Published in final edited form as:
Water Res. 2025 January 01; 268(Pt A): 122671. doi:10.1016/j.watres.2024.122671.

Epidemiological model can forecast COVID-19 outbreaks from wastewater-based surveillance in rural communities

Tyler Meadows[1,3], **Erik R. Coats**[2,3], **Solana Narum**[2,4], **Eva M. Top**[3,5,6], **Benjamin J. Ridenhour**[3,7], **Thibault Stalder**[3,5,6,8,*]

[1]Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada

[2]Department of Civil and Environmental Engineering, University of Idaho, Moscow, Idaho, USA

[3]Institute for Modeling Collaboration and Innovation (IMCI), University of Idaho, Moscow, Idaho, USA

[4]Bioinformatics and Computational Biology Graduate Program (BCB), Moscow, Idaho, USA

[5]Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA

[6]Institute for Interdisciplinary Data Sciences (IIDS), University of Idaho, Moscow, Idaho, USA

[7]Department of Mathematics and Statistical Science, University of Idaho, Moscow, Idaho, USA

[8]INSERM, CHU Limoges, RESINFIT, U1092, Univ. Limoges, F-87000, Limoges, France

Abstract

Wastewater has emerged as a crucial tool for infectious disease surveillance, offering a valuable means to bridge the equity gap between underserved communities and larger urban municipalities. However, using wastewater surveillance in a predictive manner remains a challenge. In this study, we tested if detecting SARS-CoV-2 in wastewater can forecast outbreaks in rural communities.

Under the CDC National Wastewater Surveillance program, we monitored the SARS-CoV-2 in the wastewater of five rural communities and a small city in Idaho (USA). We then used a particle filter method coupled with a stochastic susceptible-exposed-infectious-recovered (SEIR) model to infer active case numbers using quantities of SARS-CoV-2 in wastewater.

---

[*]**Corresponding Author:** thibault.stalder@unilim.fr, Full postal address: UMR 1092 RESINFIT, CBRS – 3ème étage, 1 rue du Professeur Bernard Descottes, 87025 Limoges, France.

Author contributions
B.J.R, E.R.C., E.M.T., T.M. and T.S. conceived the project; E.R.C. and T.S. designed the study; T.M. designed the math model; S.N. collected the samples and data; B.J.R, T.M. and T.S. performed data analysis; T.M. and T.S. wrote the paper. All authors helped revise the paper.

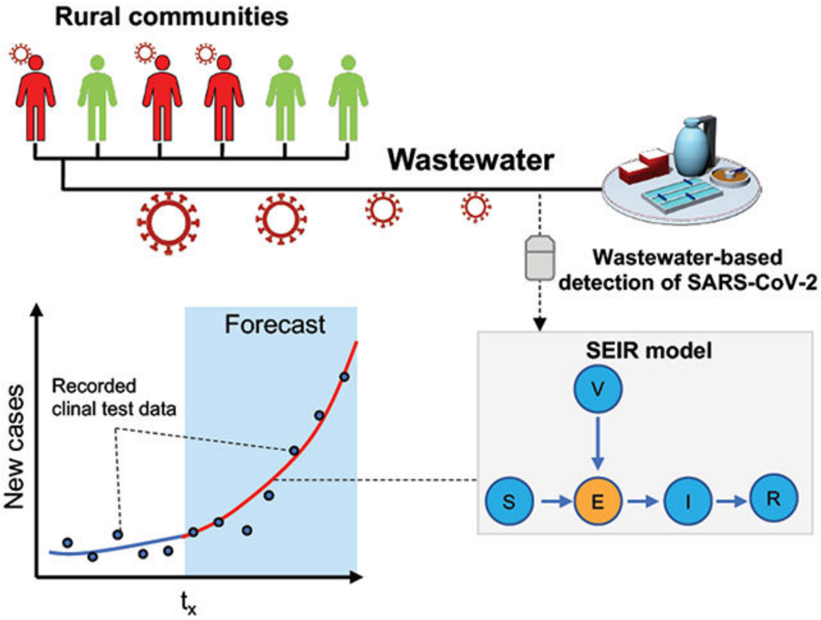
Conflict of Interest Statement
The authors do not declare any conflicts of interest.

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

Our findings revealed that while high daily variations in wastewater viral load made real-time interpretation difficult, the SEIR model successfully factored out this noise, enabling accurate forecasts of the Omicron outbreak in five of the six towns shortly after initial increases in SARS-CoV-2 concentrations were detected in wastewater. The model predicted outbreaks with a lead time of 0 to 11 days (average of 6 days +/− 4) before the surge in reported clinical cases.

This study not only underscores the viability of wastewater-based epidemiology (WBE) in rural communities—a demographic often overlooked in WBE research—but also demonstrates the potential of advanced epidemiological modeling to enhance the predictive power of wastewater data. Our work paves the way for more reliable and timely public health guidance, addressing a critical gap in the surveillance of infectious diseases in rural populations.

Graphical Abstract



At t_x, based on wastewater data, the SEIR model forecasted outbreaks with a lead time of 0 to 11 days (average of 6 days +/- 4) before the surge in reported clinical cases

Keywords

SARS-CoV-2; COVID-19; Wastewater-based epidemiology (WBE); SEIR; rural communities

Introduction.

Wastewater-based epidemiology (WBE) has rapidly emerged as a powerful tool for broad-scale surveillance of infectious diseases and antimicrobial resistance within communities. This approach is particularly effective because many pathogens, including SARS-CoV-2, poliovirus, RSV, and influenza are shed through body waste (Arts et al., 2023; Jeong et al., 2020; Parasa et al., 2020), and can be detected in wastewater (Corpuz et al., 2020; Shah et al., 2022). Unlike traditional surveillance methods, WBE can capture data from both symptomatic and asymptomatic individuals, thereby providing a more comprehensive picture of disease circulation. Importantly, WBE can improve and accelerate the early detection of infectious disease outbreaks, giving public health authorities critical time to respond. There have been successful instances where wastewater detection has triggered actionable public health responses such as the recent detection of poliovirus in the Gaza Strip (World Health Organization, n.d.). Despite such successes, significant barriers remain in translating wastewater data into actionable public health interventions, particularly in using this data predictively rather than retrospectively.

A modeling-based approach to WBE can overcome some of these barriers by providing a framework to interpret the spread and burden of infectious diseases. Such models can estimate key epidemiological parameters, including incidence (McManus et al., 2023), prevalence (Layton et al., 2022), and effective reproductive number (Huisman et al., 2022). Models, such as those based on machine learning or epidemiological models of the SEIR type, can forecast cases using wastewater data (Kanchan et al., 2024; McMahan et al., 2021). One of the primary advantages of mechanistic models is that they rely on a deep understanding of biological systems (Baker et al., 2018; Metzcar et al., 2024). Among previously published SEIR or SEIR-like models, various methodologies have been used to fit the data (Fazli et al., 2021; McMahan et al., 2021; Nourbakhsh et al., 2022; Pájaro et al., 2022; Phan et al., 2023; Polcz et al., 2023; Proverbio et al., 2022). Despite these advances, the application of WBE in small populations such as rural communities—a setting with distinct epidemiological and logistical challenges—remains underexplored.

Rural communities are particularly vulnerable to infectious diseases due to a combination of demographic factors (e.g., age), healthcare challenges (e.g., higher rates of obesity and smoking), limited resources (Cromartie et al., 2020; Cuadros et al., 2021; Kaufman et al., 2020; Lakhani et al., 2020), or reduced risk perception (Ridenhour et al., 2022). Additionally, reduced access to clinical testing in those areas can hinder timely disease detection and response. WBE offers a unique opportunity to support public health in rural areas, where traditional surveillance methods may be less effective. However, the vast majority of WBE studies have focused on urban areas and larger cities, leaving a critical gap in our understanding of how WBE can be applied to smaller, rural communities (Conway et al., 2023; D’Aoust et al., 2021; Holm et al., 2023; Jarvie et al., 2023; Toledo et al., 2022). By addressing this gap, WBE can serve as a tool for promoting health equity, ensuring that all communities – regardless of size – benefit from the latest advances in public health surveillance (Holm et al., 2023; Medina et al., 2022)

Importantly, while epidemiologists can examine wastewater data side-by-side with clinical testing to help understand what is happening, many infectious diseases are not reportable, or in the case of COVID-19, at-home self-testing replaced clinical testing (Park et al., 2023). For example, influenza and COVID-19 are not reportable diseases in Idaho (USA). Thus, epidemiologists are left with few tools to help them characterize epidemiological trends and forecasts; ultimately only wastewater data may be available to provide insight into disease burden in rural areas.

The COVID-19 pandemic has highlighted the value of WBE, particularly as a way to compensate for reduced clinical testing due to the rise of at-home tests and a general decline in clinical reporting (Fontenele et al., 2023). Trends of SARS-CoV-2 in wastewater have been shown to correlate with the number of COVID-19 cases and can even precede surveillance systems based on clinical samples (Fernandez-Cassi et al., 2021; Peccia et al., 2020; Wurtzer et al., 2020). This makes wastewater-based surveillance a promising early warning system for outbreaks (Bivins et al., 2020; Li et al., 2022; Shah et al., 2022; Weidhaas et al., 2021; Wu et al., 2021). These correlations, mostly done retrospectively, `relied` often on correlation coefficients (typically Pearson, Spearman, Kendall) and do not predict future trends. When it comes to realtime monitoring, a significant challenge is the inherent variability and uncertainty in wastewater data which hinders public health applications (Diamond et al., 2022; McClary-Gutierrez et al., 2021). This is particularly true for rural areas where data can be noisier.

To address these challenges, we introduce a new application of the SEIR model to wastewater data in small, underserved populations. Our study specifically focuses on the novel application of this model to predict the onset of SARS-CoV-2 outbreaks in rural communities. By testing the hypothesis that SARS-CoV-2 levels in wastewater can forecast the start of an outbreak, our research aims to provide a predictive framework that enhances the utility of WBE for rural communities, ultimately bridging the gap between wastewater surveillance and actionable public health guidance.

