1	Article Title: Epidemiological model can forecast COVID-19 outbreaks from wastewater-based
2	surveillance in rural communities.
3	
4	Running Title: Forecasting outbreaks in rural communities
5	
6	Authors:
7	Tyler Meadows <sup>1,3</sup> , Erik R. Coats <sup>2,3</sup> , Solana Narum <sup>2,4</sup> , Eva Top <sup>3,5,6</sup> , Benjamin J. Ridenhour <sup>3,7</sup> ,
8	Thibault Stalder <sup>3,5,6,8 *</sup>
9	
10	Authors Affiliations:
11	<sup>1</sup> Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada
12	<sup>2</sup> Department of Civil and Environmental Engineering, University of Idaho, Moscow, Idaho,
13	USA
14	<sup>3</sup> Institute for Modeling Collaboration and Innovation (IMCI), University of Idaho, Moscow,
15	Idaho, USA
16	<sup>4</sup> Bioinformatics and Computational Biology Graduate Program (BCB), Moscow, Idaho, USA
17	<sup>5</sup> Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA
18	<sup>6</sup> Institute for Interdisciplinary Data Sciences (IIDS), University of Idaho, Moscow, Idaho, USA
19	<sup>7</sup> Department of Mathematics and Statistical Science, University of Idaho, Moscow, Idaho, USA
20	<sup>8</sup> INSERM, CHU Limoges, RESINFIT, U1092, Univ. Limoges, F-87000, Limoges, France

21	* Corresponding Author: Centre de Biologie et de Recherche en Santé – CBRS – 3ème étage,
22	UMR 1092 RESINFIT, 1 rue du Professeur Bernard Descottes, 87025 Limoges, France, Phone:
23	+33 (0)5 19 56 42 60, email: <u>thibault.stalder@unilim.fr</u>
24	
25	Keyword: SARS-CoV-2, COVID-19, Wastewater-based epidemiology (WBE), SEIR, rural
26	communities
27	
28	Abstract
29	Wastewater can play a vital role in infectious disease surveillance, especially in underserved
30	communities where it can reduce the equity gap to larger municipalities. However, using
31	wastewater surveillance in a predictive manner remains a challenge. We tested if detecting
32	SARS-CoV-2 in wastewater can predict outbreaks in rural communities. Under the CDC
33	National Wastewater Surveillance program, we monitored several rural communities in Idaho
34	(USA). While high daily variations in wastewater viral load made real-time interpretation

difficult, a SEIR model could factor out the data noise and forecast the start of the Omicron

36 outbreak in five of the six cities that were sampled soon after SARS-CoV-2 quantities increased

37 in wastewater. For one city, the model could predict an outbreak 11 days before reported clinical

cases began to increase. An epidemiological modeling approach can transform how

epidemiologists use wastewater data to provide public health guidance on infectious diseases in

40 rural communities.

### 41 Main text

#### 42 Introduction.

43 Wastewater-based epidemiology (WBE) is a promising approach for broad-scale, agnostic surveillance of infectious diseases and antimicrobial resistance within and across communities. 44 Indeed, many infectious agents such as SARS-CoV-2, poliovirus, RSV, and flu shed through 45 stool (1,2), and potentially urine (3), and are thus detectable in wastewater (4,5). Additionally, 46 47 WBE assesses infectious disease circulation in both symptomatic and asymptomatic populations. Most importantly, WBE can improve and accelerate the early detection of infectious disease 48 outbreaks by public health authorities, providing actionable data for epidemiologists. However, 49 barriers remain to how epidemiologists might use or interpret this data to inform public health in 50 51 a predictive way rather than a retrospective approach. A modeling-based approach to wastewater data can provide the framework to interpret infectious disease spread and burden by estimating 52 epidemiological parameters such as incidence (6), prevalence (7–9), or effective reproductive 53 54 number (10).

WBE also presents a unique opportunity to support vulnerable and underserved communities 55 (11). In particular, rural communities are at a higher risk of severe outcomes associated with 56 57 certain infectious diseases due to demographic factors (e.g., age), underlying healthcare 58 challenges (e.g. obesity, smoking), limited resources (12–15), or reduced risk perception (16). Moreover, as rural communities often lack the resources for broader-scale clinical testing, WBE 59 can help sustain rural community health. Unfortunately, WBE has mainly focused on urban areas 60 61 and larger cities, leaving rural communities with reduced access to this critical data (11,17–20). In that regard, WBE is a tool that promotes equity (21). 62

