Online trend estimation and detection of trend deviations in sub-sewershed time series of SARS-CoV-2 RNA measured in wastewater

Katherine B. Ensor[a,*], Julia Schedler[a], Thomas Sun[a], Rebecca Schneider[c], Anthony Mulenga[c], Jingjing Wu[b], Lauren B. Stadler[b], Loren Hopkins[d]

[a] *Rice University, Department of Statistics, 6100 Main St, Houston, 77005*

[b] *Rice University, Department of Civil and Environment Engineering, 6100 Main St, Houston, 77005*

[c] *Houston Health Department, 8000 N. Stadium Dr., Houston, 7700*

[d] *Houston Health Department and Department of Statistics, Rice University, 6100 Main St, Houston, 77005*

Abstract

Wastewater surveillance has proven a key public health tool to understand a wide range of community health diseases and has proven to be especially critical to health departments throughout the SARS CoV-2 pandemic. The size of the population served by a wastewater treatment plant (WWTP) may limit the targeted insight about community disease dynamics. To investigate this concern, samples of wastewater were obtained at lift stations upstream of WWTPs within the sewer network. First, an online, semi-automatic time series model is fitted to the weekly measurements of WWTP samples to estimate the viral trend for the community and compared to the time series observations from the lift stations. Second, deviations from the WWTP trend are identified using an Exponentially Weighted Moving Average (EWMA) control chart. The analysis reveals that the lift stations display slightly different dynamics than the larger WWTP, highlighting the more granular insight gleaned from sampling sites which represent smaller populations. Discussion focuses on the use of our methods to support rapid public health decision-making based on additional, targeted samples in times of concern.

Keywords: Wastewater-based epidemiology, COVID-19, Time series data

1. Introduction

Wastewater-based epidemiology (WBE) is a cost-effective and fast way to survey the transmission of disease in populations, and it has been widely applied for the monitoring of viral pathogens, including SARS-CoV-2 (Kisand et al., 2023; Olesen et al., 2021). Multiple studies have confirmed the correlation between SARS-CoV-2 wastewater monitoring data and COVID-19 clinical testing data (Hopkins et al., 2023b; Ciannella et al., 2023; Kasprzyk-Hordern et al., 2023; Kaya et al., 2022; Mao et al., 2020; Peccia et al., 2020; Wolken et al., 2023). Routine wastewater monitoring highlights SARS-CoV-2 dynamics and provides early notice of the emergence in the population served by the wastewater catchment areas (Cao and Francis, 2021; Karthikeyan et al., 2021; Kisand et al., 2023; Li et al., 2021; Vallejo et al., 2022; Zhao et al., 2022; Kirby et al., 2022). In short, wastewater epidemiology is an effective public health tool guiding meaningful interventions (Hopkins et al., 2023a).

Wastewater monitoring for diseases has predominantly involved the collection of samples from the influent of wastewater treatment plants (WWTPs). Monitoring centralized WWTPs is typically easier for municipalities

*Katherine B. Ensor
Email address: ensor@rice.edu (Katherine B. Ensor)

to implement due to routine sampling that takes place at the treatment plants unrelated to WBE. Since these centralized WWTPs in large municipalities often serve large populations, it can be unclear where to take action when pathogens are detected.

Expanding spatial granularity of WBE surveillance can improve the accuracy and actionability of routine wastewater monitoring. Several prior studies focused on sampling sites upstream of the WWTP, such as lift stations, schools, universities, and hospitals (Castro-Gutierrez et al., 2022; Fielding-Miller et al., 2023; Gibas et al., 2021; Haak et al., 2022; Holm et al., 2022; Scott et al., 2021; Spurbeck et al., 2021). The WBE program for the City of Houston has routinely made use of sub-sewershed measurements (hou-wastewater epi.org/, 2023). Moreover, the nested sampling strategy was proposed to monitor the trend of SARS-CoV-2 in the catchment area and simultaneously identify and trace the community-level COVID-19 hotspots (Wolken et al., 2023; Wang et al., 2023; Yeager et al., 2021). The strategy highlighted the importance of sampling site selection, which can help develop a more sensitive and effective wastewater surveillance system for COVID-19 and other diseases.

Studies on the use of nested sampling strategy for wastewater monitoring remain limited, and more investigations are needed regarding the application of nested sampling on WBE to maximize the value of the data and improve the system efficiency (Acosta et al., 2022; D'Aoust et al., 2021; Holm et al., 2022; Wang et al., 2023; Yeager et al., 2021). D'Aoust et al. (2021) compared the wastewater SARS-CoV-2 surveillance results from two locations in a small, rural community with the community's COVID-19 cases, and recommended sampling from lift stations, because they observed an overall higher and more stable SARS-CoV-2 load as compared to its downstream wastewater treatment lagoon. Holm et al. (2022) observed no statistical difference between SARS-CoV-2 viral concentrations in WWTPs and several upstream sampling sites, including lift stations and manholes, and recommended sampling sites that support larger populations. Acosta et al. (2022) observed a lower correlation between COVID-19 clinic data and SARS-CoV-2 RNA in wastewater samples from neighborhoods compared to samples from WWTPs, indicating that the current smaller wastewater monitoring system may induce more heterogeneous data that can reduce the sensitivity of the surveillance. Acosta et al. (2022); Haak et al. (2022) also indicated that issues associated with sample collection may decrease the stability and representativeness of the data collected at the community or neighborhood level.

Despite the potential community-wide benefits in wastewater monitoring at the sub-sewershed level, statistical analyses on this approach have been limited in their modelling assumptions and flexibility. Several analyses assume wastewater measurements over time to be independent (Acosta et al., 2022; D'Aoust et al., 2021; Holm et al., 2022) and compare differences between sewershed and sub-sewershed measurements using a simple group-level difference of means or pairwise correlations. Failure to account for time dependence may underestimate variability and inadequately capture the temporal dynamics of the wastewater measurements, where clear trends and patterns in the data exist. Within the wide literature of statistical models for SARS-CoV-2 wastewater longitudinal trends, most typically fall under the categories of autoregressive or regression-based models (Cao and Francis, 2021; Peccia et al., 2020; Jeng et al., 2023), SEIR models (McMahan et al., 2021; Fazli et al., 2021), or machine learning approaches such as the artificial neural network (ANN) of Li et al. (2021) or the time-series based machine learning approaches in Lai et al. (2023). Application of these models to the comparison of sub-sewershed time series has been sparse to nonexistent, and many may not be appropriate for detection of structural deviations away from downstream measurements.

Previously, cubic smoothing splines have been used by the City of Houston (Hopkins et al., 2023b; Stadler et al., 2020) to identify the sometimes rapidly changing signal in the noisy wastewater time series obtained through routine monitoring of the city wastewater. This spline model captures the trend and curvature of the trend in the nonlinear time series from the noisy wastewater measurements. However, smoothing splines cannot separate the inherent variability of the time series from the measurement process.