Material & Methods.

Sites and Sample Collection

Wastewater samples were collected from wastewater treatment facilities (WWTFs) located in five rural communities serving approximately 1,000 or less inhabitants and a small city in a rural county in Idaho, USA (Table 1). These rural communities are defined as "rural" according to the 2020 U.S. Census Bureau. The county itself falls under the category of a "rural" or "nonmetropolitan" county, as classified by the U.S. Office of Management and Budget. Rural cities are abbreviated RC1 to RC5 and the small city SC. All WWTFs primarily treat domestic wastewater; the SC WWTF also receives effluent from a regional hospital.

Samples were collected three times a week from October 2021 to March 2022. Rural WWTF samples were time-composite samples collected using Teledyne ISCO 3700 Full Size Portable Sampler (Teledyne ISCO, Lincoln, NE, USA) autosamplers or homemade autosamplers constituted of a Sci-Q 323 peristaltic pump (Watson-Marlow, Falmouth, UK) controlled by an Omron H3CR timer (Omron Corporation, Kyoto, Japan) and housed in a cooler box. Sampling frequencies were comprised between 10 and 30 minutes for 24h. Approximately 3L of wastewater was collected, and subsamples were collected at the end of the 24h sampling period and transported within 6h to the laboratory, where samples were kept at 4°C until further processing. Samples from the SC WWTF were collected using a Teledyne ISCO model 3700 autosampler (Teledyne ISCO, Lincoln, NE, USA), with samples collected paced with influent flow. Sampling failed (e.g., samples were not collected because of tube failure, clogging, or not enough wastewater in the sewer pipes) for less than 10% of the total samples sampled. Samples were kept at 4°C until further processed, at most 3 days later.

Confirmed COVID-19 case counts per zip code were obtained from the Idaho Public Health District 2 website (https://idahopublichealth.com/district-2/novel-coronavirus).

Sample Processing for SARS-CoV-2 Detection and Quantification.

The detailed protocols presented below are publicly available on protocol.io (Narum et al., 2022). In brief, before concentrating the viral fraction of two replicate wastewater fractions through electronegative membrane filtration, each sample was spiked with the Bovilis® Coronavirus (BCoV) (Merck, Kenilworth, NJ, USA) as a process internal control. Subsequently, filters were inserted together with the DNA/RNA Shield™ (Zymo Research, Irvine, CA, USA) into the Lysis Bead tubes from the AllPrep® PowerViral® DNA/RNA Kit (QIAGEN, Inc., Germantown, MD, USA). Lysis was performed on a FastPrep™ (MP Biomedicals, Santa Ana, CA, USA) for 4 cycles of 20 seconds each at 4.5 m/s and the RNA was then extracted as per the kit manufacturer's protocol on a QIAcube Connect automated extraction instrument (QIAGEN, Inc., Germantown, MD, USA).

SARS-CoV-2 was quantified by dPCR using the QIAcuity Digital PCR System (QIAGEN, Inc., Germantown, MD, USA) using the GT-Digital SARS-CoV-2 Wastewater Surveillance Assay For QIAcuity® (GT Molecular, Fort Collins, CO, USA). Each 40 µl reaction contained 1x of the Qiagen QIAcuity One-Step Viral RT-PCR Kit (QIAGEN, Inc., Germantown, MD, USA), 1x of the GT Molecular N1-N2-BCoV Assay Solution, and 20 µl RNA template. RNA extraction blanks, dPCR non-template controls and positive controls were included in each dPCR run.

Data Processing and Analysis

Fluorescent thresholds were manually set based on the fluorescent level of the positive controls. Then we excluded data from samples for which (i) the recovery rate of the internal processing control BCoV was lower than 1%, or (ii) the RNA extraction process control or dPCR negative control were positive and more than 10% of the measured sample concentration.

The date of an outbreak's start was determined with a piecewise regression model using either the cumulative sum of the copies per day of the N1 target or the cumulative sum of COVID-19 clinically confirmed cases to estimate the breakpoint in a linear dataset. For each city, we subsampled the linear data around the inflection points corresponding to the dates of the main surge of N1 copies or COVID-19 reported cases in early 2022. Then we fitted a linear regression model in R using the "lm" function with cumulative copies of cases as the response (Y) and the date as the predictor (X). Finally, we fitted the piecewise regression model to the original model, estimating a breakpoint around the inflection of the line, using the segmented() function from the segmented package in R (Muggeo et al., 2014).

SARS-CoV-2 Epidemiological Model.

We constructed a compartmental model to approximate the dynamics of the epidemic in each city. Due to the small size of populations in the rural areas, we expected stochastic effects to be important, and opted to use a discrete-time discrete-state Markov process to approximate the spread of the disease. In this model, individuals in the population can be in one of four states: susceptible (S), exposed (E), infectious (I), and removed (R). Importantly, in this model exposed individuals have contracted the disease and shed the virus but are not yet infectious. The changes in the compartments are assumed to be binomially distributed: $X_t \sim Bin\left(S(t), \frac{\beta I(t)}{N(t)}\right)$ is the number of newly exposed individuals on day t, $Y_t \sim Bin\left(E(t), \frac{1}{\tau}\right)$ is the number of newly infectious individuals on day t, and $Z_t \sim Bin\left(I(t), \frac{1}{\delta}\right)$ is the number of newly recovered individuals on day t. The parameter β is the transmission rate, τ is the mean incubation period, and δ is the mean infectious period, and $N(t) = S(t) + E(t) + I(t) + R(t)$ is the total population at time t. The discrete-time Markov process is given by:

$$\Delta S(t) = -X_t, \tag{1a}$$

$$\Delta E(t) = X_t - Y_t, \tag{1b}$$

$$\Delta I(t) = Y_t - Z_t, \tag{1c}$$

$$\Delta R(t) = Z_t . \tag{1d}$$

The number of virus particles shed by exposed individuals was assumed to be log-normally distributed (Miura et al., 2021). However, the log-normal distribution is difficult to work with mathematically, so we approximated the log-normal distribution using the Gamma distribution by matching the first two moments. For simplicity, we assumed that only the individuals in the exposed class $E(t)$ shed virus in the stools. This assumption is reasonable since it has been shown that the amount of virus shed by a single individual is time-varying, with a peak occurring around symptoms onset (Puhach et al., 2023; Wu et al., 2022). If there are $E(t)$ exposed individuals, the amount of virus in the wastewater is a random variable $V(t)$ with probability density function:

$$\Phi(V; k, \theta, E) = \gamma(V; Ek, \theta), \tag{2}$$

where $\gamma(V; k, \theta)$ is the probability density function for the gamma distribution with rate $k = \frac{E_v^2}{V_v}$ and scale $\theta = \frac{V_v}{E_v}$. See Table 2 for parameter values.

We used a sequential Monte Carlo (particle filter) method to fit the collected wastewater data to the stochastic model to the collected wastewater data (Figure 1). In simulations, 50,000 particles (initial conditions) were sampled, using the normalized likelihood distribution for the initial concentrations of virus measured in the wastewater to determine the number of exposed individuals (Figure 1 top row). The initial states of the other classes were sampled uniformly from the remaining population.

Each time step, every particle evolved according to the Markov process in equation 1 (Figure 1 bottom left). On days that we have collected wastewater data, the particles were weighted according to their likelihood (Figure 1 bottom center) and resampled using a systematic sampling method to filter out the least likely particles and reinforce the most likely particles (Figure 1 bottom right).

Data and scripts of this study are available on https://github.com/Tyler-Meadows/wastewater-surveillance.

Results.

Dynamics of SARS-CoV-2 in Rural Wastewater vs. Clinically Confirmed Cases.

For the period investigated, clinically reported cases revealed that the cities experienced one or two COVID-19 outbreaks, as shown in Figure 2. The first outbreak occurred in late October 2021, but it was not detected in all cities and was relatively small compared to the second outbreak experienced by all cities in early January 2022. This second surge was driven by the Omicron variant, which emerged in the United States in early December 2021 (CDC COVID-19 Response Team, 2021).

Examining the collected wastewater data, 293 samples were retained following data processing, with each WWTF yielding 44-58 measurements over a five-month period. The daily load of SARS-CoV-2 present in wastewater varied greatly day-to-day, making interpretation of the real-time spread of COVID-19 challenging. Interestingly, the variability in the order of magnitude tended to be larger as the city population decreased, indicating a greater level of randomness (as shown in Figure 3). Additionally, the difference in variance of daily SARS-CoV-2 quantities between cities was significant, as confirmed statistically using Brown-Forsythe, Levene, Barlett, and Kligner-Killeen tests (test results provided in the Supplementary Material). These results suggest that the fluctuations in daily SARS-CoV-2 measurements in smaller cities are more stochastic than in larger cities.

Despite this stochasticity, the Omicron outbreaks resulted in a sharp increase of quantities of the virus collected at the WWTFs above the background levels (Figure 2) by 7- to 81-fold. After estimating the date of the outbreak start (vertical dashed line in Figure 2), we estimated that the SARS-CoV-2 wastewater signal tends to lead the clinically confirmed COVID-19 cases by 0 to 10 days. This supports other retrospective observations, mostly performed in larger cities, that wastewater surveillance could improve and even accelerate the early detection of infectious diseases in rural communities (Feng et al., 2021; Graham et al., 2021; Wu et al., 2022). However, the lead times were variable, and in one case, the wastewater signal was not inferred to precede the clinically reported case data. Other studies have also observed cases where wastewater signal was not preceding clinical testing (Feng et al., 2021; Xiao et al., 2022).

Epidemiological Model to Forecast a COVID-19 Outbreak from Wastewater Detection of SARS-CoV-2.

