Journal of Water & Health

IWA PUBLISHING

Journal of Water and Health Vol 22 No 3, 584 doi: 10.2166/wh.2024.345

Applying a Digital Twin and wastewater analysis for robust validation of COVID-19 pandemic forecasts: insights from Catalonia

Pau Fonseca i Casas [a,*], Joan Garcia i Subirana[a], Lluís Corominas [b] and Lluís Maria Bosch[b]

[a] Universitat Politècncia de Catalunya – Barcelona Tech, Barcelona, Catalunya 08034, Spain
[b] Catalan Institute for Water Research (ICRA)
*Corresponding author. E-mail: pau@fib.upc.edu

PFiC, 0000-0002-6747-9736

ABSTRACT

Monitoring SARS-CoV-2 spread is challenging due to asymptomatic infections, numerous variants, and population behavior changes from non-pharmaceutical interventions. We developed a Digital Twin model to simulate SARS-CoV-2 evolution in Catalonia. Continuous validation ensures our model's accuracy. Our system uses Catalonia Health Service data to quantify cases, hospitalizations, and healthcare impact. These data may be under-reported due to screening policy changes. To improve our model's reliability, we incorporate data from the Catalan Surveillance Network of SARS-CoV-2 in Sewage (SARSAIGUA). This paper shows how we use sewage data in the Digital Twin validation process to identify discrepancies between model predictions and real-time data. This continuous validation approach enables us to generate long-term forecasts, gain insights into SARS-CoV-2 spread, reassess assumptions, and enhance our understanding of the pandemic's behavior in Catalonia.

Key words: simulation, SARS-CoV-2, system validation, viruses, wastewater

HIGHLIGHTS

- The paper presents a Digital Twin model to analyze the evolution of SARS-CoV-2.
- It shows how sewage data are used to validate the model's predictions.
- It outlines how the continuous validation allows long-term forecasts.
- The model can account for the asymptomatic nature of the infections, variants, and changes in population behavior due to NPIs.
- The method enhances the reliability and accuracy of the model's forecasts.

1. INTRODUCTION

Accurate detection of the true infection rate of SARS-CoV-2 is of utmost importance for health authorities as it enables them to forecast current trends, anticipate the impact on the healthcare system, and understand the long-term effects on the population. However, under-reporting of COVID-19 cases, primarily attributed to a significant number of asymptomatic individuals, has emerged as a major challenge (Nishiura *et al.* 2020; Ma *et al.* 2021). While asymptomatic individuals may have lower infectivity compared to symptomatic individuals (Sayampanathan *et al.* 2020), it is essential to precisely calculate the true infection rate not only for assessing the spread but also for estimating the number of patients affected by secondary effects (Carfì *et al.* 2020; Lindan *et al.* 2021; Nordvig *et al.* 2021; Taquet *et al.* 2021; COVID-19 takes serious toll on heart health – a full year after recovery). To address this need, we developed in the past a Digital Twin (DT) model for monitoring the spread of SARS-CoV-2 in Catalonia (Fonseca i Casas *et al.* 2023). A DT is a virtual representation of a real-world entity or system. It mirrors an object, process, organization, person, or other abstraction and spans the lifecycle of its physical counterpart. It is updated from real-time data, and uses simulation, machine learning, and reasoning to aid in decision-making. Therefore, the DT serves as a virtual model that simulates infection dynamics based on diverse data sources and underlying assumptions. To simulate the dynamics of the pandemic, and considering the DT structure we follow, we use different sources of information. These sources include the detected infected population, the hospitalizations, the deaths and the virus load in the wastewater, which are indicators of the spread and severity of the disease. The simulation traces are composed of the number of detected cases,

which depend on the testing capacity and strategy, and the real cases, which are estimated by the model based on the infection rate and the incubation period and validated with the detected cases and the wastewater information.

The objectives of this model are twofold: (i) to forecast infection trends across different regions and scenarios and (ii) to validate model assumptions by comparing them with actual outcomes and adjusting them accordingly. By doing so, we aim to understand the factors influencing disease spread and evaluate the effectiveness of various interventions.

However, since March 28, 2022, cases of SARS-CoV-2 infection have not been reported in Catalonia, with only manifestations in elderly population and severe COVID-19 cases being registered in the healthcare system database. Nevertheless, on March 28, the infection rate IA14 (incidence within 14 days) stood at 500 cases per 100,000 people, indicating a significant outbreak requiring vigilant monitoring and control. These numbers suggest that only a fraction of COVID-19 infections have been officially registered since March 28. This becomes apparent when examining the reported cases by age range in Catalonia (Figure 1). Notably, the majority of cases were observed among children until April, but after March 28, the distribution of cases across age ranges seemingly reversed due to changes in the data acquisition process. These changes in reporting were necessary to prevent the overload of primary healthcare centers but have complicated the validation of existing SEIRD models and hindered our understanding of the quality and nature of the newly acquired synthetic data. As the reliability of reported COVID-19 cases diminishes due to under-reporting, alternative data sources that can provide a more accurate

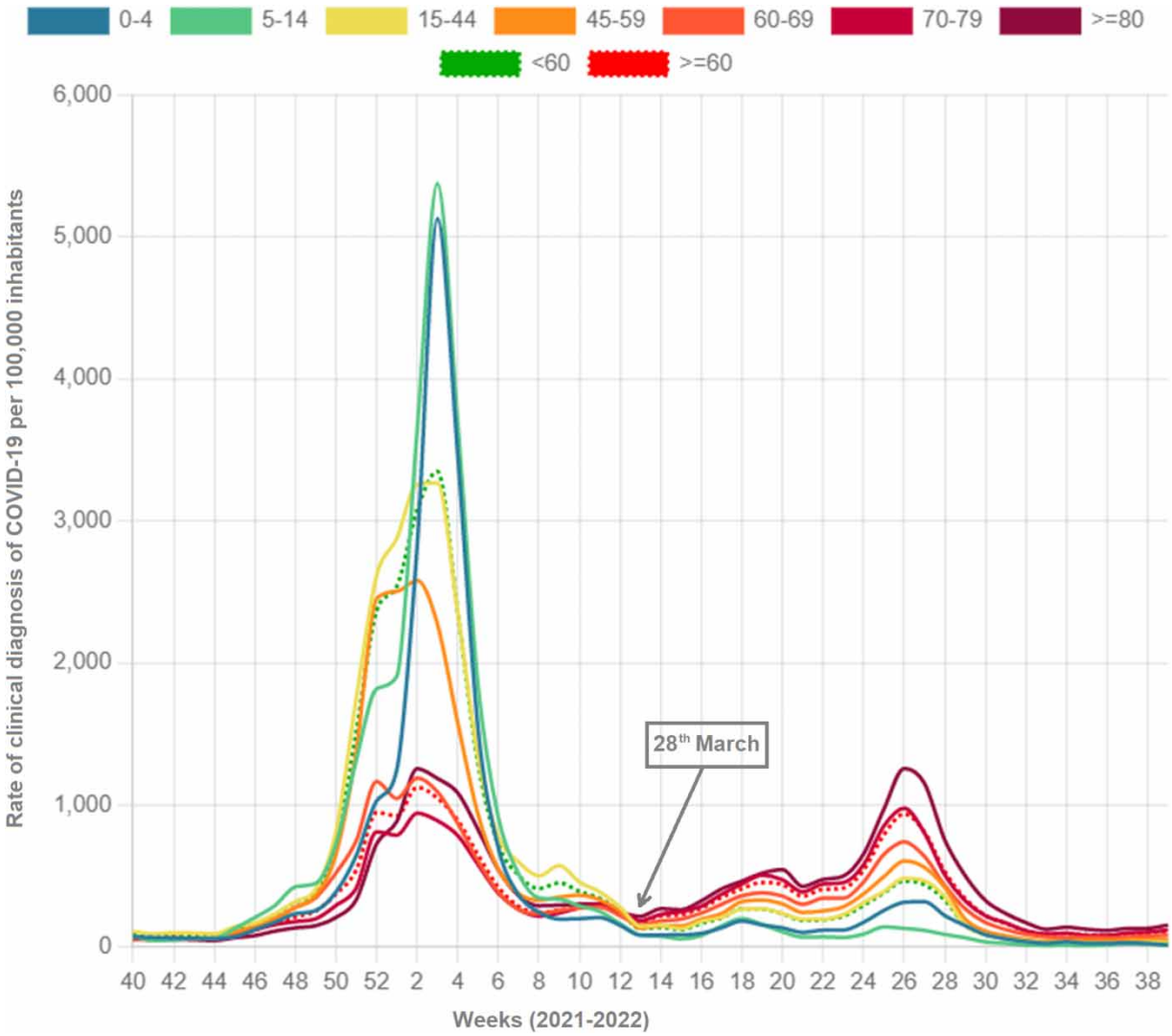

Figure 1 | Cases by 100,000 inhabitants by age range for 2021–2022, weeks. After March 28, 2022, mainly only cases of severe ill COVID-19 manifestation have been registered on the health system database.

reflection of the pandemic's true situation become indispensable. These sources include the number of deaths, critical hospitalizations, hospitalizations caused by COVID-19, and the viral load of SARS-CoV-2 detected in sewage. Sewage data, in particular, prove invaluable as it captures infections in asymptomatic or untested individuals and offers an early warning of virus circulation within a population (Medema *et al.* 2020).

