



## 24 **Abstract**

25 Wastewater-based surveillance (WBS) has been widely used as a public health tool to monitor  
26 SARS-CoV-2 transmission. However, epidemiological inference from WBS data remains  
27 understudied and limits its application. In this study, we have established a quantitative  
28 framework to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission through  
29 integrating WBS data into an SEIR-V model. We conceptually divide the individual-level viral  
30 shedding course into exposed, infectious, and recovery phases as an analogy to the compartments  
31 in population-level SEIR model. We demonstrated that the temperature effect on viral losses in  
32 the sewer can be straightforwardly incorporated in our framework. Using WBS data from the  
33 second wave of the pandemic (Oct 02, 2020 – Jan 25, 2021) in the Great Boston area, we showed  
34 that the SEIR-V model successfully recapitulates the temporal dynamics of viral load in  
35 wastewater and predicts the true number of cases peaked earlier and higher than the number of  
36 reported cases by 16 days and 8.6 folds ( $R = 0.93$ ), respectively. This work showcases a simple,  
37 yet effective method to bridge WBS and quantitative epidemiological modeling to estimate the  
38 prevalence and transmission of SARS-CoV-2 in the sewershed, which could facilitate the  
39 application of wastewater surveillance of infectious diseases for epidemiological inference and  
40 inform public health actions.

41

## 42 **1. Introduction**

43 Wastewater-based surveillance (WBS) has been used as a public health tool to monitor SARS-  
44 CoV-2 infection in the population since the beginning of the COVID-19 pandemic. So far, WBS  
45 has been widely implemented in over 67 countries (Naughton et al., 2021). The Centers for  
46 Disease Control and Prevention (CDC) also launched the National Wastewater Surveillance  
47 System in late 2020 to monitor the spread of COVID-19 in the United States (CDC, 2020).  
48 Wastewater collates SARS-CoV-2 particles excreted by infected individuals irrespective of  
49 clinical symptoms or presentation, which provides an opportunity to capture the viral shedding  
50 prior to symptoms and estimate the true magnitude of viral infections in communities (Bivins et  
51 al., 2020b; Hart and Halden, 2020; Peccia et al., 2020; Randazzo et al., 2020; Saguti et al., 2021;  
52 Wu et al., 2022b). Previous work has shown that SARS-CoV-2 concentrations in wastewater  
53 were much higher than expected from clinically reported cases and predicted clinical reported

54 data for 4-10 days (Wu et al., 2020, 2022b, Peccia et al., 2020), and up to 14 days (Krivoňáková  
55 et al., 2021; Karthikeyan et al. 2020). Furthermore, the fast turnaround time of wastewater and  
56 flexible sampling strategy enable WBS to provide a near real-time monitoring of the viral  
57 transmission in the sewershed. Finally, WBS is less resource intensive than the large-scale,  
58 individual-based clinical testing and thus can be used as a cost-efficient tool for monitor the  
59 trend of viral infection in the population and new variants when combined with next-generation  
60 sequencing (Bivins et al., 2020b; Safford et al., 2022; Wu et al., 2022a). These properties make  
61 WBS a feasible public health tool to monitor SARS-CoV-2 in an endemic, which can also be  
62 customized for future pandemics.

63 WBS has enabled researchers to estimate the total viral load in a sewershed; however, there are  
64 still limitations regarding quantifying and predicting viral transmission in a community. Few  
65 recent studies have tried to build classical susceptible-infected-removed (SIR)-type models to  
66 bridge the measured viral concentration and reported case number. For example, Proverbio et al.  
67 (2022) added a variable that keeps track of actively shedding individuals in a stochastic  
68 susceptible-exposed-infectious-recovered (SEIR) model and used a constant viral shedding rate  
69 to connect the number of infected cases to viral concentration in wastewater (Proverbio et al.,  
70 2022). Conversely, Brouwer et al. (2022) accounted for time dependent viral shedding rates by  
71 incorporating multiple subclasses with different shedding rates within each infected stage of the  
72 model to better predict viral concentrations and reported cases (Brouwer et al., 2022). A similar  
73 approach is conducted by Nourbakhsh et al. (2022), but with more sub-classification of the  
74 infected class (Nourbakhsh et al., 2022). These modeling approaches allow the modelers to  
75 connect viral concentrations in wastewater with the reported cases and predict the course of the  
76 pandemic.

77 Dynamical models in epidemiology thus far often overlook the opportunity to utilize biologically  
78 interpretable and experimentally measurable parameters in the link between infected people and  
79 the shed viral RNA in wastewater. The model structure is usually complicated with many  
80 parameters, so it is difficult to fully parametrize the models without running into issues such as  
81 model identifiability. Hence, our primary objective in this work is to leverage our understanding  
82 of the biology of SARS-CoV-2 shedding to construct a simple, mechanistic, dynamic model that  
83 connects viral load in wastewater with the total number of infected cases in the sewershed. Our

84 secondary objective is to introduce the effect of wastewater temperature into the modeling  
85 framework due to its significant impact on the viral loss (or decay) rate in the sewer (Hart and  
86 Halden, 2020).

87

## 88 **2. Materials and methods**

### 89 **2.1. Samples and wastewater data**

90 Raw, 24-hour composite wastewater samples were collected from the Deer Island wastewater  
91 treatment plant in Massachusetts from October 02, 2020 to January 25, 2021. The Massachusetts  
92 wastewater treatment plant where we obtained samples has two major influent streams, which  
93 are referred to as the “northern” and “southern” influents. The daily flow rates during the  
94 sampling period for the northern and southern influents are  $4.54e5 - 2.3e6 m^3/day$ , and  $2.16e5$   
95  $- 1.19e6 m^3/day$ , respectively. Together the two catchments represent approximately 2.3 million  
96 wastewater customers in Middlesex, Norfolk, and Suffolk counties, primarily in urban and  
97 suburban neighborhoods. There are 5,100 miles of local sewers transporting wastewater into 227  
98 miles of interceptor pipes to the wastewater treatment plant ([www.mwra.com](http://www.mwra.com)), and the typical  
99 turnaround time for the plant to treat wastewater is 24 hours. Samples were processed as they  
100 were received. Experimental method and data were reported in our previous work (Wu et al.,  
101 2022b; Xiao et al., 2022). Briefly, the samples were pasteurized at 60°C for 1 hour for  
102 disinfection, and then filtered with 0.2 µm hydrophilic polyethersulfone membrane (Millipore  
103 Sigma) to remove bacterial cells and debris. Then, 15-ml filtrate was concentrated to ~200 ul  
104 with Amicon Ultra Centrifugal Filter (30-kDa cutoff, Millipore Sigma), and lysed with Qiagen  
105 AVL buffer followed by RNA extraction with Qiagen RNeasy kit. SARS-CoV-2 concentrations  
106 were quantified by one-step reverse transcription-polymerase chain reaction (RT-PCR) with the  
107 Taqman Fast Virus 1-Step Master Mix (Thermofisher) and CDC N1 and N2 primers/probes. Ct  
108 values were transformed to copies per ml of wastewater using standard curves for N1 and N2  
109 targets established with synthetic SARS-CoV-2 RNA (Twist Bioscience) as the template. We  
110 averaged the viral concentration data on the same day in the northern and southern influents and  
111 then multiplied by the daily average flow rate to compute the total viral load in the sewershed.

### 112 **2.2. Clinical data source**

113 The clinical COVID-19 case data for Norfolk, Suffolk, and Middlesex Counties served by the  
114 Massachusetts wastewater treatment plant were downloaded from Massachusetts government  
115 website ([www.mass.gov](http://www.mass.gov)). We summed the number of clinical cases from each county to  
116 represent the total cases in the catchment of the wastewater treatment plant, which is used to  
117 compare with the modeling results. Temporal fecal viral shedding data from COVID-19 patients  
118 were kindly provided by (Wölfel et al., 2020).

