

# Optimizing Spatial Distribution of Wastewater-Based Disease Surveillance to Advance Health Equity

Maria L. Daza–Torres<sup>1,+</sup>, J. Cricelio Montesinos-López<sup>1,+</sup>, César Herrera<sup>2</sup>, Yury E. García<sup>1</sup>, Colleen C. Naughton<sup>3</sup>, Heather N. Bichel<sup>4</sup>, and Miriam Nuño<sup>1,\*</sup>

<sup>1</sup>*Department of Public Health Sciences, University of California Davis, California 95616, United States*

<sup>2</sup>*Department of Mathematics, Purdue University, Indiana 47907, United States*

<sup>3</sup>*Department of Civil and Environmental Engineering, University of California Merced, Merced, California, USA*

<sup>4</sup>*Department of Civil and Environmental Engineering, University of California Davis, Davis, California, USA*

<sup>+</sup> *co-first author*

<sup>\*</sup> *corresponding author: [mnuno@ucdavis.edu](mailto:mnuno@ucdavis.edu)*

## Abstract

In 2022, the US Centers for Disease Control and Prevention commissioned the National Academies of Sciences, Engineering, and Medicine to assess the role of community-level wastewater-based disease surveillance (WDS) beyond COVID-19. WDS is recognized as a promising mechanism for promptly identifying infectious diseases, including COVID-19 and other novel pathogens. An important conclusion drawn from this initiative is that it is crucial to maintain equity and expand access to maximize the advantages of WDS for marginalized communities. To address this need, we propose an optimization framework that focuses on the strategic allocation of wastewater monitoring resources at the wastewater treatment plant level. The framework’s purpose is to obtain a balanced spatial distribution, inclusive population coverage, and efficient representation of vulnerable communities in allocating resources for WDS. This study offers an opportunity to improve wastewater surveillance by tailoring location selection strategies to address specific priorities, improving decision-making in public health responses.

## 1 Introduction

Wastewater-based disease surveillance (WDS) is a rapidly evolving field that involves the measurement of health-related biomarkers excreted by individuals into sewer systems. WDS data offers real-time estimation of public health indicators and early warning of community disease outbreaks. Initially used to monitor poliovirus and detect pharmaceuticals and illicit drugs [1, 2], its utility greatly expanded during the Coronavirus Disease 19 (COVID-19) pandemic [3]. Widespread implementation of WDS during the pandemic proved to be a cost-effective method to monitor the dynamics of Sars Coronavirus 2 (SARS-CoV-2) community spread [4].

To enhance the nation’s ability to monitor SARS-CoV-2 effectively, the US Centers for Disease Control and Prevention (CDC) established the National Wastewater Surveillance System (NWSS) in September 2020. The NWSS expanded to encompass 1,154 testing sites nationwide as of March 2024. These sites collect samples from wastewater systems serving approximately 128 million people across the United States [5]. A critical factor behind the success of WDS is the widespread coverage of municipal wastewater collection systems, which link approximately eighty percent of households across the United States [6]. The NWSS is consolidating disparate local initiatives into a resilient and enduring national surveillance network.

WDS data can aid public health mitigation strategies by providing critical insights into the prevalence of COVID-19 within a community [5]. However, additional research is still needed to effectively implement WDS programs [7, 8, 9, 10]. A WDS program must possess sufficient spatial and temporal resolution to offer timely and dependable detection and alerts regarding disease outbreaks. Sampling sites are predominantly selected from wastewater treatment plants (WWTPs) or sewer maintenance holes, contingent on factors such as the size and demographics of the target population, the level of risk in the area, and the laboratory’s resources and capabilities.

Temporal and spatial resolution should undergo regular reassessment to ensure the system can detect significant changes with an adequate lead time to inform public health action effectively. The current distribution of sampling sites within the NWSS primarily comprises localities, tribes, and states that volunteered to participate during the pandemic emergency. This distribution might not accurately reflect the diverse demographic and geographic characteristics essential for establishing a comprehensive national network [11]. Furthermore, it may lack equity, optimal actionability, and long-term sustainability [12].

Several recent studies proposed optimal algorithms for selecting sewer monitoring sites to detect infectious diseases or hotspots at a catchment level [13, 14, 15, 16]. These algorithms aim to maximize the sensitivity of pathogen detection in the sewer network by considering factors like shedding, loss, decay, transport, and the population of the infectious agent. They utilize graph theory and optimization algorithms to identify hotspots and zero-patient locations within a sewer network. These algorithms are often complex and computationally intensive, which may limit their applicability to large-scale or national-level WDS programs.