To estimate the true infection rate of SARS-CoV-2 accurately, various methods have been developed to analyze wastewater data (Phipps *et al.* 2020; Wu *et al.* 2020; Rippinger *et al.* 2021). These methods aim to infer infection cases from the collected wastewater data. Consequently, any changes in the wastewater information can directly impact the forecasts generated by the model, making the forecast highly dependent on the quality and availability of these data. Detecting potential future turning points becomes challenging in such scenarios. To address this, it is crucial to utilize multiple data sources to feed our models and, more importantly, continuously validate the forecasts to identify discrepancies or errors in the underlying assumptions. Our assumption is that SARS-CoV-2 sewage loads serve as a valuable source of information for validating epidemic models, particularly as clinical testing and reporting decrease. Furthermore, by leveraging the model's forecasts, sewage data analysis becomes an excellent tool for understanding the nature of the newly acquired information and analyzing the assumptions that depict the effects of non-pharmaceutical interventions (NPIs) applied to the population.

Several studies have explored the use of SARS-CoV-2 loads in sewage as a means to estimate COVID-19 cases in a population, as referenced in (McMahan *et al.* 2021; Omori *et al.* 2021). In Catalonia, two specific studies (Ayala-Aldana *et al.* 2022; Joseph-Duran *et al.* 2022) have employed this approach to monitor the regional pandemic situation and provide insights into virus transmission dynamics. In contrast to previous studies that used wastewater data to infer actual COVID-19 cases, our approach employs wastewater data to validate a DT that integrates multiple data sources, enhancing the reliability of its forecasts. This paper focuses on describing the validation process using wastewater data to detect forecast inaccuracies in the models used within the Digital Twin, prompting recalibration for generating new long-term forecasts. However, it is important to note that wastewater data are not the sole source of validation. Information from the Health Service regarding confirmed cases detected through testing is also utilized. Based on our methodology, we have implemented a system that provides warnings to alert specialists when adjustments to the models are needed to accurately reflect the changing situation of the pandemic. Our main objective is to predict the true number of infection cases, which will help us understand the influence of NPIs. Ultimately, we aim to estimate the number of individuals suffering from secondary effects of COVID-19, like Long COVID.

2. MATERIALS AND METHODS

Understanding the propagation of SARS-CoV-2 is a complex task due to various factors. Firstly, different variants with varying transmission rates (parameter β as used in the epidemiological models) continue to emerge. Secondly, the prevalence of asymptomatic individuals is significant and varies depending on the variant, with a higher number of asymptomatic cases observed with the Omicron variant. Finally, citizen behavior changes, such as wearing face masks and adhering to physical distancing recommendations, also affect the transmission rate (β) (Chu *et al.* 2020). To tackle these challenges, we employed a DT approach, which combines a set of epidemiological models with a continuous validation and verification process.

The DT comprises three primary models that evolve throughout the pandemic. The first model is a System Dynamics model, conceptualized using Forrester diagrams (Forrester 1988) which represents a SEIRD model. This initial model allows us to test the initial assumptions through its implementation in InsightMaker software (Fortmann-Roe 2014). The second model is an optimization model developed in Python. Its purpose is to detect changing points and estimate the transmission values for different regimes of the infection curve, reflecting changes in NPIs implemented in the population. The third model is represented using the Specification and Description Language (SDL) (ITU-T 2019) and (Sherratt *et al.* 2015), a graphical, standardized, and unambiguous language from informatics area, but commonly used to describe environmental and social simulation models (Fonseca i Casas *et al.* 2010; Fonseca i Casas 2014; Olmos *et al.* 2014; Fonseca i Casas 2019). This SDL model extends the initial SEIRD model by incorporating different regimes based on NPI variations. Additionally, it employs cellular automata to represent different Health Regions, which are administrative divisions in Catalonia. Synthetically, a cellular automaton (CA) is a discrete model of computation that consists of a regular grid of cells, each in one of a finite number of states, such as on and off. The grid can be in any finite number of dimensions. For each cell, a set of cells called its neighborhood is defined relative to the specified cell. An initial state (time $t = 0$) is selected by assigning a state for each cell. A new generation is created (advancing t), according to some rule, or rules, that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood (Sirakoulis *et al.* 2014). While

all three models contribute to the validation process by comparing their outcomes, the SDL model, which provides the most detailed representation, is specifically used for generating forecasts. Thus, in this paper, when referring to the model or the Digital Twin, we are referring to the SDL model.

The current version of the SDL model used in the DT is version 2.11. However, this is not the first model employed in our overall project. We initially started with models 1.5–1.7, which were predominantly SEIR-like models. In the 2.X models, we incorporated CA models to enhance prediction capabilities, enabling the modeling of different Health Regions (Fonseca i Casas *et al.* 2021) and the detection of true cases and reinfections (Fonseca i Casas *et al.* 2023). Figure 2 illustrates the various SDL models developed to understand the evolution of the COVID-19 pandemic in Catalonia. It's important to note that these models are not fixed or permanent; they rely on specific assumptions and are subject to changes in NPIs over time. When new assumptions or NPIs are introduced, such as new variants, vaccination status, the model may no longer accurately reflect reality, resulting in model invalidation. In such cases, the model needs to be revised or replaced, as it becomes obsolete and cannot provide reliable predictions or explanations.

This approach requires accurate near real-time data to validate the models. The detected cases data used in this study were obtained from the Public Health Agency of Catalonia (ASPCAT). Additionally, data from the Catalan Surveillance Network of SARS-CoV-2 in Sewage (SARSAIGUA) (Guerrero-Latorre *et al.* 2022) were utilized for the continuous validation of a DT representing the pandemic evolution in Catalonia. As we mention, the DT is based on an extended SEIRD model that can forecast the true number of COVID-19-infected cases in Catalonia (Fonseca i Casas *et al.* 2023), and the overall continuous life cycle we follow for the DT is detailed in Fonseca i Casas *et al.* (2021, 2023). Our focus is not on estimating the number of cases based on SARS-CoV-2 viral loads in sewage. Instead, we aim to establish a continuous validation process to ensure the accuracy of predictions provided by the SEIRD model simulations. We specifically aim to detect discrepancies between the trends observed in sewage SARS-CoV-2 data and the simulated trends of true cases in the model. This validation process allows for the generation of long-term forecasts (several months) that can be combined with other data sources to produce more precise and reliable predictions. As we mention. our primary goal is to forecast the true number of infection cases in order to understand the impact of NPIs and eventually forecast the number of individuals experiencing secondary effects of COVID-19, such as Long COVID (Cooper *et al.* 2022; Mehandru & Merad 2022; Munblit *et al.* 2022; Davis *et al.* 2023; Landry *et al.* 2023; Mizrahi *et al.* 2023).

Figure 3 provides an overview of the components involved in our approach. The DT (DT) consists of several elements: (i) the Reference Asset, which represents the system under analysis, in this case, the pandemic evolution in Catalonia as indicated by the number of new SARS-CoV-2 infections; (ii) the Digital Shadow, which comprises digital information (raw data, processed data, and models) obtained from the system (Stark *et al.* 2017). This includes information provided by ASPCAT regarding new cases, hospitalizations, critical hospitalizations, and deaths, as well as data from SARSAIGUA on SARS-CoV-2 loads in sewage used for model validation; (iii) the Digital Master, which refers to the model or models used to represent the Reference Asset. In our case, the main Digital Master is the SDL model; and (iv) the Digital Master instances represent the implementation of the Digital Masters, in the case on the SDL model, their implementation is carried out using SDLPS software (Fonseca i Casas 2013), which facilitates the automatic coding of models represented on SDL. Finally, (v) the Traces represent synthetic data obtained from simulation models, which enable the validation of the Digital Twin.

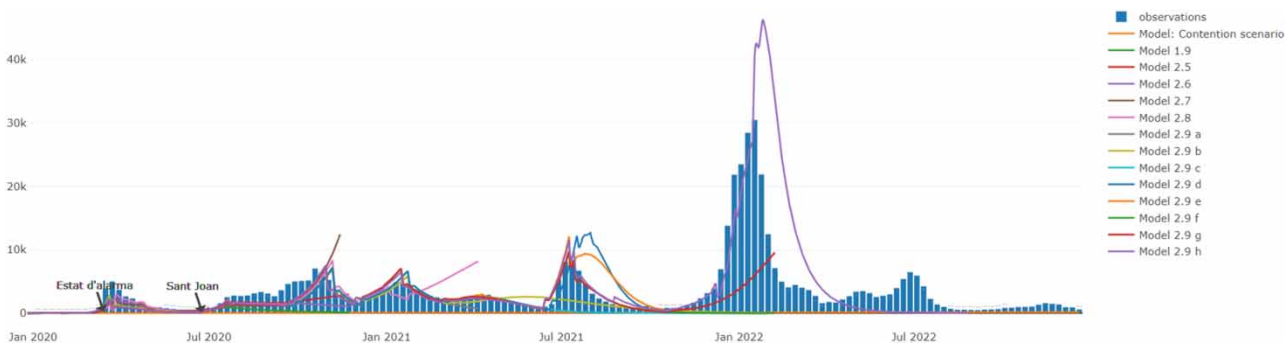

Figure 2 | History of the different SDL models we developed during the course of the pandemic. *y*-axis indicates the number of COVID-19 detected cases.

