| \approx 2.86 \times 10^{-11}, \text{ where } \|\mathfrak{X}^*\| \approx 3.1978 \times 10^{11}. \tag{24}$$

In Table 2, we present the dimensional differences between the ordinary NMPC and the $\mu\Sigma$-NMPC if the length of time window is $T = 487$ d. In Figure 2, we illustrate the computed marginal distributions of the transmission rate of the pathogen and the daily numbers of people in the different stages of the disease. The expectation for the states are presented in Plot 1 of Figure 2, which were computed through the ordinary NMPC design in Problem 1. In each of Plots 2–12 of Figures 2 and 3, the time evolution of the marginal distributions of scalar quantities are visualized, such that the shaded dark and light areas highlight the 1σ and 1σ confidence intervals, respectively. The shape of the epidemic curves clearly show the three waves of the epidemic until summer 2021.

Table 2. Computational complexity of Problems 1 and 2 illustrated through the epidemiological data assimilation case study. In this comparison, the cumulative number of recovered people $\mathbf{R}^{(\text{all})}$ as an additional state variable is *not* considered. Accordingly, the number of state variables is $n = 9$, the number of uncertain parameters is $p = 10$, and the number of inputs and the measured disturbances are $m = 1$ and $q = 1$, respectively. * The processing time was measured on a laptop PC with Intel Core i7-4710MQ CPU at 2.50 GHz and 16 GB of RAM.

Quantitative Properties of the Optimization	Problem 1	Problem 2
Total number of variables	4869	70,614
Number of variables with only lower bounds	4383	4383
Number of variables with lower and upper bounds	486	486
Total number of equality constraints	4383	70,128
Number of nonzeros in the Lagrangian Hessian	10,826	386,575
Number of iterations	212	74
Elapsed time (s) *	8	1187

As was noted in Remark 1, the daily transmission rate becomes uncertain when the number of infectious people reduces significantly. In this case, the input constraint with the computed gain K_k may be violated; therefore, we increased the input weight $\mathbf{R}_k^{\text{LQR}}$ to obtain a lower gain K_k. These heuristic operations to compute an admissible gain were relevant only when the third wave of the epidemic suddenly dropped after the end of April 2021. The scaled logarithm of the input weights on each day k are illustrated in Plot 2 of Figure 2. In Plot 3 of Figure 2, we present the reconstructed number of hospitalized patients in comparison with the official (i.e., reference) data. Plots 4, 5, and 6 of Figure 2 illustrate three derived probabilistic quantities $\hat{z}_k = h(\hat{x}_k, \hat{u}_k, \hat{\theta})$, namely the time-dependent effective reproduction number (4) ($[\hat{z}_k]_1 = R_k$), the number of all infected people ($[\hat{z}_k]_2 = \hat{\mathbf{L}}_k + \hat{\mathbf{P}}_k + \hat{\mathbf{I}}_k + \hat{\mathbf{A}}_k + \hat{\mathbf{H}}_k$), and the number of daily new infections ($[\hat{z}_k]_3 = \hat{\beta}_k\,(\hat{\mathbf{P}}_k + \hat{\mathbf{I}}_k + \hat{\delta}\,\hat{\mathbf{A}}_k)\frac{\hat{\mathbf{S}}_k}{\mathbf{N}}$). The first and third coordinates of $\hat{z}_k$ are nonlinear functions of random variables. Therefore, the mean and the variance of $\hat{z}_k$ were approximated by the first-order Taylor polynomial of function h, namely,

$$\hat{z}_k \approx \frac{\partial h}{\partial(x,u,\theta)}(\mu_k) \begin{pmatrix} \hat{x}_k \\ \hat{u}_k \\ \hat{\theta}_k \end{pmatrix} + h(\mu_k) - \frac{\partial h}{\partial(x,u,\theta)}(\mu_k)\,\mu_k \; \sim \; \mathcal{N}\left(h(\mu_k) \, , \, \frac{\partial h}{\partial(x,u,\theta)}(\mu_k)\,\Sigma_k \left(\frac{\partial h}{\partial(x,u,\theta)}(\mu_k) \right)^\top \right). \tag{25}$$