### 119 **2.3. Relationship between wastewater viral concentrations and infectious cases.**

120 Assuming we can obtain the fecal viral shedding distribution function over time, we can  
121 approximate a constant rate of fecal viral shedding over the duration of infectiousness. In this  
122 way, the viral RNA production is proportional to the number of people in the infectious  
123 compartment  $I$  of the SEIR model. That is:

$$\text{total viral load in wastewater} \approx \alpha \times \beta \times (1 - \gamma) \times I,$$

124 (1)

125 where the proportional constant is defined based on biological parameters similar to  
126 (Saththasivam et al. 2021):  $\alpha$  is the fecal load with unit g/day/person,  $\beta$  is the viral shedding rate  
127 in stool with unit viral copies/g, and  $\gamma$  is the fraction of viral loss in the sewer.

### 128 **2.4. Approximation of fecal viral shedding profile**

129 A key component of this approach is the generation of fecal viral shedding profile. Let  $f(t)$  be  
130 the function that describes the temporal fecal viral shedding profile. Upon infection, the shedding  
131 of virus in stool should be very small, then reaches a peak before decreasing to 0.  
132 Mathematically, this means  $f(0) = 0$ ,  $\lim_{t \rightarrow \infty} f(t) = 0$  and  $f(t)$  has a unique maximum for  
133 some  $t > 0$ . While beta and gamma functions are often used to represent  $f(t)$  (Wu et al., 2022a;  
134 Ferretti et al., 2020; He et al., 2020), we introduce a phenomenological function  $f(t)$  that is  
135 more tractable than the standard beta and gamma functions:

$$f(t) = \frac{\omega_1 t}{\omega_2^2 + t^2}.$$

136 (2)

137 In this form,  $\omega_1$  is a magnitude modifier parameter ( $\log_{10}$  viral RNA copy per g per day) and  
138  $\omega_2$  (day) represents the timing for peak viral shedding and influences the timing and the  
139 magnitude of the peak of the viral shedding profile. Specifically,  $f(t)$  peaks at  $\frac{\omega_1}{2\omega_2}$  when  $t = \omega_2$ .  
140 Thus, if the peak timing and magnitude of the viral shedding profile are known, then  $f(t)$  can be  
141 uniquely defined. It is necessary to mention that  $f(t)$  is the overall viral shedding into the  
142 wastewater from infected individuals; however, it mostly means fecal shedding in this work. We  
143 did not include the viral shedding from urine or other sources (sputum or saliva) because  
144 previous studies showed that no or low level of virus was detected in urine samples of typical  
145 patients despite high viral load (Wölfel et al., 2020; Jones et al., 2020), and the total amount of  
146 virus in sputum or saliva are likely to be insignificant compared to stool due to the huge  
147 difference in volume.

## 148 2.5. Simple wastewater epidemiological model

$$\begin{aligned}S' &= -\lambda IS \\E' &= \lambda IS - kE \\I' &= kE - \delta I \\V' &= \alpha\beta(1 - \gamma)I\end{aligned}$$

149 (3)

150 In this model,  $S$  denotes the susceptible population,  $E$  is the infected but yet to be infectious  
151 population, or the exposed class,  $I$  is the infectious class, and  $V$  is the cumulative viral load in  
152 wastewater. The  $R$  compartment (recovered individuals) does not contribute to the transmission  
153 dynamics in the SEIR model, hence omitted here. Susceptible people are infected by the  
154 infectious class at a rate  $\lambda I$ . Exposed individuals become infectious at a rate  $k$ . Infectious  
155 individuals recover at a rate  $\delta$  and shed virus at a rate  $\alpha \times \beta$ , where  $\alpha$  is the fecal load and  $\beta$  is  
156 the average viral shedding rate in Eq (1).  $\gamma$  is the viral degradation and loss rate in the sewer  
157 pipes, so only a fraction  $(1 - \gamma)$  of virus is detected in the wastewater sample. The expression for  
158  $V$  follows directly from Eq (1).

159 Several studies note that infectious virus is detectable in nose and throat swabs only when the  
160 total viral load is above  $10^{5-6}$  copies/mL (Killingley et al., 2022, Ke et al., 2021, Wölfel et al.,  
161 2020, Kampen et al., 2021). Since certain level of infectious viruses is required for disease

162 transmission, this implies that the infectious period does not start until the viral load (within host)  
163 reaches above  $10^{5-6}$  virus copies/mL. This agrees with previous observation that viral load  
164 above  $10^6$  copies/mL is associated with a high probability of transmission (Ke et al., 2021).  
165 Together, these observations suggest that in this SEIR epidemic model, we can separate the  
166 exposed class ( $E$ ) based on the duration before viral load reaches  $10^{5-6}$  copies/mL, and the  
167 infectious class ( $I$ ) based on the duration that viral load stays above  $10^{5-6}$  copies/mL. This  
168 results in an incubation period of about 3 days and an infectious period of 8 days based on the  
169 viral dynamics profile in the SARS-CoV-2 Human Challenge experiment in healthy young adults  
170 (Killingley et al., 2022). These estimates are within previous estimated ranges of 2-7 days for  
171 incubation periods (Li et al., 2020, Lauer et al., 2020, Guan et al., 2020) and consistent with the  
172 updated guideline from CDC where the average infectious duration is about 2 days before and 8  
173 days after symptom onset (CDC, 2022a). Thus, we fix the exposed duration to 3 days, which is  
174 equivalent to fixing  $k = \frac{1}{3}$  per day (Figure 1A). Similarly, we fix the infectious duration to 8  
175 days, which is equivalent to fixing  $\delta = \frac{1}{8}$  per day. Thus, in our model, parameters  $\lambda$ ,  $\alpha$ ,  $\beta$ , and  $\gamma$   
176 need to be estimated.

177 By fitting the model to wastewater data covering the second wave of the pandemic, specifically,  
178 from Oct 2 to Dec 16, 2020, we can approximate the susceptible (to an emerging variant) to be  
179 the entire population served by the wastewater treatment plant. For simplification, we assume  
180 that there is no infectious individuals initially ( $I(0) = 0$ ), only infected individuals ( $E(0) > 0$ ).  
181 The initial value for the virus concentration in wastewater can be taken from the first data point.  
182 Thus,  $E(0)$  is the only unknown initial condition.

183 The parameters and initial remain to be estimated are:  $\lambda$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $E(0)$ . Since the viral  
184 production rate is  $\alpha\beta(1 - \gamma)$ , and we only have viral concentration (or total viral load) data, it is  
185 impossible to estimate a unique set of values, or specific values, for  $\alpha$ ,  $\beta$ , and  $\gamma$ . For example,  
186 the product of  $\alpha = 1$ ,  $\beta = 2$ ,  $\gamma = 0.5$  is the same as when  $\alpha = 10$ ,  $\beta = 1$ ,  $\gamma = 0.9$ . This reflects  
187 the pertinent issue of model identifiability in mathematical models in biology and epidemiology  
188 (Tuncer et al., 2022; Eisenberg et al., 2013; Wu et al., 2019; Ciupe and Tuncer, 2022). Thus, an  
189 important step in our approach is the direct estimations of  $\beta$  and  $\gamma$ , which would allow us to  
190 identify  $\alpha$  uniquely. All of the parameters are listed in Table 1.

191

192

193 **Table 1. Parameters in the model.**

	Definition	Unit	Value	References
$S$	Susceptible population	People	$S(0) = 2.3 \times 10^6$ - fixed	(Wu et al., 2022b)
$E$	Exposed population	People	$E(0)$ – fitting	
$I$	Infectious population	People	$I(0) = 0$ - fixed	
$\lambda$	Transmission rate	Per day per person	fitting	
$1/k$	Exposed duration	Day	3 days	Wölfel et al., 2020; Killingley et al., 2022; Wu et al., 2022a; Van Kampen et al. 2021; Wölfel et al., 2020,
$1/\delta$	Infectious duration	Day	8 days	Killingley et al.; 2022, Wu et al., 2022a; Van Kampen et al. 2021
$\alpha$	Fecal load	Gram	51-796 g - fitting	Rose et al., 2015
$\beta$	Viral shedding in stool	Viral RNA copies per gram	fitting	
$\gamma$	Fraction of viral loss in sewer	Per day	Fitting and estimated	
$\omega_1$	Magnitude modifier	$\log_{10}$ viral RNA per g day	fitting	
$\omega_2$	Peak timing for viral shedding	Day	4 day - fixed	Killingley et al., 2022; Wu et al., 2022a.

194 Note that  $\beta$  and  $\omega_1$  are obtained from fitting to viral shedding data in stool (Wölfel et al., 2020).