63	The COVID-19 pandemic provided a unique opportunity to leverage and expand WBE while
64	concurrently establishing greater value for epidemiologists. Concentrations of SARS-CoV-2
65	found in wastewater correlate well with the number of COVID-19 cases (22–24). Thus, with the
66	right tools in place WBE can serve as an early warning for a potential COVID-19 outbreak
67	(4,25–28). Perhaps more critically for epidemiologists, the utility of wastewater-based detection
68	of SARS-CoV-2 has become even more significant with reduced COVID-19 clinical testing due
69	to home tests and general lack of clinical reporting (29). However, there remains a gap in
70	translating WBE data in a timely manner for use by public health officials. Moreover, a primary
71	obstacle for epidemiologists is the variability and uncertainty inherent in wastewater monitoring
72	(30,31).
73	Modeling SARS-CoV-2 wastewater data represents an opportunity to bridge the information
74	gap with epidemiologists while bringing enhanced and timely health monitoring to rural
75	communities. Here we present and discuss a susceptible-exposed-infectious-recovered (SEIR)

repidemiological model developed based on wastewater detection of SARS-CoV-2 for

epidemiological surveillance of COVID-19 in rural area. We specifically tested the hypothesis

that the detection and quantification of SARS-CoV-2 in wastewater forecasts the start of an

79 outbreak in rural communities.

80

# 81 Material & Methods.

## 82 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

85	county in Idaho, USA (Table 1). These rural communities are defined as "rural" according to the
86	2020 U.S. Census Bureau. The county itself falls under the category of a "rural" or
87	"nonmetropolitan" county, as classified by the U.S. Office of Management and Budget. Rural
88	cities are abbreviated RC1 to RC5 and the small city SC. All WWTFs primarily treat domestic
89	wastewater; the SC WWTF also receives effluent from a regional hospital.
90	Samples were collected three times a week from October 2021 to March 2022. Rural WWTF
91	samples were time-composite samples collected using Teledyne ISCO 3700 Full Size Portable
92	Sampler (Teledyne ISCO, Lincoln, NE, USA) autosamplers or homemade autosamplers
93	constituted of a Sci-Q 323 peristaltic pump (Watson-Marlow, Falmouth, UK) controlled by an
94	Omron H3CR timer (Omron Corporation, Kyoto, Japan) and housed in a cooler box. Sampling
95	frequencies were comprised between 10 and 30 minutes for 24h. Approximately 3L of
96	wastewater was collected, and subsamples were collected at the end of the 24h sampling period
97	and transported within 6h to the laboratory, where samples were kept at 4°C until further
98	processing. Samples from the SC WWTF were collected using a Teledyne ISCO model 3700
99	autosampler (Teledyne ISCO, Lincoln, NE, USA), with samples collected paced with influent
100	flow. Sampling failed for less than 10% of the total samples sampled. Samples were kept at 4°C
101	until further processed, at most 3 days later.
102	Confirmed COVID-19 case counts per zip code were obtained from the Idaho Public Health
103	District 2 website (https://idahopublichealth.com/district-2/novel-coronavirus).
104	Sample processing for SARS-CoV-2 detection and quantification.
105	The detailed protocols presented below are publicly available on protocol.io (32). In brief,

- 106 before concentrating the viral fraction of two replicate wastewater fractions through
- 107 electronegative membrane filtration, each sample was spiked with the Bovilis® Coronavirus

108	(BCoV) (Merck, Kenilworth, NJ, USA) as a process internal control. Subsequently, filters were
109	inserted together with the DNA/RNA Shield <sup>TM</sup> (Zymo Research, Irvine, CA, USA) into the Lysis
110	Bead tubes from the AllPrep® PowerViral® DNA/RNA Kit (QIAGEN, Inc., Germantown, MD,
111	USA). Lysis was performed on a FastPrep <sup>™</sup> (MP Biomedicals, Santa Ana, CA, USA) for 4
112	cycles of 20 seconds each at 4.5 m/s and the RNA was then extracted as per the kit
113	manufacturer's protocol on a QIAcube Connect automated extraction instrument (QIAGEN, Inc.,
114	Germantown, MD, USA).
115	SARS-CoV-2 was quantified by dPCR using the QIAcuity Digital PCR System (QIAGEN,
116	Inc., Germantown, MD, USA) using the GT-Digital SARS-CoV-2 Wastewater Surveillance
117	Assay For QIAcuity® (GT Molecular, Fort Collins, CO, USA). Each 40 µl reaction contained 1x
118	of the Qiagen QIAcuity One-Step Viral RT-PCR Kit (QIAGEN, Inc., Germantown, MD, USA),
119	1x of the GT Molecular N1-N2-BCoV Assay Solution, and 20 $\mu$ l RNA template. RNA extraction
120	blanks, dPCR non-template controls and positive controls were included in each dPCR run.
121	Data processing and analysis
122	Fluorescent thresholds were manually set based on the fluorescent level of the positive
123	controls. Then we excluded data from samples for which (i) the recovery rate of the internal
124	processing control BCoV was lower than 1%, or (ii) the RNA extraction process control or dPCR
125	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 (33).