We conduct two statistical analyses that offer flexible, semi-automatic, and online estimation of wastewater dynamics between large WWTPs and sub-sewersheds. First, we model the wastewater measurements using a dynamic linear state-space time series model. This modelling framework allows for both online and

retrospective estimation of wastewater trends accompanied with confidence bands to capture the precision of the estimates via the Kalman filter and smoother Shumway and Stoffer (2017). These estimates give broad insight into whether sub-sewershed time series give different information from large, centralized WWTP time series, as well as the ability to forecast future values. Second, we utilize tools from statistical process control (SPC) literature, namely exponentially weighted moving average (EWMA) control charts, to monitor whether even just one of the lift station measurements deviates significantly from the trend estimate for the larger WWTP, allowing for immediate online detection of community-specific spikes in SARS-CoV-2.

2. Methods

2.1. Data description

The City of Houston has 39 WWTPs, serving populations from approximately 500,000 to 600 individuals. Within the larger WWTPs, there are a number of lift station (LS) facilities where wastewater can be sampled and may serve to refine the geographic resolution provided by wastewater analysis. This work focuses on the largest WWTP that serves a population of roughly 551,150 people. Wastewater was sampled from May 24, 2021 through March 13, 2023 for four lift stations (see Figure 1) which are geographically contained within the large WWTP catchment area.

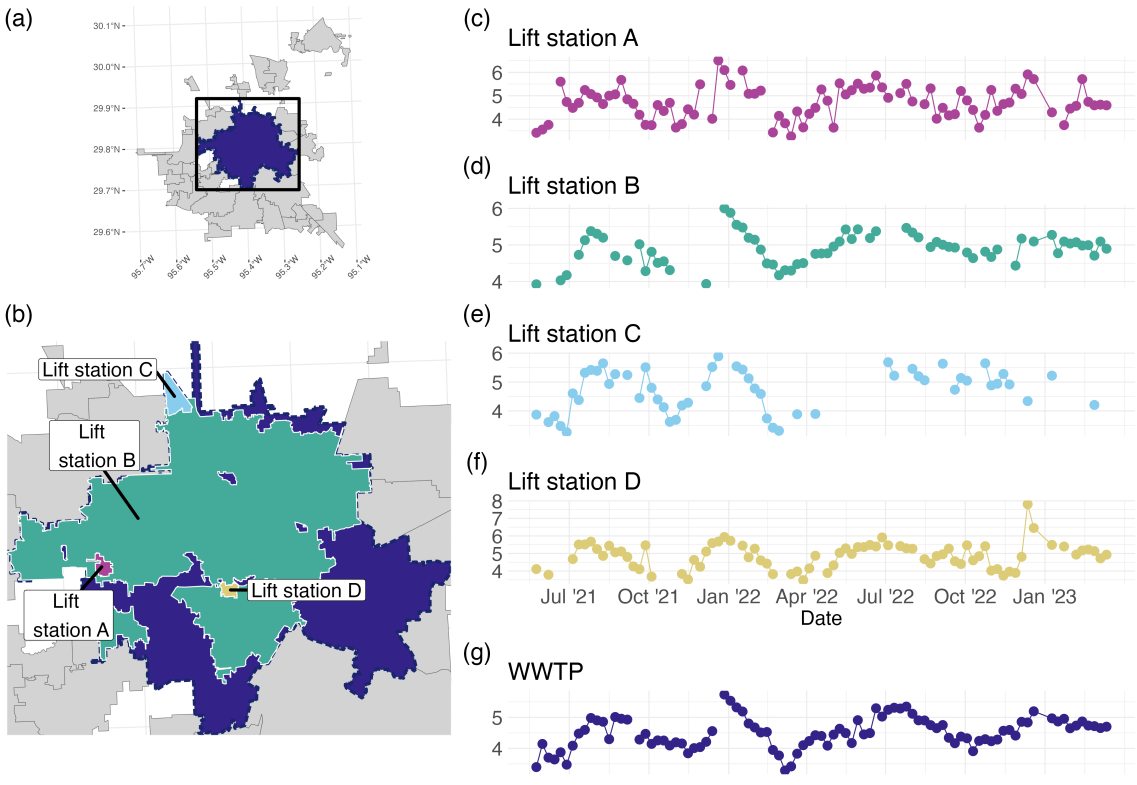

Figure 1: (a) The WWTP catchment areas for the City of Houston, with the WWTP of focus shaded. The box shows the extent of (b), the map showing the 4 lift stations considered in the analysis. (c)-(f) plot the time series of Log10 Copies/L for the WWTP and the 4 lift station facilities, referred to as Lift Station A-D.

Data on wastewater analysis results for the lift stations and the WWTP was updated on a weekly basis. For each weekly sample, we quantified SARS-CoV-2 N1 and N2 gene copies per liter of wastewater, as described previously (Hopkins et al., 2023b). We average the N1 and N2 concentrations to simplify our analysis and focus on the comparison between the WWTP and LS time series. All measurements taken in a given week

were aligned to the corresponding Monday of that week. Wastewater viral concentration data was received in units of copies per liter and was subsequently log transformed in base 10 ($\log_{10}$). Any measurements below the level of detection (LOD) were labeled as missing values. Table 1 contains the names of the 5 series considered and summary statistics for each series. Figure 1 (a) and (b) are maps of the WWTP and LS catchments for each of the series, and Figure 1 (c)-(f) plot the time series of observed values for all 5 series on the $\log_{10}$ scale.

Name	Population	Mean	St. Dev.	Min	Max	Missing
WWTP	551150	4.51	0.50	3.29	5.74	4
Lift Station A	2442	4.72	0.70	3.26	6.51	11
Lift Station B	373937	4.89	0.43	3.92	6.00	28
Lift Station C	4849	4.72	0.73	3.26	5.89	42
Lift Station D	1724	4.88	0.74	3.48	7.81	18

Table 1: Name, size of population, and summary statistics Log10 of average of replicate RNA N1 and N2 copies/L for each wastewater treatment plant (WWTP) or lift station (LS) considered. The study period spanned 93 weeks.

2.2. Hierarchical Time Series Model for Trend Estimation

When time series data are collected, the goal is often to estimate a trend, that is, whether the "typical values" are changing in time. For example, 1 (c)-(g) show the times series of viral concentration of SARS-CoV-2. Visually, it is clear that these values are changing in time, and even seem to exhibit similar behavior that may be predictable with a well-chosen model. Such a model should be able to separate out the "noise", or observation/measurement error, in these observations from the "signal", or trend. An additional constraint when modeling time series data is the presence of temporal correlation structure, i.e. the values are not independent, so models which assume independence can lead to misleading forecasts and/or conclusions about which variables are important in modeling a time series (Hyndman and Athanasopoulos, 2021). In summary, the desired model will separate sources of variability for both the trend and the observation as well as account for temporal correlation.

The state space modeling framework can accommodate both these needs. A state space model represents a time series in two levels: an unobserved trend which encodes temporal dependence structure and a noisy observed time series. In other words, it is a hierarchical model which is able to separate sources of variability as desired. In the time series literature, the levels of this model are called the state equation and the observation equation. Equations 1 and 2 display the state space model used for each series in this particular study:

$$\text{Observation equation: } y_t = \mu_t + v_t \tag{1}$$

$$\text{State equation: } (\mu_t - \mu_{t-1}) = (\mu_{t-1} - \mu_{t-2}) + w_t. \tag{2}$$

$$\text{Initial condition: } \mu_0 \sim N(c_0, m_0). \tag{3}$$

The error terms v_t and w_t are independent and normally distributed with mean zero, and variances denoted by σ_v^2 for the observation error and σ_w^2 for the state error.

