



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis

Daniele Proverbio<sup>a</sup>, Françoise Kemp<sup>a</sup>, Stefano Magni<sup>a</sup>, Leslie Ogorzaly<sup>b</sup>, Henry-Michel Cauchie<sup>b</sup>, Jorge Gonçalves<sup>a,c</sup>, Alexander Skupin<sup>a,d,e</sup>, Atte Aalto<sup>a,\*</sup>

<sup>a</sup> University of Luxembourg, Luxembourg Centre for Systems Biomedicine, 6 av. du Swing, Belvaux 4376, Luxembourg

<sup>b</sup> Luxembourg Institute of Science and Technology, Environmental Research and Innovation Department, Belvaux 4422, Luxembourg

<sup>c</sup> University of Cambridge, Department of Plant Sciences, Downing St, Cambridge CB2 3EA, UK

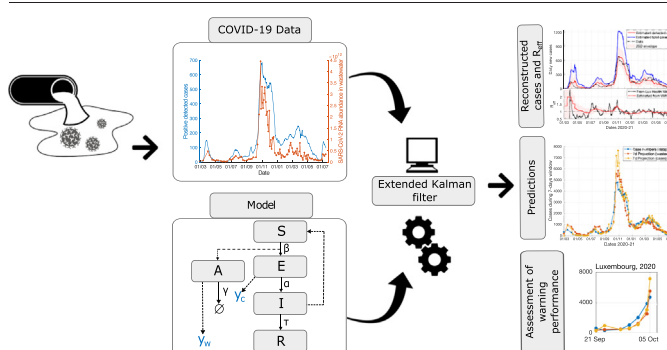
<sup>d</sup> University of Luxembourg, Department of Physics and Materials Science, 162a av. de la Faiencerie, Luxembourg 1511, Luxembourg

<sup>e</sup> University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA

## HIGHLIGHTS

- Wastewater-based epidemiology tracks the diffusion of infectious diseases.
- Suitable models are needed at the core of wastewater-based epidemiology.
- Mechanistic model-based wastewater analyser is developed.
- The method infers variables of epidemiological interest and predicts future dynamics.
- The method allows to quantify the wastewater real-time early warning performance.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 18 January 2022

Received in revised form 25 February 2022

Accepted 25 February 2022

Available online 1 March 2022

Editor: Damia Barcelo

### Keywords:

COVID-19

Wastewater-based epidemiology

Surveillance of wastewater for early epidemic

prediction (SWEEP)

Epidemiological modelling

Kalman filter

Early warning system

## ABSTRACT

Continuous surveillance of COVID-19 diffusion remains crucial to control its diffusion and to anticipate infection waves. Detecting viral RNA load in wastewater samples has been suggested as an effective approach for epidemic monitoring and the development of an effective warning system. However, its quantitative link to the epidemic status and the stages of outbreak is still elusive. Modelling is thus crucial to address these challenges. In this study, we present a novel mechanistic model-based approach to reconstruct the complete epidemic dynamics from SARS-CoV-2 viral load in wastewater. Our approach integrates noisy wastewater data and daily case numbers into a dynamical epidemiological model. As demonstrated for various regions and sampling protocols, it quantifies the case numbers, provides epidemic indicators and accurately infers future epidemic trends. Following its quantitative analysis, we also provide recommendations for wastewater data standards and for their use as warning indicators against new infection waves. In situations of reduced testing capacity, our modelling approach can enhance the surveillance of wastewater for early epidemic prediction and robust and cost-effective real-time monitoring of local COVID-19 dynamics.

\* Corresponding author at: Luxembourg Institute of Health, Department of Cancer Research, 1a-b rue Thomas Edison, Strassen 1445, Luxembourg.

E-mail address: [atte.aalto@uni.lu](mailto:atte.aalto@uni.lu) (A. Aalto).

## 1. Introduction

Effective mitigation of the COVID-19 pandemic relies on accurate estimates of the epidemic dynamics. Over the past months, vaccination programs and non-pharmaceutical interventions have managed to contain the diffusion of COVID-19 and many countries are aiming to return to normality, but the virus and its appearing variants might relapse once such measures are withdrawn. Hence, it is imperative to continue monitoring even after active RT-PCR or antigen testing is reduced. Analysing SARS-CoV-2 abundance in wastewater offers a cost-effective alternative to population-based large scale testing (Farkas et al., 2020; Larsen and Wigginton, 2020) and is largely independent of healthcare-seeking behaviors, access to clinical testing and asymptomatic cases (Peccia et al., 2020). It thus bears the potential for faster and more reliable warning indications for long-term epidemic surveillance (Weidhaas et al., 2021; Wurtzer et al., 2020; Randazzo et al., 2020). To date, more than 50 countries and 260 universities have wastewater surveillance systems in place (Naughton et al., 2021), and wastewater sampling to search for COVID-19 traces has been effectively used in multiple occasions (Ahmed et al., 2021; Quilliam et al., 2020; Reeves et al., 2021).

Despite its high potential, wastewater-based epidemiology has still several hurdles to overcome, related to experimental, data processing and modelling procedures (Daughton, 2020; Bandala et al., 2021; Lahrich et al., 2021). In particular, the usefulness of its results would greatly improve with models that can infer the size of the shedding population, account for uncertainties and infer future outcomes from current trends (Tiwari et al., 2021). In a recent call for models, Zhu et al. (2021) stress that modelling tools to reduce uncertainties and cope with noisy data are highly demanded, along with other mathematical models presenting different structures that could decently capture the correlation between viral load in wastewater and the shedding/infected population. Recently, a number of studies have proposed methods to infer the size of the shedding population, based on the viral abundance sampled in wastewaters. Some are restricted to qualitative and semi-quantitative retrospective studies of lagged correlations (Nemudryi et al., 2020; Kumar et al., 2020), others employ parametrical and non-parametrical regression models (Cao and Francis, 2021; Vallejo et al., 2021; Li et al., 2021; Hasan et al., 2021; Huisman et al., 2021); an alternative is to use an epidemiological SEIR (Susceptible-Exposed-Infectious-Recovered) model informed by estimates of individual viral trajectories (McMahan et al., 2021). These models, tested on one specific country at a time, well showed the difficulty of estimating the true number of positive cases from wastewater samples, because of noisy data (including those of detected case numbers), uncertain ratios of detected and true positive cases, and the variety of individual infection periods.

In this article, we propose an alternative, automated and causal-based approach to address the aforementioned challenges. We develop a new method that couples a Susceptible-Exposed-Infectious-Recovered (SEIR) epidemiological model (Anderson and May, 1979) to the extended Kalman filter – EKF (Kalman, 1960), a natural approach for combining noisy measurement data and modelling. The EKF deals effectively with the problem of producing model simulations that are representative of real system observations, which are in turn prone to uncertainties. It has been a standard tool in systems theory, with many applications ranging from automated control (Nagy Stovner et al., 2018) to finance (Davis and Leo, 2013), biomechanics (Marchesseau et al., 2013), time series analysis (Harvey, 1990) etc. Such an approach allows employing a validated epidemiological model (the SEIR, like in McMahan et al. (2021)) with a selected number of free parameters, fitted during the Kalman filtering steps using Bayesian methods. This way, the model estimates population-level COVID-19 diffusion without the need to rely on individual viral trajectories. It is also very flexible with respect to sampling routines and considered regional areas. Hence, including empirical measurements and tuning the model can be done in a straightforward way. In addition, the underlying SEIR model allows interpretation of the inferred infection dynamics in terms of transmitting interactions and overcomes the extrapolation limitations of correlation-based statistical approaches (Cao and Francis, 2021; Li et al., 2021). A Kalman filter was used for wastewater viral abundance data by Cluzel et al. (2022) and Courbariaux et al. (2022),

but they employed a linear first-order autoregressive model. This makes it a signal processing tool primarily aimed at reducing noise in the wastewater measurements, setting our approach apart.

After calibration, our CoWWAn (COVID-19 Wastewater Analyser) method causally and quantitatively links the results of wastewater analysis with those of population-wide testing, also accounting for the “dark number” (ratio of real and detected case numbers) through meta-parameters. This way, we are able to quantify the goodness-of-fit to observed cases (what is empirically measured) and to justify the reconstruction of the shedding population. Then, CoWWAn causally infers the shedding population, estimates the effective reproduction number  $R_{\text{eff}}$ , and provides projections of future epidemic trends. Quantifying these variables enables assessment of the epidemic status within a region and comparison between regions, and supports effective mitigation policy making. We also use this model to quantitatively assess the warning potentials of wastewater monitoring, combined with population testing or on its own. Originally, we applied CoWWAn in support to the Research Luxembourg COVID-19 taskforce of our local government. Moreover, to demonstrate its general applicability, we applied CoWWAn to public datasets from 12 regional areas from Europe and North America, associated with different population sizes and based on different wastewater data processing protocols.

## 2. Materials and methods

### 2.1. Data

Our pipeline requires at least three types of data to be calibrated: data about the COVID-19 RNA load in wastewater, detected cases associated to the area covered by the sewage system, and possibly the estimates about the ratio of detected *versus* true case numbers. In addition to its routine application to Luxembourg data, CoWWAn was tested on various datasets with different normalisation protocols for wastewater data, to show its general applicability after proper calibration. The dataset was constructed according to the following criteria. First, we employed the COVID19 Poops Dashboard (Naughton et al., 2021) to list all worldwide resources about wastewater sampling projects; among those, we focused on those having readily accessible databases. To allow proper calibration of the model, we selected time series data starting no later than beginning 2021 covering a time range of at least six months, having at least one sample per week on average and having the corresponding detected case numbers available. We rejected wastewater data with smoothing among data points to avoid introducing bias and breaking the causality of projections. Finally, if time series from multiple treatment plants were available from a single regional database, we selected two representative ones, usually with the largest population. The detected case numbers were obtained from the same publicly available databases and corresponded to officially reported numbers, confirmed with positive RT-PCR tests. We made sure that, for the majority of the time period considered, the selected countries did not have a share of positive tests exceeding 5%, which would indicate severe undertesting (according to WHO guidelines, <https://bit.ly/3dARcy1>) and would thus bias the results. When available, we also traced seroprevalence studies to better tune our pipeline to specific regional areas (e.g. Snoeck et al. (2020); Pollán et al. (2020)).

As a result, the selected datasets are: Barcelona Prat de Llobregat (Spain), Kitchener (Canada), Kranj (Slovenia), Lausanne (Switzerland), Ljubljana (Slovenia), Luxembourg, Milwaukee (USA), Netherlands, Oshkosh (USA), Raleigh (USA), Riera de la Bisbal (Spain), Zurich (Switzerland). Data from Luxembourg sewage sampling were made available by the Research Luxembourg COVID-19 initiative CORONASTEP ([researchluxembourg.lu/coronastep](https://researchluxembourg.lu/coronastep)), while detected case numbers and  $R_{\text{eff}}$  were obtained from the Luxembourg Ministry of Health website ([COVID19.public.lu/fr/graph](https://covid19.public.lu/fr/graph)). The other datasets (including both wastewater data and case numbers) were downloaded from publicly available official sources, listed in Supplementary Table 1. All datasets are updated up to August 2021. We refer to each source (see Supplementary Table 1) for details about the experimental protocols, for the units of measure of

wastewater data and for their associated equivalent population, as well as for the detected case numbers.