We used an SEIR-based model (Figure 4A) to investigate whether wastewater-based surveillance of SARS-CoV-2 could enhance the prediction of a COVID-19 outbreak. To test the model's ability to forecast upcoming trend of cases, we determined if the model could have predicted the Omicron outbreak using exclusively wastewater data. Specifically, we fit the predicted cases using the wastewater measurements up to the onset of the outbreak. This corresponded to the first measurement performed two days after the inflection point defining the start of the outbreak using wastewater data. Then we let the model forecast the upcoming trend in active cases (Figure 4B). When comparing the forecasted COVID-19 cases with the clinically confirmed cases we found that the model successfully captured the clinically active cases in five of the six cities. For the rural cities, the forecasts were within the 95% confidence interval at least up to eight days ahead; corresponding to R4 (Figure 4). While the number of predicted cases tended to be lower than the number of the clinically confirmed cases reported, they were following them closely in two rural cities and in the small city. For those two rural cities (sites R1 and R2), the model accurately predicted the number of active cases more than 14 days ahead. In city SC, the model was able to forecast the number of active cases accurately more than eight days ahead. These results show that in the majority of sewersheds surveyed the model would have confidently predicted the outbreaks even before the COVID-19 reported cases started to increase. Anticipation of the model on the COVID-19 reported cases ranged from 0 days for R5 to 11 days for R2 (see blue dashed line on Figure 4).

Although our primary objective was not to predict the peak of the outbreak, we also evaluated the model’s ability to make such predictions. To do so, we used the same approach outlined above and allowed the model to forecast active cases two days after we recorded the highest SARS-CoV-2 level in the wastewater. However, instead of predicting a decrease in cases, the model continued to forecast a sustained increase of active cases. This was in contrast to both the wastewater data and the trend observed in clinically confirmed cases (as shown in Fig. S1). Therefore, while this model may be useful during the early stages of an outbreak in rural areas, it may not be reliable for later phases.

Since our objective was primarily to test if a SEIR model could predict an outbreak occurring in rural communities, we then focused on the capacity of the model to predict an increase in the trend of active cases. To that end, we used a simulated wastewater dataset to count the number of times the model predicted upward or downward trends in cases correctly (i.e., true positive rate) or incorrectly (i.e., false positive rate). We varied the threshold used to accept predictions to create receiver operating characteristic (ROC) curves (Figure 5). Measured area under the curve (AUC) values presented in Figure 6 reflect the sensitivity and specificity of the forecast for a range of forecasted days. The model tends to predict trends better as the forecast range increases. The ROC curves for predictions made less than a week in advance were significantly lower than those made over nine days (p-values presented in Supplementary Material). After nine days, the AUC medians were above 0.7, and increased to 0.75 at 15 days. These results suggest epidemiologists could rely on these 9 to 15-day forecasts.

Discussion

Wastewater-based detection of SARS-CoV-2 has predominantly focused on large urban and metropolitan areas, leaving a significant gap in our understanding of the dynamics of viral spread in rural towns. Our study aimed to address this gap by conducting a surveillance effort of SARS-CoV-2 in the wastewater of several rural cities of the state of Idaho (USA). This effort represents one of the first comprehensive studies applying SEIR modeling to forecast COVID-19 trends in rural settings, where, as we described below, the challenges of WBE are particularly pronounced.

Novelty and Key Contributions

Our study introduces a new application of the SEIR model for WBE. We demonstrate that our SEIR model approach can reliably forecast COVID-19 outbreaks when SARS-CoV-2 loads arriving at WWTF are trending upwards in rural communities. Sensitivity and specificity assessments of the model in predicting the start of the outbreak revealed that these forecasts were more reliable when looking at forecasted trends beyond a week, with the best forecast obtained with data from a period of nine to 15 days. Similar ranges of short-term forecasts of seven to nine days to predict upcoming cases based on wastewater detection of SARS-CoV-2 were reported for larger cities using SEIR or SEIR-like approaches (Pájaro et al., 2022; Phan et al., 2023; Proverbio et al., 2022). The fact that the model was not performing well under nine days may in part be attributed to the noise of the SARS-CoV-2 quantities in the wastewater of small rural communities, which creates a sawtooth pattern in the trend (i.e., while the overall trend over a week goes up, some of the points between go down). This led us to hypothesize that the noise in the wastewater data may have made short-term trend predictions less accurate.

Addressing Variability and Noise of Wastewater-Based Detection in a Rural Context.

High daily variability in viral loads arriving in rural WWTFs was a significant challenge, making the interpretation of the data in real-time and the determination of outbreak onset in rural areas even more complicated for public health. As we observed, the variability in the order of magnitude tended to be larger as the rural sewershed size decreased (Schill et al., 2023). Indeed, larger sewersheds tend to have a more consistent activity throughout the day, reducing the variation in wastewater arriving at the WWTF (Tchobanoglous et al., 2014). In contrast, in small watersheds, each individual contributes to a higher proportion of the overall SARS-CoV-2 load at the WWTF, meaning that variations due to individual shedding differences or different stages of the infection disproportionally affect the viral signal measured (de Araújo et al., 2023; Wade et al., 2022). Similarly, missing an individual shedding SARS-CoV-2 can have significant impacts on the data. Missing individuals in rural cities can be attributed to several factors, such as mobility (e.g., commuting) and the presence of individual septic systems, which are common in rural areas. The sampling effect is also more pronounced in smaller populations, where the chance of missing a toilet flush in a smaller sewershed is higher. This increased noise in the data makes wastewater surveillance in rural communities particularly challenging, limiting its effectiveness as a standalone tool and positioning it primarily as a complementary resource to clinical surveillance. Nevertheless, as we showed in our study, our model factored out this noise, enabling accurate forecasts of the Omicron outbreak.

Enhancing Model Robustness

Even though our SEIR model has shown promise, incorporating additional strategies could further enhance its robustness and reliability. For instance, increasing sampling frequency could smooth out daily fluctuations, offering a more continuous data stream that better captures underlying trends (Ahmed et al., 2020). Similarly, extending the duration of composite sampling (e.g., from 24-hour to 48-hour or 72-hour composites) could reduce the impact of short-term variations, leading to more accurate trend detection. Data smoothing techniques could also help filter out noise from data (Rauch et al., 2022). Furthermore, normalizing viral concentrations not just by flow rate but also by population markers (e.g., human fecal indicators) could correct for variations in population dynamics affecting the measured viral load. Finally, knowing the fate and transport of targeted viruses in the sewer networks would help better understand virus decay and other unexplained fluctuations in pathogen detection at WWTF. Future research should explore the integration of these approaches to optimize model predictions.

Model Limitations and Future Directions

While our SEIR model functions well to provide advanced warning of a COVID-19 outbreak, it was less successful in predicting the outbreak peak (Fig. S1) and in estimating the number of cases accurately. Predicted cases tended to be lower than reported clinical cases for the area, contrasting with other studies on larger sewersheds that have shown that SEIR models typically estimate more COVID-19 cases than the reported number (Eikenberry et al., 2020; Phan et al., 2023; Wu et al., 2020). This limitation may be attributed, in part, to differences between the population of the city sampled and the zip code used (Table 1); the latter corresponds to the clinically recorded cases. When comparing city with zip code census, 39 to 71% of the residents of rural zip codes were not connected to the sewer system of the cities. In rural areas, zip codes often cover a larger geographical area beyond the city limits, which means that comparisons between wastewater data and reported cases in the context of rural cities should be approached with caution.

Concerning the outbreak peak predictions, several parameters may explain why we could forecast them accurately. One important factor in disease modeling is the force of infection, which essentially determines the probability of a susceptible individual encountering an infected individual and then the likelihood of a new infection occurring from this encounter. This infection force is driven by factors such as the age of individuals, the stage of infection, and their behavior (e.g., wearing masks, practicing social distancing, etc.). Current models have limitations in accurately capturing these factors due to a lack of data and the complexity of individual behaviors, which by themselves evolved during the pandemic. However, our model represents a simple first step toward creating a sufficient model.

Notably, our model is able to predict peak timing (if not as accurately as desired), whereas data-driven models (i.e., machine learning) cannot in the true sense of a disease peak.

Moreover, despite the successful application of SEIR models to wastewater surveillance of the SARS-CoV-2, there are some uncertainties on how to connect the model with wastewater data. Some authors have included the cumulative virus titer in the sewershed as a dynamic variable (Phan et al., 2023) or as a linear combination of other dynamic variables (Nourbakhsh et al., 2022). However, this approach can be problematic when measurements are sparse or there are gaps between collection periods – which would be very common in rural WWTFs. Most authors directly connect wastewater measurements to the incidence rate, or prevalence, similar to what we have done herein.

In addition, the connection between the disease compartments (the exposed ‘E’ and infected ‘I’ compartments) in the SEIR model and wastewater measurements is not well established. Contribution to viral load in wastewater can tie to the individuals in the ‘I’ compartment (McMahan et al., 2021; Nourbakhsh et al., 2022; Phan et al., 2023) or to the ‘E’ compartment; the SEIR model in this study resulted in better predictions than when connected to ‘I’. This difference could be attributed to the fact that peak virus shedding in stool may occur for a few days around the onset of symptoms (Puhach et al., 2023; Wu et al., 2022). This means that individuals contributing to the load of SARS-CoV-2 measured in wastewater may be at the transition between 'E' and 'I' compartments in the SEIR model. Some researchers have incorporated the results from wastewater measurement in both ‘E’ and ‘I’ (Fazli et al., 2021) while others created a new compartment structure to account for viral shedding dynamic in wastewater (Polcz et al., 2023; Proverbio et al., 2022). Disentangling how the different groups E, I, R (recovered ‘R’ individuals can still shed virus in their stool (Natarajan et al., 2022)) contribute to the wastewater is a challenging task that remains to be explored. Since the numbers of infectious and exposed individuals are not independent of one another, we would need to know the joint distribution in order to infer the relative contributions of the virus to the wastewater. Others have pointed out how viral shedding at different stages can indeed lead to issues with model fitting (Phan et al., 2023).