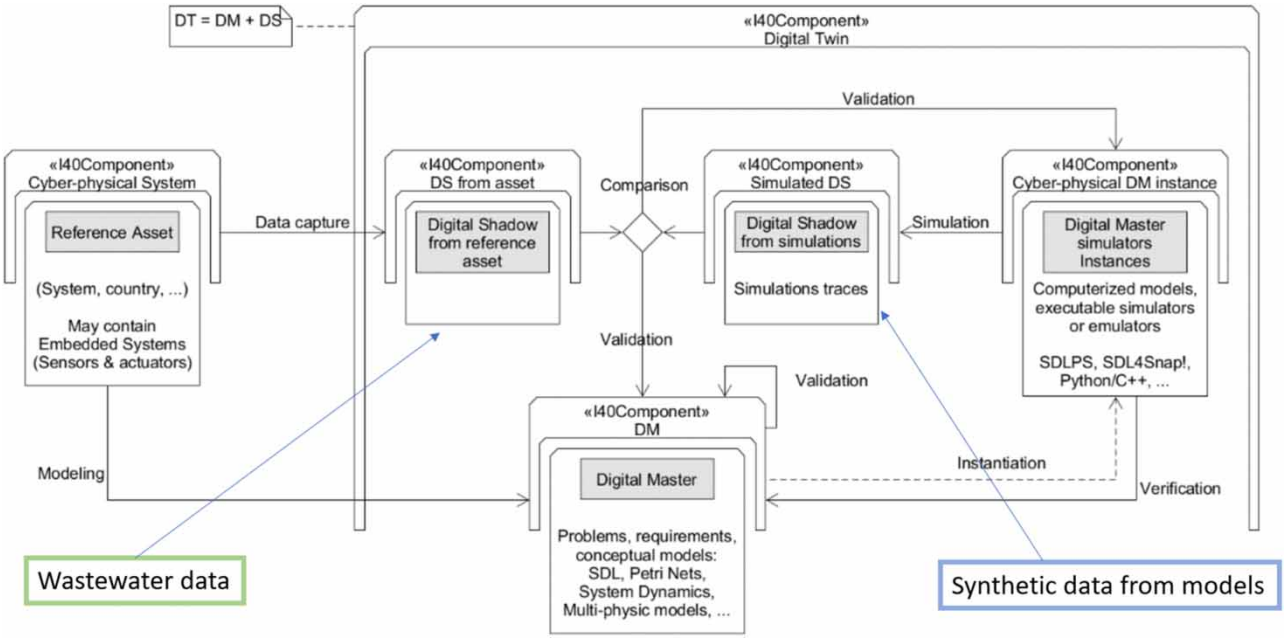

Figure 3 | Digital Twin approach used. We employ statistical analysis to enable automatic validation of the models by comparing the Digital Shadow with the Simulation Traces.

To implement the third model of the pandemic simulation, we use SDLPS software (Specification and Description Language Parallel Simulator), which is a distributed simulator that allows the definition of the models using SDL language. This way, we can directly use the Digital Master that we have conceptualized using SDL, without the need of translating it to another format. The input data (Digital Shadows) are defined in text files, which contain the time series for the different parameters that we must consider, such as the detected infected population, the hospitalizations, the deaths and the virus load in the wastewater. The simulation traces are also generated in text files, which include the number of detected cases and the real cases, estimated by the model. A Key Performance Indicator (KPI) panel written in R, uses this information to present the data in a user-friendly way, using graphs and tables, allowing us to perform validation, inspection and decision-making from the models. Figure 4 shows an example of the information that we can see on the application KPI panel, showing the curves for detected, real cases and reinfections, as well as the comparison with the reference data (Digital Shadows from the Reference Asset).

2.1. Structural and data assumption

Model parameters are implemented to account for vaccination rates, SARS-CoV-2 variants, and reinfections. It is important to note that these parameter values change over time due to the inherent evolution of the pandemic and modifications in the non-pharmaceutical interventions (NPIs) applied to the population, as well as the distribution of vaccines.

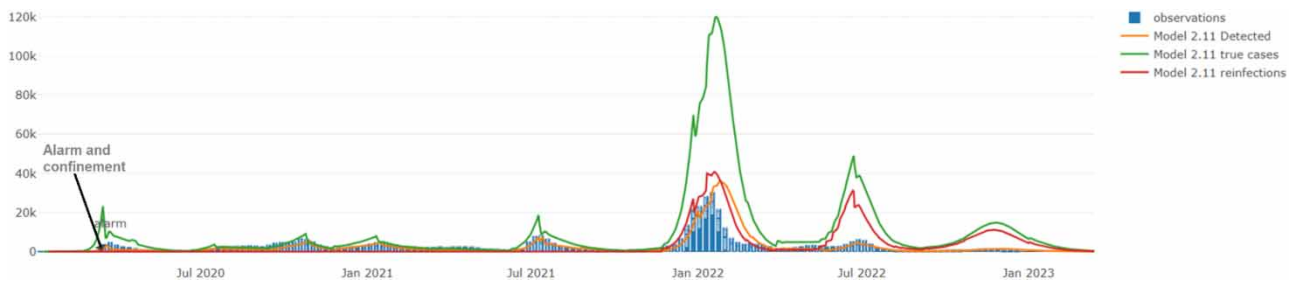

Figure 4 | The model forecast suggests that the stabilization of the pandemic during the 2023. On the *y*-axis are presented the number of cases.

Starting from the previous study (Fonseca i Casas *et al.* 2023) we need to calculate the new percentage of under-reporting due to the changes in testing protocols in Catalonia. Initially, from the beginning of the pandemic until March 28, the under-reporting rate for Catalonia was estimated to be around 60% based on a prevalence study published on several rounds (Pollán *et al.* 2020; Ministerio de Sanidad 2021a, 2021b, 2021c, 2021d). This value was derived from an assumed asymptomatic infection rate of approximately 40% (Syangtan *et al.* 2020; Ma *et al.* 2021), and around 30% for the Omicron variant (Fonseca i Casas *et al.* 2023). However, since March 28, the under-reporting rate increased due to a reduction in the number of tests conducted to assess the spread of infection across the population. To estimate the under-reporting rate, we utilized information from ASPCAT and made two assumptions.

Firstly, we assumed that the percentage of people infected by age range remains similar in the future (post-March 28, 2022) as it was in the past (during previous Omicron waves). To work with this assumption, we calculated the detection rate percentage for each age range until March 28. Then, we compared this percentage with the new percentage for each age range before March 28, assuming that the differences in detection rates reflect under-reporting. Specifically, we started the calculation using data from the Omicron wave, which showed a detection rate of approximately 30% (as mentioned in the previous study (Fonseca i Casas *et al.* 2023). However, after March 28, the detection measures changed, resulting in only severe symptomatic cases being detected. Therefore, as a second assumption, we considered that the detection rate for people over 70 years old remains unchanged. This assumption is valid since the testing approach for older people remained consistent, and the symptomatology is typically more severe in this age group, simplifying the detection process. It is important to note that since severe cases are more common in older people, the relative weight of cases in this age group increases, leading to changes in the detection rate by age range, as depicted in Figure 1. By applying these calculations, we can estimate the actual cases for the population older than 70 years old, who are more likely to be tested and reported. Then, we use the proportion of different age groups in the population to determine the number of cases that should be added to the other age groups, representing the actual cases that may be asymptomatic or untested.

Table 1 presents the calculation of under-reporting for different age ranges. The columns PW1 and PW2 represent the respective weight of detected cases for each age range. Taking into account the aforementioned assumptions, PW1/PW2 indicates the variation in the detection of each age range due to changes in monitoring methods for disease spread. Using these values, we can calculate the expected detected cases if monitoring continues for each age range, with the age range of 70+ considered invariant. It is important to note that these expected detected cases (approximately 732,890) represent the cases that would be detected from March 28 onwards, but they do not reflect the actual number of cases since the detected cases only account for about 30% of the total. To calculate the new under-reporting rate, we compare the detected cases (476,141) to the expected detected cases and determine that it represents approximately 65% of the expected value. Considering that the expected detected cases correspond to 30% of the actual cases and the detected cases now represent 65% (35% of under-reporting), we can infer that the detection rate from March 28 onwards is approximately 19%.

Table 1 | Calculus of the new under-reporting due to the relaxation of the testing applied to the population

		From 1 November 2021 to 28 March 2022		From 28 March (as previous)				From 28 March	
Age	Population rate [1]	Cases [2]	PW1[3]	Cases [2]	PW2[3]	PW2/PW1	% under-reporting [4]	Expected detected cases[5]	% under-reporting [6]
0–4	4.1	57,949.0	4.0	9,742.0	2.05	0.50	87.4	18,250.1	46.6
5–14	10.5	196,806.0	13.3	16,146.0	3.39	0.24	93.2	31,294.7	48.4
14–44	37.1	657,482.0	46.3	145,466.0	30.55	0.66	83.4	266,700.3	45.5
45–59	23.6	319,222.0	22.9	113,932.0	23.93	1.06	73.3	197,248.7	42.2
60–69	11.7	87,630.0	6.0	65,960.0	13.85	2.23	43.3	94,538.9	30.2
70–79	8.2	52,102.0	3.9	69,268.0	14.55	3.95	0.0 [1]	69,200.9	−0.1
80	5.8	41,905.0	2.7	55,657.0	11.69	3.94	0.0 [1]	55,657.0	0.0
Total	100%	1.413.096.0	100.0	476,171.0	100.00			732,890.5	35.0

(1) Data from https://www.idescat.cat/pub/?id=ep; (2) data from https://sivic.salut.gencat.cat/covid; (3) the percent of cases PW1 and PW2, detected by age range, changes because only the most serious cases are detected by the health system from 28 of March. We calculate the (4) and (5) the expected detected cases. Finally, we can calculate the increase on the under-reporting (6) in each age range. Then we can calculate an under-reporting increase which is about 35%.

Another factor considered in the model is reinfections caused by new variants. At the time of writing, we have been affected by several lineages, including the Original, Alfa, Delta, and Omicron lineages, as well as the BA.5, BA.4, BA2.75 lineages, and XBB.1. In our model, we account for the possibility of reinfections by assuming that a certain percentage of individuals who have recovered from the virus can become susceptible again. We continuously validate our model using wastewater data and detected cases reported by the Health Service to adjust this percentage accordingly. Reinfections have an impact on the intensity of infection waves, and the validation process allows us to estimate the percentage of people who become infected by different variants of the virus. Table 2 provides a summary of the reinfection percentages that we incorporate into the model to capture the effects of various variants.