## 2.2. Preliminary analysis

Among the data collected, we observed some peculiarities. First, Raleigh county reported case numbers normalised to 10,000 inhabitants and rounded to an integer value; their subsequent up-scaling induces a further uncertainty. Second, the countrywide wastewater data for Netherlands are reported as averages over a week. To improve the temporal resolution of the data, we used instead the data from all communal treatment plants, averaging over samples from the same day. Third, the wastewater data from Kitchener have a sudden jump on May 17, 2021 during a time when case numbers remain stable (Supplementary Fig. 1). Interestingly, the performance of our method increased considerably after scaling data after that date by a factor of 0.4, suggesting possible sudden changes in testing or sampling strategies. This extra analysis shows the impact of including corrections for different testing policies. Results in the main text are shown without this scaling, but we report results with and without scaling in Supplementary Table 2.

We also investigated other prominent features of case numbers and wastewater data, to inform the development of the model. Considered time series of tested positive case numbers and of RT-qPCR wastewater data, as well as their mutual relationship, are shown in Supplementary Figs. 1 and 2. The figures highlight the close but not perfect correlation between case numbers and wastewater data, stressing both the usefulness of wastewater data for epidemic monitoring, and the importance of models based on complex epidemiological dynamics. The fact that the mutual relationship between case numbers and wastewater data is not perfectly linear justifies the inclusion of a scaling parameter in the cost function used for parameter fitting (Eq. (5) in Section 2.5).

## 2.3. The SEIR stochastic model

As a basis for the Extended Kalman filter to model the epidemic dynamics, we use a modified SEIR model, which has been shown to accurately describe COVID-19 epidemic dynamics (Proverbio et al., 2021; He et al., 2020). As we aim at estimating community incidence from noisy data, we choose a simple and descriptive model rather than a complex one, which is difficult to calibrate and could suffer from identifiability issues (Roda et al., 2020; Kemp et al., 2021).

The classic, deterministic SEIR model considers Susceptible  $S(t)$ , Exposed  $E(t)$ , Infectious  $I(t)$  and Removed  $R(t)$  compartments, and population flows governed by rate parameters. We follow the standard interpretation of  $E$  compartment as the set of individuals who have been exposed and infected, but who are not yet infectious due to incubation lag (Lai et al., 2020). The mean incubation period  $\alpha^{-1}$  models the progression to becoming contagious (Kollega et al., 2021). The total community population is conserved, i.e.  $S(t) + E(t) + I(t) + R(t) = N$  (with constant  $N$ ) and we assume no possibility of re-infection during each period of infection transmission (no  $R \rightarrow S$ ). The latter assumption is supported by waning immunity being estimated in a matter of months (Goldberg et al., 2021). To model intrinsic stochasticity in transmission processes and viral shedding, we employ a stochastic version of this SEIR model, associating each transition between compartments with a random process:

$$\begin{cases} \frac{d}{dt} S(t) = \frac{-\beta(t)S(t)I(t)}{N} - \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) \\ \frac{d}{dt} E(t) = \frac{\beta(t)S(t)I(t)}{N} - \alpha E(t) + \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) - \sqrt{\alpha E(t)}w_2(t) \\ \frac{d}{dt} I(t) = \alpha E(t) - \tau I(t) + \sqrt{\alpha E(t)}w_2(t) - \sqrt{\tau I(t)}w_3(t) \\ \frac{d}{dt} R(t) = \tau I(t) + \sqrt{\tau I(t)}w_3(t) \end{cases} \quad (1)$$

where  $w_j$  are mutually independent white noise processes. See Appendix A for details. The  $\beta$ -parameter is assumed to be time-varying, reflecting

changes in social interaction, other mitigation measures (masks, vaccines, etc.), and varying infectivity of emerging viral variants.  $\beta(t)$  will as well be estimated by the Kalman filter.

In order to model viral flows into wastewater, we introduce another variable  $A(t)$  to model the effective number of active shedding cases producing virions to wastewater. Similarly to above, we incorporate stochastic processes. The dynamics of  $A$  are given by:

$$\frac{d}{dt} A(t) = \frac{\beta(t)S(t)I(t)}{N} - \gamma A(t) + \sqrt{\frac{\beta(t)S(t)I(t)}{N}}w_1(t) - \sqrt{\gamma A(t)}w_4(t). \quad (2)$$

The  $A$  compartment is parallel to  $E$ ,  $I$ , and  $R$ , that is, it still holds that  $S(t) + E(t) + I(t) + R(t) = N$ . The influx to the  $A$  compartment is the same as that to the  $E$  compartment, while the outflux lumps together the dynamics of viral production (which is known to follow some kinetic trajectory in the hosts' body (Néant et al., 2021)), the decay rate of SARS-CoV-2 RNA in water (Gundy et al., 2009; Sala-Comorera et al., 2021), and inertia in abundance dynamics due to mixing in wastewater collecting pools. Since the parameter  $\gamma$  lumps together properties of the virus and details on wastewater sampling, it is separately fitted for each region. We do not take into account delays associated with in-sewer travel time, as it was estimated to be significantly lower than the transmission time scales (median of 3.3 h versus 1 day (Kapo et al., 2017)). The  $A$  compartment thus allows to better follow the time evolution, including potential decaying inertia, and to consider explicitly the uncertainties associated to the shedding mechanism instead of the disease progression. Together, Eqs. (1) and (2) form the combined SEIR-WW system.

The outputs from the model that are compared to the real-world measurements are the number of daily detected cases and the virion abundance in wastewater. The number of detected cases on day  $t \in \mathbb{N}$  is assumed to be a share of people passing the incubation period on that day, that is,

$$y_c(t) = c_t \int_{t-1}^t \alpha E(s) ds, \quad (3)$$

where  $c_t \in [0, 1]$  is the share of detected cases out of all cases, to account for under-testing and asymptomatic cases (see Section 2.5 and Eq. (C.4) for further discussion).  $c_t$  might depend on the day of the week, since there often are some weekday-dependent fluctuations in testing. The virion abundance in wastewater is assumed to be linearly dependent on  $A$ ,

$$y_w(t) = \nu A(t), \quad (4)$$

where  $\nu$  is a tuning parameter to reflect the incubation, production and shedding of viral load from infected people (Néant et al., 2021; Wölfel et al., 2020; Miura et al., 2021) and normalisation of the wastewater data. We do not consider explicit corrections linked to precipitations or other environmental factors, as previous studies evaluated them to be poorly correlated with RT-qPCR observations (Vallejo et al., 2021; Li et al., 2021). An implicit tuning is nonetheless included in the fitting, cf. Eq. (5).

## 2.4. The complete SEIR-WW-EKF model

An Extended Kalman filter requires an underlying dynamical model (such as a SEIR-like one), its output and associated noise covariance matrices, and measurement data. The extended Kalman filter algorithm to estimate the state of the SEIR-WW system, based on different types of data, is presented in Algorithm 1. The inputs for the algorithm are the update function  $f(x)$  (implementing Eq. (1)), the observation matrices  $C(t)$  (for case numbers and wastewater data), the state noise  $Q$  (uncertainty on estimated variables), and the measurement error covariance  $U(t)$  (uncertainties on empirical data). Then, the method evaluates the set of variables of interest  $(x_{1...6}(t) = [S, E, I, A, D, \beta](t))$  and their associated uncertainty matrix  $P$ .

The algorithm is used to calculate three different estimates using only case number data, only wastewater data or using both case and wastewater data. These were then used to estimate the data that were not employed for



the state estimation, initially for calibration and reconstruction of daily cases, and then to perform the desired predictions.

For details about the numerical implementation, the characterization of inputs and outputs and the discussion of each matrix introduced in Algorithm 1, we refer to Appendix B. Our current implementation is done with custom MATLAB 2019b code (see Code Availability section).

**Algorithm 1.** The Extended Kalman filter for the SEIR-WW model with time step  $\Delta t = 1/M$  (we use  $M = 10 \text{ d}^{-1}$ ).  $J_f$  is the Jacobian of the function  $f(x)$ , obtained from the Jacobian of the reaction function by  $J_f = BJ_r$ . The algorithm is standard, but the prediction step consists in solving a time-discretised ODE. The observation matrix  $C(t)$  is chosen from the three possibilities described in (B.2). Note the resetting of  $D(t) = \tilde{x}_5$  before the prediction loop.

```

Set  $P_0 \in \mathbb{R}^{6 \times 6}$  and  $\hat{x}(0)$ ;
for  $t = 1, \dots, T$  do
    set  $\tilde{x} = \hat{x}(t-1)$  and  $\tilde{x}_5 = 0$ ;
    set  $\tilde{P} = P_{t-1}$  and  $\tilde{P}_{j,5} = \tilde{P}_{5,j} = 0$  for  $j = 1, \dots, 6$ ;
    for  $i=1, \dots, M$  do
         $\tilde{P} = (I + \Delta t J_f(\tilde{x})) \tilde{P} (I + \Delta t J_f(\tilde{x})^\top) + Q(\tilde{x})$ ;
         $\tilde{x} = \tilde{x} + \Delta t f(\tilde{x})$ ;
    end
    Measurement error covariance:  $S = C(t) \tilde{P} C(t)^\top + U(t)$ ;
    State update:  $\hat{x}(t) = \tilde{x} + \tilde{P} C(t)^\top S^{-1} (y(t) - C(t) \tilde{x})$ ;
    Error covariance update:  $P_t = \tilde{P} - \tilde{P} C(t)^\top S^{-1} C(t) \tilde{P}$ ;
end

```

## 2.5. Model parameters

Our model comes with a number of free parameters to be fixed from the data or with educated assumptions. As most time series data begin after the pandemic already diffused within a region, the initial sizes for the  $E$  and  $I$  compartments are automatically computed (cf. Appendix C for details).

Another parameter to be estimated is the average ratio of total and detected cases at day  $t$ ,  $\eta_t$ . This is necessary to link the measurements of population testing with those of wastewater analysis (ideally objective and insensitive to testing capacities). Implementing the results of early prevalence studies (Snoeck et al., 2020), we use  $\eta_t = 3$  for the first wave in Luxembourg (until June 1, 2020). Later, we use  $\eta_t = 1.8$ . This choice was cross-validated with an independent SEIR model fitted to Luxembourg data (Kemp et al., 2021). The reduction is partially due to the launch of a large scale testing campaign in Luxembourg (Wilmes et al., 2021), and partially to overall increased testing activity. Further details on parameter

values are discussed in Appendix C. For other regions, the available prevalence studies usually consider the early stages of the epidemic, but wastewater data for the corresponding period are often not available. Large changes between first and subsequent waves are expected for all regions, and therefore estimates from these prevalence studies are not usable for later stages. In the absence of additional reliable values, we maintain  $\eta_t = 1.8$  for all other regions. It is possible to further calibrate such values with further tailored prevalence studies. In principle, it is sufficient to have one estimate for  $\eta_t$  to match virion abundance in wastewater with total case numbers. Once the calibration is done, CoWWAn estimates the total number of infections, including both detected and undetected cases. In order to obtain an estimate of detected cases, a potentially time-varying estimate of  $\eta_t$  is then needed. Daily ratios  $c_t$  of detected and total cases are obtained from  $\eta_t$  modulated by a weekly testing rhythm that is automatically estimated (cf. Appendix C). The variance of the wastewater measurements  $U_w$  is estimated from data as discussed in Appendix C.