Finally, as discussed above the noise in wastewater data presents challenges in fitting the SEIR model accurately, especially in small, rural populations where the impact of individual variations is magnified. While our model was generally successful in forecasting upcoming cases, it failed in one city, indicating that the disease dynamics did not always align with the model assumptions. This discrepancy could be due to factors described above such as commuting patterns, which are more prevalent in rural areas where residents often have to commute to work in other towns. In the rural cities surveyed, between 79.4 and 94.5% of residents work outside their place of residence, versus 27.3% in the small city surveyed (data from U.S. Census Bureau Topic: Commuting – Survey: American Community Survey – 2021, ACS 5-Year Estimates Subject Tables).

Conclusion

In conclusion, our study advances the field by employing a SEIR mechanistic modeling approach to forecast COVID-19 case trends in rural communities. We demonstrate that even though SARS-CoV-2 data from rural wastewater have its own challenges, our model can reliably forecast case trends, allowing public health officials to respond more proactively. Our model can be deployed in any rural city as long as the city population census and wastewater flow data are available or can be estimated. It's important to note that inflow measurements are not always available for WWTF in small rural cities, as the influent flow does not factor into compliance with effluent water quality in the U.S. or certain European countries. Nevertheless, the predictive capability of such a SEIR model is especially crucial given the challenges in identifying outbreak onset in real-time, a problem that is exacerbated in small rural populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM104420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Data collection for this study was also made possible thanks to the State of Idaho Department of Health and Welfare under grant ## NU50CK000544 funded by the Center for Disease Control and Prevention (CDC) through the Epidemiology and Laboratory Capacity Enhancing Detection Through Coronavirus Release and Relief (CRR) Supplement Funds. The content of this project is solely the responsibility of the authors and does not necessarily represent the official views of the State of Idaho Department of Health and Welfare or the Center for Disease Control and Prevention (CDC).

Abbreviations:

WBE:	Wastewater-based epidemiology
SEIR:	susceptible-exposed-infectious-recovered
WWTF:	Wastewater treatment facilities
RC:	Rural city
SC:	Small city
ROC:	Receiver operating characteristic
AUC:	Area under the curve

References

Ahmed W, Bivins A, Bertsch PM, Bibby K, Choi PM, Farkas K, Gyawali P, Hamilton KA, Haramoto E, Kitajima M, Simpson SL, Tandukar S, Thomas K, Mueller JF, 2020. Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimisation and quality control are crucial for generating reliable public health information. Curr Opin Environ Sci Health. 10.1016/j.coesh.2020.09.003

Arts PJ, Kelly JD, Midgley CM, Anglin K, Lu S, Abedi GR, Andino R, Bakker KM, Banman B, Boehm AB, Briggs-Hagen M, Brouwer AF, Davidson MC, Eisenberg MC, Garcia-Knight M, Knight S, Peluso MJ, Pineda-Ramirez J, Diaz Sanchez R, Saydah S, Tassetto M, Martin JN, Wigginton KR, 2023. Longitudinal and quantitative fecal shedding dynamics of SARS-CoV-2, pepper mild mottle virus, and crAssphage. mSphere 0, e00132–23. 10.1128/msphere.00132-23

Baker RE, Peña J-M, Jayamohan J, Jérusalem A, 2018. Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biology Letters 14, 20170660. 10.1098/rsbl.2017.0660 [PubMed: 29769297]

Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, Bhattacharya P, Bijlsma L, Boehm AB, Brown J, Buttiglieri G, Calabro V, Carducci A, Castiglioni S, Cetecioglu Gurol Z, Chakraborty S, Costa F, Curcio S, de los Reyes FL, Delgado Vela J, Farkas K, Fernandez-Casi X, Gerba C, Gerrity D, Girones R, Gonzalez R, Haramoto E, Harris A, Holden PA, Islam Md.T., Jones DL, Kasprzyk-Hordern B, Kitajima M, Kotlarz N, Kumar M, Kuroda K, La Rosa G, Malpei F, Mautus M, McLellan SL, Medema G, Meschke JS, Mueller J, Newton RJ, Nilsson D, Noble RT, van Nuijs A, Peccia J, Perkins TA, Pickering AJ, Rose J, Sanchez G, Smith A, Stadler L, Stauber C, Thomas K, van der Voorn T, Wigginton K, Zhu K, Bibby K, 2020. Wastewater-based epidemiology: Global collaborative to maximize contributions in the fight against COVID-19. Environ. Sci. Technol 10.1021/acs.est.0c02388

CDC COVID-19 Response Team, 2021. SARS-CoV-2 B.1.1.529 (Omicron) Variant — United States, December 1–8, 2021. MMWR Morb. Mortal. Wkly Rep 70. 10.15585/mmwr.mm7050e1

Conway MJ, Kado S, Kooienga BK, Sarette JS, Kirby MH, Marten AD, Ward AS, Abel JD, King S, Billette J, Williams MR, Uzarski RL, Alm EW, 2023. SARS-CoV-2 wastewater monitoring in rural and small metropolitan communities in Central Michigan. Sci. Total Environ 894, 165013. 10.1016/j.scitotenv.2023.165013 [PubMed: 37353028]

Corpuz MVA, Buonerba A, Vigliotta G, Zarra T, Ballesteros F, Campiglia P, Belgiorno V, Korshin G, Naddeo V, 2020. Viruses in wastewater: occurrence, abundance and detection methods. Sci Total Environ 745, 140910. 10.1016/j.scitotenv.2020.140910 [PubMed: 32758747]

Cromartie J, Dobis Elizabeth A., Krumel Thomas P., McGranahan David, Pender John, 2020. Rural America at a glance: 2020 edition. United States Department of Agriculture Economic Research Service: Economic Information Bulletin 221, 6.

Cuadros DF, Branscum AJ, Mukandavire Z, Miller FD, MacKinnon N, 2021. Dynamics of the COVID-19 epidemic in urban and rural areas in the United States. Ann. Epidemiol 59, 16–20. 10.1016/j.annepidem.2021.04.007 [PubMed: 33894385]

D'Aoust PM, Towhid ST, Mercier É, Hegazy N, Tian X, Bhatnagar K, Zhang Z, Naughton CC, MacKenzie AE, Graber TE, Delatolla R, 2021. COVID-19 wastewater surveillance in rural communities: Comparison of lagoon and pumping station samples. Sci. Total Environ 801, 149618. 10.1016/j.scitotenv.2021.149618 [PubMed: 34418622]

de Araújo JC, Madeira CL, Bressani T, Leal C, Leroy D, Machado EC, Fernandes LA, Espinosa MF, Freitas GTO, Leão T, Mota VT, Pereira AD, Perdigão C, Tröger F, Ayrimoraes S, de Melo MC, Laguardia F, Reis MTP, Mota C, Chernicharo CAL, 2023. Quantification of SARS-CoV-2 in wastewater samples from hospitals treating COVID-19 patients during the first wave of the pandemic in Brazil. Sci. Total Environ 860, 160498. 10.1016/j.scitotenv.2022.160498 [PubMed: 36436622]

Diamond MB, Keshaviah A, Bento AI, Conroy-Ben O, Driver EM, Ensor KB, Halden RU, Hopkins LP, Kuhn KG, Moe CL, Rouchka EC, Smith T, Stevenson BS, Susswein Z, Vogel JR, Wolfe MK, Stadler LB, Scarpino SV, 2022. Wastewater surveillance of pathogens can inform public health responses. Nat. Med 28, 1992–1995. 10.1038/s41591-022-01940-x [PubMed: 36076085]

Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, Kostelich E, Gumel AB, 2020. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect. Dis. Model 5, 293–308. 10.1016/j.idm.2020.04.001 [PubMed: 32355904]

Fazli M, Sklar S, Porter MD, French BA, Shakeri H, 2021. Wastewater-based epidemiological modeling for continuous surveillance of COVID-19 outbreak, in: 2021 IEEE International Conference on Big Data (Big Data). Presented at the 2021 IEEE International Conference on Big Data (Big Data), pp. 4342–4349. 10.1109/BigData52589.2021.9671543

Feng S, Roguet A, McClary-Gutierrez JS, Newton RJ, Kloczko N, Meiman JG, McLellan SL, 2021. Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. ACS EST Water 1, 1955–1965. 10.1021/acsestwater.1c00160

Fernandez-Cassi X, Scheidegger A, Bänziger C, Cariti F, Tuñas Corzon A, Ganesanandamoorthy P, Lemaitre JC, Ort C, Julian TR, 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. 10.1016/j.watres.2021.117252 [PubMed: 34048984]

Fontenele RS, Yang Y, Driver EM, Magge A, Kraberger S, Custer JM, Dufault-Thompson K, Cox E, Newell ME, Varsani A, Halden RU, Scotch M, Jiang X, 2023. Wastewater surveillance uncovers regional diversity and dynamics of SARS-CoV-2 variants across nine states in the USA. Sci Total Environ 877, 162862. 10.1016/j.scitotenv.2023.162862 [PubMed: 36933724]

Graham KE, Loeb SK, Wolfe MK, Catoe D, Sinnott-Armstrong N, Kim S, Yamahara KM, Sassoubre LM, Mendoza Grijalva LM, Roldan-Hernandez L, Langenfeld K, Wigginton KR, Boehm AB, 2021. SARS-CoV-2 RNA in wastewater settled solids is associated with COVID-19 cases in a large urban sewershed. Environ. Sci. Technol 55, 488–498. 10.1021/acs.est.0c06191 [PubMed: 33283515]

Holm RH, Osborne Jelks N, Schneider R, Smith T, 2023. Beyond COVID-19: designing inclusive public health surveillance by including wastewater monitoring. Health Equity 7, 377–379. 10.1089/heq.2022.0055 [PubMed: 37351532]

Huisman JS, Scire J, Caduff L, Fernandez-Cassi X, Ganesanandamoorthy P, Kull A, Scheidegger A, Stachler E, Boehm AB, Hughes B, Knudson A, Topol A, Wigginton KR, Wolfe MK, Kohn T, Ort C, Stadler T, Julian TR, 2022. Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. Environ. Health Perspect 130, 057011. 10.1289/EHP10050 [PubMed: 35617001]