It is important to note that the emergence of variants does not occur abruptly; rather, their presence in the population increases gradually as the variant spreads. In our model, we capture this increase by defining that reinfection does not happen suddenly, but rather occurs gradually on a weekly basis. Each week, we generate an event (SDL SIGNAL) that represents the percentage of individuals who have recovered and become susceptible again. For the latest variant (XBB.1), the percentage is so small that we use a single event to represent reinfection. This approach allows us to simulate the progressive impact of variants on the population over time.

2.2. Digital shadow, wastewater treatment plant data

The Catalan Surveillance Network of SARS-CoV-2 in Sewage (SARSAIGUA) was established through the collaboration and funding of the Catalan Water Agency (ACA) and the Public Health Agency of Catalonia (ASPCAT). This network commenced its monitoring activities in July 2020 and encompassed 56 wastewater treatment plants (WWTPs) across Catalonia, covering approximately 80% of the total population of 7.5 million inhabitants (7). The initiation of sample collection and analysis took place on the 6 of July 2020, around four months after the first clinical case of COVID-19 in Catalonia was detected on the 25 of February 2020. The selected 56 WWTPs were evenly distributed throughout the region, with at least one WWTP per county, ensuring comprehensive coverage.

Sampling frequency was determined as one sample per week for 36 of the 56 selected WWTPs, while the remaining 18 were sampled biweekly, resulting in a total of 45 samples collected and analyzed per week. Notably, certain WWTPs were only monitored during the summer season to enhance surveillance in municipalities with high tourism, such as Castell-Platja d'Aro and Vilaseca-Salou. Most WWTPs collected flow-based composite samples at the entrance, which were then refrigerated at 4 °C. The 45 weekly samples were distributed among three reference laboratories with expertise in molecular diagnosis and environmental virology. Each laboratory analyzed 15 samples per week to quantify the abundance of SARS-CoV-2 genomes using optimized protocols. The analysis employed reverse transcription quantitative polymerase chain reaction (RT-qPCR) targeting a common genetic marker (N1) as well as two complementary targets, N2 and IP4 (CDC 2020; Pasteur Institute 2020). Over the course of 20 months of monitoring (from July 2020 to February 2023), SARSAIGUA analyzed approximately 3,600 samples.

For the purpose of this study, a subset of 18 WWTPs was selected to demonstrate the integration of sewage data, specifically the N1 target concentrations, into an epidemiological model that forms part of the DT representing the evolution of the pandemic in Catalonia. The selection of WWTPs was based on the following criteria: (i) inclusion of at least one WWTP per sanitary region (Catalonia has seven regions), (ii) inclusion of WWTPs with weekly monitoring frequency, and (iii) coverage of a wide range of connected populations, ranging from 2,935 to 1.4 million inhabitants. The variables obtained from the sewage data are continuous, and no significant outliers were detected in the dataset.

Table 2 | Percent of reinfection from each one of the different majority variants detected on the population

Variant	Percent of reinfection	Date
Omicron	100%	22/11/2021
BA.5	21.46%	1/05/2022–19/06/2022
BA.4	20% (increase 5% per week)	29/8/2022–19/09/2022
BA2.75	10.8% (increase 0.9% per week)	29/08/2022–14/11/2022
XBB.1	3.6%	15/01/2023

The data used in this study were sourced directly from the SARSAIGUA public repository (Corominas *et al.* 2022) which provides regular updates every Friday. The repository can be accessed through GitHub at the following URL: https://github.com/icra/sars-aigues/. It contains comprehensive data on the N1 gene copies for all collected samples, along with information on influent flows from the WWTPs. Researchers were able to retrieve the necessary data from this repository for analysis and integration into the study's methodology. The file that contains the genetic information is release.csv (https://github.com/icra/sars-aigues/blob/main/release.csv).

Analyzing the information contained on Table 3, we can see from Figure 5, that there is a huge difference between the population served by the different WWTP, being the two largest WWTP serving more than 70% of the population. The Spearman correlations between the N1 gene signal in wastewater and diagnosed COVID-19 cases, excluding the under-reporting period after April 2022, consistently fall within the favorable range of 0.5–0.8. These correlation values align well with those documented in the literature, indicating the suitability of the WWTPs under consideration for effective wastewater-based health surveillance.

2.3. Validation, time series analysis

In this study, a time series analysis was conducted as part of the validation process to assess the model's ability to capture the dynamics of the epidemic. The analysis aimed to compare the simulation traces (synthetic data) with the Digital Shadows (sewage data) and determine how well the model aligns with the observed trends over time. To reduce the impact of noise and focus on detecting significant changes, a logarithmic transformation was applied to the time series of simulated true cases and N1 concentrations. The validation process focused on a three-month window, which was deemed appropriate for evaluating the model's performance and forecasting capabilities over a few months. The first step of the time series analysis involved examining potential delays between the two time series. Cross-correlation analysis, utilizing the '*ccf*' function in R, was employed to detect any lag between the model outcomes and sewage outcomes. After analyzing various lags, it was found that the model performed well with a lag of zero, indicating that the model predicts the number of true cases without delay compared to sewage data. This lag information was then utilized to validate the key performance indicators (KPIs) obtained from the sewage data. Considering that viral shedding can occur before symptoms appear and continue for up to 10 days after symptoms start (Lavania *et al.* 2022), it is reasonable to expect that the dynamics of true cases and sewage

Table 3 | WWTPs selected for the validation. Rho indicates level of correlation between N1 and diagnosed cases

Code	WWTP	County	Province	Sanitary regions	Population	Correlation (Rho)
DPUI	PUIGCERDÀ	Cerdanya	Girona	Alt Pirineu i Aran	14.753	0.752
DBSS	BESÒS	Barcelonès	Barcelona	Barcelona	1.444.884	0.534
DGVC	GAVÀ/VILADECANS	Baix Llobregat	Barcelona	Barcelona	172.208	0.666
DPDL	PRAT DE LLOBREGAT, EL	Baix Llobregat	Barcelona	Barcelona	1.092.573	0.753
DFAL	FALSET	Priorat	Tarragona	Camp de Tarragona	2.935	0.658
DRUS	REUS	Baix Camp	Tarragona	Camp de Tarragona	110.574	0.751
DTAR	TARRAGONA	Tarragonès	Tarragona	Camp de Tarragona	98.311	0.678
DIGU	IGUALADA	Anoia	Barcelona	Catalunya Central	61.375	0.579
DMAS	MANRESA	Bages	Barcelona	Catalunya Central	86.721	0.543
DSOL	SOLSONA	Solsonès	Lleida	Catalunya Central	10.600	0.680
DBAY	BANYOLES	Pla de l'Estany	Girona	Girona	28.464	0.744
DGIR	GIRONA	Gironès	Girona	Girona	145.373	0.674
DPAM	PALAMÓS	Baix Empordà	Girona	Girona	51.673	0.761
DBAL	BALAGUER	Noguera	Lleida	Lleida	15.891	0.554
DLLE	LLEIDA	Segrià	Lleida	Lleida	117.673	0.510
DMOF	MONTFERRER	Alt Urgell	Lleida	Lleida	12.693	0.782
DAMP	AMPOSTA	Montsià	Tarragona	Terres de l'Ebre	21.083	0.748
DTOT	TORTOSA-ROQUETES	Baix Ebre	Tarragona	Terres de l'Ebre	39.332	0.640

Mean	195950,88
Standard Error	93834,49
Median	56524
Standard Deviation	398106,03
Sample Variance	1,58E+11
Kurtosis	6,69
Skewness	2,74
Range	1441949
Minimum	2935
Maximum	1444884
Sum	3527116
Count	18
Largest(1)	1444884
Smallest(1)	2935
Confidence Level (95,0%)	197973,47

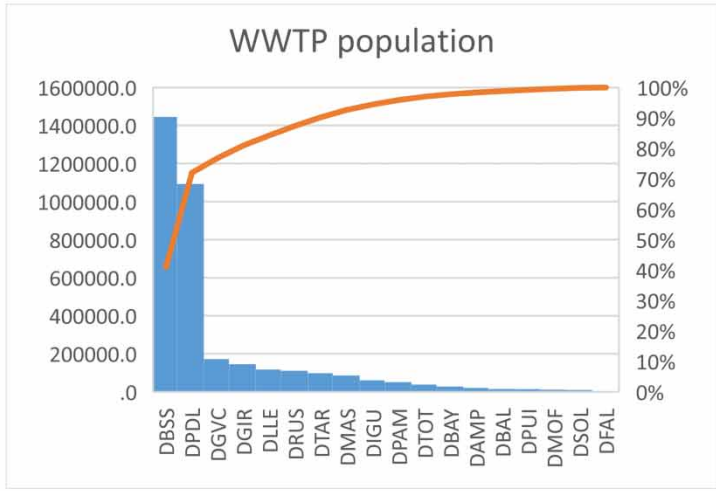

Figure 5 | Descriptive analysis of the population server by the different WWTP used for the validation in the project.

data would peak simultaneously. However, as shedding may persist beyond 10 days, the decrease in true cases after the peak is expected to be more rapid than in sewage data.

To analyze the correlation between true cases and sewage N1 data from each WWTP, Spearman correlation analysis was applied. Spearman correlation is a non-parametric test that measures the strength and direction of association between two ranked variables and is not strongly influenced by differences in variable scales. Finally, to obtain a single value indicating the validity of the model, an aggregated value was computed by weighting all WWTP correlation values based on the served population. As a convention, if this aggregated value exceeds 0.5, the model is considered to be valid. The statistical analysis, including the Spearman test and computation of aggregated values, was performed using R, utilizing the '*cor.test*' function for the Spearman test.

