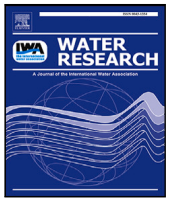
Water Research 241 (2023) 120098

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



# Wastewater-based modeling, reconstruction, and prediction for COVID-19 outbreaks in Hungary caused by highly immune evasive variants

Péter Polcz [a,*], Kálmán Tornai [a], János Juhász [a,c], György Cserey [a], György Surján [b,d], Tamás Pándics [b,e], Eszter Róka [b], Márta Vargha [b], István Z. Reguly [a], Attila Csikász-Nagy [a], Sándor Pongor [a], Gábor Szederkényi [a]

[a] National Laboratory for Health Security, Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter utca 85, Budapest, H-1083, Hungary
[b] Department of Public Health Laboratory, National Public Health Centre, Albert Flórián út 2-6, Budapest, H-1097, Hungary
[c] Institute of Medical Microbiology, Semmelweis University, Üllői út 26, Budapest, H-1085, Hungary
[d] Department of Digital Health Sciences, Semmelweis University, Üllői út 26, Budapest, H-1085, Hungary
[e] Department of Public Health Sciences, Faculty of Health Sciences, Semmelweis University, Vas utca 17, Budapest, H-1088, Hungary

ARTICLE INFO

*MSC:*
37N25
92D30

*Keywords:*
Wastewater-based epidemiology
Epidemic model
Dynamical system
Parameter estimation
Optimization

ABSTRACT

**(Motivation)** Wastewater-based epidemiology (WBE) has emerged as a promising approach for monitoring the COVID-19 pandemic, since the measurement process is cost-effective and is exposed to fewer potential errors compared to other indicators like hospitalization data or the number of detected cases. Consequently, WBE was gradually becoming a key tool for epidemic surveillance and often the most reliable data source, as the intensity of clinical testing for COVID-19 drastically decreased by the third year of the pandemic. Recent results suggests that the model-based fusion of wastewater measurements with clinical data and other indicators is essential in future epidemic surveillance.

**(Method)** In this work, we developed a wastewater-based compartmental epidemic model with a two-phase vaccination dynamics and immune evasion. We proposed a multi-step optimization-based data assimilation method for epidemic state reconstruction, parameter estimation, and prediction. The computations make use of the measured viral load in wastewater, the available clinical data (hospital occupancy, delivered vaccine doses, and deaths), the stringency index of the official social distancing rules, and other measures. The current state assessment and the estimation of the current transmission rate and immunity loss allow a plausible prediction of the future progression of the pandemic.

**(Results)** Qualitative and quantitative evaluations revealed that the contribution of wastewater data in our computational epidemiological framework makes predictions more reliable. Predictions suggest that at least half of the Hungarian population has lost immunity during the epidemic outbreak caused by the BA.1 and BA.2 subvariants of Omicron in the first half of 2022. We obtained a similar result for the outbreaks caused by the subvariant BA.5 in the second half of 2022.

**(Applicability)** The proposed approach has been used to support COVID management in Hungary and could be customized for other countries as well.

## 1. Introduction

Wastewater-based epidemiology (WBE) has gained a special interest during the COVID-19 pandemic, although it is not new in the literature (Been et al., 2014). Claro et al. (2021) suggested that WBE is a powerful tool in the long-term monitoring of COVID-19 pandemic.

Randazzo et al. (2020), and Zulli et al. (2021) demonstrated that WBE is an efficient tool for supporting decision-making on public policies during an epidemic outbreak. However, Bibby et al. (2021) drew attention to the typically uncertain dynamics of the virus secretion in feces, which may undermine the potential for WBE as an early warning system. Olesen et al. (2021) concluded that WBE can be used as a leading indicator, but the lead time should be determined with a special care for the different applications. The predictive quality of the WBE is data sensitive (Huizer et al., 2021), therefore, the processing and the normalization of the measured genome copy concentration are also essential (Zhu et al., 2022).

* Corresponding author.
*E-mail address:* polcz.peter@itk.ppke.hu (P. Polcz).

https://doi.org/10.1016/j.watres.2023.120098
Received 28 February 2023; Received in revised form 16 May 2023; Accepted 17 May 2023
Available online 25 May 2023

Radu et al. (2022) and Lee et al. (2021) used WBE to show the emergence of the Alpha variant weeks before the first clinical case detected in Austria and the United States, respectively. A similar observation is seen by Lee et al. (2022) for the rapid displacement of the Delta variant by the highly virulent Omicron variant. Morvan et al. (2022) confirmed that the first signs of a new outbreak appear approximately a week earlier than clinical testing data. Using statistical methods, Krivoňáková et al. (2021) showed a good correlation between wastewater data and clinical cases.

The dynamic relationship between the wastewater copy numbers and the clinical cases was identified by Xiao et al. (2022), and Daza-Torres et al. (2023) as a finite support discrete-time impulse response function, which constitutes the *shedding load distribution*. A similar closed-form distribution function with a detailed temperature model is used by Phan et al. (2023) to model the viral load in the dynamical model. It is interesting that the time lag between wastewater data and clinical cases may be both positive and negative when analyzing separate epidemic outbreaks, see, e.g., Xiao et al. (2022, Fig. 3., Panels C and D). Jiang et al. (2022) developed artificial neural network models for continuous epidemic monitoring using training data set including catchment, historical weather, clinical testing coverage, and vaccination rate.

Alongside the statistical and machine learning approaches, the dynamical compartmental susceptible, exposed, infected, recovered (SEIR)-type models are also promising for epidemic surveillance and prediction. Fernandez-Cassi et al. (2021) showed that the wastewater-based reconstruction fits better to a model-based epidemic curve compared to estimates based on confirmed cases using a detailed SEIR-type model introduced by Lemaitre et al. (2020). Proverbio et al. (2022) have used a wastewater-based SEIR model and the extended Kalman filter to reconstruct epidemic data and detect early-warning performance. Phan et al. (2023) developed a detailed static shedding model, and they fit an SEIR model to the measured viral load for a single epidemic wave. Pájaro et al. (2022) built a stochastic SIR model and solved it by using Monte Carlo methods to predict the evolution of the pandemics incorporating viral load data and a variable infection rate to capture the effects of mitigation policies. Fazli et al. (2021) proposed an agent-based SEIR model using viral load as a measurement for the early detection and prediction of an outbreak. The incidence was computed by McMahan et al. (2021) using a wastewater-based SEIR model, and the under-detection rate was estimated. The model was used to predict the future possibility of an outbreak. Nourbakhsh et al. (2022) used a detailed compartmental model with viral secretion to wastewater to estimate prevalence, incidence, and reproduction number using the reported clinical cases, hospital admission, and viral load in wastewater.

To analyze the epidemic spread and estimate the transmission rate, Jiang et al. (2021) used Markov processes such that the SEIR compartments are considered hidden variables, whereas, the viral load and clinical cases are observations. A likelihood-based approach made it possible to understand the internal relationship between different states and how COVID-19 actually transmits. Plateaus, rebounds, and the effects of individual behaviors in epidemics are analyzed and explained by Berestycki et al. (2021) using a specific SIR-type partial differential equation (PDE) model and high precision wastewater measurement.

During the past two years, our research group has been focusing on the dynamical analysis, data reconstruction, and prediction of the COVID-19 pandemic in Hungary using the tools of systems and control theory and high performance computation techniques. In this line of research, a detailed agent-based stochastic epidemic simulator was developed that uses realistic population structures and movement patterns in an urban environment, which was used for analysis and prediction (Reguly et al., 2022). Péni et al. (2020) proposed a methodology to design predefined discrete levels for public restrictions or interventions fulfilling complex, time-dependent, and often contradicting economic and healthcare-related requirements. The data reconstruction problem of estimating the possible real number of infections and the population of non-measurable compartments was solved by Polcz et al. (2022) using a stochastic model predictive control design. It was clearly shown by Péni et al. (2022) that the optimal design of testing intensity is essential to keep the stringency of epidemic measures on a tolerable level. Parallelly, an advanced framework for wastewater surveillance based on the detection of viral RNA was developed by Róka et al. (2021), which is still in official use in Hungary. Our earlier models used official data on hospitalizations to predict future infection event numbers. These predictions were also helpful in the early stages of the epidemic.

In this paper, we investigate how far the predictions can be improved with the incorporation of wastewater virus load data into the model. Therefore, our objective is to combine and enhance our previous models with wastewater data. We show that the additional information from wastewater-based observations efficiently improves model calibration and prediction power, which further supports epidemic monitoring and decision-making.

## 2. Materials and methods

This section is divided into four major parts. After a brief description on the methodology for collecting and processing wastewater measurements, we introduce a dynamic epidemic model in Section 2.2. Then, we propose an epidemic data reconstruction approach in two steps in Sections 2.3 and 2.4. The three subsections start with a short problem formulation.

The technical details of the methodology can be read in the Appendix, which is divided into four parts. First, a formal epidemic model is described in Appendix A, secondly, the dynamic model inversion is detailed in Appendix B. Then, a specific immunization model is presented in Appendix C, which allows to estimate additional quantities, such as the rate of immunity loss (Appendix D). Brief parameter and noise sensitive analyses of the proposed approach are given in the supplementary material.

### 2.1. Viral load measurement in wastewater

SARS-CoV-2 genome copy numbers in sewage were determined from samples obtained from the Hungarian national wastewater monitoring system. Wastewater samples have been collected weekly since July 2020 from the following locations:

- in the three wastewater treatment plants (WWTP) of Budapest,
- from the county seats ($n = 18$), and
- as a composite sample from 5 cities in the Budapest agglomeration.

This sampling scheme covers 40% of the Hungarian population. All samples were taken after a grid filter at wastewater treatment plants. The samples are still raw sewage, only the lumpy waste is filtered out. Wastewater samples were processed as published previously by Róka et al. (2021). Briefly, cell debris of a $\bar{V}_0 = 50\,\mathrm{mL}$ initial sample volume was sedimented by centrifugation, and the supernatant was concentrated by flat-sheet ultrafiltration membrane at room temperature. Viral particles were recovered from the membrane by vortexing in virus transport medium ($\bar{V}_c = 1\,\mathrm{mL}$). Nucleic acids were extracted from the concentrate by QiaAmp Viral Mini Kit according to the manufacturer's instructions. The concentrate of volume $\bar{V}_s = 140\,\mu\mathrm{L}$ was lysed and loaded on centrifuge extraction columns. Then, RNA was eluted in $\bar{V}_e = 30\,\mu\mathrm{L}$ final volume. SARS-CoV-2 genome copies were quantified by qPCR using N1 as a target gene. Finally, the value of the measurement is computed as follows:

$$C_{\mathrm{raw}} = \frac{\bar{V}_e}{\bar{V}_s} \cdot \frac{\bar{V}_c}{\bar{V}_0} \cdot \mathrm{qPCR\ result\ [GC/L]}. \tag{1}$$

We note that the raw samples were normalized using measured daily flow volumes and Enterococcus counts as described in Róka

$$
\begin{cases}
\mathbf{S}_{k+1} = \mathbf{S}_k - \overbrace{\beta_k \mathbf{S}_k(\mathbf{P}_k + \mathbf{I}_k + q_{\mathrm{A}}\mathbf{A}_k)/\mathbf{N}}^{u_k} - \nu_k \mathbf{S}_k + \omega_k \mathbf{R}_k, & \text{(2a)}\\
\mathbf{L}_{k+1} = \mathbf{L}_k + \beta_k \mathbf{S}_k(\mathbf{P}_k + \mathbf{I}_k + q_{\mathrm{A}}\mathbf{A}_k)/\mathbf{N} - \tau_{\mathrm{L},k}\mathbf{L}_k, & \text{(2b)}\\
\mathbf{P}_{k+1} = \mathbf{P}_k + \tau_{\mathrm{L},k}\mathbf{L}_k - \tau_{\mathrm{P},k}\mathbf{P}_k, & \text{(2c)}\\
\mathbf{I}_{k+1} = \mathbf{I}_k + p_{\mathrm{I},k}\tau_{\mathrm{P},k}\mathbf{P}_k - \tau_{\mathrm{I},k}\mathbf{I}_k, & \text{(2d)}\\
\mathbf{A}_{k+1} = \mathbf{A}_k + (1-p_{\mathrm{I},k})\tau_{\mathrm{P},k}\mathbf{P}_k - \tau_{\mathrm{A},k}\mathbf{A}_k, & \text{(2e)}\\
\mathbf{H}_{k+1} = \mathbf{H}_k + p_{\mathrm{H},k}\tau_{\mathrm{I},k}\mathbf{I}_k - \tau_{\mathrm{H},k}\mathbf{H}_k, & \text{(2f)}\\
\mathbf{D}_{k+1} = \mathbf{D}_k + p_{\mathrm{D},k}\tau_{\mathrm{H},k}\mathbf{H}_k, & \text{(2g)}\\
\mathbf{R}_{k+1} = \mathbf{R}_k + (1-p_{\mathrm{H},k})\tau_{\mathrm{I},k}\mathbf{I}_k + \tau_{\mathrm{A},k}\mathbf{A}_k + (1-p_{\mathrm{D},k})\tau_{\mathrm{H},k}\mathbf{H}_k + \nu_k \mathbf{S}_k - \omega_k \mathbf{R}_k. & \text{(2h)}
\end{cases}
$$

*Linear disease course model with unknown input $u_k$*

**Box I.**

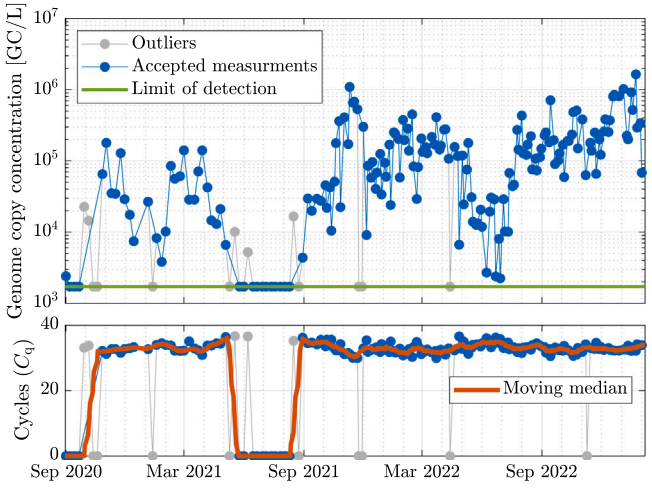



**Fig. 1.** PCR results in Szeged city: genome copy concentrations (upper) and numbers of PCR cycles (lower). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

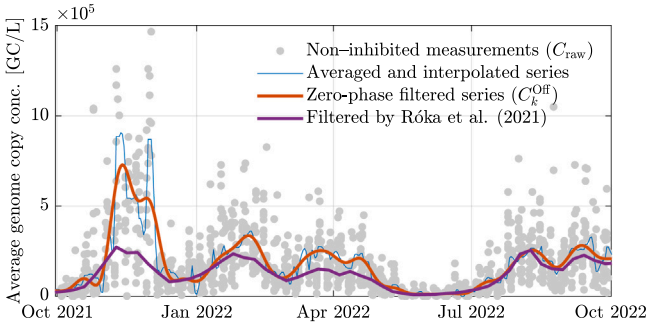



**Fig. 2.** Filtered time series alongside the raw wastewater data.

et al. (2021). In eight WWTPs, two samples were processed a week, finally resulting in 30 samples a week, which were collected between Monday and Thursday (10, 4, 12, and 4 samples, respectively). The measurements were filtered as proposed by Fernandez-Cassi et al. (2021), namely, we computed the moving median of the PCR cycles ($C_{\mathrm{q}}$) separately for all WWTPs using a centered 31 days long window. Then, we considered a PCR result inhibited if $C_{\mathrm{q}}$ was more than three cycles beyond the median. The raw PCR results, i.e., the genome copy concentrations and the numbers of cycles, obtained in Szeged city are visualized in Fig. 1, where the inhibited measurements (so-called outliers) are shaded, whereas, the red curve highlights the moving median of $C_{\mathrm{q}}$.