Jarvie MM, Reed-Lukomski M, Southwell B, Wright D, Nguyen TNT, 2023. Monitoring of COVID-19 in wastewater across the Eastern Upper Peninsula of Michigan. Environ. Adv 11, 100326. 10.1016/j.envadv.2022.100326 [PubMed: 36471702]

Jeong HW, Kim S-M, Kim H-S, Kim Y-L, Kim JH, Cho JY, Kim S-H, Kang H, Kim S-G, Park S-J, Kim E-H, Choi YK, 2020. Viable SARS-CoV-2 in various specimens from COVID-19 patients. Clin. Microbiol. Infect 26, 1520–1524. 10.1016/j.cmi.2020.07.020 [PubMed: 32711057]

Kanchan S, Ogden E, Kesheri M, Skinner A, Miliken E, Lyman D, Armstrong J, Sciglitano L, Hampikian G, 2024. COVID-19 hospitalizations and deaths predicted by SARS-CoV-2 levels in Boise, Idaho wastewater. Science of The Total Environment 907, 167742. 10.1016/j.scitotenv.2023.167742 [PubMed: 37852488]

Kaufman BG, Whitaker R, Pink G, Holmes GM, 2020. Half of rural residents at high risk of serious illness due to COVID-19, creating stress on rural hospitals. J. Rural Health 36, 584–590. 10.1111/jrh.12481 [PubMed: 32603030]

Lakhani HV, Pillai SS, Zehra M, Sharma I, Sodhi K, 2020. Systematic review of clinical insights into novel coronavirus (CoVID-19) pandemic: persisting challenges in U.S. rural population. Int. J. Environ. Res. Public Health 17, 4279. 10.3390/ijerph17124279 [PubMed: 32549334]

Layton BA, Kaya D, Kelly C, Williamson KJ, Alegre D, Bachhuber SM, Banwarth PG, Bethel JW, Carter K, Dalziel BD, Dasenko M, Geniza M, George A, Girard A-M, Haggerty R, Higley KA, Hynes DM, Lubchenco J, McLaughlin KR, Nieto FJ, Noakes A, Peterson M, Piemonti AD, Sanders JL, Tyler BM, Radniecki TS, 2022. Evaluation of a wastewater-based epidemiological approach to estimate the prevalence of SARS-CoV-2 infections and the detection of viral variants in disparate Oregon communities at city and neighborhood scales. Environ. Health Perspect 130, 067010. 10.1289/EHP10289 [PubMed: 35767012]

Li L, Mazurowski L, Dewan A, Carine M, Haak L, Guarin TC, Dastjerdi NG, Gerrity D, Mentzer C, Pagilla KR, 2022. Longitudinal monitoring of SARS-CoV-2 in wastewater using viral genetic markers and the estimation of unconfirmed COVID-19 cases. Sci. Total Environ 817, 152958. 10.1016/j.scitotenv.2022.152958 [PubMed: 35016937]

Mandrekar JN, 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J. Thorac. Oncol 5, 1315–1316. 10.1097/JTO.0b013e3181ec173d [PubMed: 20736804]

McClary-Gutierrez JS, Mattioli MC, Marcenac P, Silverman AI, Boehm AB, Bibby K, Balliet M, de los Reyes FL, Gerrity D, Griffith JF, Holden PA, Katehis D, Kester G, LaCross N, Lipp EK, Meiman J, Noble RT, Brossard D, McLellan SL, 2021. SARS-CoV-2 wastewater surveillance for public health action. Emerg. Infect. Dis 27, e210753. 10.3201/eid2709.210753 [PubMed: 34424162]

McMahan CS, Self S, Rennert L, Kalbaugh C, Kriebel D, Graves D, Colby C, Deaver JA, Popat SC, Karanfil T, Freedman DL, 2021. COVID-19 wastewater epidemiology: a model to estimate infected populations. Lancet Planet. Health 5, e874–e881. 10.1016/S2542-5196(21)00230-8 [PubMed: 34895497]

McManus O, Christiansen LE, Nauta M, Krogsgaard LW, Bahrenscheer NS, von Kappelgaard L Christiansen T, Hansen M, Hansen NC, Kähler J, Rasmussen A, Richter SR, Rasmussen LD, Franck KT, Ethelberg S, 2023. Predicting COVID-19 incidence using wastewater surveillance data, Denmark, October 2021–June 2022. Emerg. Infect. Dis 29, 1589–1597. 10.3201/eid2908.221634 [PubMed: 37486168]

Medina CY, Kadonsky KF, Roman FA, Tariqi AQ, Sinclair RG, D'Aoust PM, Delatolla R, Bischel HN, Naughton CC, 2022. The need of an environmental justice approach for wastewater based epidemiology for rural and disadvantaged communities: A review in California. Curr. Opin. Environ. Sci. Health 27, 100348. 10.1016/j.coesh.2022.100348 [PubMed: 35399703]

Metzcar J, Jutzeler CR, Macklin P, Köhn-Luque A, Brüningk SC, 2024. A review of mechanistic learning in mathematical oncology. Front. Immunol 15. 10.3389/fimmu.2024.1363144

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. 10.1016/j.scitotenv.2020.144549 [PubMed: 33477053]

Muggeo VM, Atkins DC, Gallop RJ, Dimidjian S, 2014. Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. Stat. Modelling 14, 293–313. 10.1177/1471082X13504721

Nahm FS, 2022. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J. Anesthesiol 75, 25–36. 10.4097/kja.21209 [PubMed: 35124947]

Narum S, Stalder T, Coats E, Top E, 2022. Quantification of the SARS-CoV-2 using electronegative membrane filtration and dPCR. Protocols.io.

Natarajan A, Zlitni S, Brooks EF, Vance SE, Dahlen A, Hedlin H, Park RM, Han A, Schmidtke DT, Verma R, Jacobson KB, Parsonnet J, Bonilla HF, Singh U, Pinsky BA, Andrews JR, Jagannathan P, Bhatt AS, 2022. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. Med 3, 371–387.e9. 10.1016/j.medj.2022.04.001 [PubMed: 35434682]

Nourbakhsh S, Fazil A, Li M, Mangat CS, Peterson SW, Daigle J, Langner S, Shurgold J, D'Aoust P, Delatolla R, Mercier E, Pang X, Lee BE, Stuart R, Wijayasri S, Champredon D, 2022. A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities. Epidemics 39, 100560. 10.1016/j.epidem.2022.100560 [PubMed: 35462206]

Pájaro M, Fajar NM, Alonso AA, Otero-Muras I, 2022. Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and wastewater data in small and medium-sized municipalities: A one year study. Chaos Solitons Fractals 164, 112671. 10.1016/j.chaos.2022.112671 [PubMed: 36091637]

Parasa S, Desai M, Chandrasekar VT, Patel HK, Kennedy KF, Roesch T, Spadaccini M, Colombo M, Gabbiadini R, Artifon ELA, Repici A, Sharma P, 2020. Prevalence of gastrointestinal symptoms and fecal viral shedding in patients with coronavirus disease 2019: A systematic review and meta-analysis. JAMA Netw. Open 3, e2011335–e2011335. 10.1001/jamanetworkopen.2020.11335 [PubMed: 32525549]

Park S, Marcus GM, Olgin JE, Carton T, Hamad R, Pletcher MJ, COVID-19 Citizen Science Study Consortium, 2023. Unreported SARS-CoV-2 home testing and test positivity. JAMA Netw. Open 6, e2252684. 10.1001/jamanetworkopen.2022.52684 [PubMed: 36696115]

Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, Ko AI, Malik AA, Wang D, Wang M, Warren JL, Weinberger DM, Arnold W, Omer SB, 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat. Biotech 38, 1164–1167. 10.1038/s41587-020-0684-z

Phan T, Brozak S, Pell B, Gitter A, Xiao A, Mena KD, Kuang Y, Wu F, 2023. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. Sci. Total Environ 857, 159326. 10.1016/j.scitotenv.2022.159326 [PubMed: 36220466]

Polcz P, Tornai K, Juhász J, Cserey G, Surján G, Pándics T, Róka E, Vargha M, Reguly IZ, Csikász-Nagy A, Pongor S, Szederkényi G, 2023. Wastewater-based modeling, reconstruction, and prediction for COVID-19 outbreaks in Hungary caused by highly immune evasive variants. Water Res. 241, 120098. 10.1016/j.watres.2023.120098 [PubMed: 37295226]

Proverbio D, Kemp F, Magni S, Ogorzaly L, Cauchie H-M, Gonçalves J, Skupin A, Aalto A, 2022. Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis. Sci. Total Environ 827, 154235. 10.1016/j.scitotenv.2022.154235 [PubMed: 35245552]

Puhach O, Meyer B, Eckerle I, 2023. SARS-CoV-2 viral load and shedding kinetics. Nat. Rev. Microbiol 21, 147–161. 10.1038/s41579-022-00822-w [PubMed: 36460930]

Rauch W, Schenk H, Insam H, Markt R, Kreuzinger N, 2022. Data modelling recipes for SARS-CoV-2 wastewater-based epidemiology. Environmental Research 214, 113809. 10.1016/j.envres.2022.113809 [PubMed: 35798267]

Ridenhour BJ, Sarathchandra D, Seamon E, Brown H, Leung F-Y, Johnson-Leon M, Megheib M, Miller CR, Johnson-Leung J, 2022. Effects of trust, risk perception, and health behavior on COVID-19 disease burden: Evidence from a multi-state US survey. PLoS One 17, e0268302. 10.1371/journal.pone.0268302 [PubMed: 35594254]

Rose C, Parker A, Jefferson B, Cartmell E, 2015. The characterization of feces and urine: A review of the literature to inform advanced treatment technology. Crit. Rev. Environ. Sci. Technol 45, 1827–1879. 10.1080/10643389.2014.1000761

Schill R, Nelson KL, Harris-Lovett S, Kantor RS, 2023. The dynamic relationship between COVID-19 cases and SARS-CoV-2 wastewater concentrations across time and space: Considerations for model training data sets. Sci Total Environ 871, 162069. 10.1016/j.scitotenv.2023.162069 [PubMed: 36754324]

Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, Pang J, 2022. Wastewater surveillance to infer COVID-19 transmission: A systematic review. Sci Total Environ 804, 150060. 10.1016/j.scitotenv.2021.150060 [PubMed: 34798721]

Tchobanoglous G, Stensel DH, Tsuchihashi R, Burton F, Abu-Orf M, Bowden G, Pfrang W, Metcalf & Eddy, Inc, AECOM (Eds.), 2014. Wastewater engineering: treatment and resource recovery, Fifth edition. ed. McGraw-Hill Education, New York.