In relation to the other Digital Shadows, it is important to note that even as the frequency of clinical testing decreases over time, these data still hold relevance for model validation. The validation process does not solely rely on a single dataset but instead encompasses a combination of factors, including the wastewater system, detected cases, hospitalization rates, and death rates. However, the N1 concentration observed in wastewater remains a particularly valuable and informative source of data for generating the Digital Shadows necessary to validate the model. This concentration serves as a rich indicator of the viral presence in the population and allows us to assess the consistency and accuracy of the model's predictions. By incorporating multiple data sources, including wastewater analysis, we can enhance the robustness and reliability of the model validation process because we can test several assumptions that rely on several data sources. This multi-dimensional approach provides a comprehensive understanding of the pandemic dynamics and strengthens our ability to make informed decisions based on the DT and its corresponding shadows.

3. RESULTS

The DT model as illustrated in Figure 4, accurately describes the evolution of the different waves of the pandemic in Catalonia. The SDL simulation model produces three curves that illustrate the progression of cases (refer to Figure 4). The first curve represents the detected cases, which should align with the Digital Shadows provided by health authorities. The second curve represents the real cases, accounting for under-reporting, including asymptomatic cases prevalent in the population. The third curve shows reinfections occurring in the population, considering factors such as decreased protection and the emergence of new variants. More details on this model, and how it manages reinfections or other elements (like vaccination status, variants, etc.) can be found in reference (Fonseca i Casas *et al.* 2023).

The true cases significantly outnumber the detected cases. The detected cases from the model closely match the observed data (depicted in blue), with a coefficient of correlation above 0.5. The red line represents reinfections resulting from waning

immunity and the presence of new variants. The peaks of the reinfections and true cases curves occur earlier than those of the detected cases curve, as it takes more time for detection to increase due to the need for symptom development and testing.

The website accompanying the model (https://pand.sdlps.com/) displays a time series of infection dynamics, including key indicators such as the number of real cases, reinfections, and detected cases. The website provides a user-friendly and clear presentation of this information. It is regularly updated with new data from various sources (Digital Shadows), offering valuable insights into the pandemic's evolution and the accuracy of the model's forecasts. To ensure ongoing accuracy, the website also features panels related to the continuous validation process. The validation process is crucial as it allows for the identification and correction of any errors or deviations in the model assumptions. Establishing an infrastructure that facilitates this process and enables prompt error detection and resolution is essential. So, we use wastewater data as a proxy for true infection rates. The next section shows the validation results with these data.

3.1. Continuous validation process

One of the main outcomes presented in this paper is the continuous validation process developed to monitor and verify the accuracy of the DT using wastewater data. This process is presented on a panel called WWTP presented on the website for easy validation, which provides key information obtained from the WWTPs. The panel displays the viral load measured in the wastewater and compares it with the model-estimated true cases. This allows for the evaluation of the consistency and reliability of the model's predictions and facilitates the detection of any discrepancies or anomalies in the infection dynamics. On the WWTP panel, the time series for the last 3 months is displayed for visual inspection to assess the correlation between the two time series. The choreograph is also shown to identify any potential lag between the time series. This enables easy updating of the DT as needed.

The model validation is performed on a weekly basis, reviewing the correlation indices and individual values for each WWTP. The simulator generates a panel using R scripts, which includes various KPIs and charts to represent all the necessary information for ensuring the validity of the model. The panel consists of three main areas: the time series displaying the data from the DT and the Digital Shadow, along with a correlogram to detect any lag in the time series. Although these panels are primarily intended for validation purposes, they can also be used to understand the level of confidence in the Digital Twin. The complete validation panel for December 27, 2022, can be seen at (51), and the code for the panel can be downloaded from https://github.com/PolyhedraTech/SDL-PAND. Following this process, the information from all the WWTPs is aggregated into a single measurement. This aggregation is achieved by weighting the data based on the population served by each WWTP. The aim is to create a consolidated set of KPIs that different stakeholders can utilize to make informed decisions. Figure 6 shows the panel for model validation on December 26. The value of 0.63, which is above 0.5, indicates that the discrepancy between the model (Digital Twin) and the data (Digital Shadow) is sufficiently close to accept the forecast.

Figure 6 | Validation of the model. Here, we show an overall value of 0.63, that being over 0.5 (our threshold) assumes that the model is yet valid for 26 December 2022.

If the aggregated value calculated from all the individual WWTP values considering the population, goes below 0.5 in the validation panel or if a visual inspection suggests that some WWTPs are not accurately represented, a revision is triggered to review the model assumptions. This revision can be conducted by analyzing the graphical representation of the model or examining the configuration files, if no modifications to the model structure are necessary (which is typically the case). This revision process allows us to infer the reasons behind the model's failure and enhance our understanding of how the system, in this case, how spread of SARS-CoV-2 in Catalonia, operates. It provides valuable insights into the actions that need to be taken to control the spread of the pandemic.

Additionally, the panel displays the correlations graphically for each WWTP and includes a correlogram, as shown in Figure 7. The correlogram allows for the comparison between the model's real cases and the wastewater data (N1) for two of the largest WWTP installations in Catalonia. This visual representation helps identify any discrepancies in the trend between the Digital Shadow (N1) and the synthetic data (real cases). It's worth noting that the logarithmic transformation used in the analysis simplifies the comparison of the two time series.

The identification of failures in the correlations is straightforward. As an example, Figure 8 depicts a WWTP that exhibits a failure. Both the visual inspection and the correlogram confirm this discrepancy. In this particular WWTP, the correlation value is −0.34, which falls below the acceptance threshold, indicating that the trends are opposite and therefore the assumptions must be considered. It is important to note that failures in the correlation between the Digital Shadow and the DT are more common in smaller WWTPs. Sometimes, these failures in the correlations for specific WWTP do not necessarily indicate a failure in the model forecast. The model aims to capture the overall trend of the infection rates in Catalonia, not in each WWTP. Some WWTPs serve small populations and may report values that are not representative of the whole region. This suggests that the implementation of policies should be tailored to meet the needs of each region, taking into account factors such as the size of the cities and the geographical distribution of the WWTP.

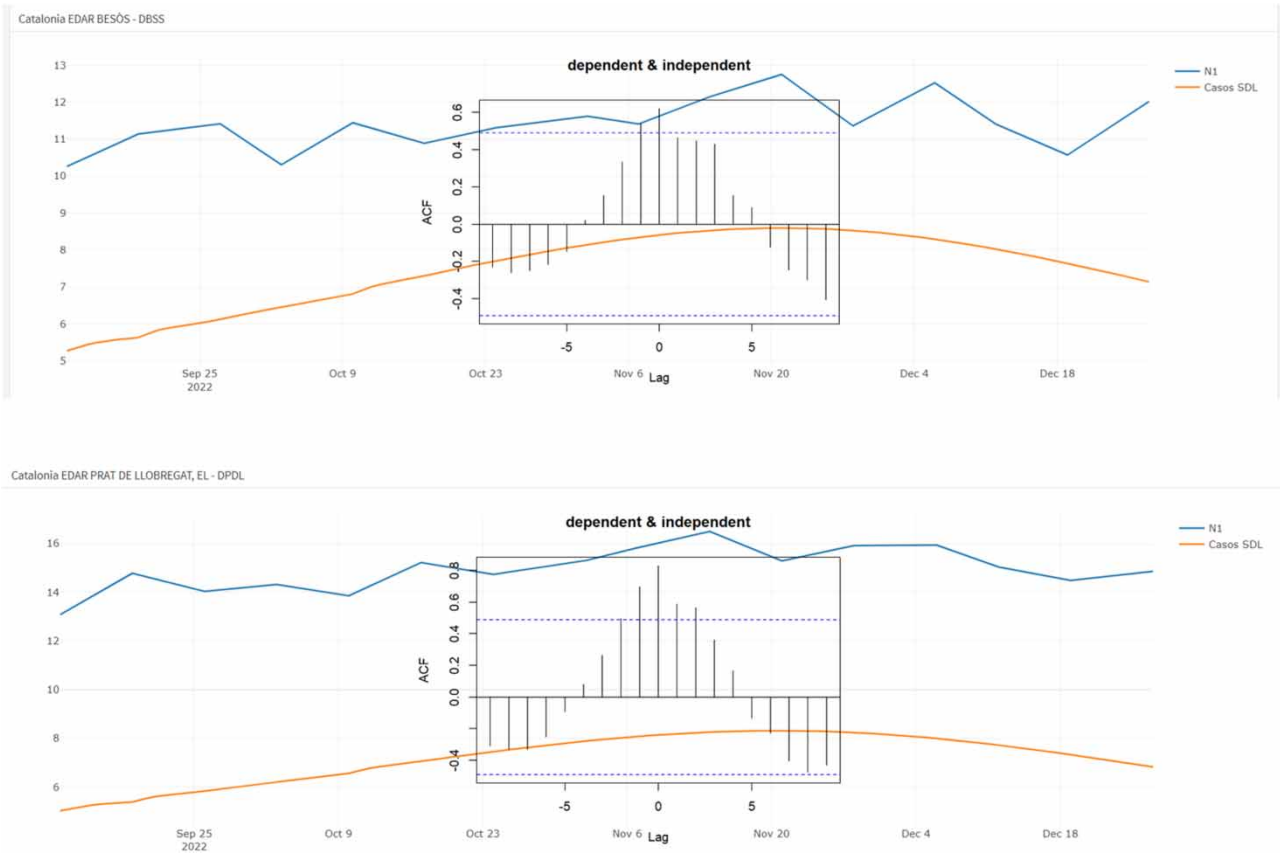

Figure 7 | Comparison of the time series for the Digital Shadow and the Synthetic data obtained from the model for the BESÒS WWTP (top) and Llobregat (bottom), two of the biggest installations of Catalonia, see (SDL-PAND 2022). The *y*-axis represents the log of the viral concentrations and the number of real cases from the model.