The yellow curve in Plot 4 illustrates the estimated reproduction number published online by Atlo Team in [74].
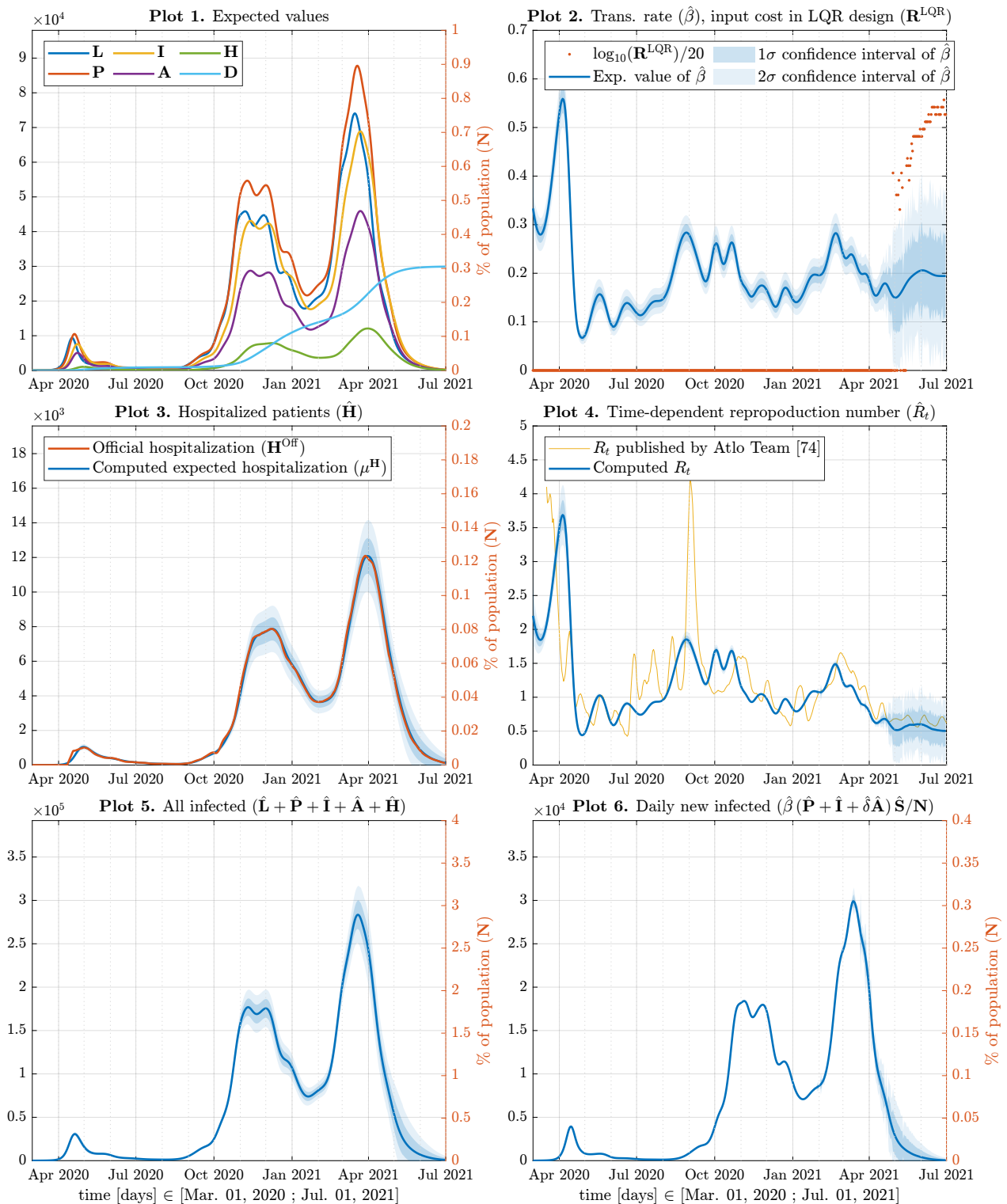
Figure 2. Reconstructed epidemiological data computed for System (1) with Gaussian model parameters: expected value of states (Plot 1), transmission rate of the pathogen (Plot 2), time-dependent reproduction number (Plot 4), number of reconstructed hospitalized patients compared to the recorded data (Plot 3), sum of all infected compartments (Plot 5), and the daily new infections (Plot 6). The gray dotted vertical grid lines show the first days of the months.
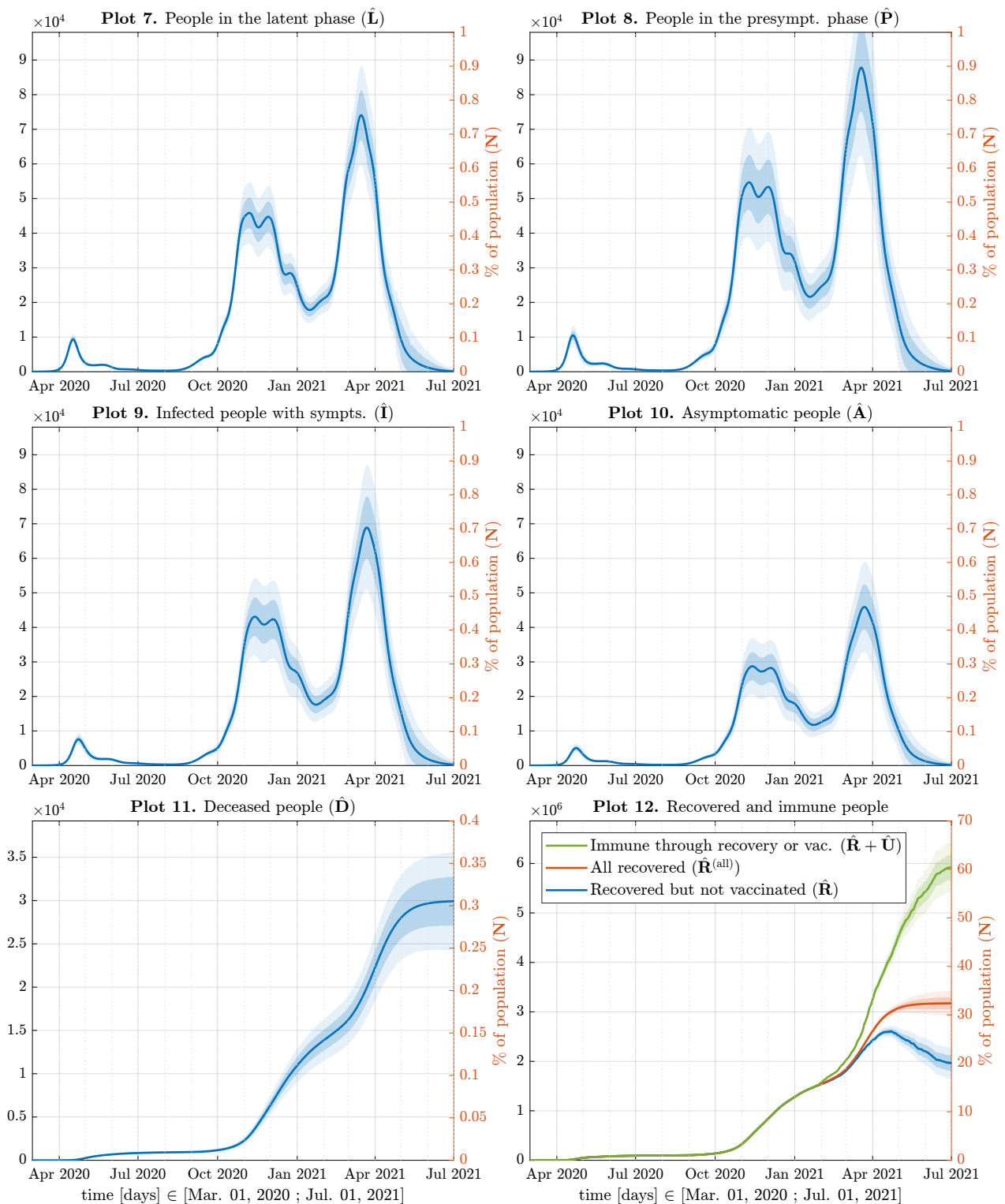
Figure 3. Unknown epidemiological data computed for System (1) with Gaussian model parameters: number of people in the different phases of the disease (Plots 7–10), number of deceased people (Plot 11), and number of recovered and/or immune people (Plot 12). The gray dotted vertical grid lines emphasize the first days of the months.