For each day, the weighted average of the measurements taken on that day was calculated, such that the weights are the population size at the WWTPs. The missing values at the weekends were computed by linear interpolation. To smooth the obtained daily time series, we used a 4th order low-pass Butterworth filter with $f_{\mathrm{co}} = (1/14)$ [1/days] cut-off frequency. We performed a *zero-phase filtering*, in the sense that the raw time series were filtered twice in both the forward and reverse directions. This type of filter is not causal, namely, the value at a given time point depends on the future values too, therefore, the averaged and interpolated values of the last 7 days are repeated once at the end of the time series. The filtered time series is denoted by $C_k^{\mathrm{Off}}$, where $k$ is the time label in days. In Fig. 2, we illustrate $C_k^{\mathrm{Off}}$ in comparison with the raw measurements. In the same figure, we present the data processed as proposed by Róka et al. (2021).

**Remark 1.** The applied filter removes the high-range frequency components with periods less than 2 weeks, which may result from an uneven sampling of the effluent over a week. Measurement errors may also contribute to high-frequency components.

### 2.2. Dynamical epidemic model

Long-term (multi-year) epidemic reconstruction is a particularly challenging task, as the pathogen has undergone significant mutations over the years. Social distancing rules, vaccination, the emergence of new immune evasive *variants of concern* (VoC), and the continuously waning immunity further add to the difficulty of epidemic modeling.

#### 2.2.1. Modeling goals

In this section, we develop a compartmental *ordinary difference equation* (ODE) model, where both the immunity loss and the clinical immunization are included, and the major model parameters are allowed to vary in time. Inspired by Péni et al. (2020), we considered intermediate states of the illness, including the hospital treatment.

The amount of detail in the model used was determined on the basis that

- the number of parameters is still computationally tractable,
- based on past experience, intermediate states are expected to be well observable from measurements.

#### 2.2.2. Compartments

We consider an improved compartmental model, where the population of **N** individuals is divided into the following groups. The

**Table 1**
Model constants used during the reconstruction.

| Parameter | Notation | Value (by virus variant) | | | | | |
|---|---|---|---|---|---|---|---|
| | | Wild | Alpha | Delta | BA.1 | BA.2 | BA.5 |
| 1: Latent period [days][a] | $1/\tau_{\mathrm{L}}$ | 3 | 2.5 | 1.7 | 1.2 | 1.2 | 1.2[c] |
| 2: Pre-sympt. period [days][a] | $1/\tau_{\mathrm{P}}$ | 3 | 3 | 2.3 | 1.8 | 1.8 | 1.8[c] |
| 3: Main seq. of sympt. [days][a] | $1/\tau_{\mathrm{I}}$ | 4 | 4 | 3 | 3 | 3 | 3[c] |
| 4: Asympt. period [days][a] | $1/\tau_{\mathrm{A}}$ | 4 | 3 | 2 | 2 | 2 | 2[c] |
| 5: Hosp. period [days] | $1/\tau_{\mathrm{H}}$ | 12[b] | 11.5[b] | 10[b] | 6[b] | 6[c] | 5[c] |
| 6: Rel. inf. of asympt.[a] | $q_{\mathrm{A}}$ | 0.75 | | | | | |
| 7: Prob. of developing sympt.[a] | $p_{\mathrm{I}}$ | 0.6 | 0.5 | 0.4 | 0.4 | 0.35[c] | 0.35[c] |
| 8: Hospitalization probability[a] | $p_{\mathrm{H}}$ | 0.076 | 0.07 | 0.05 | 0.015 | 0.015 | 0.015[c] |
| 9: Population of Hungary | **N** | 9.8 million | | | | | |

[a]Parameter values retrieved from Csutak et al. (2022).
[b]Estimated from the hospital admission numbers recorded by the National Public Health Centre.
[c]The authors' assumption.

group of susceptible people (**S**) comprises those who have never been infected or vaccinated but also any individual who are susceptible again by the waning immunity or the immune evasion of a new VoC. The group of immune, or so-called *protected* individuals (**R**) includes those who gained immunity either by recovery or vaccination. The group of infected people is categorized according to the different phases and possible outcomes of the disease. In this way, we distinguish people in the latent (**L**), presymptomatic (**P**), and the main sequence phase of the disease. Note that the incubation period corresponds to the sum of the length of the latent and presymptomatic phases. The reason, we make a difference between **L** and **P** is that people in the latent phase are not yet infectious, but they are in the presymptomatic phase.

To model the possible outcomes of the disease in the main sequence, we consider four severity scenarios, which are parameterized with three probability coefficients. First, a group of people remains asymptomatic (**A**), while others produce symptoms (**I**) with probability $p_{\mathrm{I}}$. A person who have symptoms is hospitalized (**H**) with probability $p_{\mathrm{H}}$, whereas, a hospitalized patient may decease (**D**) with probability $p_{\mathrm{D}}$.

The average residence time in compartments **L**, **P**, **I**, **A**, **H** are denoted by the inverses of time coefficients $\tau_{\mathrm{L}}$, $\tau_{\mathrm{P}}$, $\tau_{\mathrm{I}}$, $\tau_{\mathrm{A}}$, $\tau_{\mathrm{H}}$, respectively.

### 2.2.3. Time range of epidemic reconstruction

We consider a two and a half years long range of time for the reconstruction, which has the following important phases in Hungary:

1. (Wild) The first case in Hungary was detected in March 2020. Due to the strict official restrictions in Spring 2020, the first wave was drastically suppressed, but the wild SARS-CoV-2 virus strain emerged again in Autumn 2020.
2. (Alpha) In February 2021, Alpha variant emerged.
3. (Delta) Delta variant appeared in July 2021.
4. (BA.1) Omicron appeared around the end of 2021 and the beginning of 2022.
5. (BA.2) Subvariant BA.2 started to spread in March 2022.
6. (BA.5) In June 2022, subvariant BA.5 appeared.
7. (BA.5-Tp1) As a turning point in the spread of BA.5, in September 2022, multiple mutants of subvariant BA.5 have appeared.
8. (BQ.1) A rebound was observed in November 2022 caused by the emergence of a new Omicron subvariant, the BQ.1.

The milestone dates were approximated from the data of the National Public Health Centre but are also published online by Hodcroft (2023).

### 2.2.4. Available data

Our computations rely on the measured average genome copy concentration in wastewater, the number of hospitalized and deceased patients with COVID, and the number of registered vaccinations, i.e., the number of received first, second, and booster doses. The types of vaccine doses are not available in this data set, but the types of vaccines that require a single dose to obtain the full vaccination are negligible in Hungary (Atlo Team, 2021b).

The reconstruction is regularized using the registered Oxford Stringency Index $I_{\mathrm{s},k}$ (Mathieu et al., 2020). According to Hale et al. (2021), $I_{\mathrm{s},k} \in (0,1)$ quantifies the strictness of the official restrictions on the social distancing (e.g., mask wearing rules, curfew, postponed social events, etc.). During the reconstruction, we also make use of the estimated relative infectiousness of each VoC compared to the wild-type SARS-CoV-2 strain.

During the evaluation of the obtained results, we consider the estimated reproduction rate made by the Group of *Our World in Data* (OWID) (Mathieu et al., 2020). Moreover, the estimated number of new cases is compared with the scaled number of the officially detected cases.

From the authors' experience, the hospital load as a major indicator has not been reliable since the emergence of BA.2 variant, moreover, the official detected cases are not representative as a potential validation data since the last quarter of 2022 due to the continuously decreasing testing intensity. In this case study, the test positivity rate is not considered because fluctuation of testing strategies in Hungary highly influences the data, moreover, test positivity rate is not representative since the first quarter of 2022.

### 2.2.5. Assumptions

According to McEvoy et al. (2021) and Nakajo and Nishiura (2021), an asymptomatic person in the main sequence is less infectious compared to those in the presymptomatic phase or others in the main sequence having symptoms. Therefore, we make the following assumptions.

**Assumption 1** (*Infectious Population*)**.** We consider only $q_{\mathrm{A}}$ portion of asymptomatic people being infectious, but in the same rate as another symptomatic individual in the main phase of the disease. We neglect the possible infections originating from hospitalized or deceased patients.

Nourbakhsh et al. (2022) considered a detailed epidemic model, in which the *infectious people* are the major contributors to the viral load entering the sewer system. They quantified the viral load as a *linear* function of the infectious population and those in the post-infection period who have just recovered, but the hospitalized patients and the infected people in the latent phase are excluded from the shedding population. We also note that the estimates of Hewitt et al. (2022), Nourbakhsh et al. (2022), and Phan et al. (2023) made for the shedding characteristics show orders of magnitude lower shedding rates for individuals in the latent phase of the disease and about 10 days after symptoms onset. Motivated by these observations, Phan et al. (2023) did not consider recovered people as relevant contributors to the viral load.

On the other hand, Nourbakhsh et al. (2022) highlights that the viral load entering the sewer system is exposed to an exponential viral decay and is affected by the possible hydrodynamic processes (e.g., dilution,

sedimentation and resuspension) that leads to RNA degradation but also delays. In this study, we neglect the in-sewer degradation for simplicity, and assume that the actual viral load entering the sewer system is proportional to the national average genome copy concentration without delays. Considering that the measurements arrive from 22 different WWTPs, that cover about 4 million people, it is reasonable to assume that in-sewer anomalies are compensated after the averaging, filtering, and smoothing steps described in Section 2.1.

Finally, to relate the shedding population with the actual measurements, we consider a ratio denoted by $q_c$ between the number of infectious people and the national average genome copy concentration as it is formulated in the following assumption.

**Assumption 2** (*Virus Secretion*)**.** The national average genome copy concentration measured in wastewater is assumed to be proportional to the number of infectious people at that day without delay. The coefficient $q_c$, which relates the measured concentration and the infectious population is called the *virus secretion coefficient*.

The secretion coefficient $q_c$ is later calibrated using optimization for each variant. Although $q_c$ has no underlying physiological content, the comparison of the obtained values for $q_c$ can estimate the relative rate of secretion during the whole course of the disease caused by two different variants, averaged for the entire population. The measurements of Kang et al. (2022) (see, e.g., Panel (B) in Fig. 2) suggest that roughly 3 times more genome copies are shed during the Delta infection than during the Omicron wave. Therefore, we expect $q_{c,\mathrm{Delta}}$ estimated for the Delta variant to be about 3 times higher compared to $q_{c,\mathrm{Omicron}}$ obtained for Omicron variants.

For the model parameters, we make the following assumptions.

**Assumption 3** (*Model Coefficients*)**.** We assume that the relative infectiousness of asymptomatic people $q_A$ is known and constant. Moreover, the time coefficients $\tau_\bullet$ and the probability of symptomatic infection $p_I$ are considered known for all VoC emerged in Hungary. Finally, when a new VoC becomes dominant, the variant-dependent coefficients ($\tau_\bullet$, $p_I$) are changing smoothly. The presumed model coefficients are presented in Table 1.

The probability of hospitalization $p_H$ and of fatal outcome $p_D$ or the average hospitalization period $1/\tau_H$, depend not only on the properties of the dominant VoC but also on the hospitalization strategies (Kemp et al., 2021; Kozyreff, 2021). Therefore, these parameters may vary for the different countries. In this study, we focus on the epidemic waves in Hungary.

The average hospitalization length in 2021 (Row 5 in Table 1) were estimated from the hospital admission numbers recorded by the National Public Health Centre. Other constants are applied from Csutak et al. (2022).

**Assumption 4** (*Unknown Parameters*)**.** The probability coefficients $p_H$, $p_D$ of the disease course model, and the secretion coefficient $q_c$ of the pathogen are *unknown parameters*, but constant during the dominance of a given VoC. For BA.1 and BA.2 subvariants of the Omicron variant, the virus secretion coefficient is presumed identical.

**Assumption 5.** The transmission rate of the pathogen is considered smoothly time-varying.

In this work, we consider a model with vaccination, where three vaccination statuses are distinguished: not fully vaccinated, fully vaccinated, and boosted. We make the following simplifying assumptions.

**Assumption 6** (*Vaccination*)**.** Our model uses the following vaccination rules:

1. Full vaccination is ensured with two vaccine doses. (Single-dose full vaccination is negligible in Hungary.)

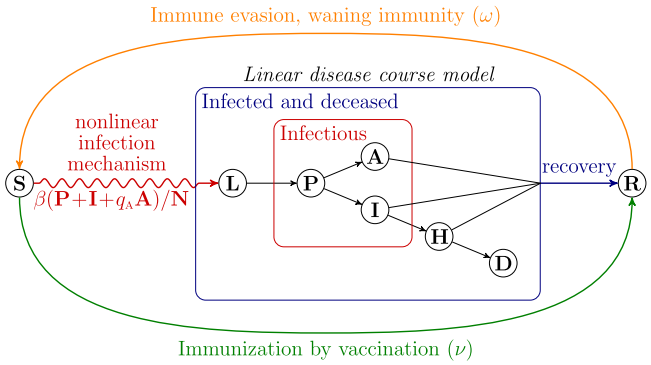



**Fig. 3.** Transition diagram of the epidemic process. Circles represent the 8 compartments detailed in Section 2.2.2, the arrows represent transitions between compartments. The capitals in the circles stand for: susceptible (**S**), latent (**L**), presymptomatic (**P**), asymptomatic (**A**), infected (**I**), hospitalized (**H**), deceased (**D**), recovered (or more precisely, protected: **R**). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. The probability that a partially vaccinated individual will receive a second dose is $p_{V_2|V_1}$.
3. The time elapsed between the first two doses is normally distributed.
4. A person who receives the full vaccine or a booster dose becomes temporarily immune with probability 1, but only after a certain period of time, which is normally distributed.
5. Infected people are not vaccinated, including those in the latent or the presymptomatic period.
6. A susceptible and a recovered individual receives the second (immunizing) vaccine dose with the same probability.
7. A fully vaccinated individual and a boosted susceptible receives a booster dose with the same probability.
8. Boosted individuals, who are still protected do not receive further booster doses but only after losing their immunity.

We note that the estimated under-detection rates (discussed later) suggest that the number of individuals who are aware of being infected by COVID is much less compared to those who are undetected infected (asymptomatic or infected with mild symptoms not attributed to COVID). This phenomenon is even more prevalent in the more advanced stages of the epidemic (e.g., during the emergence of subvariant BA.1). Considering this, points 6 and 7 of Assumption 6 are sufficiently realistic. Moreover, repeated booster doses are not excluded in Point 8 Assumption 6, but a rational administration of vaccine doses is presumed. Namely, a sufficiently long time should elapse between two booster doses. A preliminary version of our vaccination model was presented in Csutak et al. (2022).

In this study, we make the following simplifying assumption on the relative risk of infection.

**Assumption 7** (*Uniform Infection*)**.** The risk of a susceptible person becoming infected is independent of their vaccination status or previous history of infection.

Our model considers only one single immune state. A person can be susceptible to the virus (**S**) or protected (**R**), but we neglect further intermediate immune states. However, a protected individual may lose its immunity (completely) with a given rate due to a natural immune waning or due to the immune evasion capability of a new emerging virus strain. We consider the following assumption for immunity loss:

**Assumption 8** (*Uniform Immunity Waning*)**.** On a given day, a protected individual has the same probability of losing their immunity regardless of its vaccination status or previous history of infection.

Note that Assumptions 7 and 8 do not undermine the generality of the method as it can be parameterized otherwise with pre-estimated relative risk coefficients, which were measured in certain states, e.g., King County (2022).

Finally, we assume that the natural deaths and births are negligible compared to the deaths caused by the COVID.

#### 2.2.6. Dynamic equations

Finally, we divide individuals into eight disjoint groups, called the compartments, which describe the epidemic state of the population on a given day $k$. The transitions between the compartments are illustrated in Fig. 3, whereas, the transition dynamics are characterized by the discrete-time dynamical model present in the floating Eq. (2). The model parameters are explained in Table 1, where the italic letter $k$ in the subscripts denotes the time label of the respective quantities.

#### 2.2.7. Infection mechanism

The nonlinear infection process, illustrated by the red curly arrow in Fig. 3, is encoded in the first two Eqs. (2a) and (2b) (see the equation in Box I), where the transmission rate of the disease is characterized by the time-varying coefficient $\beta_k$.