135 SARS-CoV-2 Epidemiological Model.

We constructed a compartmental model to approximate the dynamics of the epidemic in each 136 city. Due to the small size of populations in the rural areas, we expected stochastic effects to be 137 important, and opted to use a discrete-time discrete-state Markov process to approximate the 138 139 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 140 141 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 142 number of newly exposed individuals on day t,  $Y_t \sim Bin\left(E(t), \frac{1}{\tau}\right)$  is the number of newly 143 infectious individuals on day t, and  $Z_t \sim Bin\left(I(t), \frac{1}{\delta}\right)$  is the number of newly recovered 144 individuals on day t. The parameter  $\beta$  is the transmission rate,  $\tau$  is the mean incubation period, 145 and  $\delta$  is the mean infectious period, and N(t) = S(t) + E(t) + I(t) + R(t) is the total 146 population at time t. The discrete-time Markov process is given by: 147

148

- 149 (1a)  $\Delta S(t) = -X_t$ ,
- 150 (1b)  $\Delta E(t) = X_t Y_t$ ,
- 151 (1c)  $\Delta I(t) = Y_t Z_t$ ,

152 (1d) 
$$\Delta R(t) = Z_t$$

153

The number of virus particles shed by exposed individuals was assumed to be log-normally 154 distributed (34). However, the log-normal distribution is difficult to work with mathematically, 155 so we approximated the log-normal distribution using the Gamma distribution by matching the 156 157 first two moments. For simplicity, we assumed that only the individuals in the exposed class 158 E(t) shed virus in the stools. This assumption is reasonable since it has been shown that the 159 amount of virus shed by a single individual is time-varying, with a peak occurring around symptoms onset (35,36). If there are E(t) exposed individuals, the amount of virus in the 160 161 wastewater is a random variable V(t) with probability density function:

162

163 (2) 
$$\Phi(V; k, \theta, E) = \gamma(V; Ek, \theta),$$

164 where  $\gamma(V; k, \theta)$  is the probability density function for the gamma distribution with rate  $k = \frac{E_v^2}{V_v}$ 165 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.

172	Each time step, every particle evolved according to the Markov process in equation 1 (Figure
173	1 bottom left). On days that we have collected wastewater data, the particles were weighted
174	according to their likelihood (Figure 1 bottom center) and resampled using a systematic sampling
175	method to filter out the least likely particles and reinforce the most likely particles (Figure 1
176	bottom right).
177	Data and scripts of this study are available on https://github.com/Tyler-Meadows/wastewater-
178	surveillance.
179	
100	Desults
190	Kesuits.
181	Dynamics of SARS-CoV-2 in rural wastewater vs. clinically confirmed cases.
182	For the period investigated, clinically reported cases revealed that the cities experienced one
183	or two COVID-19 outbreaks, as shown in Figure 2. The first outbreak occurred in late October
184	2021, but it was not detected in all cities and was relatively small compared to the second
185	outbreak experienced by all cities in early January 2022. This second surge was driven by the
186	Omicron variant, which emerged in the United States in early December 2021 (37).
187	Examining the collected wastewater data, 293 samples were retained following data
188	processing, with each WWTF yielding 44-58 measurements over a five-month period. The daily
189	load of SARS-CoV-2 present in wastewater varied greatly day-to-day, making interpretation of
190	the real-time spread of COVID-19 challenging. Interestingly, the variability in the order of
191	magnitude tended to be larger as the city population decreased, indicating a greater level of
192	randomness (as shown in Figure 3). Additionally, the difference in variance of daily SARS-CoV-
193	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 197 the virus collected at the WWTFs above the background levels (Figure 2) by 7- to 81-fold. After 198 199 estimating the date of the outbreak start (vertical dashed line in Figure 2), we estimated that the 200 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 201 wastewater surveillance could improve and even accelerate the early detection of infectious 202 203 diseases in rural communities (36,38,39). However, the lead times were variable, and in one case, the wastewater signal was not inferred to precede the clinically reported case data. Other 204 studies have also observed cases where wastewater signal was not preceding clinical testing 205 (38, 40).206

207

208 Epidemiological model to forecast a COVID-19 outbreak from wastewater detection of the
209 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 217 confirmed cases we found that the model successfully captured the clinically active cases in five 218 of the six cities. For the rural cities, the forecasts were within the 95% confidence interval at 219 least up to eight days ahead; corresponding to R4 (Figure 4). While the number of predicted 220 221 cases tended to be lower than the number of the clinically confirmed cases reported, they were 222 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. 223 In city SC, the model was able to forecast the number of active cases accurately more than eight 224 225 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 226 increase. Anticipation of the model on the COVID-19 reported cases ranged from 0 days for R5 227 to 11 days for R2 (see blue dashed line on Figure 4). 228

229 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 230 trend of active cases. To that end, we used a simulated wastewater dataset to count the number of 231 times the model predicted upward or downward trends in cases correctly (i.e., true positive rate) 232 or incorrectly (i.e., false positive rate). We varied the threshold used to accept predictions to 233 234 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 235 236 range of forecasted days. The model tends to predict trends better as the forecast range increases. 237 The ROC curves for predictions made less than a week in advance were significantly lower than 238 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.