In Plot 12 of Figure 3, we present three uncertain time series, namely the number of recovered, but not yet vaccinated people (blue), the cumulative number of recovered people (red), and the cumulative number of immune people (green).

In Figures 2 and 3, we can observe the successfully suppressed first wave with a dramatic effect of the strict lockdown introduced in March 2020. As a result, the disease was mainly confined to closed institutions such as certain hospital wards and elderly homes. This policy could not be maintained in the autumn of 2020, and therefore, a substantially larger second wave occurred, causing a huge burden on the healthcare system. Therefore, further restrictions (online education in secondary schools, closure of certain public spaces, banning of most gatherings, and curfew from 8 p.m. to 5 a.m.) had to be introduced in the first half of November 2020. These measures had the planned effect in terms of the significant reduction of the transmission rate, as is visible in Plot 2 of Figure 3. Then, from January 2021, R_t began to increase again due to the appearance of the more contagious alpha (B.1.1.7) variant, although all of the former restrictions remained in effect. The alpha variant caused the largest peak of the epidemic so far in the spring of 2021 with a maximum of 12,553 hospitalized people on 30 March 2021. Further restrictions had to be introduced on 8 March 2021, where the main component was the closing of all schools. Together with the intensive vaccination in the first half of 2021, this made the decrease in the number of infected people definitely fast. The ratio of the peaks of the estimated β in February 2021 and December 2020 was approximately 1.62, which matches well with literature reports (1.4–1.8) in the U.K. [75]. We note that we can compare these data since they were estimated under the same restriction level. Plot 12 in Figure 3 shows the estimated number of people gaining immunity by infection and/or vaccination. According to this estimation, more than 30% of the population might have gone through the COVID infection until the end of June 2021. This suggests an approximately 26% detection rate. This is significantly lower than the value of certain European countries such as Germany, Italy, or Spain, reaching or sometimes exceeding 50%, but the number of performed tests per population has also been much lower in Hungary than in the mentioned countries [76].

6. Conclusions

In this paper, we proposed an optimization-based data assimilation approach to compute the unknown inputs and states of discrete-time compartmental epidemic models with uncertain normally distributed parameters. We started from the assumption that the joint state input parameter distribution is Gaussian. Then, a deterministic mean-variance recursion was developed, which made it possible to formulate the stochastic data assimilation problem as a single model predictive control design with a nonlinear mean-variance prediction model. We noted and demonstrated that the resulting $\mu\Sigma$-NMPC is computationally intensive, but its local optimum can be well approximated by a more efficient ordinary NMPC and further closed-form variance computations.