With reference to Assumption 1, the total number of infectious people who are able to transmit the pathogen is

$$y_k = \mathbf{P}_k + \mathbf{I}_k + q_{\mathrm{A}}\mathbf{A}_k. \tag{3}$$

According to the simplified vaccination rules (Assumption 6), the numbers of people in the different phases of the disease depend only on the number of new infections within the unit time frame (24 h) caused by the $y_k$ infectious people. The number of newly infected individuals on day $k$ is denoted by $u_k$ and it can be expressed as follows:

$$u_k = \beta_k y_k \mathbf{S}_k/\mathbf{N}. \tag{4}$$

On the other hand, the new recoveries in (2h) are described by the following terms:

$$z_k = (1 - p_{\mathrm{H},k})\tau_{\mathrm{I},k}\,\mathbf{I}_k + \tau_{\mathrm{A},k}\,\mathbf{A}_k + (1 - p_{\mathrm{D},k})\tau_{\mathrm{H},k}\,\mathbf{H}_k. \tag{5}$$

#### 2.2.8. Vaccination and waning immunity

The transition between the groups of susceptible (**S**) and protected people (**R**) is affected indirectly through infection/recovery, but also by vaccination, waning immunity, or the immune evasion capability of a new VoC. The immunization effect of vaccination (green arrow in Fig. 3) is expressed by a single time-varying coefficient $\nu_k$, which denotes the proportion of susceptible people within a unit time frame who become immune by vaccination. The *immunization rate* $\nu_k$ can be estimated from the available vaccination data.

The loss of immunity is captured by a direct transition from the group of protected individuals (**R**) to susceptibles (**S**), as illustrated by the orange arrow in Fig. 3. The model does not distinguish between repeated susceptibility caused by the waning immunity or by the immune escape of a new VoC. Both cases are considered as an event when an individual becomes susceptible again, i.e., immunity is lost (completely). The proportion of protected people who lose immunity within a unit time frame is encoded in the so-called *immunity loss rate* $\omega_k$.

Immunization depends on the vaccination strategies, whereas, a sudden massive loss of immunity may be due to the emergence of a new VoC with a strong immune evasion capability. Both effects are typically time-dependent, which are encoded by linear transitions between **S** and **R** with two time-dependent rate coefficients $\nu_k$ and $\omega_k$, respectively. The outer loop of the transition diagram illustrates that the immunization and waning are different in nature than the indirect transition from **S** to **R** through infection-recovery.

#### 2.2.9. Epidemic process as an interconnected system

The direct transitions between susceptible and protected population make the epidemic dynamics uncertain as the dynamic nature of waning immunity, and hence the daily rate of immunity loss ($\omega_k$) are unknown. Moreover, immune waning makes the estimates of vaccination efficiency more uncertain. Thus, the number of registered vaccinations is useful but not sufficient data to compute the daily rate of immunization $\nu_k$.

Fortunately, the advantageous properties of the epidemic process model allow us to decouple a linear subsystem (2b)–(2f) from the overall compartmental model. Highlighted in a blue box both in (2) and in Fig. 3, this subsystem describes the course of the disease in the infected population. According to Assumption 6, the disease course model is interconnected with the immunization-waning dynamics only through the daily new infections $u_k$ (4) and recoveries $z_k$ (5). Once the sequences $u_k$ and $z_k$ are known (and fixed), the immunization-waning dynamics can be analyzed separately.

The technical details of dynamic model representation are described in Appendix A.

### 2.3. New cases, infected population, reproduction number

In this section, we use wastewater measurements and hospitalization data to solve the disease dynamics, i.e., to compute all unknown data of the isolated subsystem of (2).

#### 2.3.1. Problem and outline

The advantageous (linear) model structure of the disease course model makes it possible to perform a partial epidemic reconstruction. Through a single optimization-based dynamic data assimilation problem, we *compute* (C) the following unmeasured quantities:

- C1. the number of new cases ($u_k$),
- C2. the population in the compartments **L**, **P**, **I**, **A**,
- C3. the number of new recoveries ($z_k$),
- C4. the effective reproduction number ($R_{\mathrm{c},k}$),
- C5. the unknown variant-dependent probabilities ($p_{\mathrm{H}}$, $p_{\mathrm{D}}$)
- C6. and the secretion coefficient ($q_{\mathrm{c}}$).

Through the optimization, we compute epidemic data above such that the reconstructed quantities fit to the following *measurements* (M):

- M1. the measured genome copy concentration in wastewater ($C_k^{\mathrm{Off}}$), i.e., $y_k \approx q_{\mathrm{c},k} C_k^{\mathrm{Off}}$,
- M2. the registered number of hospitalized patients ($\mathbf{H}_k^{\mathrm{Off}}$),
- M3. the registered number of deceased patients ($\mathbf{D}_k^{\mathrm{Off}}$),

as much as possible.

**Remark 2.** The disease course dynamics have advantageous stability and observability properties. Csutak et al. (2021, 2022) proved that for fixed model coefficients the state variables ($\mathbf{L}, \mathbf{P}, \mathbf{I}, \mathbf{A}, \mathbf{D}$) are observable using only hospitalization load **H**, independently of the rates of immunity loss or medical immunization. These results suggest that the epidemic state observation formulated as an optimization problem has a solution.

#### 2.3.2. Dynamic inversion of disease course model

The unknown quantities of the disease course model can be classified as follows. The number of new cases $u_k$ (4) is considered as an *unknown disturbance input* of the disease course model. Moreover, the number of infected individuals in the different phases of the disease are *unknown states* (**L**, **P**, **I**, **A**). According to Assumption 4, coefficients $p_{\mathrm{H}}$, $p_{\mathrm{D}}$, and $q_{\mathrm{c}}$ are *piecewise constant unknown parameters*. Finally, the sequence of new recoveries $z_k$ (5) is a *computed variable*.

Using the available measurement, we are able to reconstruct the unknown input $u_k$, the unmeasured state variables (**L**, **P**, **I**, **A**), the unknown variant-dependent model constants $p_{\mathrm{H}}$, $p_{\mathrm{D}}$, $q_{\mathrm{c}}$ simultaneously

alongside further dependent variables, such as the reproduction rate or the new recoveries.

The technical details of the dynamic inversion approach are summarized in Appendix B.

### 2.4. Susceptible population, transmission rate, waning

In this section, we solve the immunization-waning loop using the computed number of new cases and recoveries, the registered stringency index, vaccination data, and an enhanced vaccination model.

#### 2.4.1. Problem and outline

To make a reliable prediction for the future evolution of the pandemic, it is essential to have appropriate estimates for the following *unknown* (U) rate coefficients:

U1. the current transmission rate of the virus ($\beta$),
U2. the rate of immunization by vaccination ($\nu$),
U3. the current rate of immunity loss ($\omega$).

Furthermore it is also necessary to estimate

U4. the susceptible population ($\mathbf{S}$).

Note that the estimated rate of immunization (U2), can be used to quantify the effectiveness of vaccination campaigns. Moreover, the registered stringency index $I_{\mathrm{S}}$ determines a prior estimate on the transmission rate (U1).

However, estimating (U3) is particularly difficult as it captures two independent processes. First, the rate of immunity loss aims to model a continuous ambient immunity waning. Secondly, the rate of immunity loss may have sudden peaks as an emerging immune evasive VoC causes a massive loss of immunity in the protected population.

At the same time, waning affects the whole process to a high degree, as it transmits the protected individuals with an arbitrary vaccination status to the susceptibles. In this way, the loss of immunity will thoroughly disrupt the susceptible population, which becomes a mixture of individuals with different vaccination status and any possible history of infection. Not to mention the fact that the "mixture of susceptibles" is affected by the nonlinear infection process, where the average rate of transmission is again uncertain as the infectiousness may depend on the immunity status of the population.

To estimate the susceptible population and the rate coefficients, we use the computed quantities (C1)–(C3), and consider the following *data* (D):

D1. the registered vaccination data,
D2. the estimated stringency index ($I_{\mathrm{S}}$),
D3. the relative infectiousness of each VoC.

Unlike the linear disease course model, the immunization-waning dynamics lack the advantage of state observability. However, the (D2) and (D3) can regulate the estimated value of the transmission rate, whereas, (D1) with an initial estimate of waning determine the rate of immunization.

The problem is still under-determined as the number of unknowns outweighs the measurements. Therefore, our strategy is to find a solution, for which

$$\textit{the rate of immunity loss is minimal}, \tag{6}$$

such that

1. the transmission rate is close to an initial estimate made on the bases of D2 and D3,
2. both the transmission and the immunity loss rates are smoothly varying in time,
3. the unknown coefficients move between reasonable bounds.

Other estimates on the reproduction number (Mathieu et al., 2020; Atlo Team, 2021a), and our previous results (Polcz et al., 2022; Csutak et al., 2022) serve as good bases to further regularize the problem.

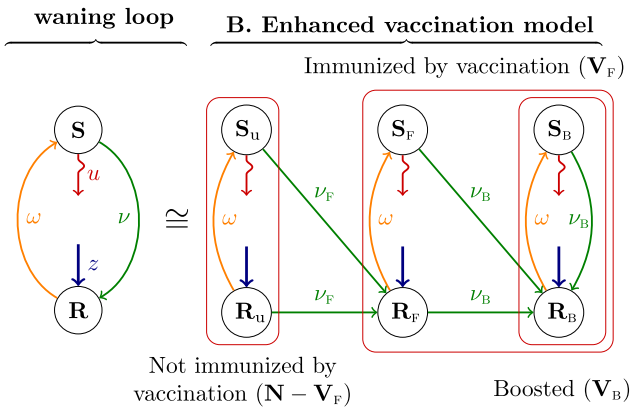



**Fig. 4.** An enhanced vaccination model. The circles are compartments, the arrows illustrate the transitions between compartments. The notations for compartments are detailed in Section 2.4.2. Variables $u$ and $z$ denote the new cases (4) and recoveries (5), $\omega$ is the immunity loss rate, $\nu$ is the immunization rate, finally, $\nu_{\mathrm{F}}$ and $\nu_{\mathrm{B}}$ constitute the immunization rates (C.8) through full vaccinations and boosters, respectively. The enhanced vaccination model is detailed in Appendix C.

#### 2.4.2. Detailed vaccination model

In model (2), it is not straightforward how to relate the immunization rate $\nu_k$ to the registered first, second, and booster vaccine doses (D1). Therefore, we consider an enhanced vaccination model with three vaccination statuses. First, we consider a group of individuals who are *fully vaccinated* and immunity has already been established after the full vaccination. Secondly, a group of people are said to be *boosted* if immunity has already been developed after receiving a booster dose. Finally, the remaining population is said to be *not fully vaccinated*. The number of fully vaccinated and boosted individuals on day $k$ are denoted by $\mathbf{V}_{\mathrm{F},k}$ and $\mathbf{V}_{\mathrm{B},k}$, respectively.

From a computational point of view, dividing all compartments according to the vaccination status is not fortunate. But observe that $\mathbf{S}$ and $\mathbf{R}$ are manipulated by the infected compartments through the number of new cases $u_k$ and new recoveries $z_k$ only, which are already computed. Therefore, we remove the disease course dynamics from the initial model as illustrated in Panel A of Fig. 4.

Then, we distinguish three subgroups of both the susceptible ($\mathbf{S} = \mathbf{S}_{\mathrm{u}} + \mathbf{S}_{\mathrm{F}} + \mathbf{S}_{\mathrm{B}}$) and the protected population ($\mathbf{R} = \mathbf{R}_{\mathrm{u}} + \mathbf{R}_{\mathrm{F}} + \mathbf{R}_{\mathrm{B}}$) as follows:

$\mathbf{S}_{\mathrm{u}}$ susceptible individuals who have not yet acquired immunity through vaccination,
$\mathbf{R}_{\mathrm{u}}$ individuals, who are not fully vaccinated, but who have acquired immunity through recovery,
$\mathbf{S}_{\mathrm{F}}$ individuals, who have been fully vaccinated but have lost their immunity due to the waning immunity or the immune evasion capability of a new VoC,
$\mathbf{R}_{\mathrm{F}}$ fully vaccinated individuals, who are still protected (possibly both by recovery or vaccination),
$\mathbf{S}_{\mathrm{B}}$ boosted individuals, who lost their immunity,
$\mathbf{R}_{\mathrm{B}}$ boosted individuals, who are still protected (possibly both by recovery or a booster dose).

The transitions between the sub-compartments are illustrated in Panel B of Fig. 4.

The enhanced vaccination model uses two rate coefficients for immunization. On day $k$, $\nu_{\mathrm{F},k}$ proportion of the not fully vaccinated population ($\mathbf{N} - \mathbf{V}_{\mathrm{F}}$) acquires immunity by full vaccination and moves to compartment $\mathbf{R}_{\mathrm{F}}$. Moreover, $\nu_{\mathrm{B},k}$ proportion of fully vaccinated individuals ($\mathbf{V}_{\mathrm{F}}$) regain or retain immunity by receiving a booster dose. Finally, $\nu_{\mathrm{B},k}$ proportion of boosted susceptibles regain immunity by receiving an additional booster dose.

To represent the indirect infection-recovery transition between each pair of sub-compartments ($\mathbf{S}_\bullet, \mathbf{R}_\bullet$), the number of new cases and recoveries are distributed appropriately according to the vaccination status of susceptible people.

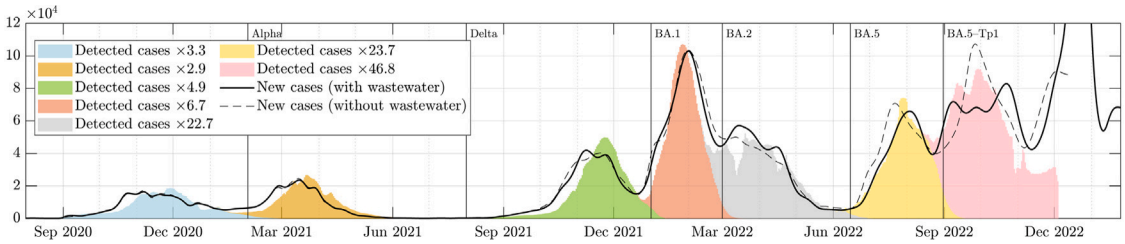



**Fig. 5.** Daily new cases computed retrospectively with and without wastewater data. The colored areas illustrate the official detected cases separately for all epidemic waves multiplied by the estimated under-detection rates ($q_u$) given in the legend as the multiplies in "Detected cases $\times q_u$". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

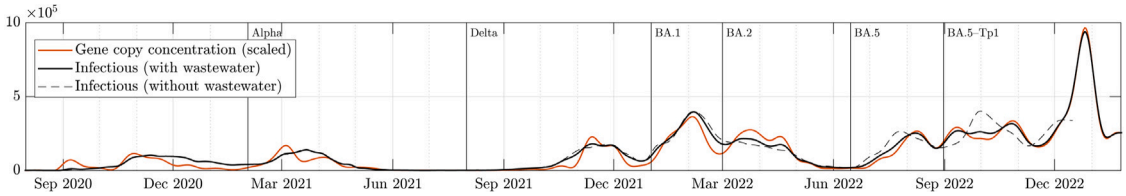



**Fig. 6.** Computed number of infectious people compared to the virus concentration in wastewater scaled by the virus secretion coefficient.

**Table 2**
Model constants computed during the retrospective reconstruction.

| Parameter | Notation | Value (by virus variant) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Wild | Alpha | Delta | BA.1 | BA.2 | BA.5 | BA.5-Tp1 |
| 1: Hosp. prob. (calibrated) | $p_H$ | 0.076 | 0.105 | 0.05 | 0.0221 | 0.019 | 0.0129 | 0.00962 |
| 2: Prob. of fatal outcome (estimated) | $p_D$ | 0.259 | 0.218 | 0.233 | 0.136 | 0.075 | 0.075 | 0.075 |
| 3: Under-detection rate | $q_u$ | 3.3 | 2.9 | 4.9 | 6.7 | 22.7 | 23.7 | 46.8 |
| 4: Virus secretion coef. (estimated) | $1/q_c$ | 2.54 | 1.2 | 3.21 | 0.923 | 0.923 | 0.967 | 0.967 |

A detailed description of the enhanced immunization model is available in Appendix C.

**Remark 3.** Until the middle of 2022, data provided by official bodies suggested that it was reasonable to separate the fully vaccinated and the boosted population. Our choices and points 1 and 4 in Assumption 6 were also confirmed by the definitions used in King County (2022). However, separating the two states does not seem to be important today, therefore, the vaccination model is planned to be reformulated in the future.