195 **2.6. Data fitting**

196 Our goal is to fit the SEIR-V model to viral concentration in wastewater data to infer the true  
 197 number of cases. Then, we compare the predicted number of cases with the daily reported case

198 data. In our model, the variable  $V$  is the cumulative viral load in wastewater. Thus, the difference  
199 of  $V$  in every 24-hour period reflects the daily measurement data of total virus concentration in  
200 wastewater. To reflect this observation, we aim to minimize the sum of square error ( $SSE_V$ )  
201 between these two quantities in our fitting. Hence, our minimization objective is:

$$SSE_V = \sum_{t_d} \left( \int_{t_{d-1}}^{t_d} V'(s) ds - \hat{V}(t_d) \right)^2.$$

202 (4)

203 Here,  $\hat{V}(t_d)$  is the total virus concentration experimentally measured on day  $t_d$ , which equals to  
204 viral RNA concentration in wastewater ( $C_{RNA}$ ) multiplied by the total flow ( $F$ ) data.  
205  $\int_{t_{d-1}}^{t_d} V'(s) ds$  is the corresponding quantity in our model. Once we obtain a reasonable fit to the  
206 data, the inferred number of true case is given by:

$$Case\ number = \frac{V(t)}{\alpha \times \beta \times (1 - \gamma)}.$$

207 (5)

208 For the minimization algorithm, we use MATLAB function *fmincon* and *multistart*. Similarly,  
209 the fecal viral shedding rate function is fitted by minimizing the objective function  $SSE_f$ :

$$SSE_f = \sum_{t_n} \left( f(t_n) - \hat{f}(t_n) \right)^2,$$

210 (6)

211 Where  $\hat{f}(t_n)$  is the fecal shedding data on day  $t_n$ .

212

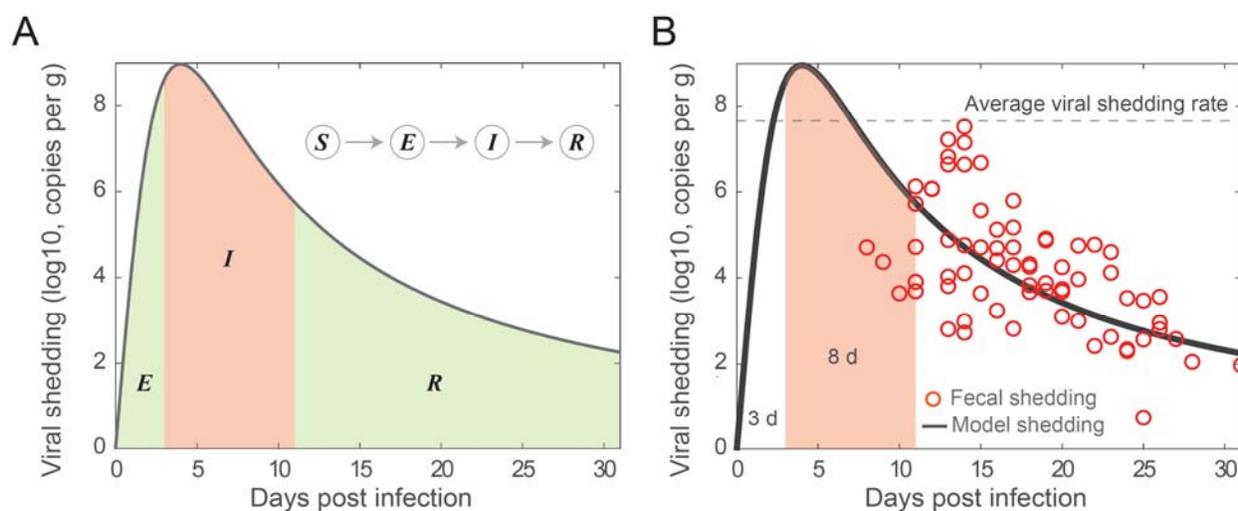
### 213 3. Result

#### 214 3.1. Determining the average fecal viral shedding rate in infectious period

215 We observed that there is a striking similarity in the viral load profiles in nose, throat, and stool  
216 for infected individuals from the time of infection to recovery qualitatively (Wölfel et al., 2020,

217 Killingley et al., 2022, Van Kampen et al. 2021). In all three cases, high viral load/shedding is  
218 associated with the infectious duration of the infection. This observation suggests that in the  
219 classical SEIR epidemic model, we can make the simplifying assumption that the infectious  
220 individuals contribute substantially to the viral pools in wastewater. As illustrated in Figure 1A,  
221 the viral shedding profile is divided into three periods shaded: Exposed (*E*), Infectious (*I*), and  
222 Recovered (*R*). With this framework, we can approximate the viral load in wastewater using the  
223 viral shedding from the infectious population. Furthermore, we can estimate the average viral  
224 shedding rate based on the viral shedding function  $f(t)$  and the fixed duration of infectiousness  
225 (see Materials and Methods).

226



227  
228 **Figure 1. Illustration and fitting fecal viral shedding dynamics.** (A) Illustration of the fecal  
229 viral shedding dynamics based on the infection progression. The viral shedding profile is divided  
230 into three periods shaded: Exposed (*E*), Infectious (*I*), and Recovered (*R*). The red-shaded region  
231 is the period of infectiousness *I*, which is corresponding to the compartment *I* in the SEIR model.  
232 (B) Fitting of the proposed viral shedding function to viral shedding in hospitalized patients'  
233 stool data from (Wolfel et al. 2020). The average viral shedding rate in stool during the  
234 infectious period (from day 3 to day 11) is  $4.48 \times 10^7$  viral RNA per g. The horizontal dash line  
235 is the average fecal viral shedding rate for infectious individuals inferred from the model. The  
236 viral shedding peak is at the 4<sup>th</sup> day post infection.

237 We fitted the fecal viral shedding function to viral shedding data. Using the best fit parameters,  
238 we constructed a fecal viral shedding profile that was used to approximate the fecal viral  
239 shedding rate for the infectious individuals. Figure 1B shows the best fit of the model to the fecal  
240 viral shedding rate data in Wolfel et al. (Wolfel et al., 2020). Based on the viral dynamics profile  
241 in the SARS-CoV-2 Human Challenge experiment in young adults (Killingley et al., 2022), the  
242 incubation period ( $E$ ) is about 3 days and the infectious period is about 8 days. We also assumed  
243 a five day from infection to symptom onset in the fecal viral shedding data, which is in range of  
244 2-14 days estimated for the general population (CDC, 2022b; Lauer et al., 2020). Furthermore,  
245 we fixed the viral peak at day four ( $\omega_2 = 4$  day). There is no well-established timing of the peak  
246 fecal viral shedding rate; however, the peak time for viral load in nose and throat is around 5  
247 days (Killingley et al., 2022) and maybe even earlier in stool (Wu et al., 2022a). The best fit  
248 parameter is  $\omega_1 = 71.97 \log_{10}$  viral RNA copy per g day. Using the best fit, we estimate the  
249 average fecal viral shedding rate for an infectious individual to be:

$$\beta = \frac{1}{11-3} \int_3^{11} f(t) dt = \frac{1}{8} \int_3^{11} \frac{71.97t}{16+t^2} dt \approx 7.65 \log_{10} \text{viral RNA per g}$$

250 (7)

251 A conversion gives:

$$\beta = 4.48 \times 10^7 \text{ viral RNA per g.}$$

252 (8)