The observation model of Equation 1 represents the model fit to the concentration of SARS-CoV-2 RNA in wastewater measured by the lab. The observation model is the underlying state μ_t plus a variance term σ_v^2 corresponding to the inherent measurement and sampling error. The state model in Equation 2 represents the true state of the viral trend derived from the measured concentration of SARS-CoV-2 RNA, for the sampled region. The noise term associated with the state equation, σ_w^2, represents the natural variability in the viral concentration in the population as measured by wastewater.

Within this framework, the state variable serves as the core component of the model, characterizing the underlying system's behavior and dynamics, in other words the trend of the virus concentration. Note

the temporal structure encoded by Equation 2: the right hand side concerns the previous two time points, while the left hand side concerns the current and past time point. In particular, Equation 2 encompasses a statistical framework that employs the concept of first difference applied twice. The first difference operation captures the change in the state variable over successive time periods, and by applying this operation twice, we gain insights into the acceleration or curvature of the trend. That is, this choice of structure for the state equation is chosen to capture the temporal dependence of the SARS-CoV-2 RNA concentration. For additional details about the state-space modeling framework and its relations to smoothing splines, see Shumway and Stoffer (2017).

Once the structure of the model is chosen, the model can be fit to the data with three goals in mind: retrospective estimates of the trend using all available data, online estimates of the trend using only past data up to a given time point, and one-step-ahead forecasts of the next time point. In the time series literature, the retrospective and online estimates are referred to as smoothers and filters, respectively. We focus on retrospective and online estimates for this paper, but provide steps for obtaining the one-step-ahead forecasts in the supplemental materials.

To estimate the online and retrospective trends, four parameters are themselves estimated: the initial state mean and variance, the variance of the measurement and sampling error (σ_v^2), and the variance of the trend (σ_w^2). Estimates are obtained through maximum likelihood estimation which is computationally fast due to use of the Kalman Filter for updating linear Gaussian systems. For the online estimates, a rolling estimation structure is used, meaning the parameters are re-estimated with each new time point. Estimation is implemented using the KFAS package in R Helske (2017), which can easily handle missing data. Once an estimate of the model is obtained, a step to check that the model fits the data is required. For the present model, an autocorrelation plot of the model's residuals can be checked for autocorrelation. If no correlation is present in the residuals, the model can be considered a good fit. Additional details of estimation and model fit checks as well as all code used for the analysis are available in the supplemental materials.

The inputs, outputs, and process of fitting the spline state space model of Equations 1 and 2 are summarized in Algorithm 1.

Algorithm 1: Variability-separating trend estimation

Input: Raw lab values

1 Process the raw data (average N1 and N2 replicates, identify non-detects, transform copies/L to Log10).
2 Initialize the model by estimating parameters using the first 10 weeks of data.
3 Compute online trend estimates and confidence limits using Equation 2 and re-estimating parameters with each additional data point.
4 Compute retrospective trend estimates and confidence limits.
5 Verify convergence of estimates and time series modeling assumptions.
6 Compute table of state and observation variances for each time series.
7 Compare visualizations of retrospective estimates of WWTP and sub-sewershed trends to determine whether a difference was present.

Output: Online and retrospective trend estimates, estimates of trend variability and measurement/sampling variability.

Visualizations of the retrospective and online estimates along with the data are provided in 2.

2.3. Detection of trend deviations

Recall the goal of determining whether sub-sewershed measurements give different information than the routinely monitored centralized WWTP measurements. Using all available data, the retrospective estimates from the model fit using Equations 1 and 2. These estimates, visualized in Figure 3, show some periods of separation, indicating that the sub-sewershed measurements do indeed give different information. However,

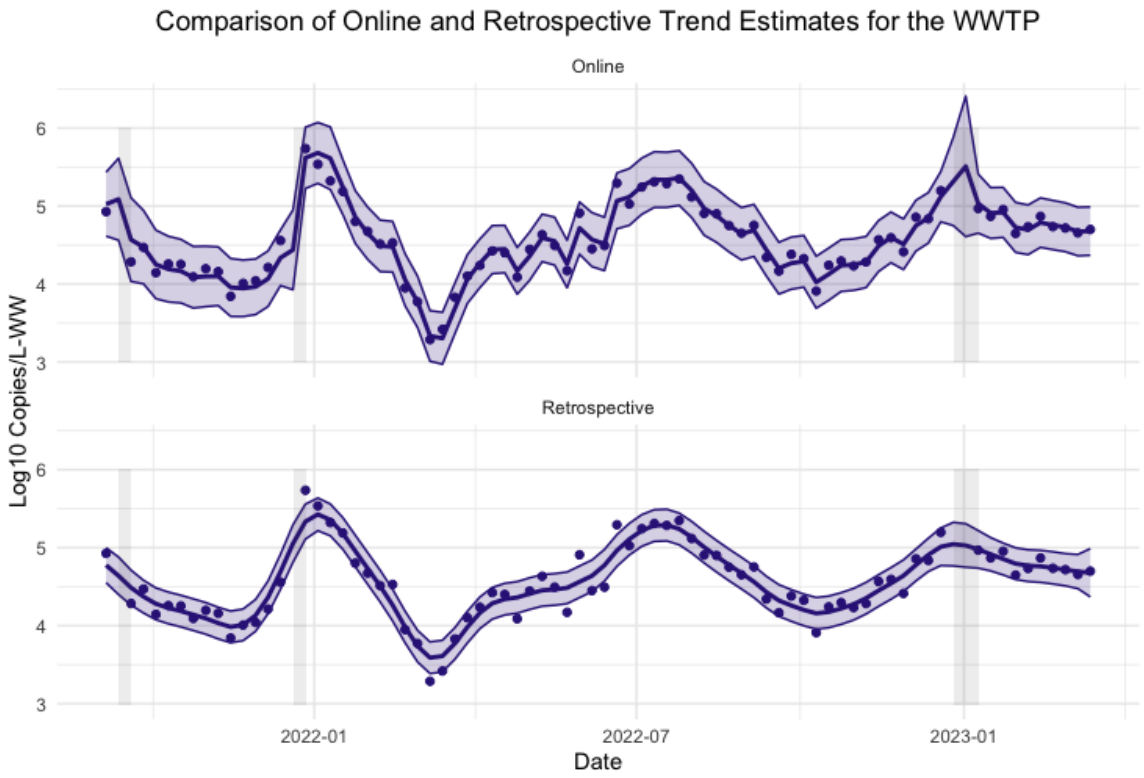

Figure 2: Retrospective and online estimates of the viral concentration trend with uncertainty quantification for the large WWTP. The vertical axis is log10 copies/liter. The shaded grey rectangles correspond to periods of missing data. Note that the online trend estimates are "noisier" and have wider uncertainty bands than the retrospective trend estimates.

if the goal is to extract actionable information from the data, the online estimates, which only use data up to the current time point, should be used. While the retrospective estimates show clear separation, the online estimates are noisier, so detecting when the sub-sewersheds may be deviating from the WWTP's trend requires more than a visual comparison of the two series. In addition, sub-sewersheds may not be sampled frequently enough to support the model described in Section 2.2, so a method which can be used with at least one sub-sewershed observation is ideal.