We proposed simple heuristics to predict appropriate feedback gains, which realize state-dependent control actions along the prediction horizon. In this way, a trade-off can be made between the uncertainty of the computed states and inputs as the predicted control action is scheduled by the deviation between the actual realization of the state and its predicted expectation. As the approach does not make a difference between the unknown parameters and inputs, the joint state observation, change detection, and parameter estimation are also possible.

Through the finite horizon predictive control computation, we estimated the unknown data of the past evolution of the COVID-19 epidemic spread within a fixed time window in Hungary. Among the unknown quantities, we considered the daily number of people in the different phases of the disease and the transmission rate of the pathogen, which highly depends on the actual social distancing rules, mobility restrictions, and virus mutations. The unknown time series were computed such that the expected value of the computed number of hospitalized patients fit the truly observed data as much as possible. Compared to our previous results [21,22], we considered an augmented and uncertain compartmental epidemiological model with normally distributed random model parameters and a simple vaccination model as well.

The main limitations of this study were the following. The length of hospital treatment was considered to be constant in the model, since no data have been published on the daily new hospital admissions with COVID-19, from which this parameter could be tracked efficiently. Moreover, no representative nationwide serological testing in Hungary has been organized since the summer of 2020. Such a result would definitely be helpful in making the estimate on the number of immune people more precise. Finally, the waning of immunity after infection or vaccination was also not taken into consideration in the model. However, such an extension does not affect the applicability of the proposed MPC-based estimation.

With a few modifications, the approach can be applied to compute multiple uncertain possibly time-dependent parameters, but also for the prediction of the future behavior of the epidemic spread. The proposed methodology is able to extract and reconstruct detailed information from the whole time horizon of the epidemic process beyond giving estimates for the cumulative number of infected and recovered people. Future work will be focused on the analysis of other European countries.

Abbreviations

The following abbreviations are used in this manuscript:

SEIR	Susceptible–exposed–infected–recovered (compartmental model)
IgG	Immunoglobulin G
CT	Continuous-time
DT	Discrete-time
ODE	Ordinary differential equation
i.i.d.	Independent identically distributed
LTI	Linear time-invariant
MPC	Model predictive controller
NMPC	Nonlinear model predictive controller
SNMPC	Stochastic nonlinear model predictive controller
LQR	Linear quadratic regulator
$\mu\Sigma$-...	Mean-variance equations/dynamics/recursion/prediction model/NMPC
$\mu\Sigma$-NMPC	Mean-variance nonlinear model predictive controller

Appendix A

In this section, we prove that the expected value of the cost:

$$J(X, U) = \sum_{k=0}^{T-1} \|C\hat{x}_{k+1} - r_{k+1}\|_Q^2 + \sum_{k=0}^{T-2} \|\hat{u}_{k+1} - \hat{u}_k\|_R^2, \qquad \text{(borrowed from (7))}$$

can be expressed as:

$$J(M, S, V) = \sum_{k=0}^{T-1} \|C\mu_{k+1}^x - r_{k+1}\|_Q^2 + \sum_{k=0}^{T-2} \|\mu_{k+1}^u - \mu_k^u\|_R^2 \qquad \text{(borrowed from (14))}$$
$$+ \sum_{k=0}^{T-1} \text{Tr}\left(Q\,C\,\Sigma_{k+1}^x\,C^\top\right) + \sum_{k=0}^{T-2} \text{Tr}\left(R\,K(\mu_{k+1}^x)\,\Sigma_{k+1}^x\,K^\top(\mu_{k+1}^x)\right)$$
$$+ \sum_{k=0}^{T-2} \text{Tr}\left(R\,K(\mu_k^x)\,\Sigma_k^x\,K^\top(\mu_k^x) - R\,\text{He}\left\{K(\mu_{k+1}^x)\,\text{Cov}(\hat{x}_{k+1}, \hat{x}_k)\,K^\top(\mu_k^x)\right\}\right).$$

The proof is given in multiple steps, but it is essentially based on a simple observation, which allows expressing the expectation of the squared weighted norm of a random variable $\hat{x}$ as follows:

$$\mathbb{E}\left(\|\hat{x}\|_Q^2\right) = \|\mu^x\|_Q^2 + \text{Tr}(Q\,\Sigma^x) \tag{A1}$$

To prove (A1), first, consider the following chain of identities:

$$Q\,\Sigma^x = Q\,\text{Var}(\hat{x}) = \text{Cov}(Q\,\hat{x}, \hat{x}) = \mathbb{E}(Q\,\hat{x}\,\hat{x}^\top) - Q\,\mu^x\,(\mu^x)^\top. \tag{A2}$$

Then, we take the trace of the quantities in (A2) to obtain:

$$\text{Tr}(Q\,\Sigma^x) = \mathbb{E}(\hat{x}^\top Q\,\hat{x}) - (\mu^x)^\top Q\,\mu^x. \tag{A3}$$