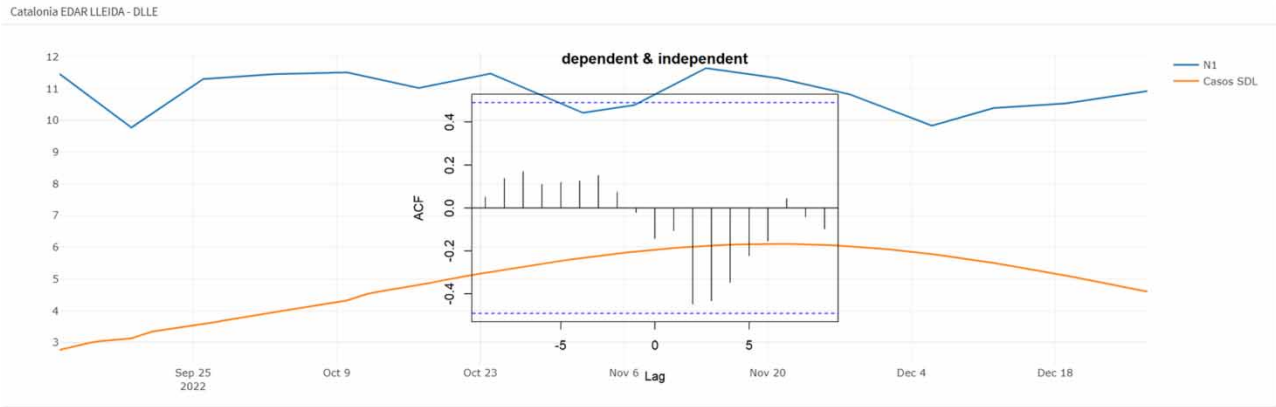

Figure 8 | Comparison of the time series for the Digital Shadow and the Synthetic data obtained from the model for the LLEIDA WWTP with the correlogram. Notice the clear discrepancy between the model and the data, therefore, the model forecast, based on this WWTP data, is suggested that the model forecast is not credible, refer SDL-PAND (2022).

3.2. Results from the continuous validation

The analysis demonstrates a strong correlation between N1 concentrations in wastewater and the number of true COVID-19 cases predicted by the model. The absence of a lag in the correlation, see Figure 7, suggests that the model accurately forecasts the future trend of the time series. Shedding of SARS-CoV-2 in feces can persist for approximately a month after infection (as indicated in reference Lavania *et al.* 2022), which implies that the viral load in wastewater may decrease at a slower rate than the reduction in cases within the population. However, this effect does not seem to impact the robustness of our method, even in the presence of different variants. This can be seen on Figure 9 that presents the historical results obtained using this method, specifically showcasing the correlations from June to October 2022. The top row displays the aggregated value for the correlations, as described earlier. The panel indicates periods where the model remains valid and no corrections are necessary. When the model becomes invalid, we review the model assumptions and calculate a new candidate transmission rate based on an optimization model and information about new variants. At this stage, only reinfections can drive a new increase in the number of cases.

As part of the continuous validation process, we have encountered some unusual outliers that require further investigation. These outliers represent cases where the viral load in wastewater and the true cases predicted by the model do not align, suggesting a possible error in the model assumptions or a change in the infection dynamics. It is essential to closely examine these outliers, understand their causes, and assess their implications for our model and forecasts. The annexes provide an example of this process, offering valuable insights into the overall behavior of the system.

4. DISCUSSION

Our study underscores the efficacy of a DT that integrates wastewater data and detected cases as Digital Shadows to generate long-term pandemic forecasts (more than two months). The forecast's reliability hinges on successful validation, ensuring alignment with the Digital Shadows. If discrepancies arise, (considering at this time only regional and no local discrepancies) we must revise the model assumptions to realign the generated data with the Digital Shadows. This methodology enables us to continually monitor and adjust our predictions accuracy. At the time of writing this paper, our forecast has remained valid for over three months without significant deviations from the Digital Shadows.

A key finding of our study is the synchronicity between the true cases estimated by the model and the viral load measured in wastewater. This synchronicity suggests a delay of less of one day between individual infections and virus detection in the sewage system. Thus, wastewater analysis provides a timely and accurate indicator of a given area's pandemic situation. The validation panel for 2022-12-27 (SDL-PAND 2022) displays the correlograms for different WWTPs, illustrating the close alignment between the viral load time series and the true cases predicted by the model. This lack of lag is attributed to our model's use of real cases rather than detected cases, which may be subject to reporting delays or testing errors. This real-time insight is invaluable for decision-making and enables the prompt detection of any changes or anomalies in the system.

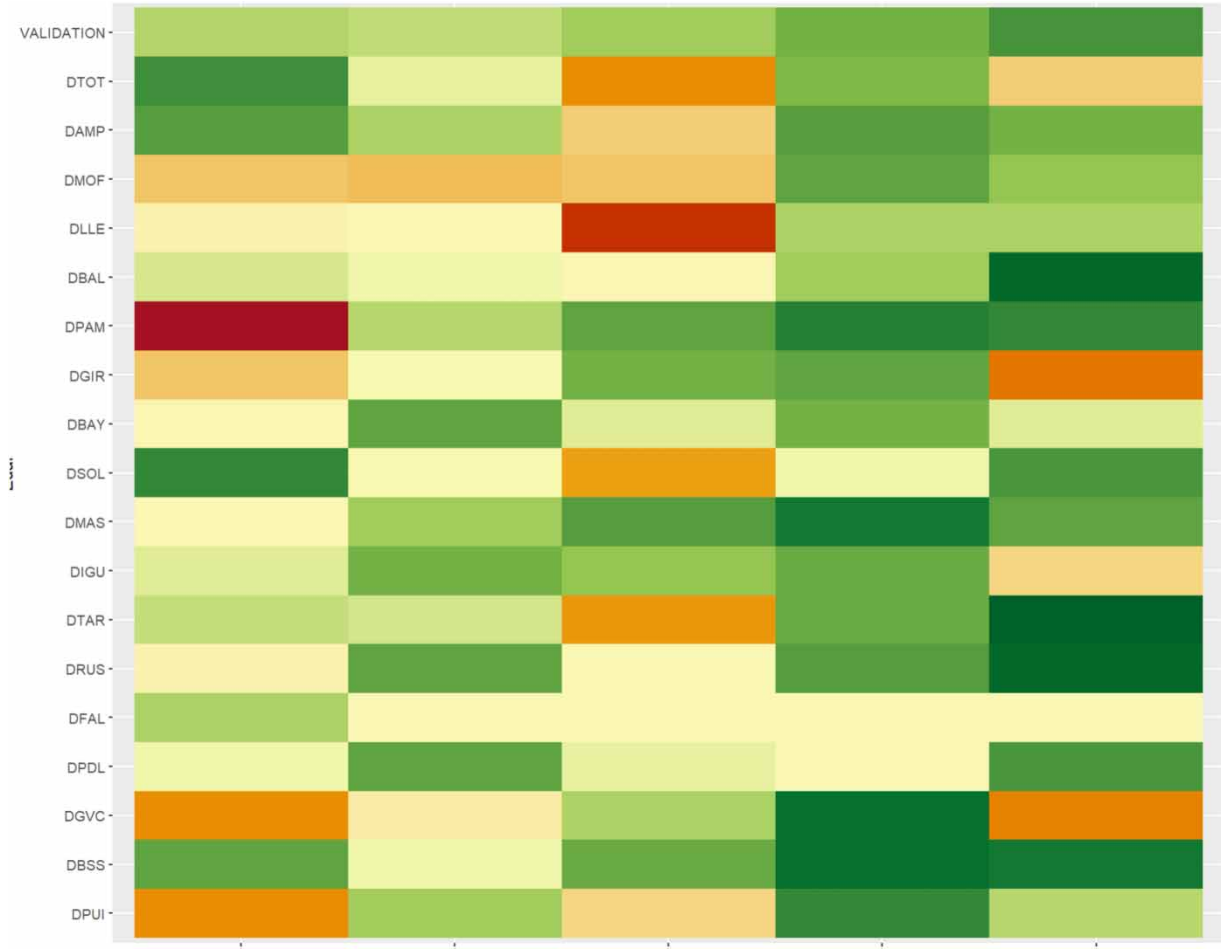

Figure 9 | History of the correlations for the WWTP from June to October of 2022. Green represents a correlation over 0.5 while other colors represent no correlation being red the worse situation. On the figure each column represents a month, from June to October.

The employed methodology aligns well with our analysis scope, focusing on the COVID-19 pandemic's evolution and maintaining a DT that accurately represents the pandemic's key indicators, including real cases, detected cases, and reinfections. The Spearman statistical analysis provides a clear view of each WWTP, with values ranging from 0 (no correlation) to 1 or −1 (perfect correlation). For analysis purposes, we set the threshold at 0.5, indicating significant differences between a specific WWTP's time series and suggesting validation failure.