The statistical process control (SPC) literature provides a framework for iterative improvement of a decision-making process based on time series data. Some examples of the traditional applications of SPC include ensuring a given percentage of on-time deliveries to a client, speed and consistency of service quality in a bank, and loading passengers onto an airplane (Montgomery, 2009). In short, SPC provides a framework for identifying when a time series of interest is "out of control" so that steps can be taken to bring that series back "in control". Although the ability to bring disease burden in a community back "in control" is limited in WBE compared to traditional applications, ideas from SPC can be borrowed to improve the actionability of the information contained in wastewater time series.

For this paper, the time series of interest is the difference between the sub-sewershed and the WWTP. If we simply subtract the observed values for each series, the resulting difference will contain the "noise", or measurement and sampling error. Instead, we use the online estimate of the trend for the WWTP obtained from Equations 1 and 2, which can be assumed to be free of observation error. Since the online estimate of the trend requires 10 weeks of data to be initialized, we use the observed (unmodeled) value(s) from the sub-sewershed directly.

Formally, using the previous notation, the standardized difference at time point t, for lift station $i = 1, \ldots, 4$ is given by:

$$d_{i,t} = \frac{y_{i,t} - \hat{\mu}_t}{\tilde{\sigma}_d}, \tag{4}$$

where $\tilde{\sigma}_d^2 = \text{Var}(y_{i,t} - \hat{\mu}_t)$. This variance is approximated by

$$\tilde{\sigma}_d^2 \approx \hat{\sigma}_{v_t}^2 + \hat{\sigma}_{w_t}^2 - 2\text{Corr}(y_i, \hat{\mu}) \cdot \hat{\sigma}_{v_t} \cdot \hat{\sigma}_{w_t}, \tag{5}$$

where $\text{Corr}(y_i, \hat{\mu})$ is the Pearson correlation coefficient between the WWTP estimated state time series and the observed copies/liter from the i^{th} lift station. If any of the sub-sewershed values $y_{i,t}$ are missing, we replace these values with the online trend estimate for the WWTP, which will yield a value of 0.

If the sub-sewershed and the WWTP are "in control", or gave equivalent information, then $d_{i,t}$ would be normally distributed with mean 0, and there would be no autocorrelation in the series. To determine whether the sub-sewershed is "out of control", or separating from the trend of the WWTP, a control chart can be constructed. Many types of control charts are available for different scenarios, for example, Shewhart (Shewhart, 1931) and CUSUM (Page, 1954) control charts. We choose an Exponentially Weighted Moving Average control chart (Roberts, 1959), which can detect small shifts in temporally correlated series such as our $d_{i,t}$ and is appropriate for use with individual observations (Montgomery, 2009). The EWMA chart is based on the following series:

$$z_{i,t} = \lambda d_{i,t} + (1 - \lambda) z_{i,t-1}, \tag{6}$$

where $z_{i,t}$ can be interpreted as a weighted average of all past values for series i, where the weighting is controlled by the value λ, for which we use the estimate of the lag 1 autocorrelation of $d_{i,t}$. In the case of a missing sub-sewershed value, the aforementioned replacement with the WWTP online estimate allows for the exponential weighting of past values to continue under the assumption of no separation. The EWMA charts are visualized for each of the 4 lift stations compared to the WWTP in Figure 4.

The dots on Figure 4 represent the values of z_t. The dotted lines are the upper and lower confidence limits. When z_t exceeds one of these confidence limits, the point is colored red, and the sub-sewershed can be considered "out of control", in other words, the sub-sewershed time series is separating from the WWTP time series which gives different information. The direction of the separation can also be determined by examining whether the point exceeds the upper limit, indicating the viral concentration is higher for the sub-sewershed, or the lower limit, indicating the sub-sewershed is lower.

We summarize the creation of the EWMA chart in Algorithm 2. For additional discussion of EWMA charts and examples of their use with correlated data, see Hunter (1986); Lucas and Saccucci (1990); Supharakonsakun et al. (2020).

Algorithm 2: Detecting deviation of sub-sewershed measurement from centralized WWTP trend estimate

Input: At least $10 + n$ WWTP observations, $n \geq 1$ sub-sewershed observations

1 Read in cleaned WWTP series and apply 1 to obtain online trend estimates through the date of the first sub-sewershed observation.
2 Replace any missing sub-sewershed observations with WWTP online trend estimate for corresponding date.
3 Create difference time series of sub-sewershed observed copies/liter ($\log 10$) - WWTP Online Trend Estimate.
4 Standardize the difference series by dividing by the standard deviation computed in Equation 5.
5 Construct EWMA chart for the standardized difference series.
6 Inspect EWMA chart for separation.

Output: EWMA chart for determining separation of sub-sewershed from centralized WWTP.

Data Availability

Given the small populations associated with some of the lift stations, real data will be made available on a case-by-case basis by contacting the corresponding author and subsequent approval by Houston Health Department. Synthetic wastewater surveillance data which preserves the statistical properties of the real data are available along with code on a GitHub repository.

Code Availability

All code used to fit the models described in this paper are available in a public GitHub repository. All code is written in the R language (R Core Team, 2023).

3. Results

3.1. Trend Estimation

The retrospective estimates depicted in Figure 3 indicates three peaks in the estimated population viral dynamics for the population served by the WWTP, with maximums that occur on January 3 and July 18, 2022, and January 9, 2023. We will refer to these peaks as PK1, PK2 and PK3, respectively. The retrospective review illustrates there are instances where the lift stations provided early information with respect to increasing or decreasing viral trends in the population measured. Again, these features are highlighted in Figure 4. A separation in the confidence intervals for each series indicates a statistically significant difference in the estimated trend for the respective series, namely the trend estimated for the WWTP and each of the lift stations.