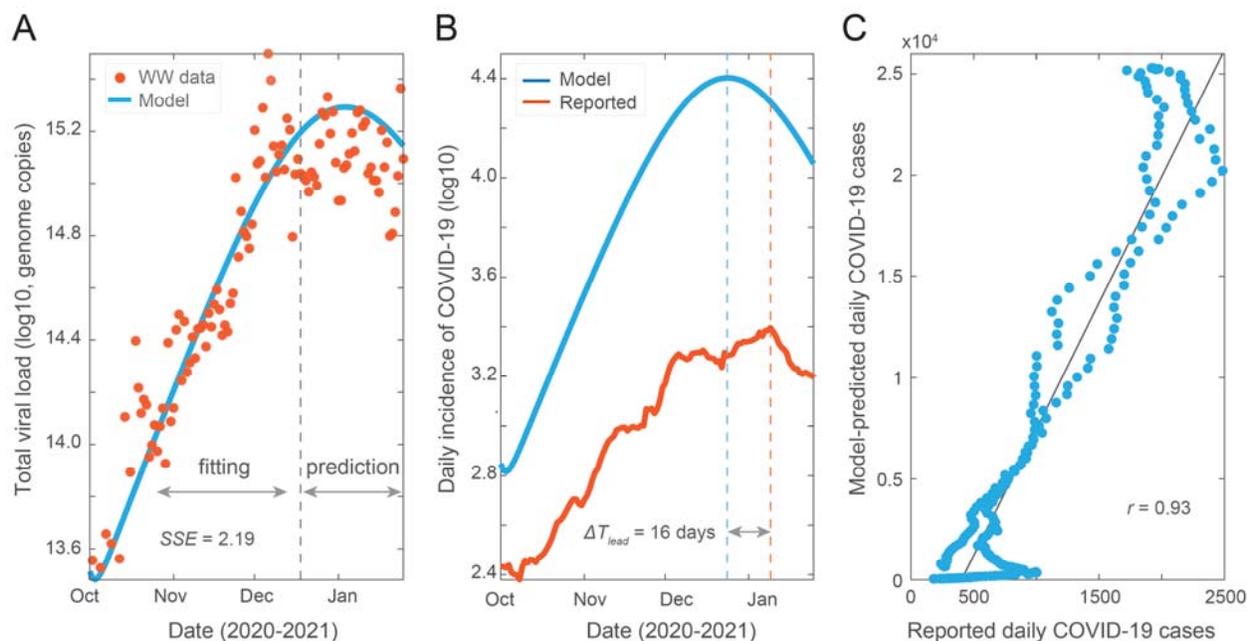
253 This number is close to the measured median viral RNA load  $10^{7.68}$  (ranging from  $10^{4.1}$ – $10^{10.27}$ )  
254 copies/ml in infected individuals in South Korea (Han et al., 2020), and the extrapolated fecal  
255 shedding rate of  $10^{7.30}$  (ranging from  $10^{5.74}$  –  $10^{8.28}$ ) copies/g of 711 infected individuals in the  
256 dormitories at University of Arizona (Schmitz et al., 2021). Thus, we fixed fecal viral production  
257 rate in our SEIR-V model to this value.

258

### 259 3.2. SEIR-V model captures the temporal dynamics of clinical COVID-19 cases

260 We developed an SEIR-V model to understand SARS-CoV-2 transmission using WBS data in  
261 the second wave of the pandemic and the computed average fecal viral shedding rate during the

262 period of infectiousness. We temporarily ignored the identifiability issue with the conversion  
263 equation in an attempt to fit the SEIR-V model to the data. Figure 2 shows the best fit and its  
264 inference. We fitted the model to total viral RNA copies in wastewater data up to the grey dashed  
265 line (December 16, 2020), then simulated the model out to January 25, 2021, see Figure 2A. The  
266 fitting region was chosen before the peak in the viral RNA data, so that we could test the model's  
267 prediction of the peak against the data. Additionally, the fitting region from October 02, 2020 to  
268 December 16, 2020 potentially limits the influence from vaccination and the emergence of the  
269 alpha variant, which began near the end of 2020.



270  
271 **Figure 2. Model fit and prediction to wastewater data covering the second wave of**  
272 **pandemic.** (A) Best fit to virus concentration data in wastewater from October 2 to December  
273 16, 2020 (dashed grey line), and model prediction to January 25, 2021. Red dots are the  
274 measured viral load in wastewater and blue curve is the modeling result. (B) Model estimation of  
275 the true number of COVID-19 cases (blue curve) and clinically reported cases (red curve). The  
276 blue and red dashed lines are dates when the two curves peak, and  $\Delta T_{lead}$  is the time difference  
277 between the two peaks. (C) Correlation between simulation cases and reported cases. Best fit  
278 parameters:  $\lambda = 9.66 \times 10^{-8} \text{ day}^{-1} \text{ person}^{-1}$ ,  $\alpha = 249 g$ ,  $\gamma = 0.08$ ).

279

280 Using the best fit parameters, we computed the number of new cases and compared it to the  
281 reported cases. As shown in Figure 2B, the model simulation recapitulates the trend of clinically  
282 reported daily new cases and predicts an earlier and higher peak than reported case data by 16  
283 days and 10.2 folds, respectively. We made a correlation plot between the model simulated cases  
284 and the reported case data (Figure 2C). The higher predicted number of cases and the high  
285 correlation coefficient ( $R = 0.93, R^2 = 0.87$ ) imply that the model accurately captures the trend  
286 of the reported case data, while accounting for the underreported rate. This indicates that the  
287 method preserves both key properties of WBS data, which is that the trend of viral concentration  
288 in wastewater leads the trend of reported cases and can be used to estimate the true prevalence  
289 without being impacted by the underreporting rate.

290 In the next step, we demonstrate how the effect of temperature on viral loss rate can be  
291 incorporated in our framework.

292

### 293 **3.3. Incorporation of wastewater temperature improves model prediction**

294 SARS-CoV-2 RNA in wastewater is subject to degradation which is affected by many factors  
295 such as temperature and travel time (Bivins et al., 2020a; McCall et al., 2022). We accounted for  
296 these factors to determine an approximate value of  $\gamma$ , the fraction of viral decay in the  
297 sewershed. The daily viral degradation rate in wastewater is described with the Arrhenius  
298 equation:

$$\eta(V, T) = \eta_0 Q_{10}^{(T-T_0)/10^\circ C} V,$$

299 (9)

300 where  $\eta_0$  is the viral genome degradation rate at ambient temperature  $T_0$  and  $Q_{10}$  is the  
301 temperature dependent rate of change (McMahan et al., 2021; Hart and Halden, 2020). Bivins  
302 and colleagues determined that, for wastewater inoculated with high titer at  $T_0 = 20^\circ \text{C}$ , the  
303 mean first-order decay rate of SARS-CoV-2 RNA is  $\eta_0 = 0.67$  per day (Bivins et al., 2020).  
304 Furthermore,  $Q_{10}$  is typically between 2 and 3 for biological systems, and assumed here to be 2.5  
305 (Běhrádek, 1930; Reyes et al., 2008). Given the relatively constant temperature from October 2

306 to December 16, 2020 (Figure S1), we used the average temperature of wastewater for the north  
307 and south systems for demonstrative purpose, and thus fix  $T = 18$  °C.

308 We used the simple exponential decay equation  $V' = -\eta(V, T)$  to estimate  $\gamma$ . Let  $\hat{\eta} =$   
309  $\eta_0 Q_{10}^{(T-T_0)/10^\circ C}$ , then solving  $V'(t) = -\eta(V, T) = \hat{\eta}V$  gives:

$$V(t) = V_0 e^{-\hat{\eta}t},$$

310 (10)

311 where  $V_0$  is the amount of viral RNA in the sewers at time  $t = 0$ . Thus, the amount of virus that  
312 arrives to the wastewater treatment plant is

$$V(t_{arrive}) = V_0 e^{-\hat{\eta}t_{arrive}},$$

313 (11)