Equality (A1) is a direct consequence of (A3). Accordingly, the squared weighted norm of the output error at time $k + 1$ can be expressed as follows:

$$\|C\hat{x}_{k+1} - r_{k+1}\|_Q^2 = \|C\mu_{k+1}^x - r_{k+1}\|_Q^2 + \text{Tr}\left(Q\,C\,\Sigma_{k+1}^x\,C^\top\right). \tag{A4}$$

Secondly, the input variation cost is expressed as follows:

$$\mathbb{E}\left(\|\hat{u}_{k+1} - \hat{u}_k\|_R^2\right) = \|\mu_{k+1}^u - \mu_k^u\|_R^2 + \text{Tr}\left(R\,\text{Var}(\hat{u}_{k+1} - \hat{u}_k)\right). \tag{A5}$$

The variance term in (A5) is further developed as follows:

$$\text{Var}(\hat{u}_{k+1} - \hat{u}_k) = \text{Var}(\hat{u}_{k+1} - K(\mu_{k+1}^x)(\hat{x}_{k+1} - \mu_{k+1}^x) - \hat{u}_k + K(\mu_k^x)(\hat{x}_k - \mu_k^x)) \tag{A6}$$
$$= \text{Var}(K(\mu_{k+1}^x)\hat{x}_{k+1} - K(\mu_k^x)\hat{x}_k) \tag{A7}$$
$$= \text{Var}(K(\mu_{k+1}^x)\hat{x}_{k+1}) + \text{Var}(K(\mu_k^x)\hat{x}_k)$$
$$- \text{He}\left\{\text{Cov}(K(\mu_{k+1}^x)\hat{x}_{k+1}, K(\mu_k^x)\hat{x}_k)\right\} \tag{A8}$$
$$= K(\mu_{k+1}^x)\,\Sigma_{k+1}^x\,K^\top(\mu_{k+1}^x) + K(\mu_k^x)\,\Sigma_k^x\,K^\top(\mu_k^x)$$
$$- \text{He}\left\{K(\mu_{k+1}^x)\,\text{Cov}(\hat{x}_{k+1}, \hat{x}_k)\,K^\top(\mu_k^x)\right\} \tag{A9}$$

Finally, the output error (A4) and the input variation error (A5) with the variance expression (A9) give the expected value (14) of the random cost (7).

References

1. Miller, I.F.; Becker, A.D.; Grenfell, B.T.; Metcalf, C.J.E. Disease and healthcare burden of COVID-19 in the United States. *Nat. Med.* **2020**, *26*, 1212–1217. [CrossRef] [PubMed]
2. Baker, S.R.; Bloom, N.; Davis, S.J.; Terry, S.J. *COVID-Induced Economic Uncertainty*; Technical Report; National Bureau of Economic Research: Cambridge, MA, USA, 2020. [CrossRef]
3. Cao, L.; Liu, Q.; Hou, W. COVID-19 modeling: A review. *arXiv* **2021**, arXiv:2104.12556
4. Shinde, G.R.; Kalamkar, A.B.; Mahalle, P.N.; Dey, N.; Chaki, J.; Hassanien, A.E. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* **2020**, *1*, 197. [CrossRef] [PubMed]

5. Biswas, M.H.A.; Paiva, L.T.; de Pinho, M. A SEIR model for control of infectious diseases with constraints. *Math. Biosci. Eng.* **2014**, *11*, 761–784. [CrossRef]

6. Brauer, F. Compartmental models in epidemiology. In *Mathematical Epidemiology*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 19–79. [CrossRef]

7. He, S.; Peng, Y.; Sun, K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn.* **2020**, *101*, 1667–1680. [CrossRef] [PubMed]

8. Röst, G.; Bartha, F.A.; Bogya, N.; Boldog, P.; Dénes, A.; Ferenci, T.; Horváth, K.J.; Juhász, A.; Nagy, C.; Tekeli, T.; et al. Early phase of the COVID-19 outbreak in Hungary and post-lockdown scenarios. *Viruses* **2020**, *12*, 708. [CrossRef]

9. Rajabi, A.; Mantzaris, A.V.; Mutlu, E.C.; Garibay, O.O. Investigating dynamics of COVID-19 spread and containment with agent-based modeling. *Appl. Sci.* **2021**, *11*, 5367. [CrossRef]

10. Reguly, I.Z.; Csercsik, D.; Juhasz, J.; Tornai, K.; Bujtar, Z.; Horvath, G.; Keomley-Horvath, B.; Kos, T.; Cserey, G.; Ivan, K.; et al. Microsimulation based quantitative analysis of COVID-19 management strategies. *PLoS Comput. Biol.* **2022**, *18*, 1–14. [CrossRef]

11. Rzadkowski, G.; Figlia, G. Logistic wavelets and their application to model the spread of COVID-19 pandemic. *Appl. Sci.* **2021**, *11*, 8147. [CrossRef]

12. Lalmuanawma, S.; Hussain, J.; Chhakchhuak, L. Applications of machine learning and artificial intelligence for COVID-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* **2020**, *139*, 110059. [CrossRef] [PubMed]

13. Satu, M.; Howlader, K.C.; Mahmud, M.; Kaiser, M.S.; M., S.I.S.; Quinn, J.M.; Alyami, S.A.; Moni, M.A. Short-term prediction of COVID-19 cases using machine learning models. *Appl. Sci.* **2021**, *11*, 4266. [CrossRef]