We also see in the retrospective review that three of the four lift stations, namely Lift station A, Lift station C, and Lift station D, exhibit a comparable trend as that estimated from the WWTP, with a few deviations

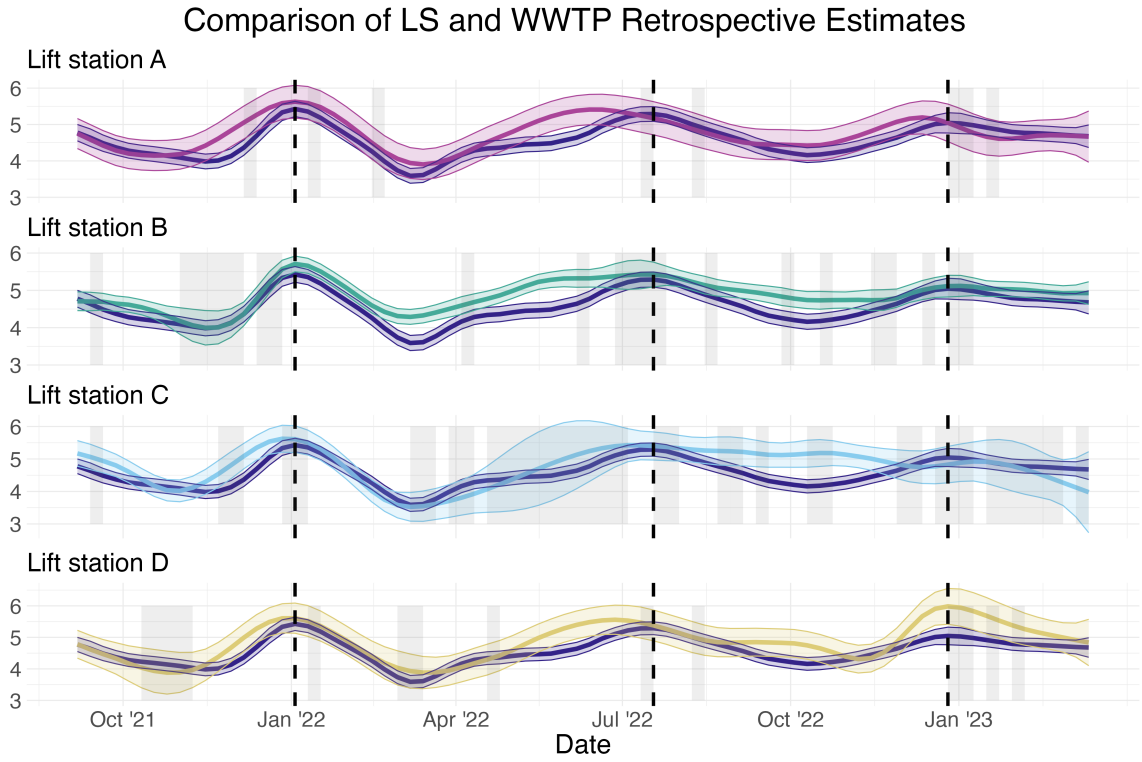

Figure 3: Retrospective estimates of the viral concentration trend with uncertainty quantification for the WWTP and each LS series, using all available information. The vertical axis is log10 copies/liter. The shaded grey rectangles correspond to periods of missing data and the dotted lines correspond to the peaks of three surges. Note that the time series model is still able to provide estimates of the trend during periods of missing data, though with greater uncertainty. Compared to Figure 1, the start date of this trend plot is later, since the first 10 weeks of data are used to initialize the model.

(see Figure 3). Lift station A indicates an early signal leading up to PK1. Lift station C separates from the WWTP trend by remaining at high levels between PK1 and PK2. It is important to note that there are several missing values in the Lift station A series during this time, as highlighted by the light grey bars in Figure 3. However, the estimated level for Lift station C does not show an increase in uncertainty due to the fact that the observed levels during this time were relatively consistent (see Figure 1). Lift station D registers a statistically higher trend in PK3. Lift station B was unique amongst the four lift stations, in that its estimated trend separated from the estimated WWTP trend following PK1 and remained higher until PK2. Lift station B also failed to drop as low between PK2 and PK3. However, the trend estimates for the WWTP and Lift station B were not significantly different in PK3. It is also worth noting that Lift station A, Lift station C, and Lift station D exhibited more variation in their trend estimates as evidenced by the width of the respective confidence intervals, than that of the WWTP.

Recall that the hierarchical trend estimation framework can separate variability associated with the trend from the "noise", or lab and sampling error. These results are summarized in Table 2. We see from this table that the measurement and sampling variation are highest for Lift station A and Lift station D, and also elevated for Lift station C. Since we expect the lab variability to be approximately constant across all measurements, the extra variation is most likely due to the lift station wastewater containing highly variable levels of SARS-CoV-2 due to the small population that it serves. Based on this observation, the sampling variation is approximately equivalent for Lift station B as it is for the WWTP. The state dynamics for each location exhibit similar variability, with slightly elevated variation for Lift station D.

Name	Sampling and Lab Variability	Trend Variability	Population
WWTP	0.0372	0.0130	551150
Lift station A	0.2798	0.0105	2442
Lift station B	0.0350	0.0130	373937
Lift station C	0.1374	0.0134	4849
Lift station D	0.2810	0.0175	1724

Table 2: Estimates of inherent variability, σ_w (state) and measurement variability σ_v (observation, lab and sampling variability) for each series.

Again, the retrospective review provides the best understanding of the dynamic population trend in virus levels for location as well as insight into the lab and sampling variability over the entire study period. However, this retrospective review is not useful in real time as it requires knowledge of the full time series, i.e. the future, to implement.

3.2. Trend Deviations

The results of Algorithm 2 applied to each lift station and the WWTP are graphically introduced in Figure 4. The information is consistent with the retrospective study in that Lift station A, Lift station C and Lift station D, all demonstrate minor perturbations from the trend estimated for the WWTP. Further, Lift station B clearly demonstrates a strong and consistent deviation from the WWTP estimated trend, between PK1 and PK2, and then again between PK2 and PK3. In Figure 4 we also include the observed standardized difference between the two measurements. You will note that the differences may be large, but they are not statistically significant based on the EWMA control chart. The control chart is used to identify a level shift in the trend, and not specific outlying events. Based on the EWMA control chart, statistically significant level shifts occurred at the red highlighted temporal locations.

4. Discussion

The objective of this paper is to highlight the additional information gleaned from samples taken within the sewer network and upstream of WWTPs, namely lift stations. We bring forward a hierarchical time