Toledo DM, Robbins AA, Gallagher TL, Hershberger KC, Barney RE, Salmela SM, Pilcher D, Cervinski MA, Nerenz RD, Szczepiorkowski ZM, Tsongalis GJ, Lefferts JA, Martin IW, Hubbard JA, 2022. Wastewater-based SARS-CoV-2 surveillance in Northern New England. Microbiol. Spectr. 10, e0220721. 10.1128/spectrum.02207-21

Wade MJ, Lo Jacomo A, Armenise E, Brown MR, Bunce JT, Cameron GJ, Fang Z, Farkas K, Gilpin DF, Graham DW, Grimsley JMS, Hart A, Hoffmann T, Jackson KJ, Jones DL, Lilley CJ, McGrath JW, McKinley JM, McSparron C, Nejad BF, Morvan M, Quintela-Baluja M, Roberts AMI, Singer AC, Souque C, Speight VL, Sweetapple C, Walker D, Watts G, Weightman A, Kasprzyk-Hordern B, 2022. Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. J. Hazard. Mater 424, 127456. 10.1016/j.jhazmat.2021.127456 [PubMed: 34655869]

Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, Hoffman K, Jamal R, Heck P, Zhang Y, Torgersen K, Laan JV, LaCross N, 2021. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. Sci Total Environ 775, 145790. 10.1016/j.scitotenv.2021.145790 [PubMed: 33618308]

Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, Schneider J, Ehmann R, Zwirglmaier K, Drosten C, Wendtner C, 2020. Virological assessment of hospitalized patients with COVID-2019. Nature 581, 465–469. 10.1038/s41586-020-2196-x [PubMed: 32235945]

World Health Organization, n.d. GPEI-Variant type 2 poliovirus isolated from sewage samples in Gaza. URL https://polioeradication.org/news-post/variant-type-2-poliovirus-isolated-from-sewage-samples-in-gaza/ (accessed 8.16.24).

Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bonneau R, Brown MA, Bushman M, Chai PR, Duvallet C, Erickson TB, Foppe K, Ghaeli N, Gu X, Hanage WP, Huang KH, Lee WL, Matus M, McElroy KA, Nagler J, Rhode SF, Santillana M, Tucker JA, Wuertz S, Zhao S, Thompson J, Alm EJ, 2022. SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. Sci Total Environ 805, 150121. 10.1016/j.scitotenv.2021.150121 [PubMed: 34534872]

Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bushman M, Chai PR, Duvallet C, Erickson TB, Foppe K, Ghaeli N, Gu X, Hanage WP, Huang KH, Lee WL, McElroy KA, Rhode SF, Matus M, Wuertz S, Thompson J, Alm EJ, 2021. Wastewater surveillance of SARS-CoV-2 across 40 U.S. states from February to June 2020. Water Res 202, 117400. 10.1016/j.watres.2021.117400 [PubMed: 34274898]

Wu F, Zhang J, Xiao A, Gu X, Lee WL, Armas F, Kauffman K, Hanage W, Matus M, Ghaeli N, Endo N, Duvallet C, Poyet M, Moniz K, Washburne AD, Erickson TB, Chai PR, Thompson J, Alm EJ, 2020. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. MSystems 5. 10.1128/mSystems.00614-20

Wurtzer S, Marechal V, Mouchel JM, Maday Y, Teyssou R, Richard E, Almayrac JL, 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. Euro Surveill. 25, 2000776. 10.2807/1560-7917.ES.2020.25.50.2000776 [PubMed: 33334397]

Xiao A, Wu F, Bushman M, Zhang J, Imakaev M, Chai PR, Duvallet C, Endo N, Erickson TB, Armas F, Arnold B, Chen H, Chandra F, Ghaeli N, Gu X, Hanage WP, Lee WL, Matus M, McElroy KA, Moniz K, Rhode SF, Thompson J, Alm EJ, 2022. Metrics to relate COVID-19 wastewater data to clinical testing dynamics. Water Res. 212, 118070. 10.1016/j.watres.2022.118070 [PubMed: 35101695]

Highlights

- Trends of SARS-CoV-2 in rural wastewater tend to foreshadow clinical cases
- Noise in viral trends in rural wastewater makes real-time interpretation difficult
- An SEIR model predicts the start of an outbreak using wastewater surveillance data
- Epidemiologists could rely on 9 to 15-day forecasts
- WBE can serve as an environmental justice tool to support rural communities

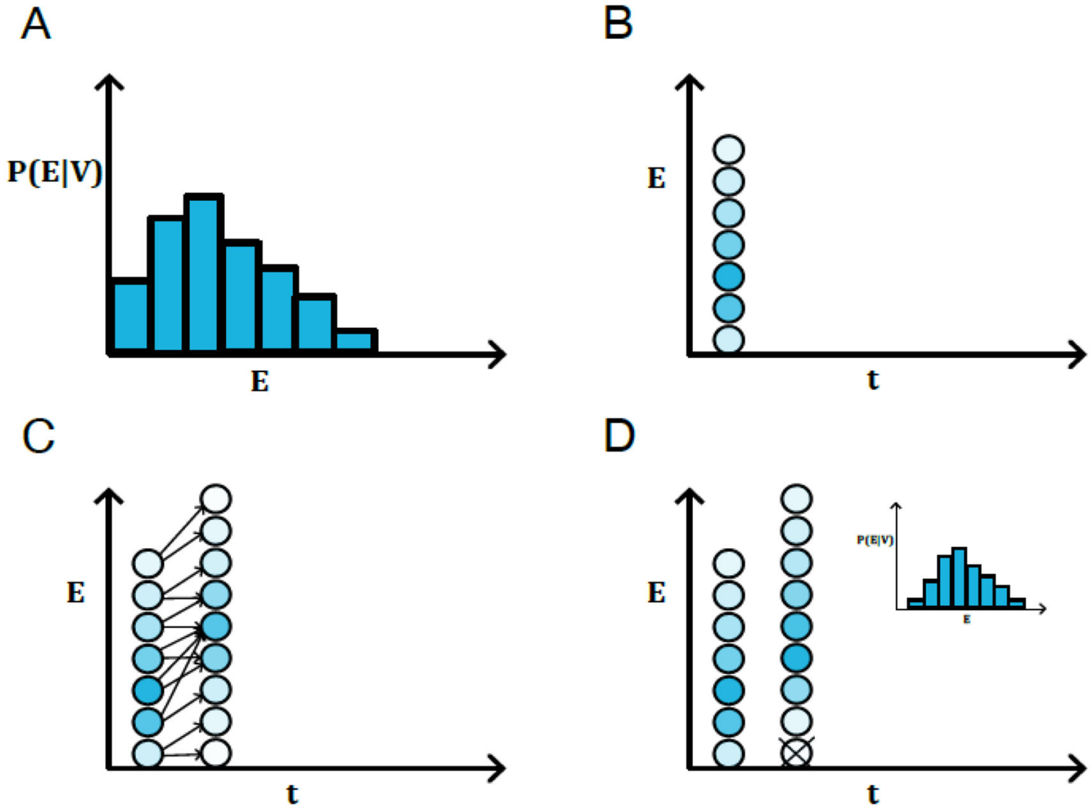


Figure 1: Diagram showing steps of the particle filter method we use to determine the number of active cases from the wastewater titers of SARS-CoV-2.

The particle filter is initialized using the first measurement of virus concentration in the wastewater. We generate a distribution of the possible number of infections in the community and sample many (50000) values from this distribution. These values are used as the possible number of exposed individuals (E) on day 1 (top right graph). Each of these values also gets a potential number of Susceptible (S), Infected (I), and Recovered (R) individuals. Each set of values (S,E,I,R) is called a particle. The darker dots in the diagram signify a higher number of particles with that value of E. We apply one step of the stochastic $SEIR$ model to each particle to predict the number of infections on the next day (bottom left graph). The measurement of the virus in the wastewater on the next measurement is used to determine which particles are more likely than others. Less likely particles are filtered out using a systematic resampling procedure and replaced with more likely particles (bottom right graph).