241

#### 242 Discussion

Wastewater-based detection of SARS-CoV-2 has been primarily focused on large urban and metropolitan areas and much less on rural towns. Here we address this gap by having conducted a surveillance effort of the SARS-CoV-2 in the wastewaters of several rural cities of the state of Idaho (USA) since October 2021. Below we discuss the barriers associated with WBE in rural areas and how SEIR modeling could help overcome some of these challenges by forecasting trends in COVID-19 cases based on wastewater-based measurements.

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 (41). 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.

However, as we experienced, the variability in viral quantities arriving at the rural WWTF per day made the determination of the start of the outbreak in real-time difficult without retrospective statistical analysis. As we observed, the variability in the order of magnitude tended to be larger as the rural sewershed size decreased (42), making interpretation of wastewaterbased detection of infectious diseases in rural areas even more complicated for public health.

This noise is typically lower in larger sewersheds that are more active over a 24-hr time compared to smaller ones (43).

SEIR epidemiological models offer a framework for epidemiologists to analyze the dynamics 262 of outbreaks using wastewater data (7,8,44–48). Specifically, in our study, we demonstrate that 263 our SEIR model can provide reliable forecasts of when case numbers are trending upwards in 264 rural communities. Sensitivity and specificity assessments of the model in predicting start of the 265 outbreak revealed that the forecasts were more reliable when looking at trends beyond a week, 266 with the best forecast being over a period of nine to 15 days. Similar ranges of short-term 267 forecasts of seven to nine days to predict upcoming cases based on wastewater detection of the 268 269 SARS-CoV-2 were reported for larger cities using SEIR or SEIR-like approaches (8,46,48). The fact that the model was not performing well under 9 days may, in part, be attributed to the noise 270 271 of the SARS-CoV-2 quantities in rural wastewater that creates a sawtooth pattern in the trend. 272 Thus, while the trend over a week goes up, there is a chance that some of the points between went down. This led us to hypothesize that the noise in the wastewater data may have made 273 short-term trend predictions less accurate. 274

275 While our SEIR model functions well to provide advanced warning of a COVID-19 276 outbreak, it failed to predict the outbreak peak (Fig. S1) and we could not assess the accuracy of 277 case number predictions. Predicted cases tended to be lower than reported clinical cases for the 278 area, contrasting with other studies on larger sewersheds that have shown that SEIR models 279 typically estimate more COVID-19 cases than the reported number (45,53,54). The underestimation of COVID-19 cases observed in this study may be attributed, in part, to 280 differences between the population of the city sampled and the zip code used (Table 1); the latter 281 corresponds to the clinically recorded cases. When comparing city with zip code census, 39 to 282

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 should be approached with
caution. It is important to note that we did not intend to use the model to predict case numbers;
the study was not designed for this purpose. Instead, it was intended to forecast an increase in
COVID-19 cases to provide an early warning that could be shared with the community.

Despite the successful application of SEIR models to wastewater surveillance of the SARS-289 CoV-2, there are some uncertainties on how to connect the model with wastewater data. Some 290 authors have included the cumulative virus titer in the sewershed as a dynamic variable (8) or as 291 292 a linear combination of other dynamic variables (45). However, this approach can be problematic when measurements are sparse or there are gaps between collection periods – which would be 293 very common in rural WWTFs. Most authors directly connect wastewater measurements to the 294 295 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 296 model and wastewater measurements is not well established. Contribution to viral load in 297 wastewater can tie to the individuals in the 'I' compartment (7,8,45) or to the 'E' compartment; 298 299 the SEIR model in this study resulted in better predictions than when connected to 'I'. This 300 difference could be attributed to the fact that peak virus shedding in stool may occur for a few days around the onset of symptoms (35,36). This means that individuals contributing to the load 301 302 of SARS-CoV-2 measured in wastewater may be at the transition between 'E' and 'I' 303 compartments in the SEIR model. Some researchers have incorporated the results from 304 wastewater measurement in both 'E' and 'I' (44) while others created a new compartment 305 structure to account for viral shedding dynamic in wastewater (47,48).