14. Ghafouri-Fard, S.; Mohammad-Rahimi, H.; Motie, P.; Minabi, M.; Taheri, M.; Nateghinia, S. Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review. *Heliyon* **2021**, *7*, e08143. [CrossRef] [PubMed]

15. Gostic, K.M.; McGough, L.; Baskerville, E.B.; Abbott, S.; Joshi, K.; Tedijanto, C.; Kahn, R.; Niehus, R.; Hay, J.A.; De Salazar, P.M.; et al. Practical considerations for measuring the effective reproductive number, Rt. *PLoS Comput. Biol.* **2020**, *16*, e1008409. [CrossRef] [PubMed]

16. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* **2007**, *2*, e758. [CrossRef] [PubMed]

17. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512. [CrossRef]

18. Kucharski, A.J.; Russell, T.W.; Diamond, C.; Liu, Y.; Edmunds, J.; Funk, S.; Eggo, R.M.; Sun, F.; Jit, M.; Munday, J.D.; et al. Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet Infect. Dis.* **2020**, *20*, 553–558. [CrossRef]

19. Koyama, S.; Horie, T.; Shinomoto, S. Estimating the time-varying reproduction number of COVID-19 with a state-space method. *PLoS Comput. Biol.* **2021**, *17*, e1008679. [CrossRef] [PubMed]

20. Tsay, C.; Lejarza, F.; Stadtherr, M.A.; Baldea, M. Modeling, state estimation, and optimal control for the US COVID-19 outbreak. *Sci. Rep.* **2020**, *10*, 10711. [CrossRef] [PubMed]

21. Péni, T.; Csutak, B.; Szederkényi, G.; Röst, G. Nonlinear model predictive control with logic constraints for COVID-19 management. *Nonlinear Dyn.* **2020**, *102*, 1965–1986. [CrossRef] [PubMed]

22. Csutak, B.; Polcz, P.; Szederkényi, G. Computation of COVID-19 epidemiological data in Hungary using dynamic model inversion. In Proceedings of the 15th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2021), Timișoara, Romania, 19–21 May 2021; pp. 91–96. [CrossRef]

23. Sereno, J.; D'Jorge, A.; Ferramosca, A.; Hernandez-Vargas, E.; González, A. Model predictive control for optimal social distancing in a type SIR-switched model. *IFAC-PapersOnLine* **2021**, *54*, 251–256. [CrossRef]

24. Phipps, S.J.; Grafton, R.Q.; Kompas, T. Robust estimates of the true (population) infection rate for COVID-19: A backcasting approach. *R. Soc. Open Sci.* **2020**, *7*, 200909. [CrossRef]

25. Rocchetti, I.; Böhning, D.; Holling, H.; Maruotti, A. Estimating the size of undetected cases of the COVID-19 outbreak in Europe: An upper bound estimator. *Epidemiol. Methods* **2020**, *9*, 20200024. [CrossRef]

26. Bartha, F.A.; Karsai, J.; Tekeli, T.; Röst, G. Symptom-based testing in a compartmental model of COVID-19. In *Analysis of Infectious Disease Problems (COVID-19) and Their Global Impact*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 357–376. [CrossRef]

27. Lemaitre, J.C.; Perez-Saez, J.; Azman, A.; Rinaldo, A.; Fellay, J. Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland. *Swiss Med. Wkly.* **2020**, *150*, w20295. [CrossRef]

28. Allgöwer, F.; Zheng, A. (Eds.) *Nonlinear Model Predictive Control*, 1st ed.; Progress in Systems and Control Theory 26; Birkhäuser Basel: Basel, Switzerland, 2000. [CrossRef]

29. Apte, A.; Jones, C.K.R.T.; Stuart, A.M.; Voss, J. Data assimilation: Mathematical and statistical perspectives. *Int. J. Numer. Methods Fluids* **2008**, *56*, 1033–1046. [CrossRef]

30. Bröcker, J. On variational data assimilation in continuous time. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 1906–1919. [CrossRef]

31. Schumann-Bischoff, J.; Parlitz, U.; Abarbanel, H.D.I.; Kostuk, M.; Rey, D.; Eldridge, M.; Luther, S. Basin structure of optimization based state and parameter estimation. *Chaos Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 053108. [CrossRef]

32. Schumann-Bischoff, J.; Parlitz, U. State and parameter estimation using unconstrained optimization. *Phys. Rev. E* **2011**, *84*, 056214. [CrossRef] [PubMed]

33. Chen, T.; Kirkby, N.F.; Jena, R. Optimal dosing of cancer chemotherapy using model predictive control and moving horizon state/parameter estimation. *Comput. Methods Programs Biomed.* **2012**, *108*, 973–983. [CrossRef]

34. Das, A. Chance-constrained optimization-based parameter estimation for Muskingum models. *J. Irrig. Drain. Eng.* **2007**, *133*, 487–494. [CrossRef]