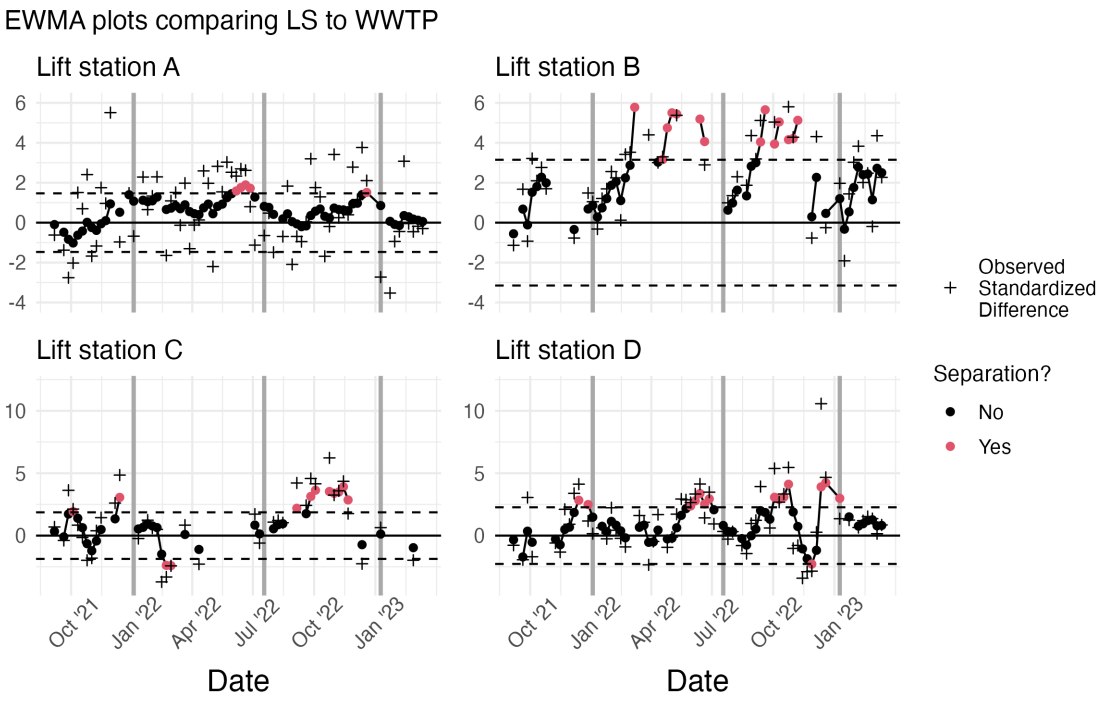

Figure 4: The EWMA chart for the observed values at each lift station compared to the WWTP online estimate. The solid dots represent the exponentially weighted standardized difference while the plus signs represent the actual standardized difference. Observations which correspond to a structural break, or exponentially weighted values beyond the dotted control limits, are colored red. The dark grey vertical lines are the approximate dates of the peaks of different surges.

series approach to capture the dynamic trend in population viral dynamics from each wastewater series. The state-space time series model is simple to implement both in a retrospective, and real-time mode, and naturally adapts to the nonlinear dynamics in the population viral trend. The EWMA control charts provide a framework for identifying when sub-sewershed measurements deviate from measurements in a larger, centralized WWTP. Note that these deviations could be due to any number of factors, for example, number of infections, dilution, degradation, inhibitors, etc.

For our system the only lift station within the large wastewater catchment area of the WTTP whose trend consistently deviated from that of the WWTP was Lift station B. Further, the measurement and sampling uncertainty for Lift station B was on par with that of the WWTP. This lift station serves 373,937 people whereas the WWTP serves 551,150 people. In other words, Lift station B serves 68% of the people in the large catchment area. Regular monitoring of Lift station B in addition to the WWTP is warranted based on this study. The Lift station B state estimate of viral load and its uncertainty, indicates that for the 68% of the population served by Lift station B, the viral load did not decrease as substantially as that of the WWTP between PK1 and PK2, and also between PK2 and PK3.

For the lift stations serving smaller populations, namely Lift station A, Lift station C and Lift station D, we see evidence of early signals through each COVID-19 peak. However, the measurement and sampling uncertainty with these smaller lift stations was substantially higher. Although routine monitoring may be prohibitively expensive, monitoring through times of high concern to public health may be warranted.

A side result of our modeling approach is the opportunity to separate the variation in the trend of the viral load from the measurement and sampling variation. In this comparison, and assuming a consistent measurement or lab variability across all samples, we find that the sampling variation for the smaller lift stations is much greater than that for the WWTP and the large Lift station B. If regular sampling at smaller lift stations, where flow may be irregular, is required, sampling strategies may need to be reviewed.

Acknowledgements

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the CDC Foundation (project no. 1085.46) and the Centers for Disease Control and Prevention (ELC-ED grant no. 6NU50CK000557-01-05 and ELC-CORE grant no. NU50CK000557). The work was also supported by the Rockefeller Foundation.

Bibliography

Acosta, N., Bautista, M.A., Waddell, B.J., McCalder, J., Beaudet, A.B., Man, L., Pradhan, P., Sedaghat, N., Papparis, C., Bacanu, A., Hollman, J., Krusina, A., Southern, D.A., Williamson, T., Li, C., Bhatnagar, S., Murphy, S., Chen, J., Kuzma, D., Clark, R., Meddings, J., Hu, J., Cabaj, J.L., Conly, J.M., Dai, X., Lu, X., Chekouo, T., Ruecker, N.J., Achari, G., Ryan, M.C., Frankowski, K., Hubert, C.R.J., Parkins, M.D., 2022. Longitudinal SARS-CoV-2 RNA wastewater monitoring across a range of scales correlates with total and regional COVID-19 burden in a well-defined urban population. Water Research 220, 118611. URL: https://doi.org/10.1016/j.watres.2022.118611, doi:10.1016/j.watres.2022.118611.

Cao, Y., Francis, R., 2021. On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. Science of The Total Environment 786, 147451. URL: https://doi.org/10.1016/j.scitotenv.2021.147451, doi:10.1016/j.scitotenv.2021.147451.

Castro-Gutierrez, V., Hassard, F., Vu, M., Leitao, R., Burczynska, B., Wildeboer, D., Stanton, I., Rahimzadeh, S., Baio, G., Garelick, H., Hofman, J., Kasprzyk-Hordern, B., Kwiatkowska, R., Majeed, A., Priest, S., Grimsley, J., Lundy, L., Singer, A.C., Di Cesare, M., 2022. Monitoring occurrence of SARS-CoV-2 in school populations: A wastewater-based approach. PLOS ONE 17, e0270168. URL: https://doi.org/10.1371/journal.pone.0270168, doi:10.1371/journal.pone.0270168.

Ciannella, S., González-Fernández, C., Gomez-Pastora, J., 2023. Recent progress on wastewater-based epidemiology for COVID-19 surveillance: A systematic review of analytical procedures and epidemiological modeling. Science of The Total Environment 878, 162953. URL: https://doi.org/10.1016/j.scitotenv.2023.162953, doi:10.1016/j.scitotenv.2023.162953.

D'Aoust, P.M., Towhid, S.T., Mercier, , Hegazy, N., Tian, X., Bhatnagar, K., Zhang, Z., Naughton, C.C., MacKenzie, A.E., Graber, T.E., Delatolla, R., 2021. COVID-19 wastewater surveillance in rural communities: Comparison of lagoon and pumping station samples. Science of The Total Environment 801, 149618. URL: https://doi.org/10.1016/j.scitotenv.2021.149618, doi:10.1016/j.scitotenv.2021.149618.

hou-wastewater epi.org/, 2023. Houston Wastewater Epidemiology. URL: https://hou-wastewater-epi.org.

Fazli, M., Sklar, S., Porter, M.D., French, B.A., Shakeri, H., 2021. Wastewater-Based Epidemiological Modeling for Continuous Surveillance of COVID-19 Outbreak, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, Orlando, FL, USA. pp. 4342–4349. URL: https://ieeexplore.ieee.org/document/9671543/, doi:10.1109/BigData52589.2021.9671543.