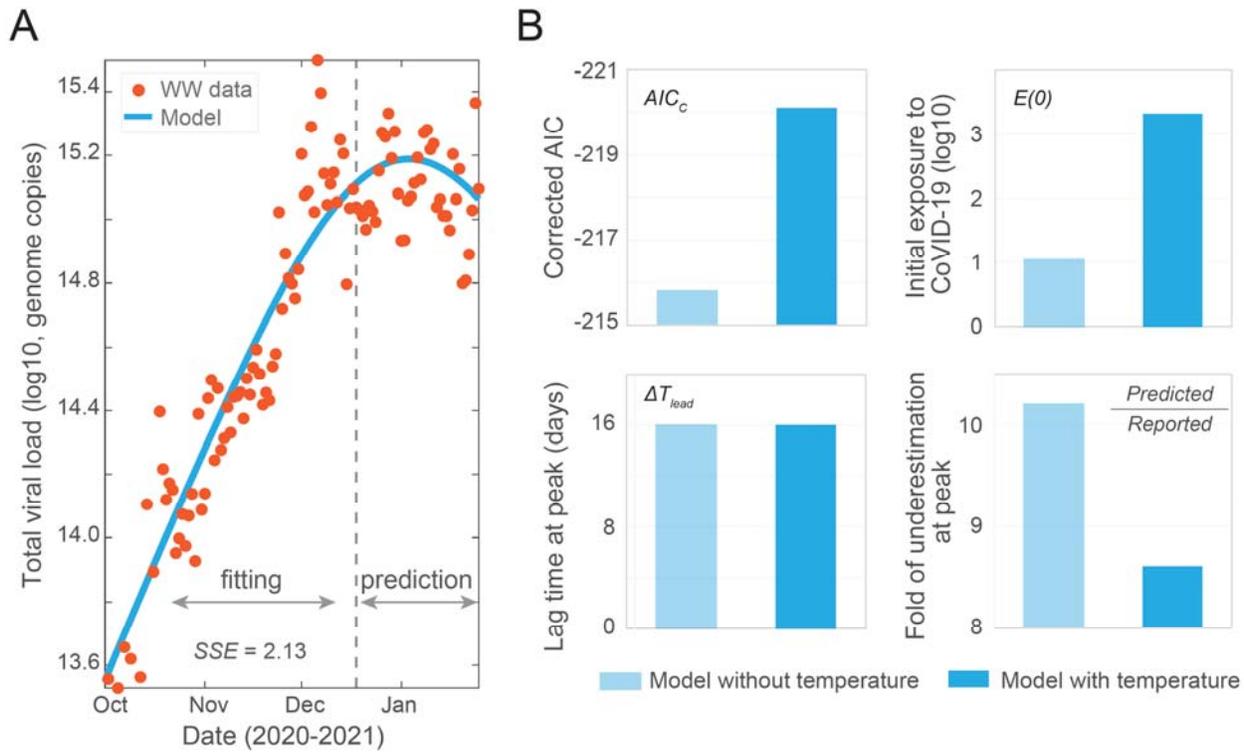
314 where  $t_{arrive}$  is the time it takes the viral RNA to travel to the wastewater treatment plant after  
315 excretion. The time  $t_{arrive}$  includes two parts: the travel time to local sewer pipes and the travel  
316 time in the interceptor pipes. Precise estimation of  $t_{arrive}$  is challenging given the varied flow  
317 rates and geographical distances to the wastewater treatment plant. Here, we assumed the  
318 average travel time is 18 hours. The amount of virus lost is given by  $V_0 - V(t_{arrive})$ . Thus, the  
319 proportion of viral RNA lost in the sewer is given by

$$\gamma = \frac{V_0 - V(t_{arrive})}{V_0} = 1 - \frac{V(t_{arrive})}{V_0} = 1 - e^{-\hat{\eta}t_{arrive}},$$

320 (12)

321 Where the last equality follows from Eq (11). In this case, our calculation yielded  $\gamma \approx 0.35$ ,  
322 which is in the ranges of previous estimations (Bivins et al., 2020; McCall et al., 2022; Hart and  
323 Halden, 2020).

324



325

326 **Figure 3. Incorporating temperature effect in the SEIR-V model.** (A) Best fit to viral  
 327 concentration data in wastewater from October 2 to December 16, 2020 (dashed grey line), and  
 328 model prediction to January 25, 2021. Red dots are the measured viral load in wastewater and  
 329 blue curve is the modeling result. (B) Comparison of the SEIR-V models with and without  
 330 incorporating temperature effect. Top left: corrected Akaike information criterion ( $AIC_c$ ) values,  
 331 the statistically significant  $AIC_c$  difference is 4.3; Top right: initial populations exposed to  
 332 SARS-CoV-2; Bottom left: wastewater lead time difference at peak, both of the  $\Delta T_{lead}$  are 16  
 333 days; Bottom right: fold of difference between the number of predicted cases and clinically  
 334 reported cases. Light blue represents the model without including temperate effect, while blue  
 335 represents the model with temperature effect. Best fit parameters when incorporating  
 336 temperature:  $\lambda = 9.13 \times 10^{-8} \text{ day}^{-1} \text{ person}^{-1}$ ,  $\alpha = 324 g$ , and  $E(0) = 2092$  people.

337 By incorporating temperature effect, the model captures the trend of clinical data with a smaller  
 338  $SSE$ , which is statistically significant based on the corrected Akaike information criterion (Figure  
 339 3A, B and S2) (Burnham and Anderson, 2004). We observe that the model simulation predicts an  
 340 earlier peak than reported case data by 16 days, which is the same as the model without  
 341 temperature effect (Figure 3B and S2A). Additionally, the model predicts the true number of

342 cases to be about 8.6 times higher than the reported number of cases as compared to a predicted  
343 factor of 10.2 without temperature effect (Figure 3B and S2A). The predicted initial exposed  
344 population is 2092 people, which is a more reasonable estimate compared to the 11 exposed  
345 individuals predicted without temperature (Figure 3B). Those results have shown that  
346 incorporating the travel time and temperature reduces the possibility of model unidentifiability  
347 and significantly improve the model performance as well as its robustness.

348

#### 349 **4. Discussion**

350 Wastewater collates viral signals excreted by infected individuals across the whole spectrum of  
351 disease symptoms from asymptomatic and subclinical-symptomatic to symptomatic (Lee et al.,  
352 2020). This inclusiveness of all virus-shedding individuals offers an opportunity to better  
353 estimate the magnitude of viral infections in communities (Hart and Halden, 2020; Sanjuán and  
354 Domingo-Calap, 2021; Wu et al., 2020). However, it is challenging to convert viral  
355 concentrations in wastewater to the number of infected cases. Our group and peers previously  
356 reported methods to estimate the infection prevalence by wastewater viral load (McMahan et al.,  
357 2021; Nourbakhsh et al., 2021; Wu et al., 2020). These efforts, however, are limited because of  
358 inconsideration of dynamic viral shedding rates during the disease course and viral degradation  
359 in wastewater.

360 In this study, we established a quantitative framework to estimate the number of infectious  
361 COVID-19 cases and predict SARS-CoV-2 transmission through integrating wastewater  
362 surveillance data and development of an SEIR-V model. As an analogy to the four compartments  
363 of the SEIR model to simulate infectious disease dynamics at the population level, the  
364 individual-level fecal viral shedding course was divided into three periods including exposed  
365 (incubation), infectious, and recovery (Figure 1A). The division is based on the observation that  
366 the temporal viral profiles in nose, mouth, and stool are strikingly similar qualitatively with high  
367 viral load associated with infectiousness (Killingley et al., 2022; Wolfel et al., 2020). In addition,  
368 the infectiousness of SARS-CoV-2 is associated with high viral load as reported by multiple  
369 studies (Killingley et al., 2022, Ke et al., 2021, Wolfel et al., 2020, Kampen et al., 2021). With  
370 this concept, we estimated the population-level average viral shedding rate during the infectious  
371 phase using clinical reported SARS-CoV-2 concentrations in hospitalized patients' stool samples

372 (Figure 1B). This estimated viral shedding rate is an average of infected individuals in the  
373 population and does not consider the heterogenous viral shedding dynamics among infected  
374 individuals (Wölfel et al., 2020; Killingley et al., 2022; Stanca and Tuncer, 2022). Thus, our  
375 model can be improved by feeding viral shedding data during the early phase of the infection and  
376 large-scale individual-level shedding dynamics data.

377 It is noteworthy to mention that the " $I$ " in the SEIR model is the "infectious" class, not the  
378 "infected" class. Hence, using the viral shedding rate in the infectious period, instead of in the  
379 whole shedding period, improves the accuracy of the SEIR model. This contrasts with the  
380 conventional approaches that use mean or median viral shedding rate in a group of tested  
381 samples regardless of the phase of the infection (Saththasivam et al., 2021; Petala et al., 2022;  
382 Schimitz et al., 2021). By focusing on the infectious population, which is also the main  
383 contributor of viral shedding in wastewater, we greatly simplify the typical complex structure of  
384 the SEIR-type models that implement WBS (Figure S3) and reduces the likelihood of model  
385 unidentifiability.