A final detail to consider when optimising the model to reproduce the observations: due to dilution, non-mixing environment and other factors, the dependency of the wastewater measurement on the number of detected cases is not perfectly linear (Vallejo et al., 2021) (see also Supplementary Figs. 1 and 2). Hence, we do a simple power transformation to the wastewater samples, for which the exponent  $\varepsilon$  is regarded as a tuning parameter of slight nonlinearity.  $\varepsilon$  and the other proportional parameters  $\gamma$  and  $\nu$  are fitted by calculating the Kalman filter state estimate using the wastewater data, and then minimising the cost function

$$\min_{\gamma, \nu, \varepsilon} \sum_{t=1}^M (y_c(t) - C_c(t) \hat{x}_w(t; \gamma, \nu, \varepsilon))^2 \quad (5)$$

such that  $\gamma \in [0.2, 4]$ ,  $\varepsilon \in [0.4, 1]$ .

This way, we minimise the error in estimating the case numbers by the EKF state estimate using only wastewater data. Model parameters, either fixed by literature or fitted from Eq. (5), are reported in Table 1. Note that, due to different wastewater data normalisations,  $\nu$  parameters are not comparable between regions. Similarly,  $\varepsilon$  parameters might depend on the used techniques. Data from different laboratories may contain significant differences (Cluzel et al., 2022).

The sensitivity of the model performance on assumed parameter values is assessed in Supplementary Fig. 18, which demonstrates the robustness of the model and justifies the current parameter choices. The sensitivity analysis was performed by varying the reference parameters up to  $\pm 50\%$  of their original value. The results are reported in Supplementary Fig. 18, using Luxembourg as a reference. For most parameters, the projections are consistent and slightly vary for values very far from the reference ones. The model is most sensitive to the parameter  $c_b$ , which is usually

**Table 1**

Model parameters: Parameter symbols, descriptions, values, and their sources. The parameter  $q_{\beta}$ , controlling the allowed change of  $\beta(t)$  in one day, is changed after 30 days. This is done to allow rapid changes in the beginning of the pandemic, when a strict lockdown quickly suppressed its propagation and to account for errors in initial  $\beta(0)$ .  $d$  stands for “days”. When the source is not indicated, the parameter values is first initiated as an educated guess and then tested with sensitivity analysis (see Supplementary Fig. 18).

Symbol	Explanation	Value	Source
$\alpha$	Rate $E \rightarrow I$	$0.44 \text{ d}^{-1}$	(Kemp et al., 2021)
$\tau$	Rate $I \rightarrow R$	$0.32 \text{ d}^{-1}$	(Kemp et al., 2021)
$\beta(0)$	Initial infectivity	$0.44 \text{ d}^{-1}$	–
$\Delta t$	Time step length	$0.1 \text{ d}$	–
$q_{\beta, 1}$	Variance of $\beta(t+1) - \beta(t)$ when $t \leq 30$	$0.05^2 \text{ d}^{-2}$	–
$q_{\beta, 2}$	Variance of $\beta(t+1) - \beta(t)$ when $t > 30$	$0.005^2 \text{ d}^{-2}$	–
$\kappa$	EKF sensitivity parameter	4	–
$N$	Population size	Regional	See Supplementary Table 1
$\gamma$	Rate $A \rightarrow \emptyset$	Regional	Fitted by Eq. (5)
$\nu$	Ratio of $y_w/A$	Regional	Fitted by Eq. (5)
$\varepsilon$	Exponent in nonlinear mapping of WW data	Regional	Fitted by Eq. (5)
$U_w$	Measurement error variance of wastewater data	Regional	Estimated by Eq. (C.5)
$E(0)$	Initial size of $E$ -compartment	Regional	Estimated by Eq. (C.3)
$I(0)$	Initial size of $I$ -compartment	Regional	Estimated by Eq. (C.3)
$\text{var}(E(0))$	Uncertainty of $E(0)$	Regional	$(E(0)/2)^2$
$\text{var}(I(0))$	Uncertainty of $I(0)$	Regional	$(I(0)/2)^2$

estimated with independent methods. The minimal error corresponds to the reference value, while deviations induce larger errors. In our pipeline, changes in  $c_t$  are normally compensated by a change in  $\nu$  by the same amount. This observation justifies the differing fitted values reported in Table 2 for each region and recalls that, the more accurate seroprevalence studies are, the smaller the error associated with projections would be. The projection error grows slower for overestimated  $c_t$  than for underestimated  $c_t$ . Therefore, in case a good estimate of the share of detected cases out of all cases is lacking, it is advised to use a possibly overestimated rather than underestimated value for short-term projections. Note, however, that this may lead to higher overshoots in long-term projections due to overestimation of the susceptible population size.

## 2.6. Analysis of model outputs

To obtain variables of epidemiological interest, we further analysed the state estimates outputted by the SEIR-EKF model. Two estimates using only wastewater data are computed: one without interpolating data between sampling days (WW) and one with linear interpolation (ipWW).

The effective reproduction number  $R_{\text{eff}}$ , the time-dependent average number of secondary infections from a single contagious case in a susceptible population (Althaus, 2014), is directly extrapolated as (Kemp et al., 2021)

$$R_{\text{eff}} = \frac{\beta(t) S(t)}{\tau N}, \quad (6)$$

where  $\beta(t)$  and  $S(t)$  are state estimates. For  $N$  and  $\tau$ , see Table 1.

Short and mid-term projections are possible at any time  $t_0$  by stopping the Kalman filtering and simulating the model forward in time, starting from the latest state estimate and keeping the infectivity parameter constant ( $\beta(t) = \beta(t_0)$ ). The effect of uncertainty in the parameter estimate  $\beta(t_0)$  can be quantified by simulating envelopes using  $\beta(t_0) \pm 2\sqrt{P_{t_0}(6,6)}$  in the simulation (for every  $t$ ,  $P(6,6)$  represents the variance associated to  $\beta$  in the Kalman filter update, as discussed in Algorithm 1). Note that other uncertainties are omitted in these simulations; therefore, the short-term uncertainty in particular is under-estimated by the envelope.

Quantifying the quality of short-term projections using either case data only, wastewater data only, or both provides more reliable estimates of the epidemic unfolding over short time horizons. At each time step when wastewater data is available, the Kalman filter state estimation is stopped, and the SEIR-WW model is simulated  $T$  days forward without taking into account any new data. The total number of observed cases from the projection is calculated and compared with the actual number of observed cases during the same time horizon. Their absolute difference constitutes the prediction error. The prediction errors are standardised by the square root of

the true number of cases, which represents the standard deviation estimate (assuming case numbers on a given time are binomially distributed). The standardised scores so obtained are then averaged over all time points on which the prediction is made, obtaining an overall average normalised error. To enable comparison between countries, the average standardised error is scaled per 100,000 equivalent inhabitants. Overall, the scaled average standardised prediction error  $\xi$  is:

$$\xi = \frac{1}{M} \sum_{i=1}^M \frac{|y_i - \hat{y}_i^j|}{\sqrt{y_i}} \frac{100,000}{N}, \quad (7)$$

where  $i$  is the index of each point in any time horizon  $[t_0, T]$  with  $M$  points in total;  $j$  is an index that considers the original type of data used for projections, i.e.  $j = \{c, w, b\}$  for case data only, wastewater data only, or both combined (note that, in the state estimate using combined data, the wastewater data are not interpolated); hatted variables are the Kalman projections while non-hatted variables correspond to measured data;  $N$  is the equivalent population of interest (cf. Table 2).

## 3. Results

### 3.1. Data integration into a SEIR model

The workflow of our CoWWAn approach, integration of empirical data into a SEIR model through the Extended Kalman filter, is illustrated in Fig. 1. The implemented SEIR model contains an additional compartment for active cases producing virions to wastewater. A detailed description of the workflow can be found in Section 2.4 (see in particular Algorithm 1). In a broad sense, our proposed Kalman filter combines a model of a dynamical system with measurements (case numbers  $y_c$ , wastewater measurement  $y_w$  or both) obtained from the real system that is being modelled. At each time step it first predicts the next state – the set of all variables – by propagating the old state estimate using the underlying model. From the predicted state estimate, the predicted measurement is calculated using the measurement model. Finally, the state estimate is updated based on the discrepancy of the true measurement and the model-predicted measurement. The model's state estimate then reflects the state of the real system, and it can be used to predict the system's dynamics in the future.

### 3.2. Reconstruction of case numbers

After appropriate calibration to test cases with parameter fitting, CoWWAn quantitatively reconstructs the time evolution of observed cases from wastewater data (Fig. 2a) by inferring the internal variables and parameters of the SEIR model. These include the susceptible, exposed and infectious population fractions, daily detected cases and time-dependent infection rate (see Section 2.4). In our case studies, full time series data were used for calibration for each region. When clear regime shifts in testing/sampling protocols are observed, the model can be re-calibrated appropriately to improve the performance, like for Kitchener (cf. Section 2.2 and Supplementary Table 2). To infer the global shedding population, the model needs additional information on the ratio of total and detected cases, typically obtained from seroprevalence studies (see Section 2.5). Thanks to the model structure, we could thus compare our results with the true number of detected cases (Fig. 2a, red and black lines), before inferring the global magnitude of the shedding population (Fig. 2a, blue line, from mean estimates). The latter is an extrapolation from CoWWAn estimates and information about the “dark number” of undetected cases (provided as a model parameter); further independent studies to estimate this quantity help fine-tune the results.

We compared our results with a linear regression model (after data curation to reduce the noise, in a similar spirit to (Vallejo et al., 2021)): CoWWAn's inferences achieve consistently higher correlation (Fig. 2b, blue and red sets), demonstrating the power of our mechanistic-based approach. These observations hold for all considered regions (Fig. 2c and

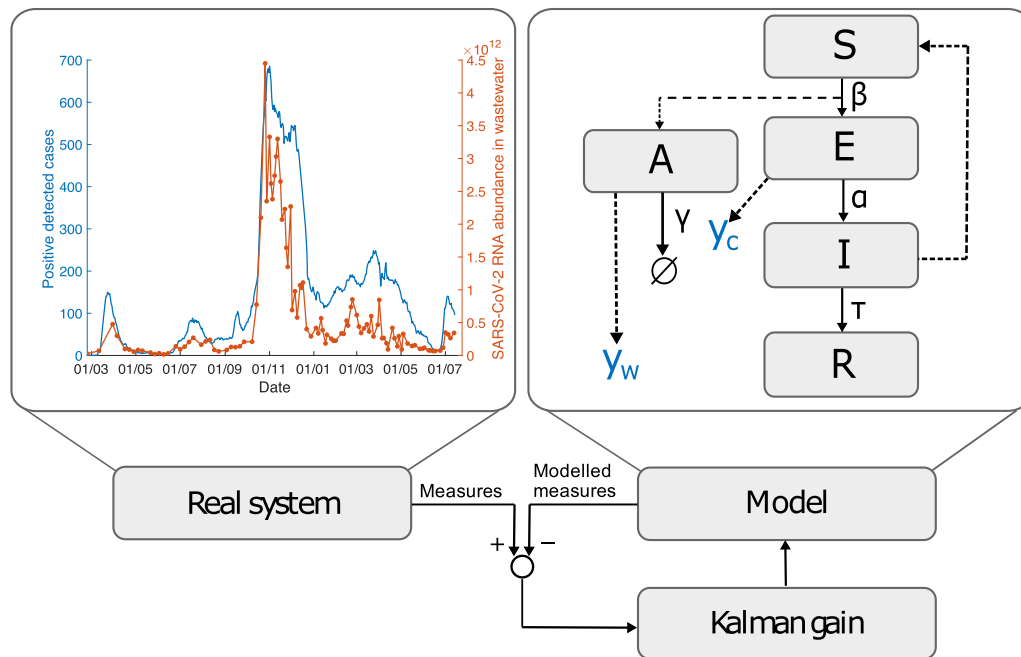
**Table 2**

Model parameters: fitted values. Region-dependent fitted parameter values. Initial values for SEIR compartments are in units of equivalent inhabitants.