306	Finally, our model was able to successfully forecast the upcoming cases in most of the cities
307	surveyed. In one city this resulted in predicting the outbreak as much as 11 days before reported
308	clinical cases started to increase. However, it failed for one city, suggesting that the disease
309	dynamics do not always follow the model assumptions. This may be because the model does not
310	include potential traveling between cities, which may impact the accuracy of predictions,
311	especially in rural areas where residents often have to commute to work. In the rural cities
312	surveyed between 79.4 and 94.5% of residents work outside their place of residence, versus
313	27.3% in the small city surveyed (data from U.S. Census Bureau Topic: Commuting – Survey:
314	American Community Survey – 2021, ACS 5-Year Estimates Subject Tables).
315	In conclusion, our study reveals that wastewater-based epidemiology (WBE) in rural
316	communities and small sewersheds in general is associated with high daily variation in SARS-
317	CoV-2 levels. This variation creates a challenge for epidemiologists who seek to monitor real-
318	time data in rural areas based solely on the raw data. However, our research also shows that the
319	SEIR modeling approach can help to decipher this data and actually predict the start of
320	outbreaks. Our model provides a suitable framework for epidemiologists to analyze the dynamics
321	of outbreaks using wastewater data.

322

#### 323 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

329	funded by the Center for Disease Control and Prevention (CDC) through the Epidemiology and
330	Laboratory Capacity Enhancing Detection Through Coronavirus Release and Relief (CRR)
331	Supplement Funds. The content of this project is solely the responsibility of the authors and does
332	not necessarily represent the official views of the State of Idaho Department of Health and
333	Welfare or the Center for Disease Control and Prevention (CDC).
334	
335	References
336	1. Parasa S, Desai M, Chandrasekar VT, Patel HK, Kennedy KF, Roesch T, et al. Prevalence of
337	gastrointestinal symptoms and fecal viral shedding in patients with coronavirus disease 2019:
338	A systematic review and meta-analysis. JAMA Netw Open. 2020 Jun 1;3(6):e2011335-
339	e2011335.
340	2. Arts PJ, Kelly JD, Midgley CM, Anglin K, Lu S, Abedi GR, et al. Longitudinal and
341	quantitative fecal shedding dynamics of SARS-CoV-2, pepper mild mottle virus, and
342	crAssphage. mSphere. 2023 Jun 20;0(0):e00132-23.
343	3. Jeong HW, Kim SM, Kim HS, Kim YI, Kim JH, Cho JY, et al. Viable SARS-CoV-2 in
344	various specimens from COVID-19 patients. Clin Microbiol Infect. 2020 Nov;26(11):1520-
345	4.
346	4. Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, Pang J. Wastewater surveillance to infer
347	COVID-19 transmission: A systematic review. Sci Total Environ. 2022 Jan 15;804:150060.

348	5. Corpuz MVA, Buonerba A, Vigliotta G, Zarra T, Ballesteros F, Campiglia P, et al. Viruses in
349	wastewater: occurrence, abundance and detection methods. Sci Total Environ. 2020 Nov
350	25;745:140910.
351	6. McManus O, Christiansen LE, Nauta M, Krogsgaard LW, Bahrenscheer NS, Kappelgaard L
352	von, et al. Predicting COVID-19 incidence using wastewater surveillance data, Denmark,
353	October 2021–June 2022. Emerg Infect Dis. 2023 Aug;29(8):1589–97.
354	7. McMahan CS, Self S, Rennert L, Kalbaugh C, Kriebel D, Graves D, et al. COVID-19
355	wastewater epidemiology: a model to estimate infected populations. Lancet Planet Health.
356	2021 Dec 1;5(12):e874-81.
357	8. Phan T, Brozak S, Pell B, Gitter A, Xiao A, Mena KD, et al. A simple SEIR-V model to
358	estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-
359	based surveillance data. Sci Total Environ. 2023 Jan 20;857:159326.
360	9. Layton BA, Kaya D, Kelly C, Williamson KJ, Alegre D, Bachhuber SM, et al. Evaluation of a
361	wastewater-based epidemiological approach to estimate the prevalence of SARS-CoV-2
362	infections and the detection of viral variants in disparate Oregon communities at city and
363	neighborhood scales. Environ Health Perspect. 2022 Jun 29;130(6):067010.
364	10. Huisman JS, Scire J, Caduff L, Fernandez-Cassi X, Ganesanandamoorthy P, Kull A, et al.
365	Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. Environ
366	Health Perspect. 2022 May;130(5):057011.