35. Das, A. Parameter estimation for Muskingum models. *J. Irrig. Drain. Eng.* **2004**, *130*, 140–147. [CrossRef]

36. Al-Hemeary, N.; Polcz, P.; Szederkényi, G. Optimal solar panel area computation and temperature tracking for a cubesat system using model predictive control. *SPIIRAS Proc.* **2020**, *19*, 564–593. [CrossRef]

37. Courtier, P.; Talagrand, O. Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus A Dyn. Meteorol. Oceanogr.* **1990**, *42*, 531–549. [CrossRef]

38. Blackmore, L.; Açıkmeşe, B.; Carson, J.M. Lossless convexification of control constraints for a class of nonlinear optimal control problems. *Syst. Control Lett.* **2012**, *61*, 863–870. [CrossRef]

39. Mao, Y.; Szmuk, M.; Açıkmeşe, B. Successive convexification of non-convex optimal control problems and its convergence properties. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 3636–3641. [CrossRef]

40. Andersson, J.; Gillis, J.; Diehl, M. User Documentation for CasADi v3.4.4. 2018. Available online: http://casadi.sourceforge.net/v3.4.4/users_guide/casadi-users_guide.pdf (accessed on 17 December 2021).

41. Andersson, J.A.E.; Gillis, J.; Horn, G.; Rawlings, J.B.; Diehl, M. CasADi: A software framework for nonlinear optimization and optimal control. *Math. Program. Comput.* **2018**, *11*, 1–36. [CrossRef]

42. Wächter, A.; Biegler, L.T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **2005**, *106*, 25–57. [CrossRef]

43. Zanelli, A.; Domahidi, A.; Jerez, J.; Morari, M. Forces NLP: An efficient implementation of interior-point methods for multistage nonlinear nonconvex programs. *Int. J. Control* **2017**, *93*, 13–29. [CrossRef]

44. Mesbah, A. Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control. Syst. Mag.* **2016**, *36*, 30–44. [CrossRef]

45. de la Penad, D.; Bemporad, A.; Alamo, T. Stochastic programming applied to model predictive control. In Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, 15 December 2005; pp. 1361–1366. [CrossRef]

46. Bernardini, D.; Bemporad, A. Stabilizing model predictive control of stochastic constrained linear systems. *IEEE Trans. Autom. Control* **2012**, *57*, 1468–1480. [CrossRef]

47. Thangavel, S.; Paulen, R.; Engell, S. Robust multi-stage nonlinear model predictive control using sigma points. *Processes* **2020**, *8*, 851. [CrossRef]

48. Thangavel, S.; Paulen, R.; Engell, S. Dual multi-stage NMPC using sigma point principles. *IFAC-PapersOnLine* **2020**, *53*, 11243–11250. [CrossRef]

49. Bonzanini, A.D.; Paulson, J.A.; Makrygiorgos, G.; Mesbah, A. Fast approximate learning-based multistage nonlinear model predictive control using Gaussian processes and deep neural networks. *Comput. Chem. Eng.* **2021**, *145*, 107174. [CrossRef]

50. Ostafew, C.J.; Schoellig, A.P.; Barfoot, T.D. Conservative to confident: Treating uncertainty robustly within learning-based control. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 421–427. [CrossRef]

51. Hewing, L.; Kabzan, J.; Zeilinger, M.N. Cautious model predictive control using Gaussian process regression. *IEEE Trans. Control. Syst. Technol.* **2020**, *28*, 2736–2743. [CrossRef]

52. Candela, J.Q.; Girard, A.; Rasmussen, C.E. *Prediction at an Uncertain Input for Gaussian Processes and Relevance Vector Machines Application to Multiple-Step Ahead Time-Series Forecasting*; Technical Report IMM-2003-18; Technical University of Denmark: Kgs. Lyngby, Denmark, 2003.

53. Deisenroth, M.P. Efficient Reinforcement Learning Using Gaussian Processes—Revised Version. Ph.D. Thesis, Faculty of Informatics, Institute for Anthropomatics, Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Karlsruhe, Germany, 2017.

54. de Moura, D.T.H.; McCarty, T.R.; Ribeiro, I.B.; Funari, M.P.; de Oliveira, P.V.A.G.; de Miranda Neto, A.A.; do Monte Júnior, E.S.; Tustumi, F.; Bernardo, W.M.; de Moura, E.G.H.; et al. Diagnostic characteristics of serological-based COVID-19 testing: A systematic review and meta-analysis. *Clinics* **2020**, *75*, e2212. [CrossRef] [PubMed]

55. Merkely, B.; Szabó, A.J.; Kosztin, A.; Berényi, E.; Sebestyén, A.; Lengyel, C.; Merkely, G.; Kárády, J.; Várkonyi, I.; Papp, C.; et al. Novel coronavirus epidemic in the Hungarian population, a cross-sectional nationwide survey to support the exit policy in Hungary. *GeroScience* **2020**, *42*, 1063–1074. [CrossRef]

56. Sedaghat, A.; Oloomi, S.A.A.; Malayer, M.A.; Band, S.; Mosavi, A.; Nadai, L. Modeling and sensitivity analysis of coronavirus disease (COVID-19) outbreak prediction. In Proceedings of the 2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary, 18–19 November 2020; pp. 000261–000266. [CrossRef]