Parameter	Barcelona	Kitchener	Kranj	Lausanne	Ljubljana	Luxembourg
$N$	2,000,000	242,000	40,000	240,000	280,000	634,730
$\gamma$	$0.20 \text{ d}^{-1}$	$4.00 \text{ d}^{-1}$	$1.43 \text{ d}^{-1}$	$3.05 \text{ d}^{-1}$	$3.21 \text{ d}^{-1}$	$1.62 \text{ d}^{-1}$
$\nu$	$4.86 \cdot 10^{-2}$	2.73	1.38	$6.67 \cdot 10^{10}$	2.77	$6.40 \cdot 10^4$
$\varepsilon$	0.40	0.40	0.40	1.00	0.526	0.613
$U_w$	656	2.68	58.5	$3.43 \cdot 10^{23}$	511	$1.75 \cdot 10^{12}$
$E(0)$	1824	379	17	156	76	8
$I(0)$	2527	525	24	203	105	11

Parameter	Milwaukee	Netherlands	Oshkosh	Raleigh	Riera	Zurich
$N$	615,934	17,178,109	68,000	460,000	100,000	450,000
$\gamma$	$4.00 \text{ d}^{-1}$	$0.368 \text{ d}^{-1}$	$4.00 \text{ d}^{-1}$	$4.00 \text{ d}^{-1}$	$1.20 \text{ d}^{-1}$	$0.547 \text{ d}^{-1}$
$\nu$	0.134	185	3.73	$2.58 \cdot 10^3$	12.7	$1.87 \cdot 10^7$
$\varepsilon$	1.00	0.500	0.866	0.434	0.400	0.789
$U_w$	1.03	$2.50 \cdot 10^{11}$	50.7	$6.00 \cdot 10^8$	$1.69 \cdot 10^3$	$3.44 \cdot 10^{17}$
$E(0)$	298	4412	166	1815	4	238
$I(0)$	413	6112	231	2514	6	330



**Fig. 1.** Model workflow. The Kalman filter combines measurements from the real system with predictions from the dynamical model, which extends a SEIR model. Empirical data are daily positive cases, shown in blue as the smoothed moving average, and wastewater sampled data, shown in orange with unit of measure of RNA copies/day/100,000 equivalent inhabitants (example for Luxembourg). Details of the SEIR blocks are described in Section 2.4.

Supplementary Figs. 3–14): the correlation coefficient  $\rho$  between inferred case numbers and true detected case numbers is typically in the range between 0.7 and 0.9 even for rather noisy data like Netherlands. Frequent sampling improves the model calibration and the subsequent reconstruction performance, like for Luxembourg with  $\rho = 0.91$  for two probes/week and Milwaukee with  $\rho = 0.95$  for two (sometimes more) probes/week compared e.g. to Barcelona with  $\rho = 0.70$  with one probe/week (Fig. 2d). The main discrepancies originate from either unnoticed changes in the share of detected cases or from changes in testing/sampling strategies (Supplementary Figs. 3–14). In addition, we notice (Fig. 2a and c) that the largest uncertainties come together with the highest case numbers, which are often associated to an augmented positivity rate (Ritchie et al., 2020). Detecting such discrepancies can provide additional evidence about potential undertesting and could guide targeted scaling of population tests. Interpolating wastewater data points before the EKF estimation can improve the reconstruction (Fig. 2b, red and yellow sets), in particular for regions with low sampling frequency like for Barcelona Prat de Llobregat (PdL) and Kranj. In general, the Extended Kalman filter improves its predictions as new data points are available, so an adequate sampling rate is recommended to improve its performance.

### 3.3. Estimation of epidemic indicators

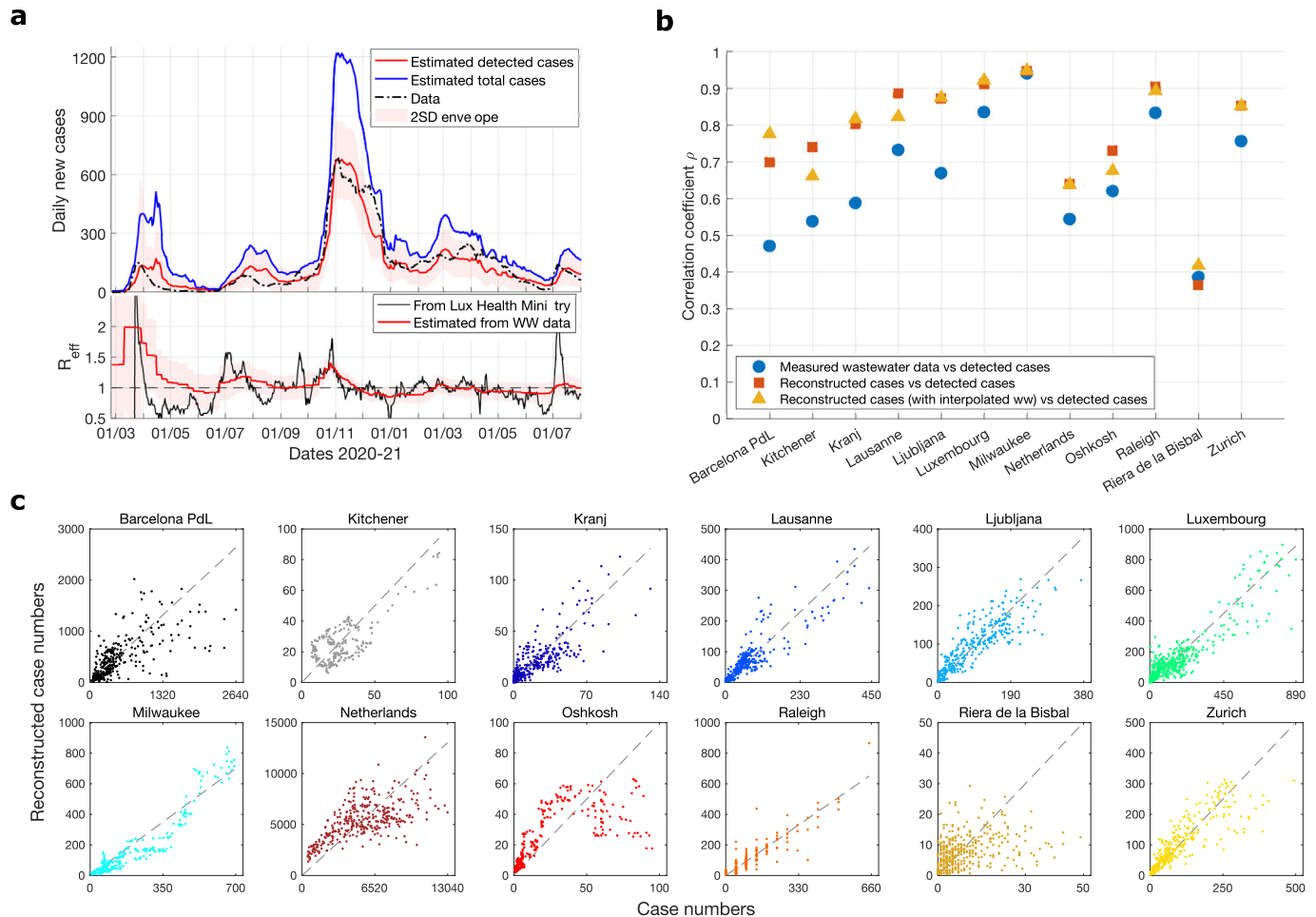
CoWWAn allows estimation of the effective reproduction number  $R_{\text{eff}}$ , an essential indicator for the trends of epidemic diffusion in a community (Huisman et al., 2021), which depends on containment measures, infectivity of viral variants, population behavior and other factors. As exemplified for Luxembourg (Fig. 2a), the  $R_{\text{eff}}$  values inferred by CoWWAn from wastewater data (according to Eq. (6)) are consistent with the indicator reported by the Ministry of Health on its website (see Section 2.1) and exhibit the same noteworthy trends: the three waves in 2020 (March, June and late October), a small rebound in March 2021 attributed to the emergence of the alpha variant and one wave in late June 2021 attributed to the emergence of the gamma and delta variants, all characterised by  $R_{\text{eff}} > 1$ . For all other considered regions as well, wastewater-based  $R_{\text{eff}}$  values (estimated with the same method and reported in Supplementary Figs. 3–14) are consistent with those estimated from case numbers using a SEIR model, and are

usually smoother due to sampling frequency and independence to testing schemes. The  $R_{\text{eff}}$  estimates reflect the trends in the development of the pandemic. On average, the  $R_{\text{eff}}$  estimates are lower in the year 2021 compared to 2020 due to vaccine rollout, but new viral variants have still caused significant waves despite increasing vaccination coverage.

### 3.4. Short-term predictions of epidemic trends

CoWWAn's underlying SEIR model permits mechanistic-based predictions of the infection dynamics, for effective monitoring of the epidemic. To predict future trends, it is possible to simulate the model forward at any desired time, starting from the latest state estimate and keeping the transmission parameter constant (cf. Section 2.6). For the epidemic dynamics in Luxembourg, Fig. 3a shows an example of such 7-days predictions for each day of wastewater sampling, where the number of detected cases (blue) is compared with the predicted numbers derived from wastewater data or from case number data. Wastewater-based short-term predictions are well correlated both with case-based projections ( $\rho = 0.95$ ) and with true case numbers ( $\rho = 0.94$ ).