#### 2.4.3. Heuristics to estimate the rate coefficients

The simultaneous reconstruction of the three smoothly time-varying rate coefficients ($\beta_k$, $\nu_k$, $\omega_k$) is not straightforward. Therefore, as the initial step, we give a preliminary estimate for the immunization rate $\nu_k$ using the registered vaccination data with the assumption that $\omega \equiv 0$.

Then, the priori estimate on $\nu_k$ allows us to formulate a simple dynamic data assimilation problem to compute the transmission rate $\beta_k$ and the immunity loss rate $\omega_k$ of the epidemic process. Using the estimated rate $\omega_k$, we can update the immunization rate $\nu_k$, which makes possible to improve the estimates for $\omega_k$ and $\beta_k$, etc.

Namely, the rate coefficients and the number of susceptibles and protected people in the three different vaccination statuses can be estimated iteratively using simple computational steps. The iterations converge and give an approximated solution to a more complex data assimilation problem.

The technical details of the optimization-based iterative approach are summarized in Appendix D.

## 3. Results and discussion

In this section, we evaluate and discuss the potential of the proposed approach for reconstruction and prediction.

### 3.1. Retrospective epidemic data reconstruction

First, we present and discuss a *retrospective reconstruction* for the entire time span of the pandemic in Hungary, using wastewater data until 1 February 2023.

#### 3.1.1. Estimated parameters and under-detection rate

The viral load in wastewater gives us information about the current epidemic state, whereas the hospitalization data gives delayed information about severe infections. These two data sets allow us to fit model parameters, which affect the course of the disease. As the incubation period and serological analyses can well approximate the average length of the illness, we had the possibility to estimate the hospitalization probability ($p_H$). The registered hospital load and the cumulative number of deaths gives the possibility to estimate the probability of a fatal outcome ($p_D$) if a person is hospitalized. The estimated probability coefficients for the different VoC are presented in Rows 1–2 of Table 2.

Due to the linear structure of the disease course model, the number of new cases and the number of infected people in the different phases of the disease can be reconstructed simply using dynamic inversion (Section 2.3). The computed number of new infections is illustrated in Fig. 5 compared with the official detected cases. The under-detection rate (denote by $q_u$) and then the scaled series of detected cases are computed separately for each epidemic wave as described in Remark 7. The scaled and "smoothly concatenated" curves of the consecutive outbreaks are also illustrated by the stacked area plot in Fig. 5. The estimated under-detection rates for the different waves are given in the legend of Fig. 5 and in Row 3 of Table 2.

We can make the following observations. The probability of developing symptoms and being hospitalized gradually decreases with the newer virus variants. The hospitalization probability peaks at the Alpha variant and then decreases gradually. This estimate corresponds to the fact that the highest peak of hospital load is attained in March 2021

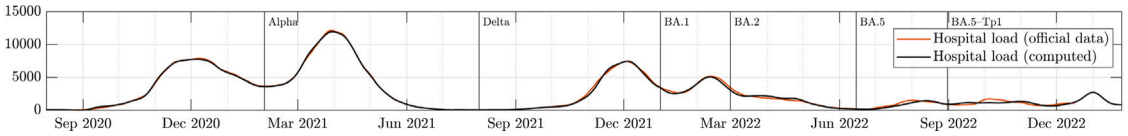



**Fig. 7.** Hospital load and its reconstruction.

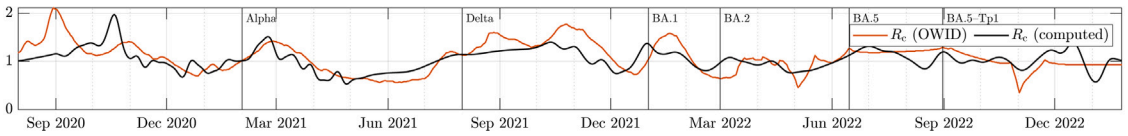



**Fig. 8.** Computed reproduction number $R_{c,k}$ (black) compared to the estimates made by the Group of Our World in Data (OWID).

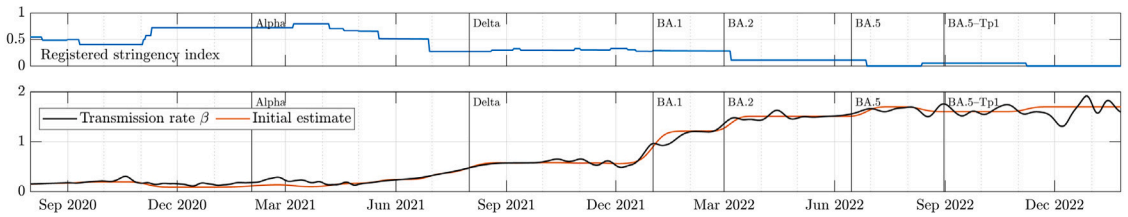



**Fig. 9.** Computed transmission rate $\beta_k$ (bottom, black) compared to the initial estimate $\beta_k^{\text{ref}}$ (bottom, red) determined by the stringency index (top).

during the third epidemic outbreak caused by the Alpha variant. In that case, the hospital load at the peak of the third wave exceeded the hospital capacity in Hungary (10 000 patients).

The estimated values of the secretion coefficient for the major variants are presented in Row 4 of Table 2. It is worth remarking that the virus secretion coefficient estimated for the Omicron variants reduced for a third compared to that obtained for Delta (Row 4 in Table 2). These estimates correspond to the results of Kang et al. (2022).

Examining the resulting values for the secretion coefficients (Row 4 of Table 2), we found that shedding is roughly 3.5 times more intensive during the Delta infection than the Omicron infection. The computed ratio is confirmed by the measurements of Kang et al. (2022). The secretion of the ancestor variant in comparison with the Delta variant seems reasonable, although, we haven't found a comparative study to confirm this result. Nevertheless, we should discuss the secretion coefficient computed for Alpha variant as it is only one 3rd of the value obtained for Delta, and fairly close to the value obtained for Omicron. The obtained secretion coefficient for the Alpha variant can be attributed to the higher hospitalization probability computed for the wave caused by the Alpha variant. Note that $q_c$ and $p_H$ are estimated simultaneously. The infectious (shedding) population in comparison with the scaled genome copy concentration is presented in Fig. 6.

The reconstructed curve of hospitalization along with the official data is presented in Fig. 7. The computed reproduction number is illustrated in Fig. 8 in alongside by the Bayesian estimate made by the Group of OWID (Mathieu et al., 2020). Finally, the heuristic approach in Section 2.4 is also able to estimate the transmission rate of the virus. Fig. 9 illustrates its evolution during the epidemic waves.

Section S1 of the supplementary document presents a brief parameter sensitivity analysis of our approach with respect to the *fixed* model constants in Table 1. Moreover, in Section S2 of the supplementary document, we analyze the uncertainty of the *estimated* parameters in Table 2 in the terms of the wastewater measurement noise.

### 3.1.2. Immunization and waning

The estimated immunization rate and the (best-case) rate of immunity loss are illustrated in Fig. 10. Knowing these two rate coefficients, we are able to compute the number of daily new immunizations and immunity loss events (Fig. 11), but also their cumulative curves (Fig. 12). When an individual loses immunity twice, e.g., first, after recovery and second, after vaccination, they are counted as two distinct events. In Fig. 12, we also present the estimated number of susceptible and protected people alongside the cumulative cases, immunizations, and immunity loss.

**Table 3**
Model-based estimates for the number of all cases, immunity loss, immunizations, and effective vaccinations during the separate epidemic outbreaks.

| % of population | Wild | Alpha | Delta | BA.1/2 | BA.5 |
|---|---|---|---|---|---|
| 1: All cases | 13% | 12% | 24% | 72% | 83% |
| 2: All immunity loss | 1% | 4% | 26% | 53% | 78% |
| 3: Clinical immunizations | – | 12% | 37% | 3% | – |
| 4: All vaccinated | – | 16% | 68% | 20% | – |
| 5: Effective vaccinations[a] | – | 78% | 54% | 14% | – |

[a]Computed as the proportion of the model-based estimates of the number of immunizations and the sum of full and booster vaccinations.

The reconstructed curves suggest that by the third quarter of 2022, the epidemic wave triggered by the emergence of variant BA.5 was rebounded multiple times, first in September 2022 possibly due to the star of school year, secondly, in November 2022 possible due to the appearance of the variant BQ.1 (Hodcroft, 2023).

In Table 3, we made estimates on the number of infections, immunity loss, and clinical immunizations. These estimates are computed separately for each epidemic outbreak. In Row 1, the estimated sum of all infections predicts that 155% of the population have been infected in 2022, and about 204% of the population have been infected during the whole time span of the COVID-19 pandemic. These estimates should be interpreted such that in average, the whole population became infected, which is not impossible considering the high immune evasive properties of the different subvariants of Omicron variant (Lyngse et al., 2022). However, as reported by Iwasaki (2021) or Sheehan et al. (2021), short-term reinfections, and hence quick immunity loss, are possible even for the wild-type or Alpha variants for individuals without any known immune disorder.

In Row 2, we estimated the number of people who have lost immunity during (and before) the outbreaks.

These estimates suggest a massive (131%) loss of immunity and/or a strong immune evasive capability of the emerging subvariants of Omicron during 2022. We highlight that these numbers are rough

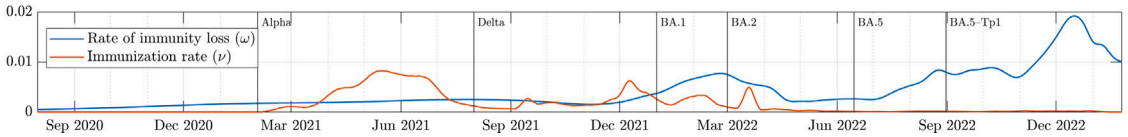



**Fig. 10.** Computed immunization rate $\nu_k$ (red) and the estimated rate of immunity loss $\omega_k$ (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

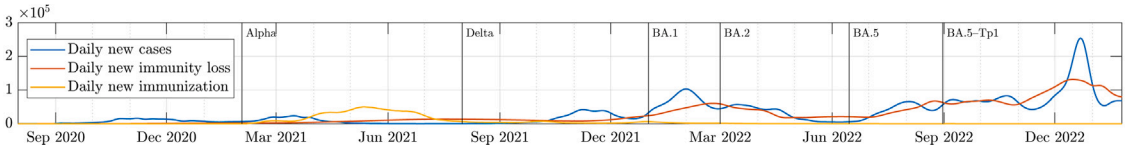



**Fig. 11.** Daily numbers.

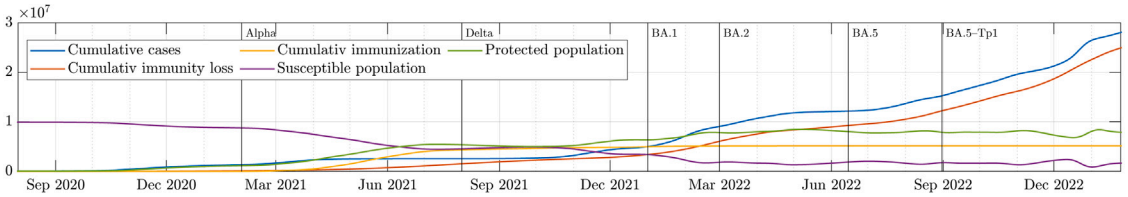



**Fig. 12.** The number of susceptible and protected individuals are illustrated in this plot. Moreover, the yellow curve illustrates the cumulative number of events, in which people have lost immunity. If an individual loses immunity twice (e.g., first, after recovery, then, after vaccination), this should be considered as two separate immunity loss events. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimates and are not appropriate for assessing the relative immune evasive capability of two virus variants. At the same time, the computed rate of immunity loss (Fig. 10) confirms the estimates of Lyngse et al. (2021) and Cocchio et al. (2022) that the Omicron BA.1 and BA.2 variants are at least four times immune evasive compared to the Delta variant.

We remark that the immunity loss rate is a time-varying coefficient, therefore, its inverse gives the expected time of being susceptible again. The computed expected time for losing immunity in specific periods of the epidemic was:

- 2021 (fall semester): 1/0.0027 [days] ≈ 1 year
- 2022 (first quarter): 1/0.008 [days] ≈ 4 months
- 2022 (Aug–Oct): 1/0.0085 [days] ≈ 4 months
- 2022 (Dec): 1/0.017 [days] ≈ 2 months

Considering that these parameters are *national estimates*, we found that they show good correlation with the findings of Stein et al. (2023) but also not too far from the more optimistic estimates of Willyard (2023).

In Row 5, the decreasing effectiveness of vaccinations is due to the fact that the majority of the vaccinated population was presumably immunized through recovery. We note that our model does not consider hybrid immunity, so a vaccine given to an immune person does not count as extra immunization. According to this interpretation, the first vaccination campaign in the 2021 wave was the most successful during and after the outbreak caused by the Alpha variant, as about 78% of the vaccinated population was not infected or had already lost immunity.

### 3.2. Prediction

Here, we illustrate the prediction potential of the proposed approach. First, we simulate short-term predictions at every day in two cases: (1) using both medical and wastewater data, (2) using only medical data. Secondly, we simulate multiple long-term (6 months long) predictions at a fixed date considering different scenarios (parameter values) but using both medical and wastewater data.

#### 3.2.1. Short-term predictions and comparison

The estimated current status of the epidemic spread allows us to simulate its future evolution. The current state and hence the future evolution of the epidemic can be estimated more accurately if the measured viral load in wastewater is used in the calculations. To illustrate the contribution of the water measurements, we performed 30-day long predictions every day starting 20 August 2020 with and without wastewater data.

The prediction at a given starting day (let us denote it by $k_0$) is simulated in three steps. First, the raw data are clipped to $k \in \{-\infty, \dots, k_0\}$, namely, the computations are performed without any measurements obtained for the future. Secondly, we perform a reconstruction to estimate the pandemic state on day $k_0$. Finally, a prediction is obtained by evaluating the dynamic recursion of the epidemic model (2).

To evaluate the quality of the predictions, the results of the reconstruction and prediction are compared with a *retrospective reconstruction* in both two cases: with and without wastewater data. Note that the reconstructions were performed using the same methodology as the *retrospective* reconstructions but with a smaller data set and with fixed coefficients $p_{\mathrm{H}}$ and $q_{\mathrm{c}}$.

**Remark 4.** The predictions both using and neglecting wastewater measurements assume that the hospitalization probability $p_{\mathrm{H}}$ and the secretion coefficient $q_{\mathrm{c}}$ are given and fixed a priori. However, the actual values of $p_{\mathrm{H}}$ and $q_{\mathrm{c}}$ considered during the current reconstructions were obtained through a retrospective reconstruction using wastewater data. Although we use the same model constants in both cases (with and without wastewater), the results suggest that wastewater measurements have the advantage to show the trends of an outbreak earlier. This observation is even more striking since the appearance of variant BA.2, where the hospitalization curve does not show the typical trends as before (cf. Figs. 6 and 7).