386 By fitting an SEIR-V model to wastewater data within our framework, we show that the method  
387 retains key advantages of using wastewater. Specifically, the inferred case data from the best fit  
388 parameters leads the reported case data by 16 days and implies a large ratio (8.6) of true  
389 prevalence to clinically reported cases, which are consistent with previous results (Wu et al.,  
390 2020; Wu et al., 2022a; Eikenberry et al., 2020; Angulo et al., 2021). We also incorporate the  
391 important effect of temperature on the viral degradation rate in a simple manner that is applicable  
392 to a larger time scale. We note that extension to incorporate time-dependent variations of the  
393 fecal viral shedding rate within this framework is straightforward, but will require careful  
394 consideration of the convergence of the numerical method. Together, our work shows the  
395 potential and flexibility of the framework to incorporate WBS in epidemic models.

396 The foundation of our framework is independent of the epidemic model formulation, yet its  
397 application depends greatly on the epidemic models for specific situations. For example, if we  
398 want to apply the framework to capture a period with significant changes to social behavior,  
399 perhaps due to the effect of a social intervention, then an appropriate change to the structure of  
400 the SEIR model to reflect these structures is necessary (Johnston and Pell, 2020; Fenichel et al.,  
401 2011; Pell et al., 2018). However, if multiple variants are of interest, then the SEIR model itself

402 needs to be extended to a multi-variant version and incorporate known biological properties of  
403 different variants (Dyson et al., 2021; Gonzalez-Parra et al., 2021). Similarly, interventions (such  
404 as vaccination) and the impact of social gatherings must first be included in the epidemic model  
405 prior to its integration within our framework (Saad-Roy et al., 2021; Giordano et al., 2021;  
406 Buckner et al., 2021; Makhoul et al., 2020).

407 Dynamical epidemic models are useful tools to track the pandemic progression and to assess the  
408 potential impact of hypothetical situations such as the stay-at-home order or the emergence of a  
409 resistant viral strain. However, sparsely reported case data with high uncertainty, due partially to  
410 the high underreporting rate, can compromise the ability of epidemic models to provide an  
411 accurate forecast of the pandemic and limit their application to retrospective studies. Hence,  
412 WBS, which bypasses both the tremendous difficulty in data collection faced by the standard  
413 clinical reporting practice and the high underreporting rate, represents a potential solution to  
414 address this challenge faced by the modeling community. WBS data also provides a leading  
415 indicator of the pandemic progression and is not limited to SARS-CoV-2, thus it can further  
416 enhance the prediction and applicability of epidemic models. Together, this aspect of our  
417 framework highlights the importance of interdisciplinary collaboration to better address public  
418 health concerns

419

## 420 **5. Conclusions**

421 In this study, we have established a quantitative framework to estimate COVID-19 prevalence  
422 and predict SARS-CoV-2 transmission by incorporating WBS data in a simple epidemic SEIR-V  
423 model. The main conclusions are:

- 424 • We constructed a simple and effective framework to incorporate WBS data to epidemic  
425 models. The developed SEIR-V model captures the temporal dynamics of clinical  
426 COVID-19 cases and preserves key advantages of WBS data over reported case data.
- 427 • We illustrated how the effect of travel time and temperature on viral decay can be  
428 incorporated within our framework to improve model performance and robustness, which  
429 is an important component to model disease transmission in real world application.

- 430       • The modeling framework is a valuable platform to integrate WBS with epidemic models  
431       to provide accurate and robust estimates of the pandemic progression and examine the  
432       potential impact of interventions to inform public health decision making.

### 433 **Declaration of Competing Interest**

434 The authors declare no competing interest.

435

### 436 **Code Availability**

437 All data and code produced in the present study are available upon reasonable request to the  
438 authors.

439

### 440 **Acknowledgement**

441 This work is supported by Faculty Startup funding from the Center of Infectious Diseases at  
442 UTHealth, the UT system Rising STARS award, and the Texas Epidemic Public Health Institute  
443 (TEPHI) to F.W. This work was also supported by Director’s postdoctoral fellowship at Los  
444 Alamos National Laboratory to T.P.; Y.K. and S.B. are partially supported by the US National  
445 Science Foundation Rules of Life program DEB -1930728 and the NIH grant 5R01GM131405-  
446 02.

447

### 448 **References**

449 Angulo, Frederick J, Lyn Finelli, and David L Swerdlow (2021). “Estimation of US SARS-CoV-  
450 2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys”.  
451 *JAMA Network Open* 4(1), pp. 2033706–2033706.

452 Běhrádek, J (1930). “Temperature coefficients in biology”. *Biological Reviews* 5(1), pp. 30–58.

453 Bivins, Aaron et al. (2020). “Persistence of SARS-CoV-2 in water and wastewater”.  
454 *Environmental Science & Technology Letters* 7(12), pp. 937–942.

- 455 Brouwer, Andrew F et al. (2022). “The role of time-varying viral shedding in modelling  
456 environmental surveillance for public health: revisiting the 2013 poliovirus outbreak in Israel”.  
457 *Journal of the Royal Society Interface* 19(190), pp. 20220006–20220006.
- 458 Buckner, Jack H, Gerardo Chowell, and Michael R Springborn (2021). “Dynamic prioritization  
459 of COVID-19 vaccines when social distancing is limited for essential workers”. *Proceedings of*  
460 *the National Academy of Sciences* 118(16), pp. 2025786118–2025786118.
- 461 Burnham, Kenneth P and David R Anderson (2004). “Multimodel inference: understanding AIC  
462 and BIC in model selection”. *Sociological Methods & Research* 33(2), pp. 261–304.
- 463 CDC (2020). “National Wastewater Surveillance System”. Centers for Disease Control and  
464 Prevention.
- 465 CDC (2022a). “Ending Isolation and Precautions for People with COVID-19: Interim  
466 Guidance”. *Centers for Disease Control and Prevention*.
- 467 CDC (2022b). “Symptoms of COVID-19”. Center for Disease Control and Prevention.
- 468 Ciupe, Stanca M., and Necibe Tuncer. (2022) "Identifiability of parameters in mathematical  
469 models of SARS-CoV-2 infections in humans." *medRxiv*.
- 470 Dyson, Louise et al. (2021). “Possible future waves of SARS-CoV-2 infection generated by  
471 variants of concern with a range of characteristics”. *Nature Communications* 12(1), pp. 1–13.
- 472 Eikenberry, Steffen E et al. (2020). “To mask or not to mask: Modeling the potential for face  
473 mask use by the general public to curtail the COVID-19 pandemic”. *Infectious Disease*  
474 *Modelling* 5, pp. 293–308.
- 475 Eisenberg, Marisa C et al. (2013). “Identifiability and estimation of multiple transmission  
476 pathways in cholera and waterborne disease”. *Journal of Theoretical Biology* 324, pp. 84–102.
- 477 Fenichel, Eli P et al. (2011). “Adaptive human behavior in epidemiological models”.  
478 *Proceedings of the National Academy of Sciences* 108(15), pp. 6306–6311.
- 479 Ferretti, Luca, et al. (2020). "Quantifying SARS-CoV-2 transmission suggests epidemic control  
480 with digital contact tracing." *Science* 368.6491: eabb6936.