Overall, for the different epidemic phases and all considered regions, the short-term predictions compare well with the real case data and with the case-based predictions (see also Supplementary Figs. 3–14). To quantify their performance, we determined the average standardised prediction error as the average discrepancy between predicted and actual case numbers in the corresponding time frame, normalised to case numbers and equivalent population Eq. (7). The performance of our wastewater-based pipeline is usually slightly lower, as it reconstructs the case numbers themselves before making the predictions, but remains similar with that of case-based predictions: all regional estimates lie within one standard deviation of the 1:1 (equal performance) line (Fig. 3b). The only exceptions are estimates for Oshkosh, probably due to under-testing during late 2020 (refer to Supplementary Fig. 2) which induced discrepancies in the detected cases fraction, and Kranj, whose low case numbers are subject to larger uncertainties (Supplementary Fig. 5). In general, the largest discrepancies are observed when case numbers plateau or decline after a rapid increase, yielding a potential overshoot of the predictions (Fig. 3a and Supplementary Figs. 3–14). This effect is associated to large changes in social activities



**Fig. 2.** Reconstruction of case numbers and inference of epidemic indicators. a: Reconstruction example for Luxembourg. Top: Comparison of case numbers, official detected data (black line), reconstructed by CoWWAn from wastewater data (red) including the 2 Standard Deviations = 95% confidence interval (shadowed region), and total positive cases inferred by CoWWAn (blue). Bottom:  $R_{eff}$  estimated by CoWWAn (red, with its associated 2 SD shadowed region) or officially reported by the Luxembourg Ministry of Health. b: Pearson's correlation coefficients  $\rho$  from linear regression between detected cases and measured wastewater data (blue),  $\rho$  between detected cases and CoWWAn-reconstructed case numbers from wastewater data (red, corresponding to correlation values from panels c), and  $\rho$  between CoWWAn-reconstructed case numbers from wastewater data (after interpolating wastewater data) and detected cases (yellow). c: Reconstruction results for all considered regional areas, compared with detected case numbers. The dashed line represents equal values.

during epidemic waves and rapid implementations of stricter restrictions, which are not explicitly included in the model but implicitly learned from the epidemic curve by the EKF with some delay.

The standardised error grows quite linearly with increasingly long prediction horizons (Fig. 3c). There, wastewater predictions are more stable (their uncertainty grows slower for longer prediction horizons) than those based on case numbers as they are usually less susceptible to daily fluctuations (Supplementary Table 2). This aspect allows quantifying and comparing the precision for different horizons.

In addition to using one type of data at a time, CoWWAn's EKF-based approach enables integrating different types of data to further improve the quality of predictions. Including both wastewater and case data slightly but systematically improves the prediction accuracy compared to case data alone (Fig. 3c and Supplementary Table 2), further suggesting that wastewater data contains independent information about the state of the epidemics, as previously put forward by Fernandez-Cassi et al. (2021).

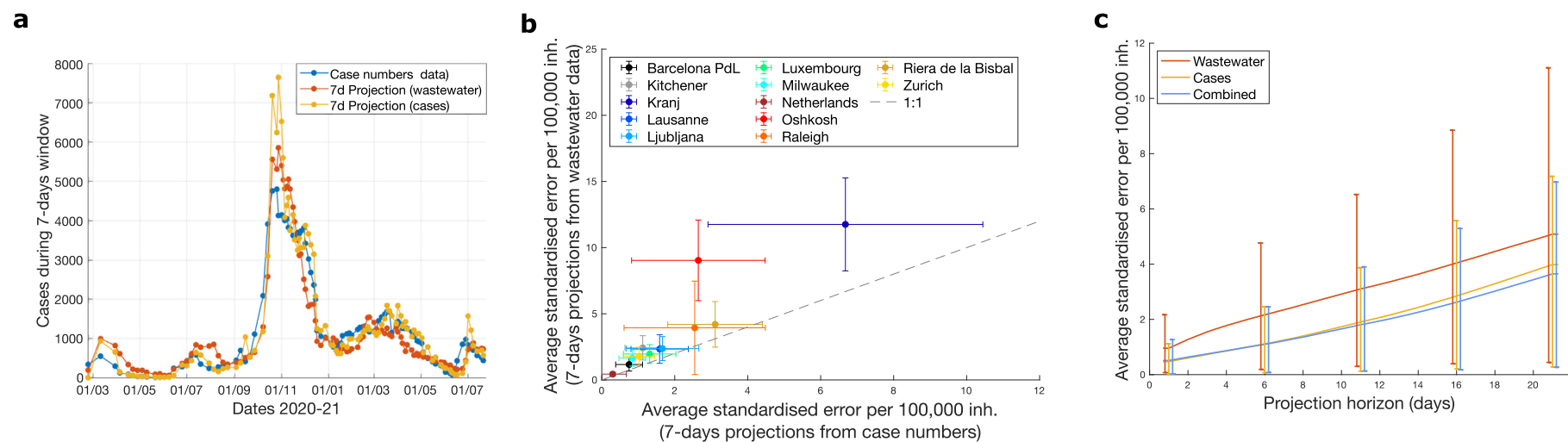
### 3.5. Long-term projections of epidemic scenarios

Due to heterogeneous and evolving adaptations of population behavior and institutional measures, epidemic forecasts are typically only

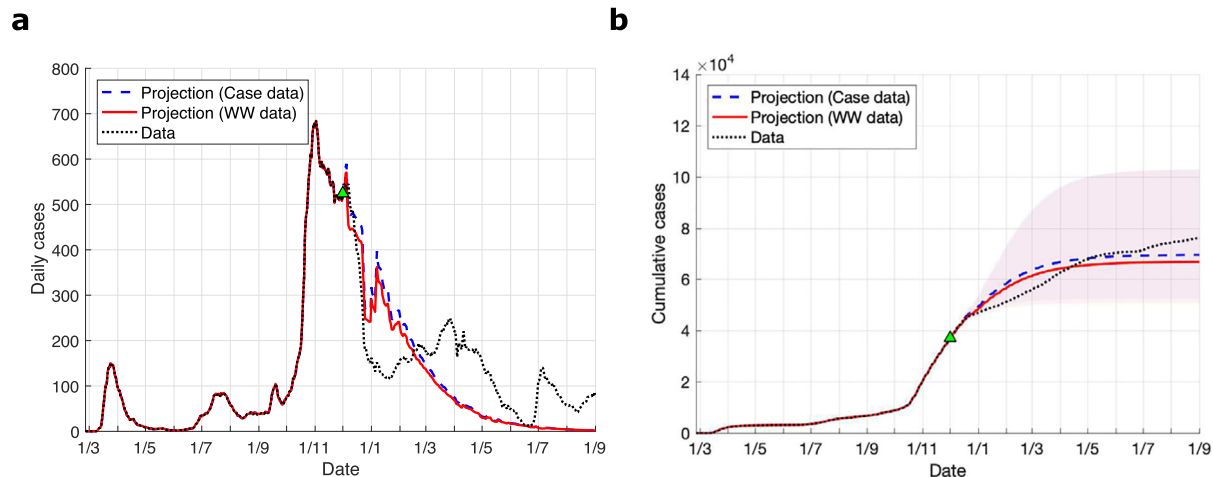
meaningful for relatively short time horizons. In fact, it is known that small uncertainties for short-term predictions are amplified over longer periods and the precision drops, similarly to what happens in weather forecasts (Petroopoulos and Makridakis, 2020). Nevertheless, long-term projections that assume no changes in infection dynamics can be useful for counterfactual analysis about the potential effects of current social or pharmaceutical measures and/or changed viral infectivity (Fig. 4a, b). They can also be used to investigate plausible scenarios, by artificially modifying the model parameters.

As for other models applied to complex systems, our projection uncertainties increase with longer time horizons (Fig. 4b), reflecting the set of potential changes of conditions. Nonetheless, projections based on case numbers or on wastewater data are consistent with each other within error bounds, therefore supporting the possibility of using wastewater data for consistent what-if analysis. In addition, our mechanistic-based model allows assessing the changes in desired precision. Models applied in quickly changing conditions are known to be uncertain (Santosh, 2020). Similarly, our projections' precision varies depending on whether they are conducted during a rapid increase of case numbers or during stable trends, calling for caution in interpreting these results as plausible projections rather than forecasts. Other examples are reported in Supplementary Fig. 15.





**Fig. 3.** Predictions of future epidemic trends using CoWWAn. a: Prediction examples for Luxembourg, comparing predictions over the 7-days ahead of each point (either estimated from case numbers or wastewater data) with the true detected cases in the same time period. b: Comparison of wastewater-based and cases-based predictions. The performance is evaluated in terms of average standardised error, normalised to equivalent population. The dashed line represents equal values. Error bars correspond to one standard deviation. c: Predictions performance for different time horizons (mean and 80th percentiles over the considered regions; outputs for single countries in Supplementary Fig. 17) for three inputs: case numbers, wastewater data, or both data combined. For all panels, “inh.” stands for inhabitants.



**Fig. 4.** Long term projections using CoWWAn. a: Long-term projected curves of daily cases compared with daily detected case numbers. b: Long-term projected curves of cumulative cases compared with cumulative detected case numbers. Blue and red ribbons represent  $\pm 2\sigma$  error bounds ( $\sigma$  corresponds to a standard deviation); note that the ribbons might overlap. Both panels a and b report examples for Luxembourg data, with projections starting at the date marked by the green triangle.

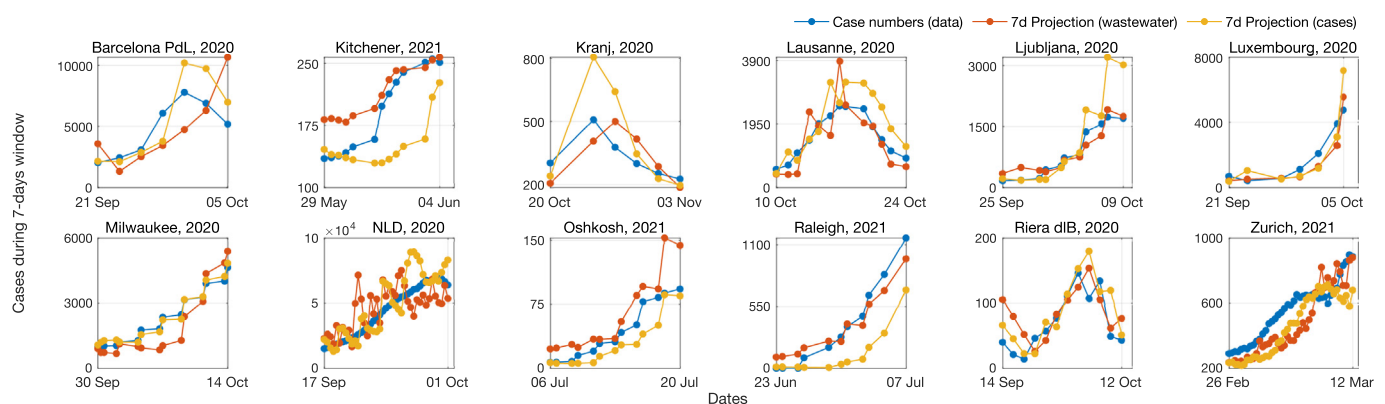
### 3.6. Modelling assesses wastewater warning performance

Among the purposes of this article is to investigate the utility of wastewater sampling to alert against new waves of infections and to inform its interpretation. It has been suggested by [Cao and Francis \(2021\)](#), [D'Aoust et al. \(2021\)](#) and [Kumar et al. \(2021\)](#) that wastewater analysis could provide early warnings for COVID-19 resurgence in a community. Using our approach, we investigate this idea beyond retrospective analysis. We recall that, for the real-time detection of impending epidemic resurgence, distinguishing between fluctuations and robust increases is crucial to optimise the true positive signals and minimise the false negatives. To evaluate the alerting power of on-line (real time) systems, it is thus not sufficient to compare two fully developed time series with a retrospective analysis. CoWWAn addresses this challenge by the EKF-based predictions, which capture robust trends in the epidemic dynamics. It thus allows to compare early warnings of COVID-19 resurgence, obtained from case-based and wastewater-based predictions. In [Fig. 5](#) and Supplementary Fig. 16, we plot the predictions about the pandemic trends, obtained from wastewater data (red) and from detected case numbers (yellow), and compared with the true observed evolution (blue). We can then compare if and when the red and yellow curves correctly track the increasing trend of the blue curve. We observe that, overall, the prediction curves accurately increase when a new COVID-19 wave is observed in a region, but the timing might slightly differ depending, e.g., on the testing frequency. This analysis

demonstrates the potential of wastewater data to detect incoming increasing trends and quantitatively verifies the recent calls by [Bibby et al. \(2021\)](#) for cautious interpretation: alerts based on wastewater analysis might be just-on-time or even lagging slightly behind the true infection waves. Nonetheless, they are often more advanced than reliable alerts based on case numbers alone, e.g. for Kitchener or Raleigh, despite counter-examples exist (e.g., Kranj). As a result, we suggest that wastewater-based monitoring could be an effective method to detect new waves of infection, but that the lead time should be carefully assessed case-by-case, according to the sampling frequency and other characteristics of the wastewater-analysis pipeline. In short, reliable warnings can be triggered, but it still remains to properly verify how “early”.