In Fig. 13, we illustrate two possible predictions on 20 distinct days simulated with (red) and without wastewater (blue). In each subplot of Fig. 13, the solid vertical red line denotes $k_0$, the actual starting day, which is also given in the title of the subplot. Each dashed line represents a reconstruction up to $k_0$, followed by a simulated prediction

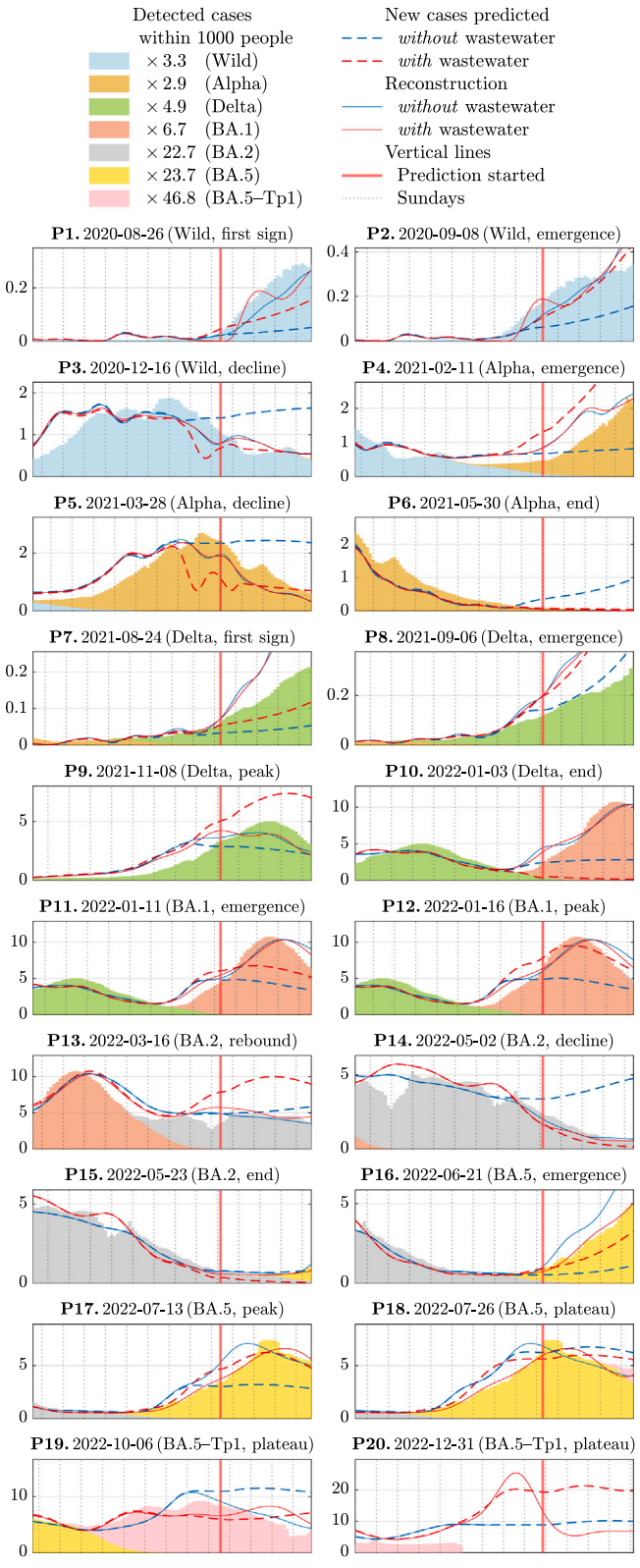



**Fig. 13.** Short-term (30-day long) prediction of new cases from different starting dates. The multipliers in the label are the estimated under-detection rates computed separately for each epidemic outbreak. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from $k_0$. In each subplot, the results of a *retrospective reconstruction* are also illustrated by solid lines in both cases: using both medical and wastewater data (red) and using only medical data (blue). The colored areas in Fig. 13 illustrate the detected cases scaled to fit the new cases computed retrospectively. Practically, the colored areas in each subplot illustrating the scaled detected cases resemble the corresponding segments of the scaled detected cases in Fig. 5. Finally, in Fig. 13, the obtained quantities were later normalized to a population of 1000 to facilitate convenient visualization of the $y$-axes.

In general, we observed that the characteristic phases of an outbreak appear earlier in the wastewater data than in hospitalization load. In Fig. 13, we illustrated a few predictions in different phases of the outbreaks, namely: first signs of an outbreak (P1, P7), emergence (P2, P4, P8, P11, P16), peaks (P9, P12, P17), decline (P3, P5,P14), end of an outbreak (P6, P10, P15), rebounds (P13), and plateaus (P18, P19, P20).

On 26 August 2020, wastewater measurements resulted in a prediction (P1), which could be obtained only 13 days later (P2) using solely hospitalization data. Then, on 16 December 2020, wastewater measurements already forecast the near end of the 1st wave (P3).

From measurements, the emergence of Alpha variant was predicted on 11 February 2021 (P4), whereas the increasing hospitalization data warned us only by the end of February. Then, we predicted the end of the outbreak on 28 March 2021 (P5), whereas the peak of the hospital occupancy was detected only 8 days later.

The first sign of the third wave caused by Delta variant was detected on 24 August 2021 (P7) using measurements and 2 days later using only hospitalization data. However, the first accurate prediction for the exponential emergence detected on 6 September 2021 (P8) was obtained only 10 days later using only hospitalization curves.

Without wastewater measurements, outbreak peaks could only be detected retrospectively in this case study, as hospital treatment data are delayed by $\tau_{\mathrm{L}}^{-1} + \tau_{\mathrm{P}}^{-1} + \tau_{\mathrm{I}}^{-1} + \tau_{\mathrm{A}}^{-1} \in [8, 14]$ [days] compared to the number of new cases. From wastewater, we detected peaks for about 10 days earlier for waves caused by variants Delta (P9), Omicron BA.1 (P12), BA.2 (P13), and BA.5 (P17).

To measure the quality of a prediction, we considered the Euclidean distance between the sequences obtained for the new cases through the current prediction and the retrospective reconstruction within the prediction horizon ($N = 30$ [days]). In this case study, underestimating forecasts are penalized against those that overestimated the severity of the outbreak. Correspondingly, we consider the distances in the logarithmic scale ($\log_{10}$). Formally, a prediction made on day $k$ is qualified by the *square root* of the following measure:

$$\sum_{i=1}^{N} \Big( \log_{10}(y_{k+i} \text{ predicted}) - \log_{10}(y_{k+i} \text{ reconstructed}) \Big)^2. \tag{7}$$

Additionally, a prediction made *without* wastewater data is compared to the reconstruction made *without* the knowledge of wastewater data.

In terms of (7), the maximum prediction error obtained without wastewater is 25% higher compared to that obtained using wastewater data. The two worst predictions are illustrated in P6 and P10 of Fig. 13. On 30 May 2021, the medical data suggested a rebound of the past wave caused by the Alpha variant (P6). However, a new wave emerged only three months later, by early September 2021 (P7). On the other hand, the virus secretion coefficient estimated for the Omicron variants reduced to a third compared to that of Delta (Row 4 in Table 2). Therefore, the emergence of the Omicron variant could not be detected yet on 3 January 2022 (P10) but only 8 days later (P11).

**Remark 5.** We note that, in 53% of days, wastewater-based predictions resulted in a higher prediction error in the terms of (7). It is also worth mentioning that the Euclidean distance of the logarithm is not the best quality measure as it does not qualify the difference between the shapes of two curves.

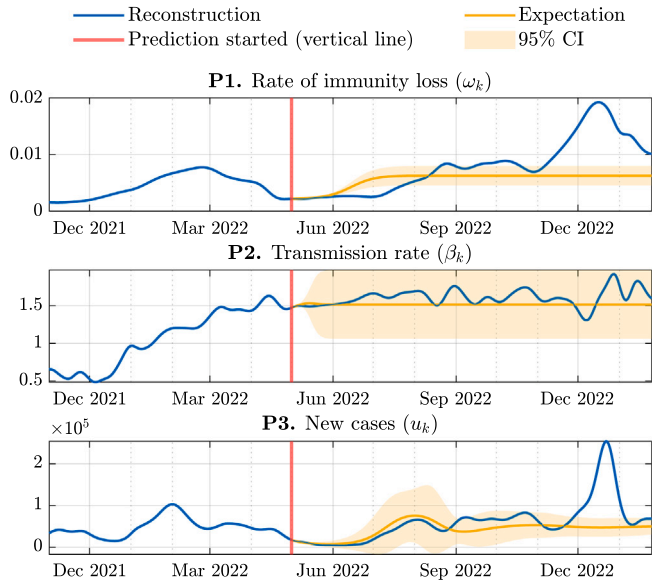



**Fig. 14.** Retrospective reconstruction (blue) and possible future scenarios (yellow) illustrated by their mean and the 95% credibility interval (CI) for the rate of immunity loss (P1), transmission rate (P2), and new cases (P3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.2.2. Long-term prediction

Considering a starting time, 1 May 2022 as the time label of the last available data, we simulate the epidemic process model (2) to estimate the future evolution of the pandemic. By that time, the fourth wave caused by the Omicron BA.1 and BA.2 variants ended, but a new immune evasive variant (BA.5) was expected to emerge. In that situation, we made the following assumptions:

1. The immune evasive capability of BA.5 is similar to that estimated for BA.1/BA.2, i.e., $\omega = 0.0062$ in average between 1 January and 15 April 2022. Therefore, the future immunity loss rate is estimated to be $\omega = 0.0062 \pm 25\%$ (P1 in Fig. 14).
2. The estimated transmission rate is extrapolated with a possible $\pm 30\%$ deviation from its last computed value (P2 in Fig. 14).
3. As for the fixed (Table 1) and computed model constants (Table 2), we used the available values, namely, those available or estimated for variant BA.2.
4. We assume no significant vaccination activity in the near future.

To make a statistical prediction, we simulated 126 scenarios using different $\beta$ and $\omega$ values. Plot P3 in Fig. 14 illustrates the average and the standard deviation of the predicted curves for the estimated new cases. In all figures, we illustrate the actual curves obtained later through a retrospective reconstruction.

## 4. Conclusions

As the number and relevance of clinical tests for COVID-19 are declining, WBE is becoming the primary source of information on disease transmission in a community. Recognizing this, both WHO and the European Union considers WBE a promising public health tool, even beyond the current pandemic.

Motivated by these international trends, we proposed a novel approach combining wastewater-based epidemic surveillance with detailed dynamical modeling to improve epidemic monitoring and prediction. We developed a novel compartmental epidemic model with a 2-phase vaccination dynamics and waning immunity. Immune evasion and waning are described by a single compartmental transition from the recovered people to the susceptibles but with a time-dependent transition rate.

Using time-dependent model parameters (relevant to the dominant virus variants), the model allows for us to reconstruct the unknown epidemiological data during the whole time span of the pandemic using the registered number of hospitalized patients and the measured genome copy concentration in the wastewater. Alongside the popular epidemic indicators (incidence, prevalence, reproduction number, transmission rate), a full epidemic state reconstruction is possible in each time instant, including the susceptible and protected population in the different vaccination status and the infected population in the different phases of the disease.

It is shown that wastewater-based viral information can be efficiently used to track the transmission rate of the virus, the rate of immunization by vaccination, and the rate of losing immunity within a time frame as time-varying parameters. In the computations, we exploit the registered stringency index and the relative infectiousness of the dominant virus variant, which provides a starting point in estimating the transmission rate. The estimated immunization rate allows us to quantify the effectiveness of vaccination by comparing the estimated new immunizations with the received vaccine doses within a given time interval.

The detailed reconstruction provides a reliable estimate of the current epidemic situation, which allows us to give a prediction with an extrapolated transmission rate and waning characteristics.

To promote computational tractability, the long-term epidemic surveillance framework (reconstruction and prediction) was implemented by multiple efficient optimization steps.

The effectiveness of the approach is illustrated through a case study of pandemic reconstruction and prediction in Hungary. The results suggest that wastewater measurements make predictions more reliable as the characteristic moments of an outbreak appear about a week earlier in wastewater data. The computed model-based predictions were shared with the Hungarian National Public Health Center to support decision-making on public policies in Hungary.

In this work, we successfully demonstrated that wastewater-based epidemic modeling substantially contributes to the translation of environmental surveillance data to outbreak management decisions.

For convenience, a MATLAB implementation of the proposed approach is made available online in a public repository (Polcz, 2021).

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data

## Acknowledgments

The study was funded by the National Research, Development, and Innovation Office in Hungary (RRF-2.3.1-21-2022-00006). Moreover, project no. TKP2021-NKTA-66 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. The authors acknowledge the support of the Hungarian Academy of Sciences through the Grant POST-COVID2021-64.

## Appendix A. Dynamical model representation

In this section, we provide a few technical details and further remarks on the epidemic process model introduced in Section 2.2.

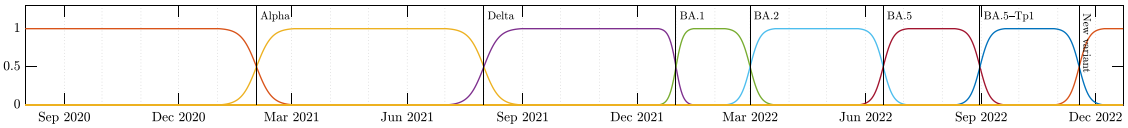



**Fig. A.15.** Variant dominance pattern.

### A.1. Model decomposition

The special model structure of the epidemic process model allows to reformulate it as an interconnection of a linear parameter-varying model and a scalar nonlinear model. First, we consider a dynamical model for the course of the disease of the infected population as a subsystem of (2) as follows:

$$x_{k+1} = F(\theta_k)\, x_k + B\, u_k, \quad y_k = C\, x_k, \tag{A.1}$$

where $x = \mathrm{vec}(\mathbf{L}, \mathbf{P}, \mathbf{I}, \mathbf{A}, \mathbf{H}, \mathbf{D}) \in \mathbb{R}^{n_x}$ is the state, $\theta = \mathrm{vec}\left(\tau_L, \tau_P, \tau_I, \tau_A, \tau_H, p_I, p_H, p_D\right) \in \mathbb{R}^{n_p}$ is the parameter, the daily new cases $u$ appears as an input, the hospitalized $\mathbf{H}$ and deceased $\mathbf{D}$ patients are measured states, the number of infectious people $y$ is an indirectly measured output of the system, furthermore, the coefficient matrices are given below:

$$F(\theta) = \begin{pmatrix} -\tau_L & 0 & 0 & 0 & 0 & 0 \\ \tau_L & -\tau_P & 0 & 0 & 0 & 0 \\ 0 & p_I\,\tau_P & -\tau_I & 0 & 0 & 0 \\ 0 & (1-p_I)\,\tau_P & 0 & -\tau_A & 0 & 0 \\ 0 & 0 & p_H\,\tau_I & 0 & -\tau_H & 0 \\ 0 & 0 & 0 & 0 & p_D\,\tau_H & 0 \end{pmatrix},$$

$$B = \mathrm{vec}\,(1,0,0,0,0,0)\,, \qquad C = \mathrm{row}\,\left(0,1,1,q_A,0,0\right). \tag{A.2}$$

System (A.1) with matrices (A.2) constitutes the *disease course model* or simply the *disease model*, which is highlighted in the compartmental model (2).

As the overall population $\mathbf{N}$ was assumed invariant, the number of immune people can be expressed as $\mathbf{R} = \mathbf{N} - \mathbf{K} - \mathbf{S}$,

where $\mathbf{K} = \mathbf{L} + \mathbf{P} + \mathbf{I} + \mathbf{A} + \mathbf{H} + \mathbf{D}$ denotes the sum of infected and deceased people. Accordingly, the time evolution of the susceptible individuals can be described by a single recursion as follows:

$$\mathbf{S}_{k+1} = (1 - \nu_k - \omega_k)\,\mathbf{S}_k + \omega_k\left(\mathbf{N} - \mathbf{K}_k\right) - u_k. \tag{A.3}$$

System (A.1) and (A.3) are interconnected through the feedback $u_k = \beta_k y_k \mathbf{S}_k/\mathbf{N}$ (4), where $\beta$ is unknown.

**Remark 6.** System (A.3) is manipulated indirectly by $\beta$ and directly both by the rate of immunity loss $\omega$ and the immunization rate $\nu$. The two rate coefficients $\omega$ and $\nu$ can be considered as a dual pair of each other, however, their roles and nature are different. Unlike $\nu$, which is well-defined and can be manipulated by vaccination strategies, the rate of immunity loss $\omega$ is unknown and (in an ideal world) cannot be manipulated by humans but is primarily due to mutation of the virus. In this way, the transition dynamics between $\mathbf{S}$ and $\mathbf{R}$ can also be considered as a differential game (Fisac et al., 2019), in which the two opponents (humans and virus) are struggling to win over as many people as possible in compartments $\mathbf{R}$ and $\mathbf{S}$, respectively. Coefficients $\nu$ and $\omega$ are manipulated exclusively by both humans and virus, respectively. Whereas, the transmission rate $\beta$ of the pathogen can be manipulated both by social distancing strategies and the increasing infectiousness of the virus.

### A.2. Variant-dependent model constants and their smooth transitions

The model coefficients $\tau_\bullet$, $p_\bullet$, and $q_c$ are variant-dependent such that their values are smoothly changing as a new VoC gains dominance. To characterize the piece-wise constant parameter functions, which vary smoothly between two constant levels, we introduce specific characteristic functions, that highlights the dominance period of a given VoC. When a new VoC emerges and overcomes another variant, the proportion of the new variant compared to the sum of both variants follows a sigmoid-like curve (Radu et al., 2022).