- 481 Giordano, Giulia et al. (2021). “Modeling vaccination rollouts, SARS-CoV-2 variants and the  
482 requirement for non-pharmaceutical interventions in Italy”. *Nature Medicine* 27(6), pp. 993–998.
- 483 Gonzalez-Parra, Gilberto, David Martínez-Rodríguez, and Rafael J Villanueva-Micó (2021).  
484 “Impact of a new SARS-CoV-2 variant on the population: A mathematical modeling approach”.  
485 *Mathematical and Computational Applications* 26(2), pp. 25–25.
- 486 Guan, Wei-Jie et al. (2020). “Clinical characteristics of coronavirus disease 2019 in China”. *New*  
487 *England Journal of Medicine* 382(18), pp. 1708–1720.
- 488 Hart, O E and R U Halden (2020). “Computational analysis of SARS-CoV-2/COVID-19  
489 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy,  
490 opportunities and challenges”. *Science of the Total Environment* 730, pp. 138875–138875.
- 491 He, Xi et al. (2020). “Temporal dynamics in viral shedding and transmissibility of COVID-19”.  
492 *Nature Medicine* 26(5), pp. 672–675.
- 493 Johnston, Matthew D and Bruce Pell (2020). “A dynamical framework for modeling fear of  
494 infection and frustration with social distancing in COVID-19 spread”. *Mathematical Biosciences*  
495 *and Engineering* 17(6), pp. 7892–7915.
- 496 Jones, David L., et al. (2020). "Shedding of SARS-CoV-2 in feces and urine and its potential role  
497 in person-to-person transmission and the environment-based spread of COVID-19." *Science of*  
498 *the Total Environment* 749: 141364.
- 499 Karthikeyan, Smruthi, et al. (2022). "Wastewater sequencing reveals early cryptic SARS-CoV-2  
500 variant transmission." *Nature*: 1-4.
- 501 Ke, Ruian et al. (2021). “In vivo kinetics of SARS-CoV-2 infection and its relationship with a  
502 person’s infectiousness”. *Proceedings of the National Academy of Sciences* 118(49).
- 503 Kampen, Van et al. (2021). “Duration and key determinants of infectious virus shedding in  
504 hospitalized patients with coronavirus disease-2019 (COVID-19)”. *Nature Communications*  
505 12(1), pp. 1–6.
- 506 Killingley et al. (2022). “Safety, tolerability and viral kinetics during SARS-CoV-2 human  
507 challenge in young adults”. *Nature Medicine* 28(5), pp. 1031–1041.

- 508 Krivonůřáková, Nad'a et al. (2021). "Mathematical modeling based on RT-qPCR analysis of  
509 SARS-CoV-2 in wastewater as a tool for epidemiology". *Scientific Reports* 11(1), pp. 1–10.
- 510 Lauer, Stephen A et al. (2020). "The incubation period of coronavirus disease 2019 (COVID-19)  
511 from publicly reported confirmed cases: estimation and application". *Annals of Internal*  
512 *Medicine* 172(9), pp. 577–582.
- 513 Lee, S et al. (2020). "Clinical course and molecular viral shedding among asymptomatic and  
514 symptomatic patients with SARS-CoV-2 infection in a community treatment center in the  
515 Republic of Korea". *JAMA Internal Medicine* 180, pp. 1447–1452.
- 516 Li, Qun et al. (2020). "Early transmission dynamics in Wuhan, China, of novel coronavirus-  
517 infected pneumonia". *New England Journal of Medicine*.
- 518 Makhoul et al. (2020). "Epidemiological impact of SARS-CoV-2 vaccination: Mathematical  
519 modeling analyses". *Vaccines* 8, no. 4: 668.
- 520 McCall, Camille et al. (2022). "Modeling SARS-CoV-2 RNA degradation in small and large  
521 sewersheds". *Environmental Science: Water Research & Technology* 8(2), pp. 290–300.
- 522 McMahan, CS et al. (2021). "COVID-19 wastewater epidemiology: a model to estimate infected  
523 populations". *The Lancet Planetary Health*. 2021; 5: e874–e881.
- 524 Naughton, Colleen C et al. (2021). "Show us the data: global COVID-19 wastewater monitoring  
525 efforts, equity, and gaps". *MedRXiv*.
- 526 Nourbakhsh, Shokoofeh et al. (2022). "A wastewater-based epidemic model for SARS-CoV-2  
527 with application to three Canadian cities". *Epidemics* 39, pp. 100560–100560.
- 528 Peccia, Jordan et al. (2020). "Measurement of SARS-CoV-2 RNA in wastewater tracks  
529 community infection dynamics". *Nature Biotechnology* 38(10), pp. 1164–1167.
- 530 Pell, Bruce et al. (2018). "Simple multi-scale modeling of the transmission dynamics of the 1905  
531 plague epidemic in Bombay". *Mathematical Biosciences* 301, pp. 83–92.
- 532 Petala, Maria et al. (2022). "Relating SARS-CoV-2 shedding rate in wastewater to daily positive  
533 tests data: A consistent model based approach". *Science of the Total Environment* 807, pp.  
534 150838–150838.

- 535 Proverbio, Daniele et al. (2022). “Model-based assessment of COVID-19 epidemic dynamics by  
536 wastewater analysis”. *Science of the Total Environment* 827, pp. 154235–154235.
- 537 Randazzo, W et al. (2020). “SARS-CoV-2 RNA in wastewater anticipated COVID-19  
538 occurrence in a low prevalence area”. *Water Research* 181, pp. 115942–115942.
- 539 Reyes, Bryan A, Julie S Pendergast, and Shin Yamazaki (2008). “Mammalian peripheral  
540 circadian oscillators are temperature compensated”. *Journal of Biological Rhythms* 23(1), pp.  
541 95–98.
- 542 Rose, C et al. (2015). “The characterization of feces and urine: a review of the literature to  
543 inform advanced treatment technology”. *Critical Reviews in Environmental Science and*  
544 *Technology* 45(17), pp. 1827–1879.
- 545 Saad-Roy et al. (2021). “Epidemiological and evolutionary considerations of SARS-CoV-2  
546 vaccine dosing regimes”. *Science* 372(6540), pp. 363–370.
- 547 Saguti, Fredy et al. (2021). “Surveillance of wastewater revealed peaks of SARS-CoV-2  
548 preceding those of hospitalized patients with COVID-19”. *Water Research* 189, pp. 116620–  
549 116620.
- 550 Sanjuán, R and P Domingo-Calap (2021). “Reliability of wastewater analysis for monitoring  
551 COVID-19 incidence revealed by a long-term follow-up study”. *Frontiers in Virology* 1.
- 552 Saththasivam et al. (2021). “COVID-19 (SARS-CoV-2) outbreak monitoring using wastewater-  
553 based epidemiology in Qatar”. *Science of The Total Environment* 774, pp. 145608–145608.
- 554 W Schmitz, Bradley et al. (2021). “Enumerating asymptomatic COVID-19 cases and estimating  
555 SARS-CoV-2 fecal shedding rates via wastewater-based epidemiology”. *Science of the Total*  
556 *Environment* 801, pp. 149794–149794.
- 557 Tuncer, Necibe et al. (2022). “Parameter identifiability and optimal control of an SARS-CoV-2  
558 model early in the pandemic”. *Journal of Biological Dynamics* 16(1), pp. 412–438.
- 559 Wölfel, Roman et al. (2020). “Virological assessment of hospitalized patients with COVID-  
560 2019”. *Nature* 581(7809), pp. 465–469.

561 Wu, F, W L Lee, et al. (2022b). “Making waves: Wastewater surveillance of SARS-CoV-2 in an  
562 endemic future”. *Water Research*, pp. 118535–118535.

563 Wu, Fuqing et al. (2020). “SARS-CoV-2 titers in wastewater are higher than expected from  
564 clinically confirmed cases”. *mSystems*. 2020;5: e00614-20.

565 Wu, Fuqing, Amy Xiao, et al. (2022a). “SARS-CoV-2 RNA concentrations in wastewater  
566 foreshadow dynamics and clinical presentation of new COVID-19 cases”. *Science of the Total  
567 Environment* 805, pp. 150121–150121.

568 Wu, Zhimin et al. (2019). “Predictability and identifiability assessment of models for prostate  
569 cancer under androgen suppression therapy”. *Mathematical Biosciences and Engineering* 16(5),  
570 pp. 3512–3536.

571 Xiao, Amy et al. (2022). “Metrics to relate COVID-19 wastewater data to clinical testing  
572 dynamics”. *Water Research*, pp. 118070–118070.