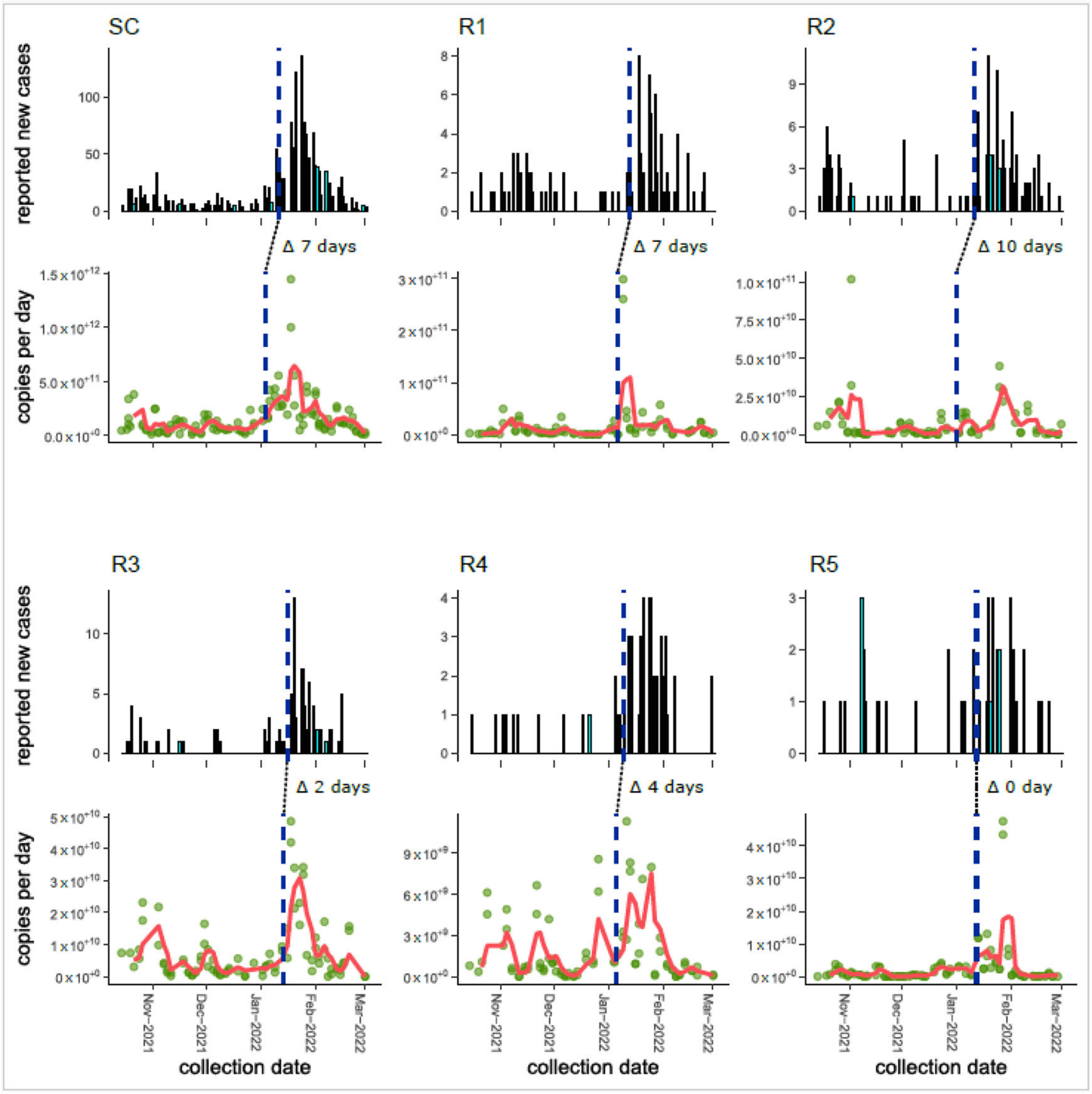


Figure 2: Trend in SARS-CoV-2 in wastewater mirrors the dynamic of the COVID-19 outbreak in rural areas.

Each panel represents a city. In each panel, the bar graph shows the time series of the COVID-19 clinically confirmed cases at the specimen collection dates and the second graph shows the measured concentration of SARS-CoV-2 (green dots) with the 7-day moving average (red line). Vertical dash lines represent the estimated start of the outbreak using either the cumulative sum of the copies per day of the N1 target or the cumulative sum of COVID-19 clinically confirmed cases, determined by the Piecewise regression model. Delta shows the difference of days between predicted dates from wastewater-based detection of SARS-CoV-2 and clinically confirmed COVID-19 cases. Cities are ordered by population size (largest on the top left and smallest on the bottom right).

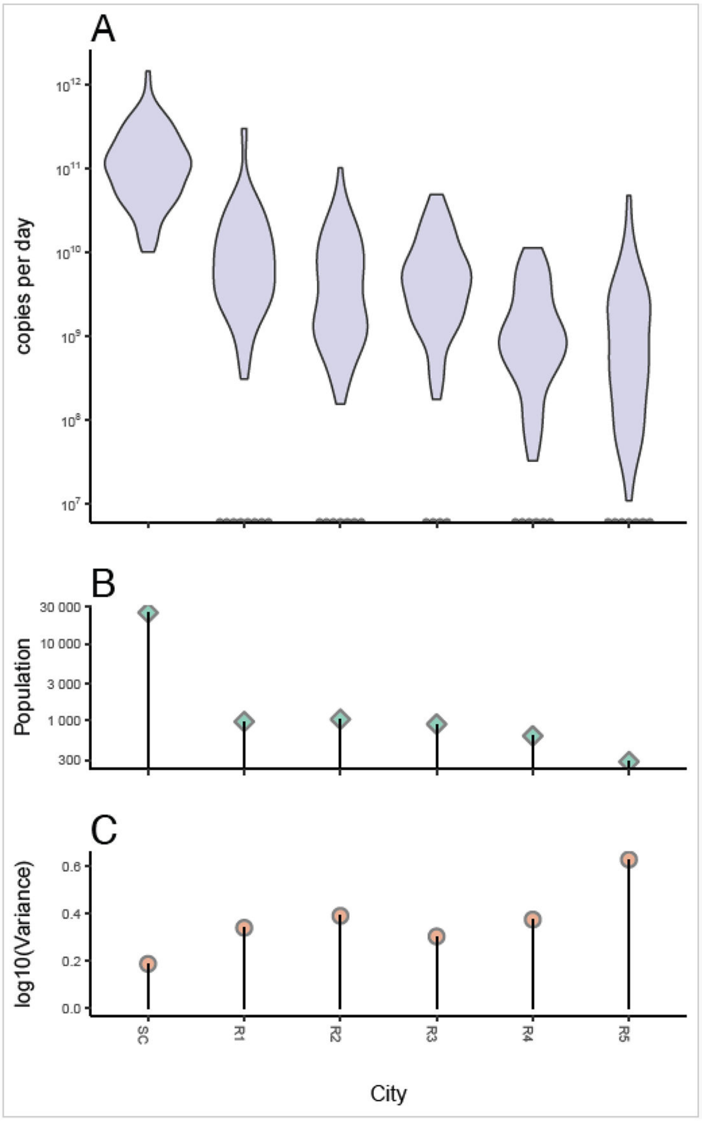


Figure 3: Daily quantities of SARS-CoV-2 tend to be more spread as the city population get smaller.

Panel A) shows the distribution of the copies per day of the SARS-CoV-2 on the log scale over the sampling period at each site ordered by city size, detailed in panel B. Note: bin width = 1/30. Dots on the x axis show the samples where N1 was under the detection limit (SC: n = 0, R1: n = 8, R2: n = 8, R3: n = 4, R4: n = 6, R5: n = 7). Panels B) shows the population size of the cities sampled, and C) shows the log scale sample variance measured for each city calculated using the log10 of copies per day of the SARS-CoV-2. This essentially shows that the magnitude of the estimate is less consistent as population size gets smaller (i.e., more stochasticity).

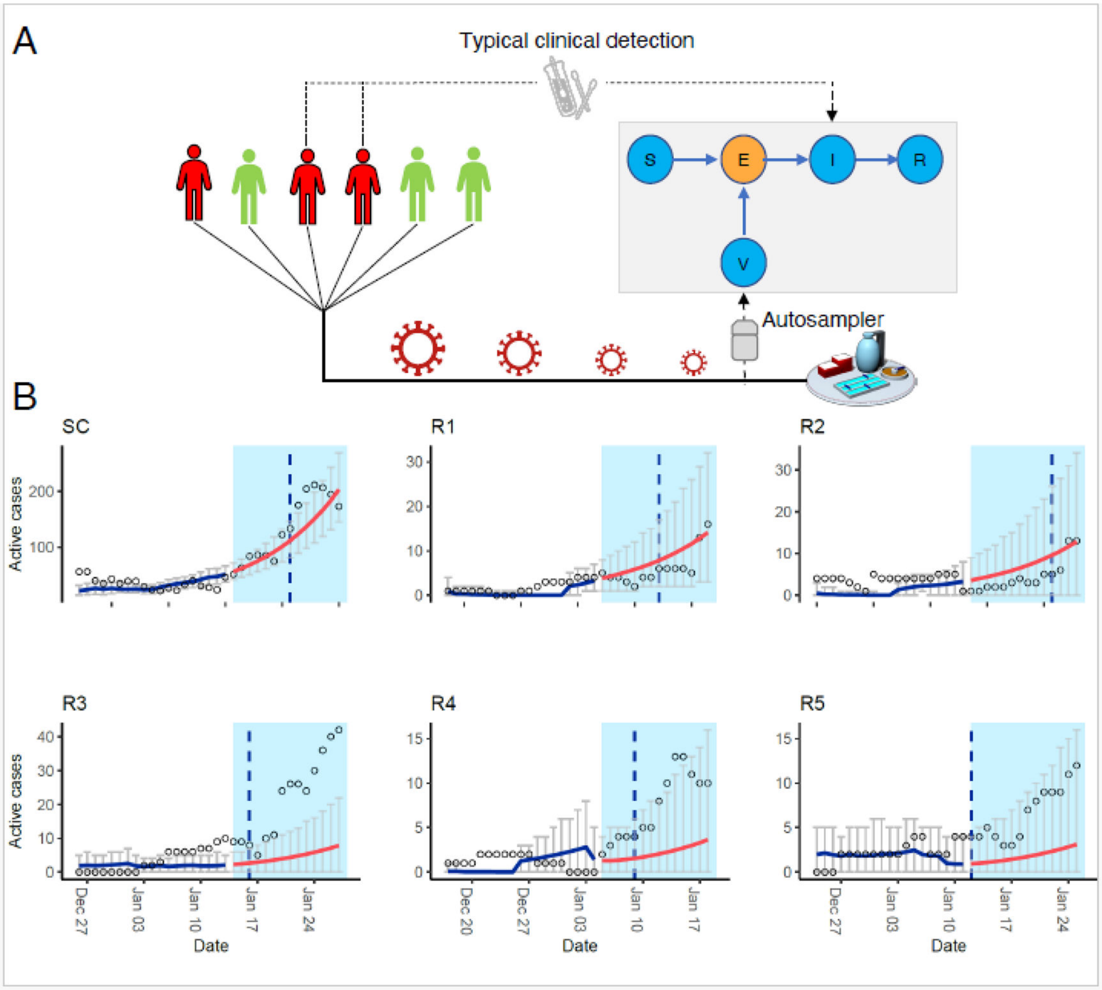


Figure 4: Susceptible-exposed-infectious-recovered model can forecast cases in the early stage of a COVID-19 outbreak.