When a model parameter $\theta$ changes from value $\theta^{(0)}$ to $\theta^{(1)}$ such that the new VoC emerges at time $k_1 - K$, it takes dominance at time $k_1$, and the old variant vanishes at time $k_1 + K$, then, we consider the following smooth transition between the two parameter values:

$$\theta_k = \theta^{(0)} + (\theta^{(1)} - \theta^{(0)})\,\bar{\sigma}\Big(\frac{k-k_1}{K}\Big) \quad \text{for all } k \in \{k_1 - K, \dots, k_1, \dots, k_1 + K\}, \tag{A.4}$$

where $\bar{\sigma} : [-1, 1] \to [0, 1]$ constitutes the following normalized sigmoid-type function:

$$\bar{\sigma}(t) = \frac{\sigma(t) - \sigma(-1)}{\sigma(1) - \sigma(-1)}, \text{ and } \sigma(t) = (1 + e^{-s_0 t})^{-1}. \tag{A.5}$$

In (A.5), we use a scaling factor $s_0 = 5$, which provides a sufficiently small slope at the two ends of the parameter transition.

Assume that a given VoC emerges and takes dominance at time $k_1$ in a duration of $2K_1 + 1$ [days], then, another variant overcomes it in a duration of $2K_2$+1 around time $k_2$. Then, using function $\bar{\sigma}$, we introduce the characteristic function $\chi_{\text{VoC}}$ for that variant which highlights its dominance pattern, as follows:

$$\chi_{\text{VoC},k} = \begin{cases} 0 & \text{if } k < k_1 - K_1, \\ \bar{\sigma}\Big(\frac{k-k_1}{K_1}\Big) & \text{if } k_1 - K_1 \le k \le k_1 + K_1, \\ 1 & \text{if } k_1 + K_1 < k < k_2 - K_2, \\ 1 - \bar{\sigma}\Big(\frac{k-k_2}{K_2}\Big) & \text{if } k_2 - K_2 \le k \le k_2 + K_2, \\ 0 & \text{if } k_2 + K_2 < k, \end{cases}$$

where, as described in Section 2.2.3, the variants of concern, that emerged in Hungary are as follows:

$$\text{VoC} \in \{\text{Wild, Alpha, Delta, BA.1, BA.2, BA.5}\}. \tag{A.6}$$

We note that the sum of the characteristic functions gives identically one, namely, $\sum_{\text{VoC}} \chi_{\text{VoC}} \equiv 1$. In Fig. A.15, we illustrate the characteristic functions correspondingly to the variant dominance observed in Hungary.

If we fix the characteristic curves for all dominant variants in (A.6), we are able to estimate all model coefficients by piece-wise constant functions with smooth transitions as follows:

$$p_{\bullet,k} = \sum_{\text{VoC}} p_{\bullet,\text{VoC}}\,\chi_{\text{VoC},k}, \tag{A.7}$$

$$\tau_{\bullet,k} = \sum_{\text{VoC}} \tau_{\bullet,\text{VoC}}\,\chi_{\text{VoC},k}, \quad q_{c,k} = \sum_{\text{VoC}} q_{c,\text{VoC}}\,\chi_{\text{VoC},k},$$

where, according to Assumption 3, the virus secretion coefficient corresponding to variants BA.1 and BA.2 are identical, i.e., $q_{c,\text{BA.1}} = q_{c,\text{BA.2}}$.

**Remark 7** (*Under-detection rate*)**.** The characteristic curves allow the estimation of the number of all infections within an epidemic wave such that the series of new infections are multiplied by the characteristic curve of an epidemic wave and the resulting series are summarized over the time. E.g., the new cases during the spread of Delta variant can be computed as follows:

$$\sum_{k=0}^{T-1} \chi_{\text{Delta},k}\ \underbrace{[\textit{computed}\ \text{new cases on day}\ k]}_{u_k}. \tag{A.8}$$

The number of *detected cases* can be computed similarly:

$$\sum_{k=0}^{T-1} \chi_{\text{Delta},k} \underbrace{[\textit{detected} \text{ new cases on day } k]}_{\text{Denoted by } u_k^{\text{Off}}}. \tag{A.9}$$

Therefore, the under-detection rate during the wave caused by Delta can be computed as the proportion of the two quantities above, namely,

$$q_{\text{u,Delta}} = \frac{\sum_k \chi_{\text{Delta},k}\, u_k}{\sum_k \chi_{\text{Delta},k}\, u_k^{\text{Off}}} = 4.9. \tag{A.10}$$

Finally, the scaled series of the detected cases (denoted by $\bar{u}_k^{\text{Off}}$) are computed such that each segment of the series $u^{\text{Off}}$ characterized by $q_{\text{u,VoC}}$ is scaled by the corresponding multiplier $q_{\text{u,VoC}}$, then, the resulting series are accumulated for all outbreaks. Formally:

$$\bar{u}_k^{\text{Off}} = \sum_{\text{VoC}} q_{\text{u,VoC}}\ \chi_{\text{VoC},k} \ll u_k^{\text{Off}}. \tag{A.11}$$

Sequence $\bar{u}_k^{\text{Off}}$ is illustrated in Fig. 5 by the colored area.

## Appendix B. Dynamic inversion to compute infected people, undetected cases, and unknown model coefficients

In this section, we formalize the optimization-based dynamic inversion approach introduced in Section 2.3.

The dynamic inversion of the linear disease course dynamics (A.1) constitutes in the computation of the unknown input $u_k$ (4) using the measured output $q_{\text{c},k} C_k^{\text{Off}}$ M1 and states $\mathbf{H}_k^{\text{Off}}$ and $\mathbf{D}_k^{\text{Off}}$. Practically, the *dynamic inversion* or the *output tracking problem* has three reference outputs M1, M2, M3 to track by manipulating one single variable C1, the number of new infections $u_k$. In this construction, the problem is over-determined, therefore, new degrees of freedom have to be introduced to the optimization task by relaxing a few model coefficients as follows:

1. We consider $q_{\text{c,VoC}}$ as free variables C6.
2. The probability levels of being hospitalized $p_{\text{H,VoC}}$, and of fatal outcome $p_{\text{D,VoC}}$ are searched as free variables within a given reasonable domain C5.

The cost function of the multi-objective optimization problem is constructed as follows:

$$J = \underbrace{\sum_{k=1}^{T} w_{\text{H}}^{\text{ref}} |\mathbf{H}_k - \mathbf{H}_k^{\text{Off}}|^2}_{\text{tracking error } (\mathbf{H}^{\text{Off}})} + \underbrace{\sum_{k=1}^{T} w_{\text{D}}^{\text{ref}} |\mathbf{D}_k - \mathbf{D}_k^{\text{Off}}|^2}_{\text{tracking error } (\mathbf{D}^{\text{Off}})} + \underbrace{\sum_{k=1}^{T} w_{\text{c}}^{\text{ref}} |y_k - q_{\text{c},k} C_k^{\text{Off}}|^2}_{\text{tracking error } (C^{\text{Off}})} + \underbrace{\sum_{k=0}^{T-2} w_{\text{u}}^{\text{s}} |u_{k+1} - u_k|^2}_{\text{smoothly varying new cases}}, \tag{B.1}$$

where $y_k$ is defined in the output Eq. (3), and $w_{\text{H}}^{\text{ref}}$, $w_{\text{D}}^{\text{ref}}$, $w_{\text{c}}^{\text{ref}}$, $w_{\text{u}}^{\text{s}}$ are weight constants, which scale the different cost terms into a common order of magnitude. The last term in (B.1) promotes a smooth solution for $u_k$, which does not make abrupt changes in time. Differently from variables $\mathbf{H}_k$, $\mathbf{D}_k$, $y_k$, which are initialized in $k = 0$, the input $u_k$ is unknown in $k = 0$ and its value is irrelevant in $k = T$. This explains the range of indices in the last summation of (B.1).

Finally, the optimization problem can be formulated as follows:

**Problem 1** (*Optimization-based dynamic inversion*)**.** Given the dynamical disease course model (A.1) with initial condition $x_0 = 0$, fixed model constants $q_{\text{A}}$ and $\tau_{\bullet,\text{VoC}}$, corresponding to each dominant VoC (A.6). There are given M1, M2, M3 as references to track with $k = 1, \dots, T$. We are looking for

1. $p_{\text{H,VoC}}$ and $p_{\text{D,VoC}} \in (0, 1)$, the unknown probability levels,
2. $q_{\text{c,VoC}}$, the unknown secretion coefficients,
3. $u_k \geq 0$, the daily new cases as a sequence of unknown inputs, $k = 0, \dots, T-1$,
4. $x_k \geq 0$, the number of infected people in the 5 different states of the disease, $k = 1, \dots, T$,

which solve the state recursion (A.1) and minimize the cost function (B.1).

**Remark 8.** The number of infectious people $y_k$ can also be added as a free instrumental variable ruled by the output Eq. (3). In this way, further degrees of freedom are introduced in Problem 1, which promotes finding a feasible solution to the nonlinear gradient-based optimization task.

After the optimization of Problem 1, we are able to compute other important epidemiological quantities, such as the number of infectious people (3), the new recoveries (5), as well as the reproduction number of the disease. These quantities can be inferred without knowing the nature of waning immunity and the effects of vaccination.

The time-dependent effective reproduction number of the disease can be given as follows:

$$R_{\text{c},k} = \beta_k \left( \frac{1}{\tau_{\text{P},k}} + \frac{p_{\text{I},k}}{\tau_{\text{I},k}} + \frac{q_{\text{A}}(1 - p_{\text{I},k})}{\tau_{\text{A},k}} \right) \frac{\mathbf{S}_k}{\mathbf{N}}. \tag{B.2}$$

In this formulation, $R_{\text{c},k}$ requires the knowledge of $\mathbf{S}_k$ and $\beta_k$, which are still unknown after the solution of Problem 1. However, the term $\beta_k \mathbf{S}_k$ can be expressed from $u_k = \beta_k y_k \mathbf{S}_k / \mathbf{N}$ (4). Then, an alternative formula for $R_{\text{c}}$ can be given as follows:

$$R_{\text{c},k} = \frac{u_k}{y_k} \left( \frac{1}{\tau_{\text{P},k}} + \frac{p_{\text{I},k}}{\tau_{\text{I},k}} + \frac{q_{\text{A}}(1 - p_{\text{I},k})}{\tau_{\text{A},k}} \right). \tag{B.3}$$

During the optimization, we used the following weight constants in the cost function: $w_{\text{H}}^{\text{ref}} = 1$, $w_{\text{D}}^{\text{ref}} = w_{\text{c}}^{\text{ref}} = 10^{-5}$, $w_{\text{u}}^{\text{s}} = 10^{-4}$. The values of the parameters are available in Table 1.

## Appendix C. Immunization model

Based on the descriptions of Section 2.4.2, here we derive a dynamical model describing immunity waning, infection, and vaccination in three different vaccination statuses. Moreover, we propose a simple model for the dynamic effects of the delayed immunization through vaccination.

In this section, the cumulative number of first, second, and booster doses delivered are denoted by $\mathbf{V}_{\text{first}}^{\text{Off}}$, $\mathbf{V}_{\text{second}}^{\text{Off}}$, $\mathbf{V}_{\text{boosted}}^{\text{Off}}$, respectively (see the equation in Box II).

### C.1. Delayed effect of vaccination

A susceptible individual does not acquire immunity immediately after vaccination, but a certain amount of time is needed for the vaccine to take effect. In this section, we present the points 2, 3, and 4 of Assumption 6 in more detail.

**Assumption 9.** The delayed immunization is modeled as follows:

1. The probability that a partially vaccinated individual will be fully vaccinated can be well approximated as follows:

$$p_{\text{v}_2|\text{v}_1} = \mathbf{V}_{\text{second},\infty}^{\text{Off}} \left( \mathbf{V}_{\text{first},\infty}^{\text{Off}} \right)^{-1} \approx 0.967, \tag{C.2}$$

   where $\mathbf{V}_{\text{first},\infty}^{\text{Off}}$ and $\mathbf{V}_{\text{second},\infty}^{\text{Off}}$ denote the last available vaccination data.
2. The time elapsed between the first two doses is normally distributed with expected value $t_{12} = 21$ [days] and standard deviation $\sigma_{12} = 7$ [days].

$$
\begin{cases}
\mathbf{S}_{\mathrm{u},k+1} = \mathbf{S}_{\mathrm{u},k} - u_k\,\mathbf{S}_{\mathrm{u},k}/\mathbf{S}_k + \omega_k\,\mathbf{R}_{\mathrm{u},k} - \nu_{\mathrm{F},k}\,\mathbf{S}_{\mathrm{u},k}, & \text{(C.1a)}\\
\mathbf{S}_{\mathrm{F},k+1} = \mathbf{S}_{\mathrm{F},k} - u_k\,\mathbf{S}_{\mathrm{F},k}/\mathbf{S}_k + \omega_k\,\mathbf{R}_{\mathrm{F},k} - \nu_{\mathrm{B},k}\,\mathbf{S}_{\mathrm{F},k}, & \text{(C.1b)}\\
\mathbf{S}_{\mathrm{B},k+1} = \mathbf{S}_{\mathrm{B},k} - u_k\,\mathbf{S}_{\mathrm{B},k}/\mathbf{S}_k + \omega_k\,\mathbf{R}_{\mathrm{B},k} - \nu_{\mathrm{B},k}\,\mathbf{S}_{\mathrm{B},k}, & \text{(C.1c)}\\
\mathbf{R}_{\mathrm{u},k+1} = \mathbf{R}_{\mathrm{u},k} + z_k\,\mathbf{S}_{\mathrm{u},k-t_{0,k}}/\mathbf{S}_{k-t_{0,k}} - \omega_k\,\mathbf{R}_{\mathrm{u},k} - \nu_{\mathrm{F},k}\,\mathbf{R}_{\mathrm{u},k}, & \text{(C.1d)}\\
\mathbf{R}_{\mathrm{F},k+1} = \mathbf{R}_{\mathrm{F},k} + z_k\,\mathbf{S}_{\mathrm{F},k-t_{0,k}}/\mathbf{S}_{k-t_{0,k}} - \omega_k\,\mathbf{R}_{\mathrm{F},k} + \nu_{\mathrm{F},k}\,(\mathbf{S}_{\mathrm{u},k} + \mathbf{R}_{\mathrm{u},k}) - \nu_{\mathrm{B},k}\,\mathbf{R}_{\mathrm{F},k}, & \text{(C.1e)}\\
\mathbf{R}_{\mathrm{B},k+1} = \mathbf{R}_{\mathrm{B},k} + z_k\,\mathbf{S}_{\mathrm{B},k-t_{0,k}}/\mathbf{S}_{k-t_{0,k}} - \omega_k\,\mathbf{R}_{\mathrm{B},k} + \nu_{\mathrm{B},k}\,(\mathbf{S}_{\mathrm{F},k} + \mathbf{R}_{\mathrm{F},k} + \mathbf{S}_{\mathrm{B},k}), & \text{(C.1f)}
\end{cases}
$$

**Box II.**

3. After receiving a the full vaccine or a booster dose, a person becomes temporarily immune after a certain period of time, which is normally distributed with expected value $t_{\mathrm{e}} = 14$ [days] and standard deviation $\sigma_{\mathrm{e}} = 4$ [days].

In accordance with Assumption 9, an individual gains immunity after the first vaccine dose with probability $p_{\mathrm{v}_2|\mathrm{v}_1}$ but only after a normally distributed delay with expected value $t_{12\mathrm{e}} = t_{12} + t_{\mathrm{e}} = 51$ [days] and standard deviation $\sigma_{12\mathrm{e}} = \sqrt{\sigma_{\mathrm{e}}^2 + \sigma_{12}^2} \approx 8$ [days].