Fielding-Miller, R., Karthikeyan, S., Gaines, T., Garfein, R.S., Salido, R.A., Cantu, V.J., Kohn, L., Martin, N.K., Wynn, A., Wijaya, C., Flores, M., Omaleki, V., Majnoonian, A., Gonzalez-Zuniga, P., Nguyen, M., Vo, A.V., Le, T., Duong, D., Hassani, A., Tweeten, S., Jepsen, K., Henson, B., Hakim, A., Birmingham, A., De Hoff, P., Mark, A.M., Nasamran, C.A., Rosenthal, S.B., Moshiri, N., Fisch, K.M., Humphrey, G., Farmer, S., Tubb, H.M., Valles, T., Morris, J., Kang, J., Khaleghi, B., Young, C., Akel, A.D., Eilert, S., Eno, J., Curewitz, K., Laurent, L.C., Rosing, T., Knight, R., Baer, N.A., Barber, T., Castro-Martinez, A., Chacón, M., Cheung, W., Crescini, E.S., Eisner, E.R., Franco Vargas, L., Hakim, A., Hobbs, C., Lastrella, A.L., Lawrence, E.S., Matteson, N.L., Gangavarapu, K., Ngo, T.T., Seaver, P., Smoot, E.W., Tsai, R., Xia, B., Aigner, S., Anderson, C., Belda-Ferre, P., Sathe, S., Zeller, M., Andersen, K.G., Yeo, G.W., Kurzban, E., 2023. Safer at school early alert: an observational study of wastewater and surface monitoring to detect COVID-19 in elementary schools. The Lancet Regional Health - Americas 19, 100449. URL: https://doi.org/10.1016/j.lana.2023.100449, doi:10.1016/j.lana.2023.100449.

Gibas, C., Lambirth, K., Mittal, N., Juel, M.A.I., Barua, V.B., Roppolo Brazell, L., Hinton, K., Lontai, J., Stark, N., Young, I., Quach, C., Russ, M., Kauer, J., Nicolosi, B., Chen, D., Akella, S., Tang, W., Schlueter, J., Munir, M., 2021. Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. Science of The Total Environment 782, 146749. URL: https://doi.org/10.1016/j.scitotenv.2021.146749, doi:10.1016/j.scitotenv.2021.146749.

Haak, L., Delic, B., Li, L., Guarin, T., Mazurowski, L., Dastjerdi, N.G., Dewan, A., Pagilla, K., 2022. Spatial and temporal variability and data bias in wastewater surveillance of SARS-CoV-2 in a sewer system. Science of The Total Environment 805, 150390. URL: https://doi.org/10.1016/j.scitotenv.2021.150390, doi:10.1016/j.scitotenv.2021.150390.

Helske, J., 2017. KFAS: Exponential Family State Space Models in R. Journal of Statistical Software 78, 1–39. URL: https://doi.org/10.18637/jss.v078.i10, doi:10.18637/jss.v078.i10.

Holm, R.H., Mukherjee, A., Rai, J.P., Yeager, R.A., Talley, D., Rai, S.N., Bhatnagar, A., Smith, T., 2022. SARS-CoV-2 RNA abundance in wastewater as a function of distinct urban sewershed size. Environmental Science: Water Research & Technology 8, 807–819. URL: https://doi.org/10.1039/D1EW00672J, doi:10.1039/D1EW00672J. publisher: The Royal Society of Chemistry.

Hopkins, L., Ensor, K.B., Stadler, L., Johnson, C.D., Schneider, R., Domakonda, K., McCarthy, J.J., Septimus, E.J., Persse, D., Williams, S.L., 2023a. Public Health Interventions Guided by Houston's Wastewater Surveillance Program During the COVID-19 Pandemic. Public Health Reports , 00333549231185625URL: http://journals.sagepub.com/doi/10.1177/00333549231185625, doi:10.1177/00333549231185625.

Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., Stadler, L.B., 2023b. Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves. Science of The Total Environment 855, 158967. URL: https://doi.org/10.1016/j.scitotenv.2022.158967, doi:10.1016/j.scitotenv.2022.158967.

Hunter, J.S., 1986. The Exponentially Weighted Moving Average. Journal of Quality Technology 18, 203–210. URL: https://doi.org/10.1080/00224065.1986.11979014, doi:10.1080/00224065.1986.11979014.

Hyndman, R.J., Athanasopoulos, G., 2021. Forecasting: Principles and Practice. 3rd edition ed., OTexts, Melbourne, Australia. URL: https://otexts.com/fpp3/expsmooth.html.

Jeng, H.A., Singh, R., Diawara, N., Curtis, K., Gonzalez, R., Welch, N., Jackson, C., Jurgens, D., Adikari, S., 2023. Application of wastewater-based surveillance and copula time-series model for COVID-19 forecasts. Science of The Total Environment 885, 163655. URL: https://linkinghub.elsevier.com/retrieve/pii/S0048969723022751, doi:10.1016/j.scitotenv.2023.163655.

Karthikeyan, S., Ronquillo, N., Belda-Ferre, P., Alvarado, D., Javidi, T., Longhurst, C.A., Knight, R., 2021. High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of Community Infection Dynamics in San Diego County. mSystems 6, e00045–21. URL: https://doi.org/10.1128/msystems.00045-21, doi:10.1128/mSystems.00045-21.

Kasprzyk-Hordern, B., Sims, N., Farkas, K., Jagadeesan, K., Proctor, K., Wade, M.J., Jones, D.L., 2023. Wastewater-based epidemiology for comprehensive community health diagnostics in a national surveillance study: Mining biochemical markers in wastewater. Journal of Hazardous Materials 450, 130989. URL: https://doi.org/10.1016/j.jhazmat.2023.130989, doi:10.1016/j.jhazmat.2023.130989.

Kaya, D., Falender, R., Radniecki, T., Geniza, M., Cieslak, P., Kelly, C., Lininger, N., Sutton, M., 2022. Correlation between Clinical and Wastewater SARS-CoV-2 Genomic Surveillance, Oregon, USA. Emerging Infectious Diseases 28, 1906–1908. URL: https://doi.org/10.3201/eid2809.220938, doi:10.3201/eid2809.220938.

Kirby, A.E., Welsh, R.M., Marsh, Z.A., Yu, A.T., Vugia, D.J., Boehm, A.B., Wolfe, M.K., White, B.J., Matzinger, S.R., Wheeler, A., Bankers, L., Andresen, K., Salatas, C., New York City Department of Environmental Protection, Gregory, D.A., Johnson, M.C., Trujillo, M., Kannoly, S., Smyth, D.S., Dennehy, J.J., Sapoval, N., Ensor, K., Treangen, T., Stadler, L.B., Hopkins, L., 2022. Notes from the Field: Early Evidence of the SARS-CoV-2 B.1.1.529 (Omicron) Variant in Community Wastewater — United States, November–December 2021. MMWR. Morbidity and Mortality Weekly Report 71, 103–105. URL: http://www.cdc.gov/mmwr/volumes/71/wr/mm7103a5.htm?s_cid=mm7103a5_w, doi:10.15585/mmwr.mm7103a5.

Kisand, V., Laas, P., Palmik-Das, K., Panksep, K., Tammert, H., Albreht, L., Allemann, H., Liepkalns, L., Vooro, K., Ritz, C., Hauryliuk, V., Tenson, T., 2023. Prediction of COVID-19 positive cases, a nation-wide SARS-CoV-2 wastewater-based epidemiology study. Water Research 231, 119617. URL: https://doi.org/10.1016/j.watres.2023.119617, doi:10.1016/j.watres.2023.119617.