367	11. Holm RH, Osborne Jelks N, Schneider R, Smith T. Beyond COVID-19: designing inclusive
368	public health surveillance by including wastewater monitoring. Health Equity. 2023
369	Jun;7(1):377–9.
370	12. Lakhani HV, Pillai SS, Zehra M, Sharma I, Sodhi K. Systematic review of clinical insights
371	into novel coronavirus (CoVID-19) pandemic: persisting challenges in U.S. rural population.
372	Int J Environ Res Public Health. 2020 Jun 15;17(12):4279.
373	13. Kaufman BG, Whitaker R, Pink G, Holmes GM. Half of rural residents at high risk of
374	serious illness due to COVID-19, creating stress on rural hospitals. J Rural Health. 2020 Jun
375	30;36(4):584–90.
376	14. Cuadros DF, Branscum AJ, Mukandavire Z, Miller FD, MacKinnon N. Dynamics of the
377	COVID-19 epidemic in urban and rural areas in the United States. Ann Epidemiol. 2021
378	Jul;59:16–20.
379	15. Cromartie J, Dobis, Elizabeth A., Krumel, Thomas P., McGranahan, David, Pender, John.
380	Rural America at a glance: 2020 edition. United States Department of Agriculture Economic
381	Research Service: Economic Information Bulletin. 2020 Dec;221:6.
382	16. Ridenhour BJ, Sarathchandra D, Seamon E, Brown H, Leung FY, Johnson-Leon M, et al.
383	Effects of trust, risk perception, and health behavior on COVID-19 disease burden: Evidence
384	from a multi-state US survey. PLoS One. 2022 May 20;17(5):e0268302.
385	17. Conway MJ, Kado S, Kooienga BK, Sarette JS, Kirby MH, Marten AD, et al. SARS-CoV-2
386	wastewater monitoring in rural and small metropolitan communities in Central Michigan. Sci
387	Total Environ. 2023 Oct 10;894:165013.

388	18. D'Aoust PM, Towhid ST, Mercier É, Hegazy N, Tian X, Bhatnagar K, et al. COVID-19
389	wastewater surveillance in rural communities: Comparison of lagoon and pumping station
390	samples. Sci Total Environ. 2021 Dec 20;801:149618.
391	19. Toledo DM, Robbins AA, Gallagher TL, Hershberger KC, Barney RE, Salmela SM, et al.
392	Wastewater-based SARS-CoV-2 surveillance in Northern New England. Microbiol Spectr.
393	2022 Apr 27;10(2):e0220721.
394	20. Jarvie MM, Reed-Lukomski M, Southwell B, Wright D, Nguyen TNT. Monitoring of
395	COVID-19 in wastewater across the Eastern Upper Peninsula of Michigan. Environ Adv.
396	2023 Apr;11:100326.
397	21. Medina CY, Kadonsky KF, Roman FA, Tariqi AQ, Sinclair RG, D'Aoust PM, et al. The
398	need of an environmental justice approach for wastewater based epidemiology for rural and
399	disadvantaged communities: A review in California. Curr Opin Environ Sci Health. 2022

400 Jun;27:100348.

22. Wurtzer S, Marechal V, Mouchel JM, Maday Y, Teyssou R, Richard E, et al. 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. 2020 Dec
17;25(50):2000776.

23. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al.
Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics.
Nat Biotech. 2020 Oct;38(10):1164–7.

408	24.	Fernandez-Cassi X, Scheidegger A, Bänziger C, Cariti F, Tuñas Corzon A,
409		Ganesanandamoorthy P, et al. Wastewater monitoring outperforms case numbers as a tool to
410		track COVID-19 incidence dynamics when test positivity rates are high. Water Res. 2021 Jul
411		15;200:117252.
412	25.	Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, et al. Wastewater-based
413		epidemiology: Global collaborative to maximize contributions in the fight against COVID-
414		19. Environ Sci Technol. 2020 Apr 12;
415	26.	Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, et al.
416		Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in
417		sewersheds. Sci Total Environ. 2021 Jun 25;775:145790.
418	27.	Li L, Mazurowski L, Dewan A, Carine M, Haak L, Guarin TC, et al. Longitudinal
419		monitoring of SARS-CoV-2 in wastewater using viral genetic markers and the estimation of
420		unconfirmed COVID-19 cases. Sci Total Environ. 2022 Apr 15;817:152958.
421	28.	Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, et al. Wastewater surveillance of
422		SARS-CoV-2 across 40 U.S. states from February to June 2020. Water Res. 2021 Sep
423		1;202:117400.
424	29.	Fontenele RS, Yang Y, Driver EM, Magge A, Kraberger S, Custer JM, et al. Wastewater
425		surveillance uncovers regional diversity and dynamics of SARS-CoV-2 variants across nine
426		states in the USA. Sci Total Environ. 2023 Jun 15;877:162862.