Then, the expected fully vaccinated population on day $k$ can be approximated as follows:

$$
\mathbf{V}_{\mathrm{F},k} \approx p_{\mathrm{v}_2|\mathrm{v}_1} \sum_{i=-2\sigma_{12\mathrm{e}}}^{2\sigma_{12\mathrm{e}}} \mathbf{V}^{\mathrm{Off}}_{\mathrm{first},k-t_{12\mathrm{e}}+i} \cdot q_i(\sigma_{12\mathrm{e}}). \tag{C.3}
$$

whereas, the boosted population is:

$$
\mathbf{V}_{\mathrm{B},k} \approx \sum_{i=-2\sigma_{\mathrm{e}}}^{2\sigma_{\mathrm{e}}} \mathbf{V}^{\mathrm{Off}}_{\mathrm{boost},k-t_{\mathrm{e}}+i} \cdot q_i(\sigma_{\mathrm{e}}). \tag{C.4}
$$

In (C.3) and (C.4), the weight coefficients $q_i(\cdot)$ are derived from the cumulative distribution function $\Phi$ of the standard normal distribution:

$$
q_i(\sigma) = \left(\Phi\left(\tfrac{2i+1}{2\sigma}\right) - \Phi\left(\tfrac{2i-1}{2\sigma}\right)\right) \cdot \left(2\Phi\left(\tfrac{4\sigma+1}{2\sigma}\right) - 1\right)^{-1}. \tag{C.5}
$$

**Remark 9.** Note that the number of vaccinated people in (C.3) and (C.4) are practically computed by a convolution of the official data with a Gaussian window. Therefore, the normally distributed delay of the effect of the first and booster doses has a beneficial smoothing effect on the official data.

### C.2. Distribution of new cases and recoveries

In accordance with Assumption 7, the distribution of new cases by vaccination status on day $k$ is determined by the distribution $(\mathbf{S}_{\mathrm{u},k}, \mathbf{S}_{\mathrm{F},k}, \mathbf{S}_{\mathrm{B},k})$ of the susceptible population at that time, namely,

$$
u_k = u_{\mathrm{u},k} + u_{\mathrm{F},k} + u_{\mathrm{B},k}, \text{ such that} \tag{C.6}
$$
$$
(u_{\mathrm{u},k}, u_{\mathrm{F},k}, u_{\mathrm{B},k}) = \left(u_k \frac{\mathbf{S}_{\mathrm{u},k}}{\mathbf{S}_k}, u_k \frac{\mathbf{S}_{\mathrm{F},k}}{\mathbf{S}_k}, u_k \frac{\mathbf{S}_{\mathrm{B},k}}{\mathbf{S}_k}\right).
$$

Moreover, the distribution of new recoveries by vaccination status on day $k$ is approximated by the distribution of the susceptible population on day $k - t_{0,k}$, namely,

$$
z_k = z_{\mathrm{u},k} + z_{\mathrm{F},k} + z_{\mathrm{B},k}, \text{ such that} \tag{C.7}
$$
$$
(z_{\mathrm{u},k}, z_{\mathrm{F},k}, z_{\mathrm{B},k}) = \left(z_k \frac{\mathbf{S}_{\mathrm{u},k-t_0}}{\mathbf{S}_{k-t_0}}, z_k \frac{\mathbf{S}_{\mathrm{F},k-t_0}}{\mathbf{S}_{k-t_0}}, u_k \frac{\mathbf{S}_{\mathrm{B},k-t_0}}{\mathbf{S}_{k-t_0}}\right),
$$

where $t_{0,k} = \left\lfloor \frac{1}{\tau_{\mathrm{L},k}} + \frac{1}{\tau_{\mathrm{P},k}} + \frac{p_{\mathrm{I},k}}{\tau_{\mathrm{I},k}} + \frac{1-p_{\mathrm{I},k}}{\tau_{\mathrm{A},k}} + \frac{p_{\mathrm{H},k}}{\tau_{\mathrm{H},k}} \right\rfloor$ constitutes the expected duration of recovery.

### C.3. State recursion with vaccination and waning model

Finally, a state recursion for the dynamics of the susceptible and protected population is given in the floating Eq. (C.1). where $\nu_{\mathrm{F},k}$ and $\nu_{\mathrm{B},k}$ denote. the rates of immunization triggered by the full vaccinations and booster doses.

The immunization rates can be expressed as follows:

$$
\nu_{\mathrm{F},k} = \frac{\mathbf{V}_{\mathrm{F},k} - \mathbf{V}_{\mathrm{F},k-1}}{\mathbf{S}_{\mathrm{u},k} + \mathbf{R}_{\mathrm{u},k}}, \quad \nu_{\mathrm{B},k} = \frac{\mathbf{V}_{\mathrm{B},k} - \mathbf{V}_{\mathrm{B},k-1}}{\mathbf{S}_{\mathrm{F},k} + \mathbf{R}_{\mathrm{F},k} + \mathbf{S}_{\mathrm{B},k}}. \tag{C.8}
$$

Note that the two rates in (C.8) corresponding to the received vaccine doses are a fine-tuning of the single immunization rate $\nu_k$ in the 8-compartmental model (2). To retain the overall rate of immunization of susceptibles through vaccination (in general) the value of $\nu_k$ can be expressed as follows:

$$
\nu_k = \left(\nu_{\mathrm{F},k}\,\mathbf{S}_{\mathrm{u},k} + \nu_{\mathrm{B},k}(\mathbf{S}_{\mathrm{F},k} + \mathbf{S}_{\mathrm{B},k})\right)/\mathbf{S}_k. \tag{C.9}
$$

**Remark 10.** Assuming no waning immunity ($\omega \equiv 0 \Rightarrow \mathbf{S}_{\mathrm{F}} \equiv 0$ and $\mathbf{S}_{\mathrm{B}} \equiv 0$), the overall rate of immunization is

$$
\nu_k = \nu_{\mathrm{F},k} = \frac{\mathbf{V}_{\mathrm{F},k} - \mathbf{V}_{\mathrm{F},k-1}}{\mathbf{V}_{\mathrm{F},k} - \mathbf{K}_k}. \tag{C.10}
$$

The subgroups of the susceptible and protected people and the transitions between the subgroups are presented in Fig. 4.

## Appendix D. Heuristics to estimate the rate of transmission of the pathogen and the waning immunity

Here, we describe the iterative approach to estimate the three time-varying rate coefficients of the epidemic process. The basic idea of the heuristics was introduced briefly in Section 2.4.3. Namely, the rates of immunization and immunity loss are computed step-by-step, one after another, by fixing one and computing the other. And then vice versa.

### D.1. Fix $\omega$, compute $\nu$

The recursion model (C.1) makes it possible to simulate the time evolution of susceptibles in the three different vaccination statuses for a given rate of immunity loss. This algebraic problem can be formulated as follows:

**Problem 2** (*Estimate the rate of immunization through vaccination*)**.** Consider the following pre-computed quantities:

1. $u_k$, the pre-computed number of new cases,
2. $z_k$, the pre-computed number of new recoveries,
3. $\omega_k$, an estimated rate of immunity loss (first initialized by $\omega \equiv 0$),
4. $\mathbf{V}_{\mathrm{F},k}$ and $\mathbf{V}_{\mathrm{B},k}$, pre-processed official vaccination data,

where $k = 0, \dots, T$. We are looking for a solution to recursion (C.1), with initial condition $\mathbf{S}_{u,0} = \mathbf{N}$.

In Problem 2, the new cases ($u_k$) and new recoveries ($z_k$) were pre-computed through the dynamic inversion described in Problem 1. Whereas, the rate of immunity loss $\omega_k$ is first initialized by zero, i.e., no waning immunity is assumed. Then, a solution to the recursion (C.1) allows computing the overall rate of immunization $\nu_k$ as described in (C.9). This value to $\nu_k$ allows us to give an estimate for the rate of immunity loss $\omega_k$ as described in the following subsection.

### D.2. Fix $\nu$, compute $\omega$

A pre-computed immunization rate allows us to formulate a simple optimization problem to estimate both the immunity loss and the transmission rates.

The stringency index D2 and the approximated infectiousness $\beta_{\text{VoC}}$ D3 of each VoC allows us to determine a prior estimate for the actual transmission rate of the virus as follows:

$$\beta_k^{\text{ref}} = \sum_{\text{VoC}} \chi_{\text{VoC},k} \cdot \beta_{\text{VoC}} \cdot I_{s,k} \tag{D.1}$$

We presume that $\beta_k^{\text{ref}}$ is a good estimate to use as a flexible reference for the optimized smooth transmission rate $\beta_k$.

During the optimization, the immunity loss rate $\omega_k$ is searched as a smooth time-varying sequence (not producing abrupt changes), such that its value is included into the objective function with a given weight. In this way, the unknown coefficients are kept between reasonable bounds, and the transmission rate sequence is regularized by the measured stringency index and the relative infectiousness of the emerging variants. With these considerations, we finally define the following cost function to minimize:

$$\begin{aligned} J_2 = & \underbrace{\sum_{k=0}^{T-1} w_{\text{u}}^{\text{ref}} |u_k - \hat{u}_k|^2}_{\text{tracking error (new cases)}} + \underbrace{\sum_{k=0}^{T-1} w_{\text{b}}^{\text{ref}} |\beta_k^{\text{ref}} - \beta_k|^2}_{\text{tracking error } (\beta)} \\ & + \underbrace{\sum_{k=0}^{T-2} w_{\text{b}}^{\text{sm}} |\beta_{k+1} - \beta_k|^2}_{\text{smooth transmission rate}} + \underbrace{\sum_{k=0}^{T-2} w_{\text{w}}^{\text{sm}} |\omega_{k+1} - \omega_k|^2}_{\text{smooth immunity loss rate}} \\ & + \underbrace{\sum_{k=0}^{T-1} w_{\omega}^{c} |\omega_k|^2}_{\text{immunity loss cost}} \end{aligned} \tag{D.2}$$

Note that variables $\hat{u}_k$, $\beta_k$, $\omega_k$ are unknown in $k = 0, \dots, T$, however, their values in $k = T$ are not needed as they are not required to compute the state in $k = T$. Similarly, the distance between $\beta_T$ and $\beta_{T-1}$ or between $\omega_T$ and $\omega_{T-1}$ are not relevant. These considerations explain the ranges of summations in (D.2).

The optimization problem to find $\beta_k$, $\omega_k$, and an updated number of susceptibles $\mathbf{S}_k$ is formulated as follows:

**Problem 3** (*Estimate the susceptible population and the rate of immunity loss*)**.** Consider the following pre-computed quantities:

1. $u_k$, the pre-computed number of new cases within 24 h,
2. $y_k$, the number of infectious people,
3. $\mathbf{K}_k$, the number of infected and deceased people,
4. $\beta_k^{\text{ref}}$, a reference transmission rate of the pathogen inferred from the stringency index,
5. $\nu_k$, an estimated rate of immunization triggered by vaccination,

where $k = 0, \dots, T$. We are looking for

1. $\mathbf{S}_k \in [0, \mathbf{N}]$, the number of susceptible individuals,
2. $\omega_k \geq 0$, the rate of immunity loss,
3. $\beta_k \in [\underline{\beta}, \overline{\beta}]$, the actual rate of transmission of the pathogen,
4. $\hat{u}_k$, an instrumental variable reconstructing the number of new cases,

which solve $\mathbf{S}_{k+1} = (1 - \nu_k - \omega_k)\mathbf{S}_k + \omega_k(\mathbf{N} - \mathbf{K}_k) - \hat{u}_k$ (A.3), satisfy $\hat{u}_k = \beta_k y_k \mathbf{S}_k / \mathbf{N}$ (4), and minimize (D.2).

**Remark 11.** The instrumental variable $\hat{u}_k$ introduces a further degree of freedom in Problem 3, which makes the nonlinear gradient-based optimization task computationally more tractable, namely, it promotes finding an initial feasible solution.

### D.3. Alternating iterations

Observe that Problem 3 makes use of the immunization rate $\nu_k$ but provides an estimated immunity loss rate $\omega_k$. Whereas, in Problem 2, we solve the recursion (C.1) to find the immunization rate for a given sequence of immunity loss rate. Therefore, Problems 2 and 3 can be solved multiple times one after another, such that

(A ) Problem 2 is solved first with no waning assumed $\omega \equiv 0$,

or

(B) Problem 3 is solved first with a simple vaccination model (C.10).

Scenario (A) and (B) are equivalent as the solution of Problem 2 with $\omega \equiv 0$ generates $\nu$ as given in (C.10). We note that the proposed iteration of the two problems in succession converged in all our experiments.

**Remark 12.** The alternating iterations considered here were inspired by the iterative solution of Cisneros and Werner (2020) for nonlinear model predictive control problems. The main similarity between the approaches is the existence of a *precomputed parameter trajectory*, in our case the immunization rate $\nu$, which can be initialized (e.g., by 0) and updated by a closed-form recursion (C.1) for any feasible values of the unknown variables, the immunity loss rate $\omega$. Moreover, the unknown variables are computed by an optimization (Problem 3), which is parameterized by the parameter trajectory. A significant difference is the type of the optimization problem: Cisneros and Werner (2020) presumes a convex quadratic program, whereas, Problem 3 is bilinear. Therefore, the existing preliminary convergence results of Cisneros and Werner (2020), Hespe and Werner (2021), Morato et al. (2021) and Cisneros (2021) are not applicable here, and considering the difficulty of the problem, it will be a target of future work to give a proof of convergence for this iteration.

## Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.watres.2023.120098.

## References


Atlo Team, 2021a. Koronamonitor: Detailed diagrams of the coronavirus outbreak. https://atlo.team/koronamonitor-reszletesadatok (Accessed on 11/11/2021).

Atlo Team, 2021b. Koronamonitor: Hungarian status of coronavirus vaccination. https://atlo.team/vakcinacio (Accessed on 11/11/2021).

Been, F., Rossi, L., Ort, C., Rudaz, S., Delémont, O., Esseiva, P., 2014. Population normalization with ammonium in wastewater-based epidemiology: Application to illicit drug monitoring. Environ. Sci. Technol. 48 (14), 8162–8169. http://dx.doi.org/10.1021/es5008388.

Berestycki, H., Desjardins, B., Heintz, B., Oury, J.-M., 2021. Plateaus, rebounds and the effects of individual behaviours in epidemics. Sci. Rep. 11 (1), http://dx.doi.org/10.1038/s41598-021-97077-x.

Bibby, K., Bivins, A., Wu, Z., North, D., 2021. Making waves: Plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. Water Res. 202, 117438. http://dx.doi.org/10.1016/j.watres.2021.117438.

Cisneros, P.S.G., 2021. Quasi-linear model predictive control: Stability, modelling and implementation (Ph.D. thesis). Hamburg University of Technology, http://dx.doi.org/10.15480/882.3574.

Cisneros, P.S.G., Werner, H., 2020. Nonlinear model predictive control for models in quasi-linear parameter varying form. Internat. J. Robust Nonlinear Control 30 (10), 3945–3959. http://dx.doi.org/10.1002/rnc.4973.

Claro, I.C.M., Cabral, A.D., Augusto, M.R., Duran, A.F.A., Graciosa, M.C.P., Fonseca, F.L.A., Speranca, M.A., de Freitas Bueno, R., 2021. Long-term monitoring of SARS-COV-2 RNA in wastewater in Brazil: A more responsive and economical approach. Water Res. 203, 117534. http://dx.doi.org/10.1016/j.watres.2021.117534.

Cocchio, S., Zabeo, F., Facchin, G., Piva, N., Venturato, G., Marcon, T., Saia, M., Tonon, M., Mongillo, M., Re, F.D., Russo, F., Baldo, V., 2022. Differences in immunological evasion of the Delta (B.1.617.2) and Omicron (B.1.1.529) SARS-CoV-2 variants: A retrospective study on the veneto region's population. Int. J. Environ. Res. Public Health 19 (13), 8179. http://dx.doi.org/10.3390/ijerph19138179.

Csutak, B., Polcz, P., Szederkényi, G., 2021. Computation of COVID-19 epidemiological data in Hungary using dynamic model inversion. In: 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics. SACI, pp. 91–96. http://dx.doi.org/10.1109/SACI51354.2021.9465563.

Csutak, B., Polcz, P., Szederkényi, G., 2022. Model-based epidemic data reconstruction using feedback linearization. In: 2022 International Conference on Electrical, Computer and Energy Technologies. ICECET, pp. 1–6. http://dx.doi.org/10.1109/ICECET55527.2022.9873061.

Daza-Torres, M.L., Montesinos-López, J.C., Kim, M., Olson, R., Bess, C.W., Rueda, L., Susa, M., Tucker, L., García, Y.E., Schmidt, A.J., Naughton, C.C., Pollock, B.H., Shapiro, K., Nuño, M., Bischel, H.N., 2023. Model training periods impact estimation of COVID-19 incidence from wastewater viral loads. Sci. Total Environ. 858, 159680. http://dx.doi.org/10.1016/j.scitotenv.2022.159680.