Lai, M., Cao, Y., Wulff, S.S., Robinson, T.J., McGuire, A., Bisha, B., 2023. A time series based machine learning strategy for wastewater-based forecasting and nowcasting of COVID-19 dynamics. Science of The Total Environment 897, 165105. URL: https://linkinghub.elsevier.com/retrieve/pii/S0048969723037282, doi:10.1016/j.scitotenv.2023.165105.

Li, X., Kulandaivelu, J., Zhang, S., Shi, J., Sivakumar, M., Mueller, J., Luby, S., Ahmed, W., Coin, L., Jiang, G., 2021.

Data-driven estimation of COVID-19 community prevalence through wastewater-based epidemiology. Science of The Total Environment 789, 147947. URL: https://doi.org/10.1016/j.scitotenv.2021.147947, doi:10.1016/j.scitotenv.2021.147947.

Lucas, J.M., Saccucci, M.S., 1990. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. Technometrics 32, 1–12. URL: http://www.tandfonline.com/doi/abs/10.1080/00401706.1990.10484583, doi:10.1080/00401706.1990.10484583.

Mao, K., Zhang, K., Du, W., Ali, W., Feng, X., Zhang, H., 2020. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. Current Opinion in Environmental Science & Health 17, 1–7. URL: https://doi.org/10.1016/j.coesh.2020.04.006, doi:10.1016/j.coesh.2020.04.006.

McMahan, C.S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J.A., Popat, S.C., Karanfil, T., Freedman, D.L., 2021. COVID-19 wastewater epidemiology: a model to estimate infected populations. The Lancet Planetary Health 5, e874–e881. URL: https://linkinghub.elsevier.com/retrieve/pii/S2542519621002308, doi:10.1016/S2542-5196(21)00230-8.

Montgomery, D.C., 2009. Introduction to statistical quality control. 6th ed ed., Wiley, Hoboken, N.J.

Olesen, S.W., Imakaev, M., Duvallet, C., 2021. Making waves: Defining the lead time of wastewater-based epidemiology for COVID-19. Water Research 202, 117433. URL: https://doi.org/10.1016/j.watres.2021.117433, doi:10.1016/j.watres.2021.117433.

Page, E.S., 1954. Continuous Inspection Schemes. Biometrika 41, 100–115. URL: https://www.jstor.org/stable/2333009, doi:10.2307/2333009. publisher: [Oxford University Press, Biometrika Trust].

Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D.M., Arnold, W., Omer, S.B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nature Biotechnology 38, 1164–1167. URL: https://doi.org/10.1038/s41587-020-0684-z, doi:10.1038/s41587-020-0684-z.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Roberts, S.W., 1959. Control Chart Tests Based on Geometric Moving Averages. Technometrics 1, 239–250. URL: https://www.jstor.org/stable/1266443, doi:10.2307/1266443. publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

Scott, L.C., Aubee, A., Babahaji, L., Vigil, K., Tims, S., Aw, T.G., 2021. Targeted wastewater surveillance of SARS-CoV-2 on a university campus for COVID-19 outbreak detection and mitigation. Environmental Research 200, 111374. URL: https://doi.org/10.1016/j.envres.2021.111374, doi:10.1016/j.envres.2021.111374.

Shewhart, W.A., 1931. Economic Quality Control of Manufactured Product. Bell Syst Tech J 9, 364–389.

Shumway, R.H., Stoffer, D.S., 2017. Time Series Analysis and Its Applications: With R Examples. Springer Texts in Statistics, Springer International Publishing, Cham. URL: http://link.springer.com/10.1007/978-3-319-52452-8, doi:10.1007/978-3-319-52452-8.

Spurbeck, R.R., Minard-Smith, A., Catlin, L., 2021. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. Science of The Total Environment 789, 147829. URL: https://doi.org/10.1016/j.scitotenv.2021.147829, doi:10.1016/j.scitotenv.2021.147829.

Stadler, L., Ensor, K., Clark, J., Kalvapalle, P., LaTurner, Z.W., Mojica, L., Terwilliger, A., Zhuo, Y., Ali, P., Avadhanula, V., Bertolusso, R., Crosby, T., Hernandez, H., Hollstein, M., Weesner, K., Zong, D., Persse, D., Piedra, P., Maresso, A., Hopkins, L., 2020. Wastewater Analysis of SARS-CoV-2 as a Predictive Metric of Positivity Rate for a Major Metropolis. preprint. Epidemiology. URL: http://medrxiv.org/lookup/doi/10.1101/2020.11.04.20226191, doi:10.1101/2020.11.04.20226191.

Supharakonsakun, Y., Areepong, Y., Sukparungsee, S., 2020. The performance of a modified EWMA control chart for monitoring autocorrelated PM2.5 and carbon monoxide air pollution data. PeerJ 8, e10467. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7747693/, doi:10.7717/peerj.10467.

Vallejo, J.A., Trigo-Tasende, N., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, , Barbeito, I., Vaamonde, M., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B.K., Nasser-Ali, M., Cid, , Veiga, M., Acevedo, A., Lamora, C., Bou, G., Cao, R., Poza, M., 2022. Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. Science of The Total Environment 811, 152334. URL: https://doi.org/10.1016/j.scitotenv.2021.152334, doi:10.1016/j.scitotenv.2021.152334.

Wang, Y., Liu, P., VanTassell, J., Hilton, S.P., Guo, L., Sablon, O., Wolfe, M., Freeman, L., Rose, W., Holt, C., Browning, M., Bryan, M., Waller, L., Teunis, P.F., Moe, C.L., 2023. When case reporting becomes untenable: Can sewer networks tell us where COVID-19 transmission occurs? Water Research 229, 119516. URL: https://doi.org/10.1016/j.watres.2022.119516, doi:10.1016/j.watres.2022.119516.

Wolken, M., Sun, T., McCall, C., Schneider, R., Caton, K., Hundley, C., Hopkins, L., Ensor, K., Domakonda, K., Kalvapalle, P., Persse, D., Williams, S., Stadler, L.B., 2023. Wastewater surveillance of SARS-CoV-2 and influenza in preK-12 schools shows school, community, and citywide infections. Water Research 231, 119648. URL: https://doi.org/10.1016/j.watres.2023.119648, doi:10.1016/j.watres.2023.119648.

Yeager, R., Holm, R.H., Saurabh, K., Fuqua, J.L., Talley, D., Bhatnagar, A., Smith, T., 2021. Wastewater Sample Site Selection to Estimate Geographically Resolved Community Prevalence of COVID-19: A Sampling Protocol Perspective. GeoHealth 5, e2021GH000420. URL: https://doi.org/10.1029/2021GH000420, doi:10/gnqvjt. zSCC: 0000001 Publisher: John Wiley & Sons, Ltd.

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R.E., Withington, S., McFarlane, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A.W.U., Norton, J., Xagoraraki, I., 2022. Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. Science of The Total Environment 844, 157040. URL: https://doi.org/10.1016/j.scitotenv.2022.157040, doi:10.1016/j.scitotenv.2022.157040.