57. Sedaghat, A.; Oloomi, S.A.A.; Malayer, M.A.; Band, S.; Rezaei, N.; Mosavi, A.; Nadai, L. Coronavirus (COVID-19) outbreak prediction using epidemiological models of Richards Gompertz Logistic Ratkowsky and SIRD. In Proceedings of the 2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), Budapest, Hungary, 18–19 November 2020; pp. 000289–000298. [CrossRef]

58. Tartof, S.Y.; Slezak, J.M.; Fischer, H.; Hong, V.; Ackerson, B.K.; Ranasinghe, O.N.; Frankland, T.B.; Ogun, O.A.; Zamparo, J.M.; Gray, S.; et al. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: A retrospective cohort study. *Lancet* **2021**, *398*, 1407–1416. [CrossRef]

59. Hu, X.; Lindquist, A. *Geometric Control Theory*; Royal Institute of Technology: Stockholm, Sweden, 2012.

60. González Cisneros, P.S.; Werner, H. Nonlinear model predictive control for models in quasi-linear parameter varying form. *Int. J. Robust Nonlinear Control* **2020**, *30*, 3945–3959. [CrossRef]

61. Bemporad, A.; Ricker, N.L.; Morari, M. Model Predictive Control Toolbox™User's Guide (R2019b). MathWorks. 2019. Available online: https://www.mathworks.com/help/pdf_doc/mpc/mpc_ug.pdf (accessed on 17 December 2021).

62. Data on COVID-19 Vaccination in the EU/EEA. Available online: https://www.ecdc.europa.eu/en/publications-data/data-covid-19-vaccination-eu-eea (accessed on 17 December 2021).

63. Atlo Team. Koronamonitor: Hungarian Status of Coronavirus Vaccination. 2021. Available online: https://atlo.team/vakcinacio (accessed on 11 November 2021).

64. Salath, M.; Althaus, C.L.; Neher, R.; Stringhini, S.; Hodcroft, E.; Fellay, J.; Zwahlen, M.; Senti, G.; Battegay, M.; Wilder-Smith, A.; et al. COVID-19 epidemic in switzerland: On the importance of testing, contact tracing and isolation. *Swiss Med. Wkly.* **2020**. [CrossRef]

65. Steinbrook, R. Contact tracing, testing, and control of COVID-19—Learning from Taiwan. *JAMA Intern. Med.* **2020**, *180*, 1163–1164. [CrossRef]

66. Data on Hospital and ICU Admission Rates and Current Occupancy for COVID-19. 2021. Available online: https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19 (accessed on 11 November 2021).

67. Mesbah, A.; Streif, S.; Findeisen, R.; Braatz, R.D. Stochastic nonlinear model predictive control with probabilistic constraints. In Proceedings of the 2014 IEEE American Control Conference, Portland, OR, USA, 4–6 June 2014. [CrossRef]

68. De Larminat, P. *Analysis and Control of Linear Systems*; John Wiley & Sons: Hoboken, NJ, USA, 2013.

69. Lorenzen, M.; Dabbene, F.; Tempo, R.; Allgöwer, F. Constraint-tightening and stability in stochastic model predictive control. *IEEE Trans. Autom. Control* **2017**, *62*, 3165–3177. [CrossRef]

70. MathWorks. Control System Toolbox™Reference (R2021b). 2021. Available online: https://www.mathworks.com/help/pdf_doc/control/control_ref.pdf (accessed on 17 December 2021).

71. Amestoy, P.; Duff, I.S.; Koster, J.; L'Excellent, J.Y. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.* **2001**, *23*, 15–41. [CrossRef]

72. Amestoy, P.; Buttari, A.; L'Excellent, J.Y.; Mary, T. Performance and scalability of the block low-rank multifrontal factorization on multicore architectures. *ACM Trans. Math. Softw.* **2019**, *45*, 2:1–2:26. [CrossRef]

73. Polcz, P. Epidemiological Data Reconstruction for Hungary Using Stochastic Nonlinear MPC Computations. GitHub Repository. 2021. Available online: https://github.com/ppolcz/MPC-monitoring-for-COVID-19 (accessed on 17 December 2021).

74. Atlo Team. Koronamonitor: Detailed Diagrams of the Coronavirus Outbreak. 2021. Available online: https://atlo.team/koronamonitor-reszletesadatok (accessed on 11 November 2021).

75. Volz, E.; Mishra, S.; Chand, M.; Barrett, J.C.; Johnson, R.; Geidelberg, L.; Hinsley, W.R.; Laydon, D.J.; Dabrera, G.; O'Toole, Á.; et al. Transmission of SARS-CoV-2 lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv* **2021**. [CrossRef]

76. Institute for Health Metrics and Evaluation. COVID-19 Results Briefing, European Union 1 July 2021. 2021. Available online: https://www.healthdata.org/sites/default/files/files/Projects/COVID/2021/4743_briefing_European_Union_23.pdf (accessed on 17 December 2021).