The methodology outlined in our study holds potential for application across various regions, extending its utility beyond the current geographical scope. The first step in this process involves defining a model that accurately represents the real cases of infection within the target region. This model should be tailored to the specific characteristics of the region, taking into account factors such as population density, population behavior, healthcare infrastructure, and local pandemic response measures. Once a suitable model has been established, the next step is to obtain data from the WWTPs within the region. These data form the basis for the validation process, which ensures that the model's predictions align with the actual situation on the ground. It's important to note that the quality and reliability of the WWTP data can significantly impact the accuracy of the model's validation process. Therefore, establishing robust data collection and processing protocols is crucial. Upon successful validation, the model can then be used to monitor the evolution of the pandemic within the selected area. This involves regularly updating the model with the latest WWTP data and adjusting the model parameters as necessary to reflect changing conditions. In this way, the model serves as a dynamic tool for tracking the pandemic's progression, providing valuable insights that can inform public health decision-making.

Future research can be pursued along two primary avenues. The first involves determining the minimum population size that a WWTP must serve to effectively monitor the evolution of a pandemic or to function as an early warning system. This

analysis would entail a comprehensive study of various WWTPs serving different population sizes, examining the accuracy and timeliness of data derived from each. The goal would be to establish a population threshold below which the effectiveness of a WWTP as a monitoring tool significantly diminishes. The second line of inquiry involves the integration of additional data sources into the validation process. For instance, incorporating data on critical hospitalizations could provide a more nuanced understanding of the pandemic's severity at any given time. Similarly, data on secondary effects observed in the population, such as long-term health impacts or societal changes, could offer insights into the broader implications of the pandemic. These additional data sources could also enhance the model's predictive capabilities. By correlating these data with the real cases inferred from wastewater analysis, we could potentially reconstruct the historical trajectory of the pandemic. This retrospective analysis would not only validate the model's predictions but also contribute to a more comprehensive understanding of the long-term effects of the pandemic. These proposed lines of research underscore the potential of our approach to evolve and adapt in response to new data and insights, ultimately enhancing its utility as a tool for pandemic monitoring and management.

5. CONCLUSIONS

We demonstrated the efficacy of a DT that integrates wastewater data and detected cases as Digital Shadows to generate long-term COVID-19 pandemic forecasts.

We considered the potential impact of reinfections caused by the XBB.1 variant (Arora *et al.* 2023) and assumed that the reinfection rate would not exceed that observed for the BA.5 or BQ.1 variants. Based on this assumption, we projected Catalonia's future pandemic course, which indicated a stable trend with minor fluctuations, suggesting a stabilization of the pandemic situation.

Our analysis reveals that the COVID-19 case detection rate dropped to 19% from March 28, 2022 onwards, as the health system database only recorded cases with severe symptoms. Before that date, the detection rate was around 30% for Omicron waves and 40% for previous variants. From our analysis, we inferred that approximately 90% of Catalonia's population had been infected or reinfected with SARS-CoV-2. While this value seems high, it aligns with a study conducted in the USA in February 2022 (Clarke *et al.* 2022), nearly a year earlier. This finding is understandable considering Catalonia's dense population.

We acknowledge that both the DT and the Shadows, which indicate the N1 concentration in wastewater, perform better for larger populations. The effectiveness of this approach diminishes for smaller populations, an area we intend to investigate further.

Finally, we emphasize that the work presented in this paper is not merely an academic proposal but has been implemented as a product accessible through the website https://pand.sdlps.com. A permanent capture of the published website for January 2, 2023, is available (Fonseca i Casas & Garcia Subirana 2023).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Arora, P., Cossmann, A., Schulz, S. R., Ramos, G. M., Stankov, M. v, Jäck, H.-M., Behrens, G. M. N., Pöhlmann, S. & Hoffmann, M. 2023 Neutralisation sensitivity of the SARS-CoV-2 XBB.1 lineage. *The Lancet Infectious Diseases* 0(0). https://doi.org/10.1016/S1473-3099(22)00831-3.

Ayala-Aldana, N., Monleon-Getino, A. & Canela-Soler, J. 2022 Regresión lineal para sars-cov-2 en aguas residuales y la dinámica infecciosa de COVID-19 en el baix llobregat, España. *Ciencia Latina Revista Científica Multidisciplinar* **6** (6), 250–261. doi:10.37811/cl_rcm.v6i6.3486.

Carfì, A., Bernabei, R. & Landi, F. 2020 Persistent symptoms in patients after acute COVID-19. *JAMA* **324** (6), 603. doi:10.1001/jama.2020.12603.

CDC. 2020 CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel For Emergency Use Only Instructions for Use. Atlanta.

Chu, D. K., Akl, E. A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H. J., Chu, D. K., Akl, E. A., El-harakeh, A., Bognanni, A., Lotfi, T., Loeb, M., Hajizadeh, A., Bak, A., Izcovich, A., Cuello-Garcia, C. A., Chen, C., Harris, D. J., Borowiack, E. & … Schünemann, H. J. 2020 Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis. *The Lancet* **395** (10242), 1973–1987. doi:10.1016/S0140-6736(20)31142-9.

Clarke, K. E. N., Jones, J. M., Deng, Y., Nycz, E., Lee, A., Iachan, R., Gundlapalli, A. v., Hall, A. J. & MacNeil, A. 2022 Seroprevalence of infection-induced SARS-CoV-2 antibodies – United States, September 2021–February 2022. *MMWR. Morbidity and Mortality Weekly Report* **71** (17), 606–608. doi:10.15585/mmwr.mm7117e3.

Cooper, S., Tobar, A., Konen, O., Orenshtein, N., Kropach, N., Landau, Y., Mozer-Glassberg, Y., Bar-Lev, M. R., Shaoul, R., Shamir, R. & Waisbourd-Zinman, O. 2022 Long COVID-19 liver manifestation in children. *Journal of Pediatric Gastroenterology & Nutrition*. doi:10.1097/MPG.0000000000003521.

Corominas, L., Collado, N., Guerrero-Latorre, L., Abasolo-Zabalo, N., Anfruns-Estrada, E., Anzaldi-Varas, G., Bofill-Mas, S., Bosch, A., Bosch-Lladó, L., Caimari-Palou, A., Canela-Canela, N., Chavarria-Miró, G., Bas-Prior, J. M. del, Espiñeira-Robaina, Y., Forés-Gil, E., Fuentes, C., Rosina Gironés, S. G., Hundesa, A., Marta Itarte, R. M.-C. & … Borrego, C. 2022 Catalan Surveillance Network of SARS-CoV-2 in Sewage. Zenodo. doi:10.5281/zenodo.4554720.

COVID-19 takes serious toll on heart health – a full year after recovery | Science | AAAS. Available from: https://www.science.org/content/article/covid-19-takes-serious-toll-heart-health-full-year-after-recovery. (accessed 10 February 2022).

Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. 2023 Long COVID: Major findings, mechanisms and recommendations. *Nature Reviews Microbiology*. doi:10.1038/s41579-022-00846-2.

Fonseca i Casas, P. 2013 Co-simulation using specification and description language. In: *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in A Complex World*, pp. 4022–4023.

Fonseca i Casas, P. 2014 Using specification and description language to formalize multiagent systems. *Applied Artificial Intelligence* **28** (5), 504–531. doi:10.1080/08839514.2014.905820.

Fonseca i Casas, P., 2019 Towards a representation of cellular automaton using specification and description language. In: *System Analysis and Modeling. Languages, Methods, and Tools for Industry 4.0.* (Fonseca i Casas, P., Sancho, M.-R. & Sherratt, E., eds). Springer, Munich. pp. 163–179. doi:10.1007/978-3-030-30690-8_10.

Fonseca i Casas, P. 2023 A continuous process for validation, verification, and accreditation of simulation models. *Mathematics* **11** (4), 845. doi:10.3390/MATH11040845.

Fonseca i Casas, P. & Garcia Subirana, J. 2023 Screenschots for https://pand.sdlps.com for 2023/01/02. Figshare. Available from: https://figshare.com/articles/figure/Screenschots_for_https_pand_sdlps_com_for_strong_2023_01_02_strong_/23536509 (accessed 17 June 2023).

Fonseca i Casas, P., Colls, M. & Casanovas, J. 2010 Towards a representation of environmental models using specification and description language – From the Fibonacci model to a wildfire model. In: *KEOD 2010 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pp. 343–346.

Fonseca i Casas, P., Garcia I Subirana, J., García I Carrasco, V. & Pi I Palomés, X. 2021 SARS-CoV-2 spread forecast dynamic model validation through digital twin approach, Catalonia case study. *Mathematics* **9** (14), 1–17. doi:10.3390/math9141660.

Fonseca i Casas, P., Garcia i Subirana, J. & Garcia i Carrasco, V. 2023 Modeling SARS-CoV-2 true infections in Catalonia through a digital twin. *Advanced Theory and Simulations* **6**, 2200917. doi:10.1002/ADTS.202200917.

Forrester, J. W. 1988 *Principles of Systems*. Productivity Pr, Michigan.

Fortmann-Roe, S. 2014 Insight maker: A general-purpose tool for web-based modeling & simulation. *Simulation Modelling Practice and Theory* **47**, 28–45. doi:10.1016/j.simpat.2014.03.013.

Guerrero-Latorre, L., Collado, N., Abasolo, N., Anzaldi, G., Bofill-Mas, S., Bosch, A., Bosch, L., Busquets, S., Caimari, A., Canela, N., Carcereny, A., Chacón, C., Ciruela, P., Corbella, I., Domingo, X., Escoté, X., Espiñeira, Y., Forés, E., Gandullo-Sarró, I., Garcia-Pedemonte, D., Girones, R., Guix, S., Hundesa, A., Itarte, M., Mariné-Casadó, R., Anna Martínez, A., Martínez-Puchol, S., Mas-Capdevila, A., Mejías-Molina, C., Moliner i Rafa, M., Munné, A., Pintó, R. M., Pueyo-Ros, J., Robusté-Cartró, J., Rusiñol, M., Sanfeliu, R., Teichenné, J., Torrell, H., Corominas, L. & Borrego, C. M. 2022 The Catalan surveillance network of SARS-CoV-2 in sewage: Design, implementation, and performance. *Scientific Reports* **12** (1), 1–10. doi:10.1038/s41598-022-20957-3.