## 4. Discussion

CoWWAn combines two powerful approaches to process wastewater data in an automated and mechanistic-based manner: an epidemiological SEIR model and an extended Kalman filter, to fit the model parameters adequately and to provide predictions of epidemic trends. This allows for new avenues for wastewater-based epidemic monitoring. In situations of reduced population testing, our approach allows to enhance the performances and robustness of real-time surveillance in a cost-effective manner. Our model can support the reconstruction of the infection curves from wastewater data and allows projections of future trends, in particular



**Fig. 5.** Zoom into the epidemic resurgences visually recognised in the considered regions. Short-term projections used to identify robust trends in epidemic resurgence, for different examples (one per region; other examples in Supplementary Fig. 16). We compare 7-days projections from case numbers and from wastewater data with the true detected case numbers.

close to epidemic resurgence. Since hospital admission is downstream of the susceptible-exposed-infectious flow (Kemp et al., 2021), healthcare management as well (D'Aoust et al., 2021; Saguti et al., 2021) can obtain crucial information from an early detection of increasing case numbers supported by quantitative models that account for noise. We recall that our approach provides information on a community level but does not single out the infected individuals, hence it does not enable contact tracing nor does it reveal detailed information like age distribution of cases or infection clusters. Nonetheless, as already proven in our applications for the Luxembourg government, it proves useful to track the evolution of the pandemic, as a complement or even supplement to widespread testing. As a consequence, our results can enhance the SWEEP (Surveillance of Wastewater for Early Epidemic Prediction) framework recently proposed by Tiwari et al. (2021) for implementation of wastewater-based epidemiology.

CoWWAn can be easily applied and extended to different areas, and it overcomes some of the limitations of previous studies. In addition to employing a mechanistic underlying model or the Kalman filter, it allows improving  $R_{\text{eff}}$  estimation. In fact, Huisman et al. (2021) estimated  $R_{\text{eff}}$  from wastewater samples in Zurich, deconvoluting smoothed wastewater data using a kernel based on the shedding load distribution. For some reason, these estimate drops quite early in October 2020 while the case numbers are still steeply increasing. Our estimates do not exhibit such a shortcoming and is overall comparable over the remaining periods, see Supplementary Fig. 14c. However, we acknowledge the limitations of our approach, to be further improved in future studies. To begin with, the reconstruction of case numbers depends on the mean-field SEIR approximation: although meaningful when concentrating on average epidemic trends (Kollega et al., 2021), it might yield uncertainties in case of heterogeneous behaviors like clusters. In addition, we observe that tailoring region-specific model parameters is recommended to fine-tune the performance and reduce the uncertainties over the estimates. The parameters can usually be estimated with independent methods or educated prior information, in particular concerning seroprevalence, so we acknowledge that the current set of proposed parameters might not be complete for all countries. As observed in the sensitivity analysis, the projections are somewhat sensitive to the ratio of total and detected cases. This is a shortcoming of every model-based projection. Long-term projections are more influenced by the choice of this parameter, due to potential errors in the estimated level of natural immunity in the population. Short-term projections are less sensitive, since any error in the estimated size of the susceptible population is compensated by the infectivity parameter estimate. In particular, we recommend using reliable estimates for the ratio of true *versus* detected cases during the model calibration. The best estimates originate from seroprevalence studies that are able to distinguish between antibodies from previous infection and vaccination, that is, studies that detect antibodies against other parts of the virus than just the spike protein (Suhandynata et al., 2021). Future discrepancies between wastewater-based estimates and detected cases might be used as indications about changes in the share of detected cases and could be used to trigger a warning against potential undertesting. As for what concern predictions, we observe a close relationship between data and prediction quality: escalating the sampling precision and rates is essential to improve the model estimations. Finally, although we have paved the way for the assessment of on-line warnings of epidemic waves using model-based predictions, we suggest that future studies further quantify the lead time and the precision/recall. Future analysis may also concentrate on optimising the desired performance and the costs associated, or might focus on expanding the current methodology to other epidemic contexts.

## 5. Conclusion

Sampling and analysing SARS-CoV-2 fluxes in wastewater has been suggested as an efficient, non-invasive and cost-effective complement or alternative to testing routines. Our study leverages the potential of wastewater analysis to provide quantitative information for monitoring, alerting and decision-making. By introducing an effective coupling of causal-based

models and wastewater sampling, our approach goes beyond statistical methods. In fact, it allows immediate interpretation of its outputs and enables counterfactual analysis, to estimate plausible epidemic scenarios. Overall, the flexibility of our freely available approach, its ease of implementation and its performance make it an important tool for long-term monitoring and support of epidemic mitigation.

## Data availability

The wastewater and case numbers data that support the findings of this study are available from the websites listed in Section 2.1 and Supplementary Table 1. Luxembourg data for this study are available at [gitlab.lcsb.uni.lu/SCG/cowwan](https://gitlab.lcsb.uni.lu/SCG/cowwan).

## Code availability

CoWWAn's implementation for Matlab 2019b is available at [gitlab.lcsb.uni.lu/SCG/cowwan](https://gitlab.lcsb.uni.lu/SCG/cowwan).

## CRediT authorship contribution statement

**Daniele Proverbio:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Roles/Writing - original draft, Writing - review & editing. **Françoise Kemp:** Methodology, Writing - review & editing. **Stefano Magni:** Methodology, Writing - review & editing. **Leslie Ogorzaly:** Data curation, Funding acquisition, Writing - review & editing. **Henry-Michel Cauchie:** Data curation, Funding acquisition. **Jorge Gonçalves:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Alexander Skupin:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Atte Aalto:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Roles/Writing - original draft, Writing - review & editing.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Atte Aalto reports financial support was provided by Luxembourg National Research Fund (FNR). Daniele Proverbio reports financial support was provided by Luxembourg National Research Fund (FNR). Françoise Kemp reports financial support was provided by Luxembourg National Research Fund (FNR). Stefano Magni reports financial support was provided by Luxembourg National Research Fund (FNR). Leslie Ogorzaly reports financial support was provided by Luxembourg National Research Fund (FNR). Henri-Michel Cauchie reports financial support was provided by Luxembourg National Research Fund (FNR). Jorge Gonçalves reports financial support was provided by 111 Project on Computational Intelligence and Intelligent Control.

## Acknowledgments

D.P. and S.M. are supported by the Luxembourg National Research Fund (FNR) through PRIDE15/10907093/CriTiCS and F.K. by the FNR project PRIDE17/12244779/PARK-QC. A.A. is supported by the FNR through CORE19/13684479/DynCell. L.O. and H.M.C. are supported by the FNR through the COVID-19-FT2/14806023/Coronastep+. J.G. is partly supported by the 111 Project on Computational Intelligence and Intelligent Control, ref. B18024. The authors want to thank the Research Luxembourg COVID-19 Task Force for general support and collaborative spirit.

## Appendix A. Derivation of the SEIR stochastic model

The SEIR model is based on the assumption that each susceptible person has probability  $\beta(t)I(t)/Ndt$  to become infected on an infinitesimal time interval  $[t, t + dt)$ , and that infection events are independent. The number of

new infections at  $[t, t + dt]$  is then a random variable from the binomial distribution  $\mathcal{B}(n, p)$  with  $n = S(t)$  and  $p = \beta(t)I(t)/Ndt$ . Assuming high enough number of cases and stationary rate parameters over a time interval  $\Delta t = 1$  day (Gillespie, 2000), the binomial distribution can be well approximated by the normal distribution with mean  $\beta(t)S(t)I(t)/Ndt$  and variance  $\beta(t)I(t)/Ndt(1 - \beta(t)I(t)/Ndt)S(t) = \beta(t)S(t)I(t)/Ndt + \mathcal{O}(dt^2)$ . The same reasoning can be repeated for all other transitions between compartments. The stochastic SEIR model is then given in Eq. (1).

## Appendix B. Developing the complete SEIR-WW-EKF model

To embed the SEIR dynamical system in the Extended Kalman filter, we formulate a time-discretised state-space version of the dynamical system Eq. (1) by explicit Euler method:

$$x(t + \Delta t) = x(t) + \Delta t f(x(t)) + w(t). \quad (\text{B.1})$$

To obtain the number of daily new infections from the model on a given day, an additional auxiliary state variable  $D(t)$  is defined, whose dynamics are given by

$$\begin{cases} D(t) = 0, \text{ for } t \in \mathbb{N}, \\ \frac{d}{dt} D(t) = \alpha E(t), \end{cases}$$

that is,  $D(t)$  is the differential counterpart of  $y_c(t)$  and is reset every day to keep track of new infections on the current day.

Including the auxiliary variable, the state space is 6-dimensional with variables  $x_{1...6}(t) = [S, E, I, A, D, \beta](t)$ . Due to conservation of  $N$ ,  $R(t)$  is redundant and is therefore omitted. Eq. (B.1) is complemented with the resetting of  $x_5(t)$  to zero once per day. The function  $f(x)$  can be represented by a reaction function  $r(x)$  which is multiplied by the stoichiometric matrix  $B$ :

$$f(x) = \begin{bmatrix} -x_1 x_3 x_7 / N \\ x_1 x_3 x_7 / N - \alpha x_2 \\ \alpha x_2 - \tau x_3 \\ x_1 x_3 x_7 / N - \gamma x_4 \\ \alpha x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 x_3 x_7 / N \\ \alpha x_2 \\ \tau x_3 \\ \gamma x_4 \end{bmatrix} = B r(x).$$

As argued in the previous section, the state noise  $w(t)$  can be well approximated as normally distributed with mean zero and covariance

$$Q(x) = \kappa^2 \Delta t B \text{diag}(r(x)) B^T + \Delta t Q_\beta$$

arising from the stochastic model Eq. (1); note that each white noise process  $w_j$  for  $j = 1, \dots, 4$  in (1) corresponds to its respective reaction  $r_j(x)$ . The coefficient  $\kappa$  is used to account for modelling errors. In particular, the SEIR model implicitly assumes a homogeneous and perfectly mixed population. This assumption leads to a rather small uncertainty. The coefficient  $\kappa$  can also be interpreted as a sensitivity tuning parameter. Lower  $\kappa$  leads to higher sensitivity but noisy estimates. Higher  $\kappa$  decreases sensitivity but increases robustness against noise. The parameter  $\beta$  has no dynamics through  $f(x)$ , but it is updated by the Kalman filter. The matrix  $Q_\beta$  is otherwise zero, except for the element (6,6) being  $q_\beta$ , which acts as a tuning parameter controlling the magnitude of change of  $\beta(t)$  in one day.