427	30.	Diamond MB, Keshaviah A, Bento AI, Conroy-Ben O, Driver EM, Ensor KB, et al.
428		Wastewater surveillance of pathogens can inform public health responses. Nat Med. 2022
429		Oct;28(10):1992–5.
430	31.	McClary-Gutierrez JS, Mattioli MC, Marcenac P, Silverman AI, Boehm AB, Bibby K, et al.
431		SARS-CoV-2 wastewater surveillance for public health action. Emerg Infect Dis. 2021
432		Sep;27(9):e210753.
433	32.	Narum S, Stalder T, Coats E, Top E. Quantification of the SARS-CoV-2 using
434		electronegative membrane filtration and dPCR. Protocols.io [Internet]. 2022 Mar 28 [cited
435		2023 Apr 18]; Available from: https://www.protocols.io/view/quantification-of-the-sars-cov-
436		2-using-electronega-b6udres6
437	33.	Muggeo VM, Atkins DC, Gallop RJ, Dimidjian S. Segmented mixed models with random
438		changepoints: a maximum likelihood approach with application to treatment for depression
439		study. Stat Modelling. 2014 Aug 1;14(4):293-313.
440	34.	Miura F, Kitajima M, Omori R. Duration of SARS-CoV-2 viral shedding in faeces as a
441		parameter for wastewater-based epidemiology: Re-analysis of patient data using a shedding
442		dynamics model. Sci Total Environ. 2021 May 15;769:144549.
443	35.	Puhach O, Meyer B, Eckerle I. SARS-CoV-2 viral load and shedding kinetics. Nat Rev
444		Microbiol. 2023 Mar;21(3):147-61.
445	36.	Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, et al. SARS-CoV-2 RNA
446		concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-
447		19 cases. Sci Total Environ. 2022 Jan 20;805:150121.

448 3'	7. CDC COVID	-19 Response Team.	SARS-CoV-2 B.1.1.529	(Omicron)	) Variant — United
--------	--------------	--------------------	----------------------	-----------	--------------------

- 449 States, December 1–8, 2021. MMWR Morb Mortal Wkly Rep. 2021 Dec;70.
- 450 38. Feng S, Roguet A, McClary-Gutierrez JS, Newton RJ, Kloczko N, Meiman JG, et al.
- 451 Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2
- 452 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. ACS
- 453 EST Water. 2021 Aug 13;1(8):1955–65.
- 454 39. Graham KE, Loeb SK, Wolfe MK, Catoe D, Sinnott-Armstrong N, Kim S, et al. SARS-CoV-
- 455 2 RNA in wastewater settled solids is associated with COVID-19 cases in a large urban
- 456 sewershed. Environ Sci Technol. 2021 Jan 5;55(1):488–98.
- 457 40. Xiao A, Wu F, Bushman M, Zhang J, Imakaev M, Chai PR, et al. Metrics to relate COVID-
- 45819 wastewater data to clinical testing dynamics. Water Res. 2022 Apr 1;212:118070.
- 459 41. Park S, Marcus GM, Olgin JE, Carton T, Hamad R, Pletcher MJ, et al. Unreported SARS-
- 460 CoV-2 home testing and test positivity. JAMA Netw Open. 2023 Jan 25;6(1):e2252684.
- 461 42. Schill R, Nelson KL, Harris-Lovett S, Kantor RS. The dynamic relationship between
- 462 COVID-19 cases and SARS-CoV-2 wastewater concentrations across time and space:
- 463 Considerations for model training data sets. Sci Total Environ. 2023 May 1;871:162069.
- 464 43. Tchobanoglous G, Stensel DH, Tsuchihashi R, Burton F, Abu-Orf M, Bowden G, et al.,
- 465 editors. Wastewater engineering: treatment and resource recovery. Fifth edition. New York:
- 466 McGraw-Hill Education; 2014. 2018 p.

467	44.	Fazli M, Sklar S, Porter MD, French BA, Shakeri H. Wastewater-based epidemiological
468		modeling for continuous surveillance of COVID-19 outbreak. In: 2021 IEEE International
469		Conference on Big Data (Big Data). 2021. p. 4342–9.
470	45.	Nourbakhsh S, Fazil A, Li M, Mangat CS, Peterson SW, Daigle J, et al. A wastewater-based
471		epidemic model for SARS-CoV-2 with application to three Canadian cities. Epidemics. 2022
472		Jun 1;39:100560.
473	46.	Pájaro M, Fajar NM, Alonso AA, Otero-Muras I. Stochastic SIR model predicts the
474		evolution of COVID-19 epidemics from public health and wastewater data in small and
475		medium-sized municipalities: A one year study. Chaos Solitons Fractals. 2022
476		Nov;164:112671.
477	47.	Polcz P, Tornai K, Juhász J, Cserey G, Surján G, Pándics T, et al. Wastewater-based
478		modeling, reconstruction, and prediction for COVID-19 outbreaks in Hungary caused by
479		highly immune evasive variants. Water Res. 2023 Aug 1;241:120098.
480	48.	Proverbio D, Kemp F, Magni S, Ogorzaly L, Cauchie HM, Gonçalves J, et al. Model-based
481		assessment of COVID-19 epidemic dynamics by wastewater analysis. Sci Total Environ.
482		2022 Jun 25;827:154235.
483	49.	Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological
484		assessment of hospitalized patients with COVID-2019. Nature. 2020 May;581(7809):465-9.
485	50.	Rose C, Parker A, Jefferson B, Cartmell E. The characterization of feces and urine: A review
486		of the literature to inform advanced treatment technology. Crit Rev Environ Sci Technol.
487		2015 Sep 2;45(17):1827–79.