ITU-T. 2019 ITU-T 2019 Specification and Description Language – Overview of SDL- 2010. ITU-T Recommendation Z.100.

Joseph-Duran, B., Serra-Compte, A., Sàrrias, M., Gonzalez, S., López, D., Prats, C., Català, M., Alvarez-Lacalle, E., Alonso, S. & Arnaldos, M. 2022 Assessing wastewater-based epidemiology for the prediction of SARS-CoV-2 incidence in Catalonia. *Scientific Reports* **12** (1), 1–11. doi:10.1038/s41598-022-18518-9.

Landry, M., Bornstein, S., Nagaraj, N., Sardon, G. A., Castel, A., Vyas, A., McDonnell, K., Agneshwar, M., Wilkinson, A. & Goldman, L. 2023 Postacute sequelae of SARS-CoV-2 in university setting. *Emerging Infectious Diseases Journal – CDC* **29** (3), 519–527. doi:10.3201/EID2903.221522.

Lavania, M., Joshi, M. S., Ranshing, S. S., Potdar, V. A., Shinde, M., Chavan, N., Jadhav, S. M., Sarkale, P., Mohandas, S., Sawant, P. M., Tikute, S., Padbidri, V., Patwardhan, S. & Kate, R. 2022 Prolonged shedding of SARS-CoV-2 in feces of COVID-19 positive patients: Trends in genomic variation in first and second wave. *Frontiers in Medicine* **9**, 436. doi:10.3389/fmed.2022.835168.

Lindan, C. E., Mankad, K., Ram, D., Kociolek, L. K., Silvera, V. M., Boddaert, N., Stivaros, S. M., Palasis, S., Akhtar, S., Alden, D., Amonkar, S., Aouad, P., Aubart, M., Bacalla, J. A., Barbosa, A. A., Basmaci, R., Berteloot, L., Blauwblomme, T., Brun, G. & … Vézina, G. 2021

Neuroimaging manifestations in children with SARS-CoV-2 infection: A multinational, multicentre collaborative study. *The Lancet Child and Adolescent Health* **5** (3), 167–177. doi:10.1016/S2352-4642(20)30362-X.

Ma, Q., Liu, J., Liu, Q., Kang, L., Liu, R., Jing, W., Wu, Y. & Liu, M. 2021 Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis. *JAMA Network Open* **4** (12), 2137257. doi:10.1001/jamanetworkopen.2021.37257.

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. *The Lancet Planetary Health* **5** ((12), e874–e881. doi:10.1016/S2542-5196(21)00230-8.

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. 2020 Presence of SARS-coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in The Netherlands. *Environmental Science and Technology Letters* **7** (7), 511–516. doi:10.1021/ACS.ESTLETT.0C00357/ASSET/IMAGES/LARGE/EZ0C00357_0002.JPEG.

Mehandru, S. & Merad, M. 2022 Pathological sequelae of long-haul COVID. *Nature Immunology* **23** (2), 194–202. doi:10.1038/S41590-021-01104-Y.

Ministerio de Sanidad. 2021a ESTUDIO ENE-COVID: CUARTA RONDA. ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA. Available from https://www.mscbs.gob.es/gabinetePrensa/notaPrensa/pdf/15.12151220163348113.pdf (accessed 14 July 2021).

Ministerio de Sanidad. 2021b ESTUDIO ENE-COVID: INFORME FINAL. ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA. Available from: https://www.mscbs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_INFORME_FINAL.pdf (accessed 14 July 2021).

Ministerio de Sanidad 2021c ESTUDIO ENE-COVID19: SEGUNDA RONDA. ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA. Available from: https://www.mscbs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_SEGUNDA_RONDA_INFORME_PRELIMINAR.pdf (accessed 14 July 2021).

Ministerio de Sanidad 2021d ESTUDIO ENE-COVID19: PRIMERA RONDA. ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA. Available from: https://www.mscbs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_PRIMERA_RONDA_INFORME_PRELIMINAR.pdf (accessed 14 July 2021).

Mizrahi, B., Sudry, T., Flaks-Manov, N., Yehezkelli, Y., Kalkstein, N., Akiva, P., Ekka-Zohar, A., ben David, S. S., Lerner, U., Bivas-Benita, M. & Greenfeld, S. 2023 Long COVID outcomes at one year after mild SARS-CoV-2 infection: Nationwide cohort study. *BMJ (Clinical Research ed.)* **380**, e072529. doi:10.1136/bmj-2022-072529.

Munblit, D., Simpson, F., Mabbitt, J., Dunn-Galvin, A., Semple, C. & Warner, J. O. 2022 Legacy of COVID-19 infection in children: Long-COVID will have a lifelong health/economic impact. *Archives of Disease in Childhood* **107** (3), E2. doi:10.1136/archdischild-2021-321882.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S. M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A. R. & Linton, N. M. 2020 Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases* **94**, 154–155. doi:10.1016/j.ijid.2020.03.020.

Nordvig, A. S., Fong, K. T., Willey, J. Z., Thakur, K. T., Boehme, A. K., Vargas, W. S., Smith, C. J. & Elkind, M. S. V. 2021 Potential neurologic manifestations of COVID-19. *Neurology: Clinical Practice* **11** (2), e135–e146. doi:10.1212/CPJ.0000000000000897.

Olmos, J. L., Fonseca i Casas, P. & Rebull, J. O. 2014 *Modeling a Chilean Hospital Using Specification and Description Language*, Vol. 1. doi:10.4018/978-1-4666-6339-8.ch023.

Omori, R., Miura, F. & Kitajima, M. 2021 Age-dependent association between SARS-CoV-2 cases reported by passive surveillance and viral load in wastewater. *Science of the Total Environment* **792**, 148442. doi:10.1016/j.scitotenv.2021.148442.

Pasteur Institute, Paris 2020 Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2.

Phipps, S. J., Grafton, R. Q. & Kompas, T. 2020 Robust estimates of the true (population) infection rate for COVID-19: A backcasting approach. *Royal Society Open Science* **7** (11), 200909. doi:10.1098/rsos.200909.

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 Larrea, Fernández, Molina, N., Rodríguez-Cabrera, M., Martín, F., Merino-Amador, M., León Paniagua, P., Muñoz-Montalvo, J., Blanco, J. F., Yotti, F., Gutiérrez Fernández, R. & Vázquez de la Villa, R. A. 2020 Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A nationwide, population-based seroepidemiological study. *The Lancet* **396** (10250), 535–544. doi:10.1016/S0140-6736(20)31483-5.

Port Barcelona Estadístiques 2022 Estadístiques. Available from: https://www.portdebarcelona.cat/ca/web/autoritat-portuaria/estadisticas (accessed 26 October 2022).

Port de Barcelona 2020 Estadístiques de tràfic del Port de Barcelona. Dades acumulades desembre 2019.

Rippinger, C., Bicher, M., Urach, C., Brunmeir, D., Weibrecht, N., Zauner, G., Sroczynski, G., Jahn, B., Mühlberger, N., Siebert, U. & Popper, N. 2021 Evaluation of undetected cases during the COVID-19 epidemic in Austria. *BMC Infectious Diseases* **21** (1), 1–11. doi:10.1186/s12879-020-05737-6.

Sayampanathan, A. A., Heng, C. S., Pin, P. H., Pang, J., Leong, T. Y. & Lee, V. J. 2020 Infectivity of asymptomatic versus symptomatic COVID-19. *The Lancet* **6736** (20), 32651. doi:10.1016/S0140-6736(20)32651-9.

SDL-PAND Digital Twin validation panel for wastewater Digital Shadow data on 2022-12-27. Available from: https://figshare.com/s/3513906d70a092cc536f (accessed 4 March 2023).

Sherratt, E., Ober, I., Gaudin, E., Fonseca i Casas, P. & Kristoffersen, F. 2015 SDL – The IoT Language. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, Vol. 9369. Springer, Berlin, pp. 27–41. doi:10.1007/978-3-319-24912-4_3.

Stark, R., Kind, S. & Neumeyer, S. 2017 Innovations in digital modelling for next generation manufacturing system design. *CIRP Annals – Manufacturing Technology* **66** (1), 169–172. doi:10.1016/j.cirp.2017.04.045.

Syangtan, G., Bista, S., Dawadi, P., Rayamajhee, B., Shrestha, L. B., Tuladhar, R. & Joshi, D. R. 2020 Asymptomatic SARS-CoV-2 carriers: A systematic review and meta-analysis. *Frontiers in Public Health* **8**, 1–10. 2021. doi:10.3389/fpubh.2020.587374.

Syangtan, G., Bista, S., Dawadi, P., Rayamajhee, B., Shrestha, L. B., Tuladhar, R. & Joshi, D. R. 2020 Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* **11** (1). doi:10.1038/s41467-020-18272-4.

Taquet, M., Geddes, J. R., Husain, M., Luciano, S. & Harrison, P. J. 2021 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: A retrospective cohort study using electronic health records. *The Lancet. Psychiatry*. doi:10.1016/S2215-0366(21)00084-5.

Wąs, J., Sirakoulis, G. Ch. & Bandini, S. 2014 Cellular Automata. In: Wąs, J., Sirakoulis, G. Ch. & Bandini, S., Eds. Vol. 8751. Springer International Publishing. https://doi.org/10.1007/978-3-319-11520-7

First received 14 November 2023; accepted in revised form 16 January 2024. Available online 9 February 2024