The measurements from the model are either detected cases on a given day and/or wastewater sampling. To this end, we define possible observation matrices:

$$C_c(t) = [0000c_t0], \quad C_w = [000\nu00], \text{ and } C_b(t) = \begin{bmatrix} C_c(t) \\ C_w \end{bmatrix}, \quad (\text{B.2})$$

where the sub-indices refer to cases ( $c$ ), wastewater ( $w$ ), and both ( $b$ ). We recall that  $c_t$  is the share of detected cases on a day  $t$ . It is a coefficient that reflects the testing strategy, which often depends on the day (reduced testing on weekends and on public holidays). The empirical measurements

are assumed to be noisy, with an additive, normally distributed noise with mean zero and covariance  $U(t) = \text{diag}(U_c(t), U_w)$  (or just  $U(t) = U_c(t)$  or  $U(t) = U_w$  if only one of the measurements is available). The variance of observed cases,  $U_c(t)$ , is obtained by assuming that cases are detected independently with probability  $c_t$ . This leads again to a Binomial distribution for detected cases with mean  $c_t D(t)$ , where  $D(t)$  is the number of new infections on day  $t$ . This is unknown to us, and we use a smoothed estimate  $\bar{D}(t) = \bar{y}_c(t)/\bar{c}_t$  (barred variables stand for 7-days moving averages). The variance of the Binomial distribution is given by  $U_c(t) = D(t)c_t(1 - c_t)$ . For Raleigh,  $23^2$  is added to the variance  $U_c(t)$  to account for the (independent) uncertainty due to the aforementioned rounding of the case numbers, where 23 is the largest possible rounding error ( $N/20,000$ ).

Algorithm 1 is used to calculate three different state estimates:  $\hat{x}_c(t)$  using only case number data ( $C(t) = C_c(t)$ );  $\hat{x}_w(t)$  using only wastewater data ( $C(t) = C_w$  on days when wastewater sampling is done, otherwise Kalman update is skipped);  $\hat{x}_b(t)$  using both case and wastewater data ( $C(t) = C_b(t)$  on days when wastewater sampling is done,  $C(t) = C_c(t)$  otherwise). These were then used to estimate the data that were not employed for the state estimation, that is, we calculated  $\hat{y}_w(t) := C_w(t)\hat{x}_c(t)$  and  $\hat{y}_c(t) := C_c(t)\hat{x}_w(t)$  ( $C_i$  are the Kalman filter observation matrices, Eq. (B.2)).

The Kalman filter is complemented with a simple outlier saturation for the wastewater data. The model-predicted value for a wastewater measurement is given by  $C_w \tilde{x}$ , with prediction error variance  $C_w \tilde{P} C_w^T + U_w$ . If the measurement differs from the model-prediction by more than four standard deviations, the measurement is replaced by the saturated value  $C_w \tilde{x} \pm 4(C_w \tilde{P} C_w^T + U_w)^{1/2}$ .

## Appendix C. Estimate of the model parameters

As discussed in the main text, we estimate the free parameters of our model from available data, to calibrate them appropriately.

The initial sizes for the  $E$  and  $I$  compartments are directly automatically estimated from the data by

$$E(0) = \frac{\eta_0}{\alpha} \left( 1 + \sum_{i=1}^5 \frac{y_c(i)}{5} \right) \quad \text{and} \quad I(0) = \frac{\eta_0}{\tau} \left( 1 + \sum_{i=1}^5 \frac{y_c(i)}{5} \right), \quad (\text{C.3})$$

where  $\alpha$  and  $\tau$  are the transition rates  $E \rightarrow I$  and  $I \rightarrow R$ , respectively, whose inverses are the average duration an infected person remains in  $E$  and  $I$  compartments.  $\eta_t$  is the average ratio of total and detected cases at day  $t$ . Considering 5 data points is a trade-off between approximating values on the first day and sensitivity to noise. The model is little sensitive to this choice, cf. Supplementary Fig. 18. For Luxembourg, the data starts from the very beginning of the epidemic, when testing was not performed as actively as in the later stages, hence the values discussed in Section 2.5.

The daily ratios  $c_t$  of detected and total cases are obtained as follows. Initially, a weekly rhythm for case numbers is identified by averaging first over five weeks, and then by a moving average over three weeks:

$$\tilde{c}_t = \begin{cases} \frac{35 \sum_{j=0}^4 y_c(\text{mod}(t-1, 7) + 1 + 7j)}{5 \sum_{s=1}^{35} y_c(s)} & \text{for } t \leq 35, \\ \frac{21 y_c(t-7) + y_c(t-14) + y_c(t-21)}{3 \sum_{s=t-20}^t y_c(s)} & \text{for } t > 35. \end{cases}$$

Then, these values are normalised by the weekly moving average:

$$c_t = \frac{\tilde{c}_t}{\eta_t \sum_{s=t-6}^t \tilde{c}_s}. \quad (\text{C.4})$$

Note that the procedure for the first five weeks is not causal, but some data is anyway needed for model calibration. The later values  $c_t$  are causally determined from data. To obtain final values on public non-weekend holidays,  $c_t$  is reduced by a factor of 4 from the value given by Eq. (C.4) to account for reduced testing. In case the weekly rhythm is not regular,



manual tuning could help improving performance (or estimating  $c_t$  based on number of performed tests, for example).

The variance of the wastewater measurements  $U_w$  is estimated from data by

$$U_w = K \left( \text{median}_{|y_w(t_j) - \frac{1}{5} \sum_{i=j-2}^{j+2} y_w(t_i)|} \right)^2, \quad (\text{C.5})$$

where each  $t_i$  is the time point when wastewater sampling is done. The scaling factor  $K$  is either 1/10 when wastewater data is used alone and  $K = 1$  when both case and wastewater data are used, as well as for the outlier detection. In the plots of wastewater data reconstruction,  $K = 1$  is used for plotting the uncertainty envelope.