A) SEIR model framework depicting a population in green with infected people in red. The SARS-CoV-2 shed by a fraction of the exposed population is measured in the wastewater collected at the WWTF. This titer is integrated into a Susceptible (S), Exposed (E), Infected (I), and Recovered (R) model to estimate the number of exposed individuals E. **B)** Left white side contains known data at the time of the forecast where the blue lines show the fitted predicted active cases from wastewater up to the beginning of the outbreak, and the blue shade shows the data not yet observed at the time of forecast whereas the red lines are active cases forecasted. Vertical dashed lines represent the estimated start of the outbreak based on clinically confirmed COVID-19 cases. Using the wastewater data, the model forecasted the start of the outbreak between 0 to 11 days earlier than the onset of the increase in clinical confirmed cases. 95% confidence intervals are shown by the gray bars. Dots show the active cases determined as the 11-day moving sum of the clinically confirmed cases. Since the mean infectious period from fitting data was 10.88 days, we determined the actual active cases as the 11-day moving sum of new clinically confirmed cases. Breakpoints between fitted and forecast values were chosen to be two days after the start of the outbreak, determined by the Piecewise regression model. Cities are ordered by population size (largest on the top left and smallest on the bottom right).

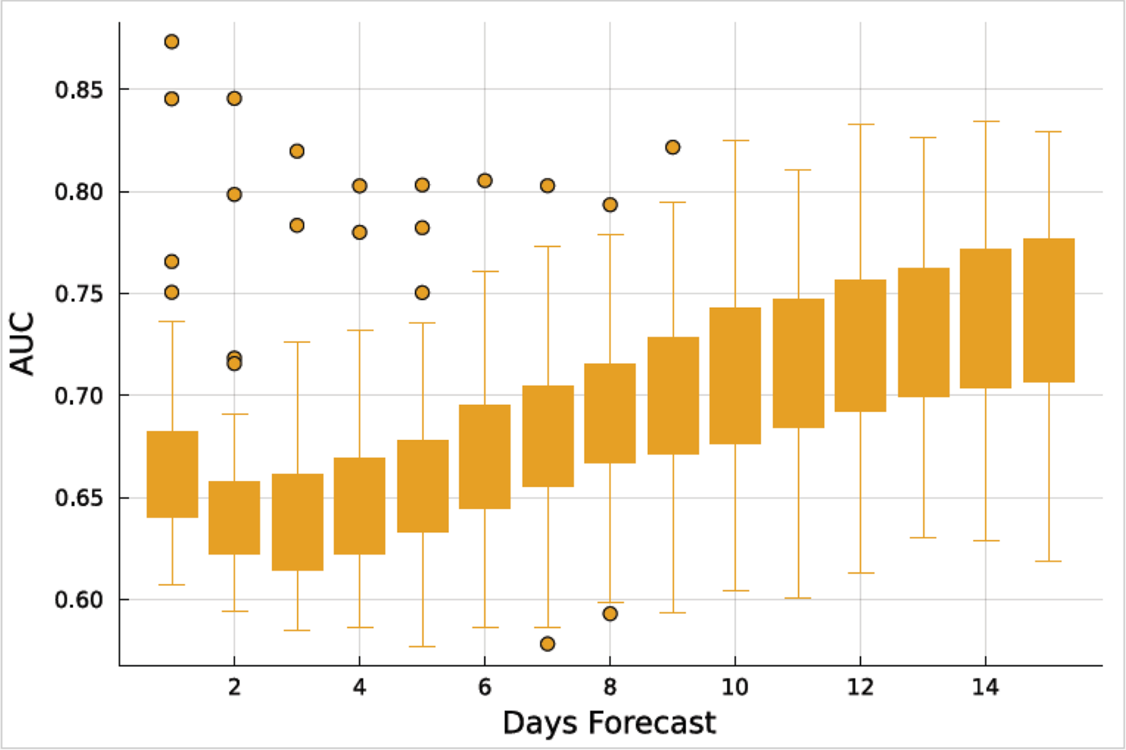


Figure 5: Receiver operating characteristic (ROC) curves for a synthetic data set.

A plot showing the number of true increases against false increases for predicted case counts 1,3,5,7,9,11,13, and 15 days beyond the current measurements. A true increase is counted when there was an increase in cases and the model predicted a greater than α probability of an increase. A false increase is counted when there was no increase in cases but the model predicted a greater than α probability of increase.

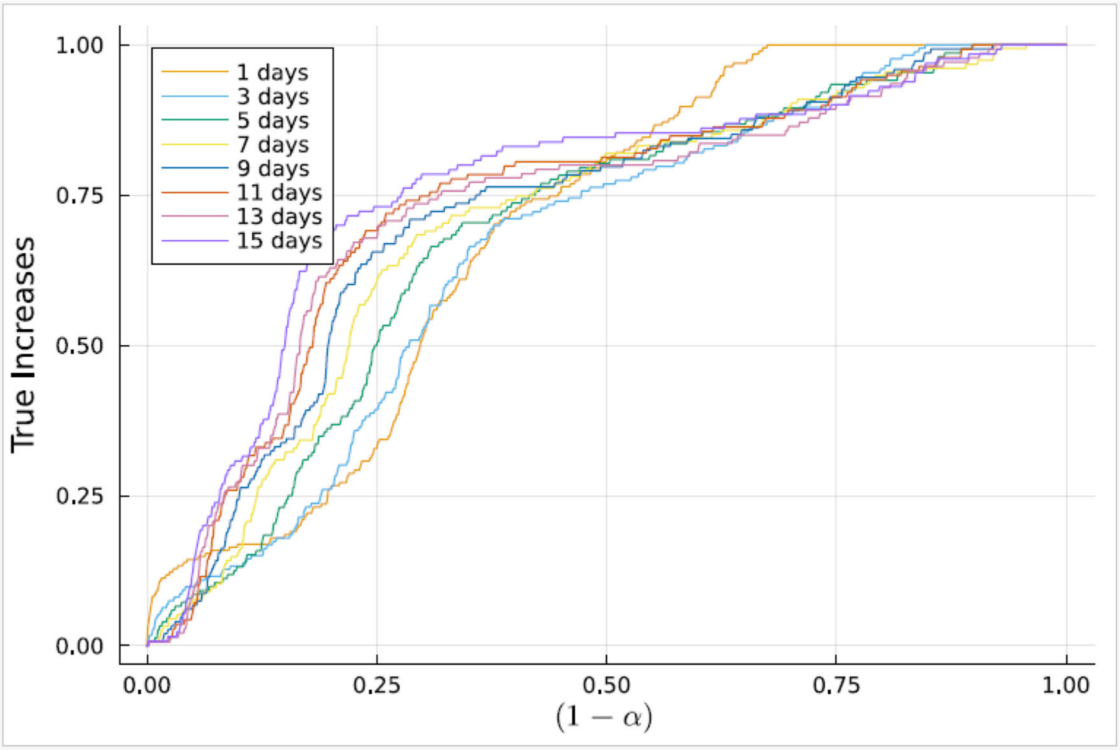


Figure 6: Evaluating SEIR model predictability for an emerging COVID-19 outbreak.

Box plot showing the distribution of measured area under the curve (AUC) when computing 50 receiver operating characteristic (ROC) curves when true positive rate is plotted as function of the false positive rate for prediction forecasted from one to 15 days. A random classifier, which represents the outcome if the model randomly picks predictions, has an AUC of 0.5. The further away the curve is from the one of the random classifiers, the higher the AUC and the better it illustrates the ability of the model to forecast a trend, with the 1 representing the highest accuracy corresponding to 100% positive rate and 0% false negatives. In general, for a diagnostic test to be able to discriminate patients with and without a disease, the AUC must be above 0.5. Values between 0.7 and 0.8 are considered to be ‘fair’ or acceptable (Mandrekar, 2010; Nahm, 2022).

Table 1:

Site characteristics and outbreak detection from the wastewater data and clinically confirmed cases.

Start of outbreaks was measured using a Piecewise regression model.

City	City census[*]	ZIP Code population[*]	Outbreak Start WW	Outbreak Start Cases	Δ Outbreak Start (Day)	Fold change in SARS-CoV-2
SC	25,435	26,739	2022-01-04	2022-01-11	7	7.4
R1	1,030	1,701	2022-01-06	2022-01-13	7	15.2
R2	763+196[†]	2,115	2022-01-01	2022-01-11	10	53.9
R3	890	2,015	2022-01-14	2022-01-16	2	21.7
R4	624	1,167	2022-01-05	2022-01-09	4	42.5
R5	288	985	2022-01-12	2022-01-12	0	81.2

[*] Data from the 2020 Decennial Census obtained from https://data.census.gov/

[†] Wastewater treatment facility collects effluents from a second city.

Table 2:

Parameters used in model fitting.

Parameter	Description	Value	Reference
E_v	Mean virus copies shed by a single individual	4.49 x 10^7 gc/l	(Phan et al., 2023)
V_v	Variance of virus shed by individuals	2 x 10^7 $(gc/l)^2$	(Wölfel et al., 2020)
τ	Incubation period	3 days	(Phan et al., 2023)
δ	Infectious period	8 days	(Phan et al., 2023)
β	Force of infection	Fit from data	
Q	Flow rate of sewershed	Various	Individual treatment plants
	Average faeces produced per day	128 g	(Rose et al., 2015)