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

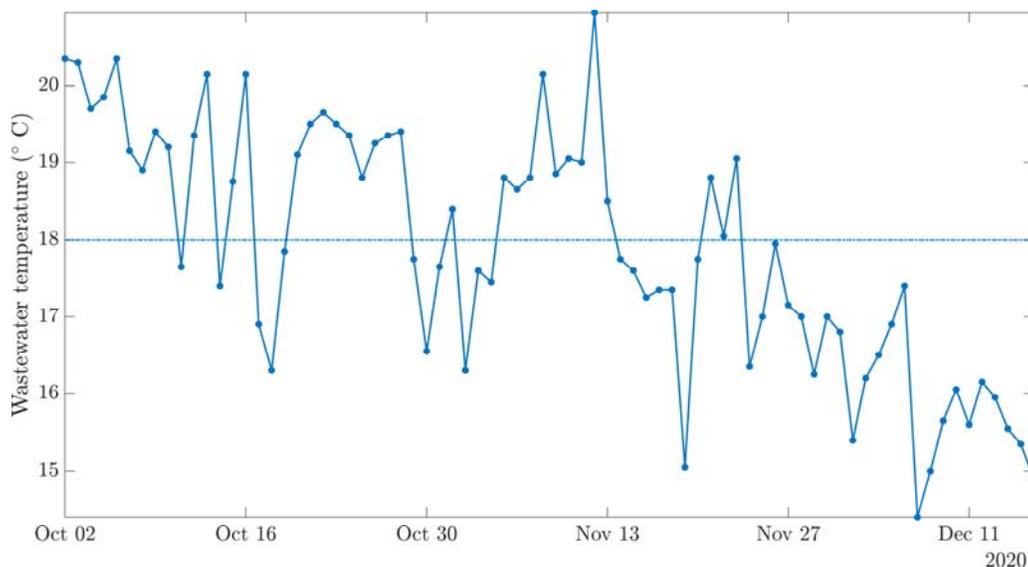
588

589

590

591 **Supplementary Figures**

592

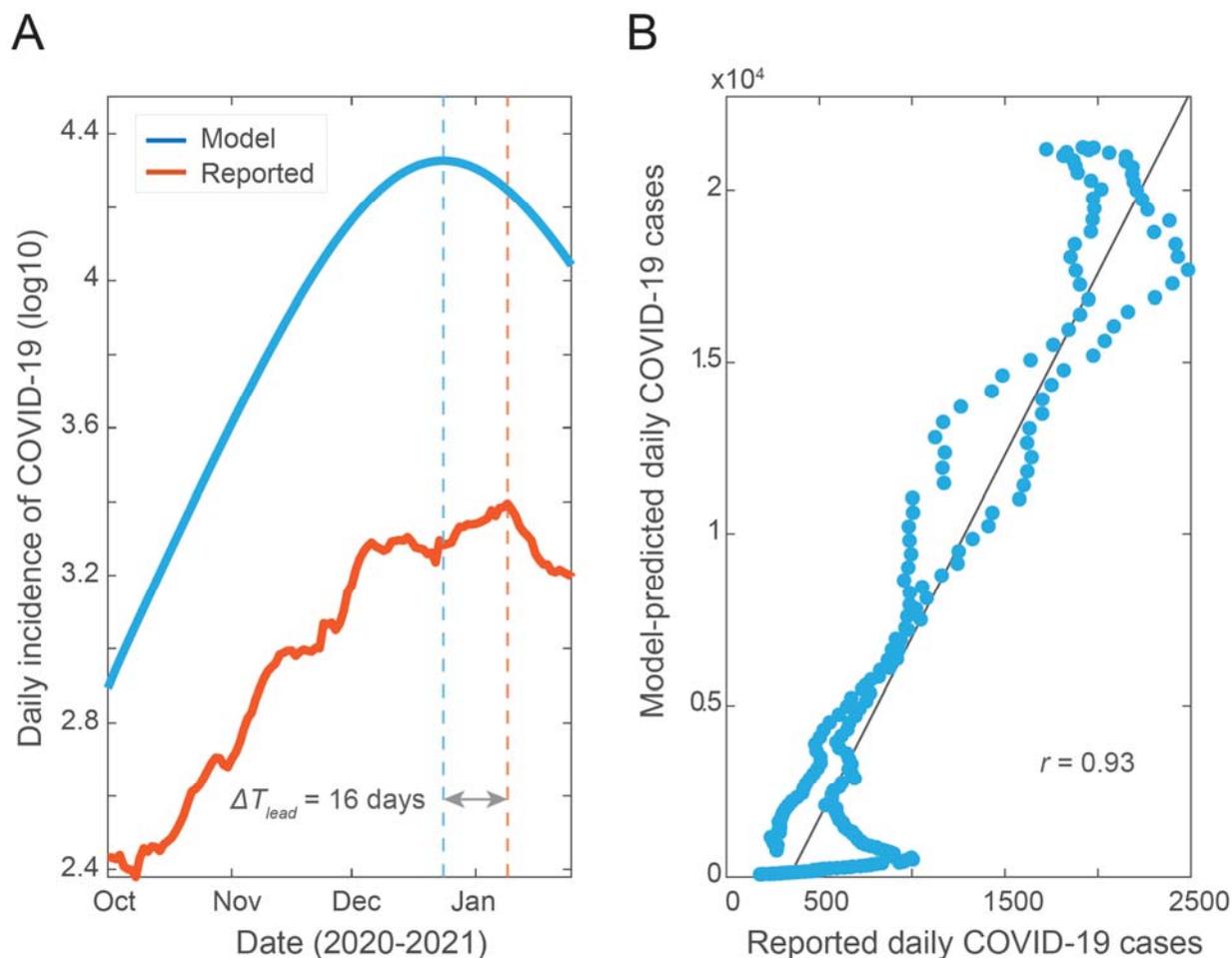


593

594 **Figure S1. Wastewater temperature data during the model fitting period.** Over the fitting  
595 duration (from Oct 02 to Dec 16, 2020), the average temperature is relatively constant at around  
596 18 degrees Celsius.

597

598

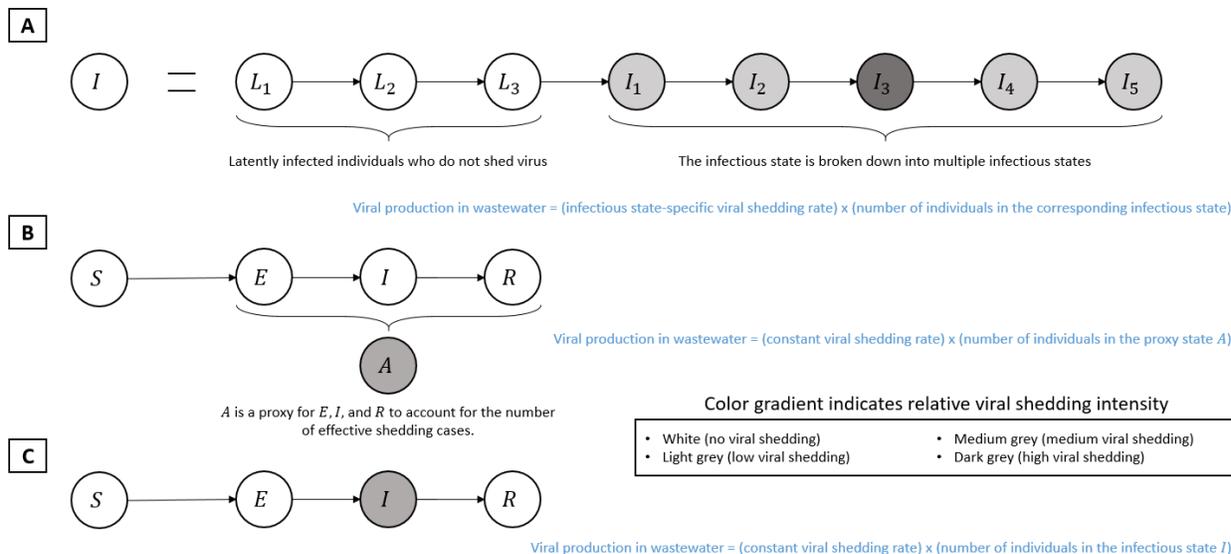


599

600 **Figure S2. Predicted cases and correlation plot for the SEIR-V model with temperature**  
601 **effect.** The wastewater data covers the second wave of pandemic (October 2, 2020 to January 25,

602 2021). (A) Model simulation of the true number of COVID-19 cases and clinically reported

603 cases. (B) Correlation between the number of model-predicted cases and reported cases (daily).



604

605 **Figure S3. Comparison of underlying frameworks for modeling of wastewater surveillance data.** Two general  
 606 methods have been reported to connect viral concentration in wastewater to epidemic model. (A) is representative of  
 607 Brouwer et al. and Nourbakhsh et al. (Brouwer et al., 2022; Nourbakhsh et al., 2022), (B) is representative of  
 608 Proverbio et al. (Proverbio et al., 2022), and (C) is the proposed model framework.