#### Appendix D. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.154235>.

#### References

- Ahmed, W., Bivins, A., Simpson, S.L., Bertsch, P.M., Ehret, J., Hosegood, I., Metcalfe, S., Smith, W.J., Thomas, K.V., Tynan, J., et al., 2021. Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. *Environ. Int.* 158, 106938. <https://doi.org/10.1016/j.envint.2021.106938>.
- Althaus, C.L., 2014. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Curr.* 6. <https://doi.org/10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288>.
- Anderson, R.M., May, R.M., 1979. Population biology of infectious diseases: part I. *Nature* 280 (5721), 361–367. <https://doi.org/10.1038/280361a0>.
- Bandalá, E.R., Kruger, B.R., Cesarino, I., Leao, A.L., Wijesiri, B., Goonetilleke, A., 2021. Impacts of COVID-19 pandemic on the wastewater pathway into surface water: a review. *Sci. Total Environ.* 774, 145586. <https://doi.org/10.1016/j.scitotenv.2021.145586>.
- 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. <https://doi.org/10.1016/j.watres.2021.117438>.
- Cao, Y., Francis, R., 2021. On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Sci. Total Environ.* 786, 147451. <https://doi.org/10.1016/j.scitotenv.2021.147451>.
- Cluzel, N., Courbariaux, M., Wang, S., Moulin, L., Wurtzer, S., Bertrand, I., Laurent, K., Monfort, P., Gantzer, C., Le Guyader, S., et al., 2022. A nationwide indicator to smooth and normalize heterogeneous SARS-CoV-2 RNA data in wastewater. *Environ. Int.* 158, 106998. <https://doi.org/10.1016/j.envint.2021.106998>.
- Courbariaux, M., Cluzel, N., Wang, S., Maréchal, V., Moulin, L., Wurtzer, S., Mouchel, J.-M., Maday, Y., Nuel, G., Bertrand, Obépine Consortium, 2022. A flexible smoother adapted to censored data with outliers and its application to SARS-CoV-2 monitoring in wastewater. *Front. Appl. Math. Stat.* 8, 836349. <https://doi.org/10.3389/fams.2022.836349>.
- D'Aoust, P.M., Graber, T.E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A.T., Mayne, J., Zhang, X., Alain, T., et al., 2021. Catching a resurgence: increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Sci. Total Environ.* 770, 145319. <https://doi.org/10.1016/j.scitotenv.2021.145319>.
- Daughton, C.G., 2020. Wastewater surveillance for population-wide COVID-19: the present and future. *Sci. Total Environ.* 736, 139631. <https://doi.org/10.1016/j.scitotenv.2020.139631>.
- Davis, M., Leo, S., 2013. Black-Litterman in continuous time: the case for filtering. *Quant. Financ. Lett.* 1 (1), 30–35. <https://doi.org/10.1080/21649502.2013.803794>.
- Farkas, K., Hillary, L.S., Malham, S.K., McDonald, J.E., Jones, D.L., 2020. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Curr. Opin. Environ. Sci. Health* 17, 14–20. <https://doi.org/10.1016/j.coesh.2020.06.001>.
- 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. <https://doi.org/10.1016/j.watres.2021.117252>.
- Gillespie, D., 2000. The chemical Langevin equation. *J. Chem. Phys.* 113 (1), 297–306. <https://doi.org/10.1063/1.481811>.
- Goldberg, Y., Mandel, M., Bar-On, Y.M., Bodenheimer, O., Freedman, L., Haas, E.J., Milo, R., Alroy-Preis, S., Ash, N., Huppert, A., 2021. Waning immunity after the BNT162b2 vaccine in Israel. *N. Eng. J. Med.* 385 (24), e85. <https://doi.org/10.1056/NEJMoa2114228>.
- Gundy, P.M., Gerba, C.P., Pepper, I.L., 2009. Survival of coronaviruses in water and wastewater. *Food Environ. Virol.* 1 (1), 10–14. <https://doi.org/10.1007/s12560-008-9001-6>.
- Harvey, A.C., 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Hasan, S.W., Ibrahim, Y., Daou, M., Kannout, H., Jan, N., Lopes, A., et al., 2021. Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: surveillance of COVID-19 epidemic in the United Arab Emirates. *Sci. Total Environ.* 764, 142929. <https://doi.org/10.1016/j.scitotenv.2020.142929>.
- He, S., Peng, Y., Sun, K., 2020. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynam.* 101 (3), 1667–1680. <https://doi.org/10.1007/s11071-020-05743-y>.
- Huisman, J.S., Scire, J., Caduff, L., Fernandez-Cassi, X., Ganesanandamoorthy, P., Kull, A., Scheidegger, A., Stachler, E., Boehm, A.B., Hughes, B., 2021. Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. <https://doi.org/10.1101/2021.04.29.21255961> medRxiv, 2021.04.29.21255961.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82, 35–45. <https://doi.org/10.1115/1.3662552>.
- Kapo, K.E., Paschka, M., Vamshi, R., Sebasky, M., McDonough, K., 2017. Estimation of US sewer residence time distributions for national-scale risk assessment of down-the-drain chemicals. *Sci. Total Environ.* 603, 445–452. <https://doi.org/10.1016/j.scitotenv.2017.06.075>.
- Kemp, F., Proverbio, D., Aalto, A., Mombaerts, L., d'Herouel, 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. Theor. Biol.* 530, 110874. <https://doi.org/10.1016/j.jtbi.2021.110874>.
- Kollepara, P.K., Siegenfeld, A.F., Bar-Yam, Y., 2021. Modeling complex systems: A case study of compartmental models in epidemiology. *arXiv arXiv:2110.02947*.
- Kumar, M., Patel, A.K., Shah, A.V., Raval, J., Rajpara, N., Joshi, M., Joshi, C.G., 2020. First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2. *Sci. Total Environ.* 746, 141326. <https://doi.org/10.1016/j.scitotenv.2020.141326>.
- Kumar, M., Joshi, M., Patel, A.K., Joshi, C.G., 2021. Unravelling the early warning capability of wastewater surveillance for COVID-19: a temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environ. Res.* 196, 110946. <https://doi.org/10.1016/j.envres.2021.110946>.
- Lahrich, S., Laghrib, F., Farahi, A., Bakasse, M., Saqrane, S., El Mhammedi, M., 2021. Review on the contamination of wastewater by COVID-19 virus: impact and treatment. *Sci. Total Environ.* 751, 142325. <https://doi.org/10.1016/j.scitotenv.2020.142325>.
- Lai, S., Ruktanonchai, N.W., Zhou, L., Prosper, O., Luo, W., Floyd, J.R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., et al., 2020. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 585 (2), 410–413. <https://doi.org/10.1038/s41586-020-2293-x>.
- Larsen, D.A., Wigginton, K.R., 2020. Tracking COVID-19 with wastewater. *Nat. Biotechnol.* 38 (10), 1151–1153. <https://doi.org/10.1038/s41587-020-0690-1>.
- Li, X., Kundaivelu, J., Zhang, S., Shi, J., Sivakumar, M., Mueller, J., Luby, S., Ahmed, W., Coin, L., Jiang, G., 2021. Data-driven estimation of COVID-19 community prevalence through wastewater-based epidemiology. *Sci. Total Environ.* 789, 147947. <https://doi.org/10.1016/j.scitotenv.2021.147947>.
- Marchesseau, S., Delingette, H., Sermesant, M., Cabrera-Lozoya, R., Tobon-Gomez, C., Moireau, P., i Ventura, R.F., Lekadir, K., Hernandez, A., Garreau, M., et al., 2013. Personalization of a cardiac electromechanical model using reduced order unscented Kalman filtering from regional volumes. *Med. Im. An.* 17 (7), 816–829. <https://doi.org/10.1016/j.media.2013.04.012>.
- McMahan, C.S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J.A., Papat, S.C., Karanfil, T., et al., 2021. Covid-19 wastewater epidemiology: a model to estimate infected populations. *Lancet Planet. Health* 5 (12), e874–e881. [https://doi.org/10.1016/j.s2542-5196\(21\)00230-8](https://doi.org/10.1016/j.s2542-5196(21)00230-8).
- Miura, F., Kitajima, M., Omori, R., 2021. Duration of SARS-CoV-2 viral shedding in faeces as a parameter for wastewater-based epidemiology: re-analysis of patient data using a shedding dynamics model. *Sci. Total Environ.* 769, 144549. <https://doi.org/10.1016/j.scitotenv.2020.144549>.
- Nagy Stovner, B., Johansen, T., Fossen, T., Schjølberg, I., 2018. Attitude estimation by multiplicative exogenous Kalman filter. *Automatica* 95, 347–355. <https://doi.org/10.1016/j.automatica.2018.05.038>.
- Naughton, C.C., Roman, F.A., Alvarado, A.G.F., Tariqi, A.Q., Deeming, M.A., Bibby, K., Bivins, A., Rose, J.B., Medema, G., Ahmed, W., 2021. Show us the data: Global COVID-19 wastewater monitoring efforts, equity, and gaps. <https://doi.org/10.1101/2021.03.14.21253564> medRxiv, 2021.03.14.21253564.
- Néant, N., Lingas, G., Le Hingrat, Q., Ghosn, J., Engelmann, I., Lepiller, Q., Gaymard, A., Plantier, J.-C., Cédric Hartard, V.F., 2021. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proc. Natl. Acad. Sci. USA* 118 (8). <https://doi.org/10.1073/pnas.2017962118> e2017962118.
- Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K.K., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. *Cell Rep. Med.* 1 (6), 100098. <https://doi.org/10.1016/j.xcrm.2020.100098>.
- Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik, A.A., Wang, D., Wang, M., et al., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* 38 (10), 1164–1167. <https://doi.org/10.1038/s41587-020-0684-z>.
- Petropoulos, F., Makridakis, S., 2020. Forecasting the novel coronavirus COVID-19. *PLoS one* 15 (3), e0231236. <https://doi.org/10.1371/journal.pone.0231236>.
- Pollán, M., Pérez-Gómez, B., Pastor-Barriuso, R., Oteo, J., Hernán, M.A., Pérez-Olmeda, M., Sanmartín, J.L., Fernández-García, A., Cruz, I., de la Torre, N.F., et al., 2020. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet* 396 (10250), 535–544. [https://doi.org/10.1016/S0140-6736\(20\)31483-5](https://doi.org/10.1016/S0140-6736(20)31483-5).
- Proverbio, D., Kemp, F., Magni, S., Husch, A., Aalto, A., Mombaerts, L., Skupin, A., Gonçalves, J., Ameijeiras-Alonso, J., Ley, C., 2021. Dynamical SPOEIR model assesses the effectiveness of non-pharmaceutical interventions against COVID-19 epidemic outbreaks. *Plos one* 16 (5), 1–21. <https://doi.org/10.1371/journal.pone.0252019>.
- Quilliam, R.S., Weidmann, M., Moresco, V., Purshouse, L., O'Hara, Z., Oliver, D.M., 2020. COVID-19: the environmental implications of shedding SARS-CoV-2 in human faeces. *Environ. Int.* 140, 105790. <https://doi.org/10.1016/j.envint.2020.105790>.
- Randazzo, W., Cuevas-Ferrando, E., Sanjuán, R., Domingo-Calap, P., Sánchez, G., 2020. Metropolitan wastewater analysis for COVID-19 epidemiological surveillance. *Int. J. Hyg. Environ. Health* 230, 113621. <https://doi.org/10.1016/j.ijheh.2020.113621>.

- Reeves, K., Liebig, J., Feula, A., Saldi, T., Lasda, E., Johnson, W., Lilienfeld, J., Maggi, J., Pulley, K., Wilkerson, P.J., et al., 2021. High-resolution within-sewer SARS-CoV-2 surveillance facilitates informed intervention. *Water Res.* 204, 117613. <https://doi.org/10.1016/j.watres.2021.117613>.
- Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., Roser, M., 2020. Coronavirus pandemic (COVID-19). Our World in Data. <https://ourworldindata.org/coronavirus>. (Accessed 1 January 2022).
- Roda, W.C., Varughese, M.B., Han, D., Li, M.Y., 2020. Why is it difficult to accurately predict the COVID-19 epidemic? *Inf. Dis. Mod.* 5, 271–281. <https://doi.org/10.1016/j.idm.2020.03.001>.
- Saguti, F., Magnil, E., Enache, L., Churqui, M.P., Johansson, A., Lumley, D., et al., 2021. Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Res.* 189, 116620. <https://doi.org/10.1016/j.watres.2020.116620>.
- Sala-Comorera, L., Reynolds, L.J., Martin, N.A., O'Sullivan, J.J., Meijer, W.G., Fletcher, N.F., 2021. Decay of infectious SARS-CoV-2 and surrogates in aquatic environments. *Water Res.* 201, 117090. <https://doi.org/10.1016/j.watres.2021.117090>.
- Santosh, K., 2020. Covid-19 prediction models and unexploited data. *J. Med. Syst.* 44 (9), 1–4. <https://doi.org/10.1007/s10916-020-01645-z>.
- Snoeck, C.J., Vaillant, M., Abdelrahman, T., Satagopam, V.P., Turner, J.D., Beaumont, K., Gomes, C.P.C., Fritz, J.V., Schröder, V.E., Kaysen, A., et al., on behalf of the CON-VINCE study group, 2020. Prevalence of SARS-CoV-2 infection in the Luxembourgish Population – The CON-VINCE Study. <https://doi.org/10.1101/2020.05.11.20092916> medRxiv. 2020.05.11.20092916.
- Suhandynata, R.T., Bevins, N.J., Tran, J.T., Huang, D., Hoffman, M.A., Lund, K., Kelner, M.J., McLawhon, R.W., Gonias, S.L., Nemazee, D., Fitzgerald, R.L., 2021. SARS-CoV-2 serology status detected by commercialized platforms distinguishes previous infection and vaccination adaptive immune responses. *J. Appl. Lab. Med.* 6 (5), 1109–1122. <https://doi.org/10.1093/jalm/jfab080>.
- Tiwari, S.B., Gahlot, P., Tyagi, V.K., Zhang, L., Zhou, Y., Kazmi, A., Kumar, M., 2021. Surveillance of Wastewater for Early Epidemic Prediction (SWEEP): environmental and health security perspectives in the post COVID-19 Anthropocene. *Environ. Res.* 195, 110831. <https://doi.org/10.1016/j.envres.2021.110831>.
- Vallejo, J.A., Trigo, N., Rumbo-Feal, S., Conde-Pérez, K., Lopez-Oriona, Á., Barbeito, I., Vaamonde, M., Tarro-Saavedra, J., Reif, R., Ladra, S., et al., 2021. Modeling the number of people infected with SARS-CoV-2 from wastewater viral load in Northwest Spain. *Sci. Total Environ.* 811, 152334. <https://doi.org/10.1016/j.scitotenv.2021.152334>.
- Weidhaas, J., Aanderud, Z.T., Roper, D.K., VanDerslice, J., Gaddis, E.B., Ostermiller, J., Hoffman, K., Jamal, R., Heck, P., Zhang, Y., et al., 2021. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Sci. Total Environ.* 775, 145790. <https://doi.org/10.1016/j.scitotenv.2021.145790>.
- Wilmes, P., Zimmer, J., Schulz, J., Glod, F., Veiber, L., Mombaerts, L., Rodrigues, B., Aalto, A., Pastore, J., Snoeck, C.J., et al., 2021. SARS-CoV-2 transmission risk from asymptomatic carriers: results from a mass screening programme in Luxembourg. *Lancet Reg. Health-Eur.* 4, 100056. <https://doi.org/10.1016/j.lanepe.2021.100056>.
- Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al., 2020. Virological assessment of hospitalized patients with COVID-2019. *Nature* 581 (7809), 465–469. <https://doi.org/10.1038/s41586-020-2196-x>.
- Wurtzer, S., Marechal, V., Mouchel, J., Maday, Y., Teyssou, R., Richard, E., Almayrac, J., Moulin, L., 2020. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Eurosurveillance* 25 (50), 2000776. <https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000776>.
- Zhu, Y., Oishi, W., Maruo, C., Saito, M., Chen, R., Kitajima, M., Sano, D., 2021. Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Sci. Total Environ.* 767, 145124. <https://doi.org/10.1016/j.scitotenv.2021.145124>.