# 489 Table 1: Site characteristics and outbreak detection from the wastewater data and

490 clinically confirmed cases. Start of outbreaks was measured using a Piecewise regression

## 491 model.

City	City	ZIP Code	Outbreak	Outbreak	∆ Outbreak	Fold change in
	census*	population*	Start WW	Start Cases	Start (Day)	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 <sup>†</sup>	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 <u>https://data.census.gov/</u>						
<sup>†</sup> Wastewater treatment facility collects effluents from a second city.						

492

493

# 494 **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 <sup>7</sup> gc/l	(8)
$V_{v}$	Variance of virus shed by individuals	2 x 10 <sup>7</sup> (gc/l) <sup>2</sup>	(49)
τ	Incubation period	3 days	(8)
δ	Infectious period	8 days	(8)
β	Force of infection	Fit from data	
Q	Flow rate of sewershed	Various	Individual treatment plants
	Average faeces produced per day	128 g	(50)

495

#### 496 FIGURE LEGENDS



Figure 1: Diagram showing steps of the particle filter method we use to determine the 498 number of active cases from the wastewater titers of SARS-CoV-2. The particle filter is 499 500 initialized using the first measurement of virus concentration in the wastewater. We generate a 501 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 502 503 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 504 a particle. The darker dots in the diagram signify a higher number of particles with that value of 505 E. We apply one step of the stochastic SEIR model to each particle to predict the number of 506 507 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 508

509 likely particles are filtered out using a systematic resampling procedure and replaced with more



510 likely particles (bottom right graph).



512

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
- 516 moving average (red line). Vertical dash lines represent the estimated start of the outbreak using
- 517 either the cumulative sum of the copies per day of the N1 target or the cumulative sum of
- 518 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-
- 520 CoV-2 and clinically confirmed COVID-19 cases. Cities are ordered by population size (largest
- 521 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

- 529 calculated using the log10 of copies per day of the SARS-CoV-2. This essentially shows that the
- 530 magnitude of the estimate is less consistent as population size gets smaller (i.e., more
- 531 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 538 539 lines show the fitted predicted active cases from wastewater up to the beginning of the outbreak, and the blue shade shows the data not vet observed at the time of forecast whereas the red lines 540 are active cases forecasted. Vertical dashed lines represent the estimated start of the outbreak 541 based on clinically confirmed COVID-19 cases. Using the wastewater data, the model forecasted 542 the start of the outbreak between 0 to 11 days earlier than the onset of the increase in clinical 543 confirmed cases. 95% confidence intervals are shown by the gray bars. Dots show the active 544 cases determined as the 11-day moving sum of the clinically confirmed cases. Since the mean 545 infectious period from fitting data was 10.88 days, we determined the actual active cases as the 546 11-day moving sum of new clinically confirmed cases. Breakpoints between fitted and forecast 547 values were chosen to be two days after the start of the outbreak, determined by the Piecewise 548 549 regression model. Cities are ordered by population size (largest on the top left and smallest on the bottom right). 550





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  $\alpha$  probability of an

increase. A false increase is counted when there was no increase in cases but the model predicted

557 a greater than  $\alpha$  probability of increase.



Figure 6: Evaluating SEIR model predictability for an emerging COVID-19 outbreak. Box 559 plot showing the distribution of measured area under the curve (AUC) when computing 50 560 561 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 562 represents the outcome if the model randomly picks predictions, has an AUC of 0.5. The further 563 away the curve is from the one of the random classifier, the higher the AUC and the better it 564 illustrates the ability of the model to forecast a trend, with the 1 representing the highest 565 accuracy corresponding to 100% positive rate and 0% false negatives. In general, for a 566 diagnostic test to be able to discriminate patients with and without a disease, the AUC must be 567 above 0.5. Values between 0.7 and 0.8 are considered to be 'fair' or acceptable (34,35). 568