Fazli, M., Sklar, S., Porter, M.D., French, B.A., Shakeri, H., 2021. Wastewater-based epidemiological modeling for continuous surveillance of COVID-19 outbreak. In: 2021 IEEE International Conference on Big Data (Big Data). IEEE, http://dx.doi.org/10.1109/bigdata52589.2021.9671543.

Fernandez-Cassi, X., Scheidegger, A., Bänziger, C., Cariti, F., Corzon, A.T., Ganesanandamoorthy, P., Lemaitre, J.C., Ort, C., Julian, T.R., Kohn, T., 2021. Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high. Water Res. 200, 117252. http://dx.doi.org/10.1016/j.watres.2021.117252.

Fisac, J.F., Akametalu, A.K., Zeilinger, M.N., Kaynama, S., Gillula, J., Tomlin, C.J., 2019. A general safety framework for learning-based control in uncertain robotic systems. IEEE Trans. Automat. Control 64 (7), 2737–2752. http://dx.doi.org/10.1109/TAC.2018.2876389.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H., 2021. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). Nat. Hum. Behav. 5 (4), 529–538. http://dx.doi.org/10.1038/s41562-021-01079-8.

Hespe, C., Werner, H., 2021. Convergence properties of fast quasi-LPV model predictive control. In: 2021 60th IEEE Conference on Decision and Control (CDC). IEEE, http://dx.doi.org/10.1109/cdc45484.2021.9683612.

Hewitt, J., Trowsdale, S., Armstrong, B.A., Chapman, J.R., Carter, K.M., Croucher, D.M., Trent, C.R., Sim, R.E., Gilpin, B.J., 2022. Sensitivity of wastewater-based epidemiology for detection of SARS-CoV-2 RNA in a low prevalence setting. Water Res. 211, 118032. http://dx.doi.org/10.1016/j.watres.2021.118032.

Hodcroft, E., 2023. CoVariants: SARS-CoV-2 mutations and variants of interest. URL https://covariants.org/per-country?country=Hungary.

Huizer, M., ter Laak, T.L., de Voogt, P., van Wezel, A.P., 2021. Wastewater-based epidemiology for illicit drugs: A critical review on global data. Water Res. 207, 117789. http://dx.doi.org/10.1016/j.watres.2021.117789.

Iwasaki, A., 2021. What reinfections mean for COVID-19. Lancet Infect. Dis. 21 (1), 3–5. http://dx.doi.org/10.1016/s1473-3099(20)30783-0.

Jiang, S., Maggard, K., Shakeri, H., Porter, M.D., 2021. An application of the partially observed Markov process in the analysis of transmission dynamics of COVID-19 via wastewater. In: 2021 Systems and Information Engineering Design Symposium (SIEDS). IEEE, http://dx.doi.org/10.1109/sieds52267.2021.9483793.

Jiang, G., Wu, J., Weidhaas, J., Li, X., Chen, Y., Mueller, J., Li, J., Kumar, M., Zhou, X., Arora, S., Haramoto, E., Sherchan, S., Orive, G., Lertxundi, U., Honda, R., Kitajima, M., Jackson, G., 2022. Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. Water Res. 218, 118451. http://dx.doi.org/10.1016/j.watres.2022.118451.

Kang, S.-W., Kim, J.Y., Park, H., Lim, S.Y., Kim, J., Chang, E., Bae, S., Jung, J., Kim, M.J., Chong, Y.P., Lee, S.-O., Choi, S.-H., Kim, Y.S., Park, M.-S., Kim, S.-H., 2022. Comparison of secondary attack rate and viable virus shedding between patients with SARS-CoV-2 Delta and Omicron variants: A prospective cohort study. J. Med. Virol. 95 (1), http://dx.doi.org/10.1002/jmv.28369.

Kemp, F., Proverbio, D., Aalto, A., Mombaerts, L., d'Hérouël, A.F., Husch, A., Ley, C., Gonçalves, J., Skupin, A., Magni, S., 2021. Modelling COVID-19 dynamics and potential for herd immunity by vaccination in Austria, Luxembourg and Sweden. J. Theoret. Biol. 530, 110874. http://dx.doi.org/10.1016/j.jtbi.2021.110874.

King County, 2022. COVID-19 outcomes by vaccination status. www.kingcounty.gov/covid/data/vaxoutcomes (Accessed on 14/01/2023).

Kozyreff, G., 2021. Hospitalization dynamics during the first COVID-19 pandemic wave: SIR modelling compared to Belgium, France, Italy, Sswitzerland and New York City data. Infect. Dis. Model. 6, 398–404. http://dx.doi.org/10.1016/j.idm.2021.01.006.

Krivoňáková, N., Šoltýsová, A., Tamáš, M., Takáč, Z., Krahulec, J., Ficek, A., Gál, M., Gall, M., Fehér, M., Krivjanská, A., Horáková, I., Belišová, N., Bímová, P., Škulcová, A.B., Mackuľak, T., 2021. Mathematical modeling based on RT-qPCR analysis of SARS-CoV-2 in wastewater as a tool for epidemiology. Sci. Rep. 11 (1), http://dx.doi.org/10.1038/s41598-021-98653-x.

Lee, W.L., Armas, F., Guarneri, F., Gu, X., Formenti, N., Wu, F., Chandra, F., Parisio, G., Chen, H., Xiao, A., Romeo, C., Scali, F., Tonni, M., Leifels, M., Chua, F.J.D., Kwok, G.W., Tay, J.Y., Pasquali, P., Thompson, J., Alborali, G.L., Alm, E.J., 2022. Rapid displacement of SARS-CoV-2 variant Delta by Omicron revealed by allele-specific PCR in wastewater. Water Res. 221, 118809. http://dx.doi.org/10.1016/j.watres.2022.118809.

Lee, W.L., Imakaev, M., Armas, F., McElroy, K.A., Gu, X., Duvallet, C., Chandra, F., Chen, H., Leifels, M., Mendola, S., Floyd-O'Sullivan, R., Powell, M.M., Wilson, S.T., Berge, K.L.J., Lim, C.Y.J., Wu, F., Xiao, A., Moniz, K., Ghaeli, N., Matus, M., Thompson, J., Alm, E.J., 2021. Quantitative SARS-CoV-2 Alpha variant B.1.1.7 tracking in wastewater by allele-specific RT-qPCR. Environ. Sci. Technol. Lett. 8 (8), 675–682. http://dx.doi.org/10.1021/acs.estlett.1c00375.

Lemaitre, J.C., Perez-Saez, J., Azman, A.S., Rinaldo, A., Fellay, J., 2020. Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland. Swiss Med. Wkly. http://dx.doi.org/10.4414/smw.2020.20295.

Lyngse, F.P., Kirkeby, C.T., Denwood, M., Christiansen, L.E., lbak, K.M., ller, C.H.M., Skov, R.L., Krause, T.G., Rasmussen, M., Sieber, R.N., Johannesen, T.B., Lillebaek, T., Fonager, J., Fomsgaard, A., ller, F.T.M., Stegger, M., Overvad, M., Spiess, K., Mortensen, L.H., 2022. Household transmission of SARS-CoV-2 Omicron variant of concern subvariants BA.1 and BA.2 in Denmark. Nature Commun. 13 (1), http://dx.doi.org/10.1038/s41467-022-33498-0.

Lyngse, F.P., Mortensen, L.H., Denwood, M.J., Christiansen, L.E., Mø ller, C.H., Skov, R.L., Spiess, K., Fomsgaard, A., Lassaunière, M.M., Rasmussen, M., Stegger, M., Nielsen, C., Sieber, R.N., Cohen, A.S., Mø ller, F.T., Overvad, M., Mø lbak, K., Krause, T.G., Kirkeby, C.T., 2021. SARS-CoV-2 Omicron VOC transmission in Danish households. http://dx.doi.org/10.1101/2021.12.27.21268278, MedRxiv.

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., Roser, M., 2020. Coronavirus pandemic (COVID-19). https://ourworldindata.org/coronavirus or https://ourworldindata.org/covid-stringency-index (Oxford stringency index). Accessed on 2022 November.

McEvoy, D., McAloon, C., Collins, A., Hunt, K., Butler, F., Byrne, A., Casey-Bryars, M., Barber, A., Griffin, J., Lane, E.A., Wall, P., More, S.J., 2021. Relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: A rapid scoping review. BMJ Open 11 (5), e042354. http://dx.doi.org/10.1136/bmjopen-2020-042354.

McMahan, C.S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J.A., Popat, S.C., Karanfil, T., Freedman, D.L., 2021. COVID-19 wastewater epidemiology: a model to estimate infected populations. Lancet Planetary Health 5 (12), e874–e881. http://dx.doi.org/10.1016/s2542-5196(21)00230-8.

Morato, M.M., Normey-Rico, J.E., Sename, O., 2021. On the robustness properties of gain-scheduled unconstrained MPC for LPV systems. In: Procedings Do XV SimpÓSio Brasileiro de AutomaÇãO Inteligente. SBA Sociedade Brasileira de Automática, http://dx.doi.org/10.20906/sbai.v1i1.2585.

Morvan, M., Jacomo, A.L., Souque, C., Wade, M.J., Hoffmann, T., Pouwels, K., Lilley, C., Singer, A.C., Porter, J., Evens, N.P., Walker, D.I., Bunce, J.T., Engeli, A., Grimsley, J., O'Reilly, K.M., Danon, L., 2022. An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. Nature Commun. 13 (1), http://dx.doi.org/10.1038/s41467-022-31753-y.

Nakajo, K., Nishiura, H., 2021. Transmissibility of asymptomatic COVID-19: Data from Japanese clusters. Int. J. Infect. Dis. 105, 236–238. http://dx.doi.org/10.1016/j.ijid.2021.02.065.

Nourbakhsh, S., Fazil, A., Li, M., Mangat, C.S., Peterson, S.W., Daigle, J., Langner, S., Shurgold, J., D'Aoust, P., Delatolla, R., Mercier, E., Pang, X., Lee, B.E., Stuart, R., Wijayasri, S., Champredon, D., 2022. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. Epidemics 39, 100560. http://dx.doi.org/10.1016/j.epidem.2022.100560.

Olesen, S.W., Imakaev, M., Duvallet, C., 2021. Making waves: Defining the lead time of wastewater-based epidemiology for COVID-19. Water Res. 202, 117433. http://dx.doi.org/10.1016/j.watres.2021.117433.

Pájaro, M., Fajar, N.M., Alonso, A.A., Otero-Muras, I., 2022. Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and wastewater data in small and medium-sized municipalities: A one year study. Chaos Solitons Fractals 164, 112671. http://dx.doi.org/10.1016/j.chaos.2022.112671.

Péni, T., Csutak, B., Bartha, F.A., Röst, G., Szederkényi, G., 2022. Optimizing symptom based testing strategies for pandemic mitigation. IEEE Access 1. http://dx.doi.org/10.1109/ACCESS.2022.3197587.

Péni, T., Csutak, B., Szederkényi, G., Röst, G., 2020. Nonlinear model predictive control with logic constraints for COVID-19 management. Nonlinear Dynam. 102 (4), 1965–1986. http://dx.doi.org/10.1007/s11071-020-05980-1.

Phan, T., Brozak, S., Pell, B., Gitter, A., Xiao, A., Mena, K.D., Kuang, Y., Wu, F., 2023. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. Sci. Total Environ. 857, 159326. http://dx.doi.org/10.1016/j.scitotenv.2022.159326.

Polcz, P., 2021. Wastewater-based modeling, reconstruction, and prediction for COVID-19 outbreaks in Hungary using optimization. GitHub Repos. https://github.com/ppolcz/WBE-monitoring-for-COVID-19.git.

Polcz, P., Csutak, B., Szederkényi, G., 2022. Reconstruction of epidemiological data in Hungary using stochastic model predictive control. Appl. Sci. 12 (3), 1113. http://dx.doi.org/10.3390/app12031113.

Proverbio, D., Kemp, F., Magni, S., Ogorzaly, L., Cauchie, H.-M., Gonçalves, J., Skupin, A., Aalto, A., 2022. Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis. Sci. Total Environ. 827, 154235. http://dx.doi.org/10.1016/j.scitotenv.2022.154235.

Radu, E., Masseron, A., Amman, F., Schedl, A., Agerer, B., Endler, L., Penz, T., Bock, C., Bergthaler, A., Vierheilig, J., Hufnagl, P., Korschineck, I., Krampe, J., Kreuzinger, N., 2022. Emergence of SARS-CoV-2 Alpha lineage and its correlation with quantitative wastewater-based epidemiology data. Water Res. 215, 118257. http://dx.doi.org/10.1016/j.watres.2022.118257.

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., Sánchez, G., 2020. SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. Water Res. 181, 115942. http://dx.doi.org/10.1016/j.watres.2020.115942.

Reguly, I.Z., Csercsik, D., Juhász, J., Tornai, K., Bujtár, Z., Horváth, G., Keömley-Horváth, B., Kós, T., Cserey, G., Iván, K., Pongor, S., Szederkényi, G., Röst, G., Csikász-Nagy, A., 2022. Microsimulation based quantitative analysis of COVID-19 management strategies. PLoS Comput. Biol. 18 (1), 1–14. http://dx.doi.org/10.1371/journal.pcbi.1009693.

Róka, E., Khayer, B., Kis, Z., Kovács, L.B., Schuler, E., Magyar, N., Málnási, T., Oravecz, O., Pályi, B., Pándics, T., Vargha, M., 2021. Ahead of the second wave: Early warning for COVID-19 by wastewater surveillance in Hungary. Sci. Total Environ. 786, 147398. http://dx.doi.org/10.1016/j.scitotenv.2021.147398.

Sheehan, M.M., Reddy, A.J., Rothberg, M.B., 2021. Reinfection rates among patients who previously tested positive for coronavirus disease 2019: A retrospective cohort study. Clin. Infect. Dis. 73 (10), 1882–1886. http://dx.doi.org/10.1093/cid/ciab234.

Stein, C., Nassereldine, H., Sorensen, R.J.D., Amlag, J.O., Bisignano, C., Byrne, S., Castro, E., Coberly, K., Collins, J.K., Dalos, J., Daoud, F., Deen, A., Gakidou, E., Giles, J.R., Hulland, E.N., Huntley, B.M., Kinzel, K.E., Lozano, R., Mokdad, A.H., Pham, T., Pigott, D.M., Jr., R.C.R., Vos, T., Hay, S.I., Murray, C.J.L., Lim, S.S., 2023. Past SARS-CoV-2 infection protection against re-infection: A systematic review and meta-analysis. Lancet 401 (10379), 833–842. http://dx.doi.org/10.1016/S0140-6736(22)02465-5.

Willyard, C., 2023. How quickly does COVID immunity fade? What scientists know. Nature http://dx.doi.org/10.1038/d41586-023-00124-y.

Xiao, A., Wu, F., Bushman, M., Zhang, J., Imakaev, M., Chai, P.R., Duvallet, C., Endo, N., Erickson, T.B., Armas, F., Arnold, B., Chen, H., Chandra, F., Ghaeli, N., Gu, X., Hanage, W.P., Lee, W.L., Matus, M., McElroy, K.A., Moniz, K., Rhode, S.F., Thompson, J., Alm, E.J., 2022. Metrics to relate COVID-19 wastewater data to clinical testing dynamics. Water Res. 212, 118070. http://dx.doi.org/10.1016/j.watres.2022.118070.

Zhu, Y., Oishi, W., Maruo, C., Bandara, S., Lin, M., Saito, M., Kitajima, M., Sano, D., 2022. COVID-19 case prediction via wastewater surveillance in a low-prevalence urban community: A modeling approach. J. Water Health 20 (2), 459–470. http://dx.doi.org/10.2166/wh.2022.183.

Zulli, A., Pan, A., Bart, S.M., Crawford, F.W., Kaplan, E.H., Cartter, M., Ko, A.I., Sanchez, M., Brown, C., Cozens, D., Brackney, D.E., Peccia, J., 2021. Predicting daily COVID-19 case rates from SARS-CoV-2 RNA concentrations across a diversity of wastewater catchments. FEMS Microb. 2, http://dx.doi.org/10.1093/femsmc/xtab022.