In this study, we introduce an optimization framework designed to strategically allocate WDS resources at the WWTP level to guide a WDS program. The framework takes into account a range of factors that may influence the effectiveness and efficiency of WDS. These include spatial distribution, population coverage, the presence of vulnerable communities, population density, and variability in wastewater signals. The optimization objective is to

maximize a weighted combination of these factors while minimizing the total number of WWTPs actively monitoring wastewater. We address this challenge as an integer optimization problem and utilize the simulated annealing (SA) algorithm for an efficient solution. We assess various scenarios by adjusting the weightings for each factor to optimize the configurations of WWTPs in the WDS system. The primary objective is to minimize the total number of WWTPs actively involved in WDS.

This approach optimizes resource allocation in WDS by offering tailored location selection strategies that effectively address specific priorities and scenarios. By streamlining WDS efforts and reducing the number of WWTPs collecting samples, we can expand WDS to new communities, especially those at high risk, thereby broadening the scope of our efforts. This not only enhances the efficiency and cost-effectiveness of WDS but also ensures equitable support for informed decision-making across all communities.

## 2 Data description

We built the analytical framework using publicly accessible data in California, USA. We obtained data from the California Department of Public Health (CDPH) for 84 WWTPs that have been monitored for SARS-CoV-2 [17]. We include the following variables in our analysis: 1) *Geographic coordinates*, which include longitude and latitude of each WWTP based on its zip code location, as reported in [17]; this geo-location data is utilized to potentially reduce the selection of WWTPs close to each other, promoting a more geographically dispersed distribution, 2) *Population served*, representing the estimated number of individuals served by each sampling site; the incorporation of this term in our analysis prioritizes WWTPs that serve larger populations [17], 3) *Population density* measures individuals per square mile for each city where a WWTP is located [18]; we use population density data to prioritize areas with high population density, which may be more susceptible to disease transmission, 4) *Social Vulnerability Index*, which uses 16 US census variables in a relative measure of a population's risk during a public health due to external stressors on health such as socioeconomic status and demographic factors [19]. We incorporate the social vulnerability index (SVI) at a county level to ensure the representation of vulnerable populations in the optimization problem. This factor is essential, particularly in regions with insufficient infrastructure and limited resources, which tend to be disproportionately affected during pandemics.

**SARS-CoV-2 RNA concentrations.** We analyzed wastewater concentration data from 19 of the 84 WWTPs that monitored wastewater in California between March 21, 2022, and May 21, 2023. We selected this period because it corresponds with the availability of

published protocols for sampling, analysis, and data processing, as outlined in the Supplementary Material (SM). Wastewater data were obtained from CPDH and WastewaterSCAN [17, 20]. The concentration of SARS-CoV-2 RNA measured in each wastewater sample was normalized by the concentration of pepper mild mottle virus (PMMoV) RNA measured in the same sample. PMMoV is a fecal indicator and laboratory process control that is commonly measured alongside SARS-CoV-2 to help mitigate noise arising from analytical processes, population size, and wastewater flow variations. Normalization results in the dimensionless metric  $N/PMMoV$ , which we refer to as the normalized wastewater concentration.

Based on this normalized wastewater concentration data, we construct a dissimilarity matrix that measures the difference between the wastewater signals of different WWTPs based on their patterns, such as trends, peaks, and variations.

Figure 1A illustrates the spatial distribution of 84 WWTPs monitoring for SARS-CoV-2, along with the population they serve and SVI at the county level. The subset of 19 WWTPs utilized in the optimization problem involving the factor of signal dissimilarity is illustrated in Figure 4A.

### 3 Methods

This study aims to inform WDS planning in California by considering spatial distributions of WWTPs, population coverage, representation of vulnerable populations, and dissimilarity between wastewater signals. We normalize the factors to a common scale to facilitate comparisons, and we assign weights to the factors to demonstrate a range of strategies for balancing their relative importance. We then maximize the weighted combinations of factors while satisfying pre-defined constraints to minimize the number of WWTPs. This approach acknowledges practical considerations and constraints associated with wastewater monitoring capabilities.

#### 3.1 Optimization problem formulation

Let  $I = \{1, \dots, n\}$  represent the set of WWTPs to be optimized. The objective function, denoted as  $f(\mathbf{x})$ , seeks to maximize a weighted combination of various factors:

$$f(\mathbf{x}) = w_{PS} \frac{\sum_{i \in I} PS_i x_i}{S_{PS}} + w_{PD} \frac{\sum_{i \in I} PD_i \cdot x_i}{S_{PD}} + w_{dist} \frac{\sum_{i \in I} \sum_{j \in I} dist(i, j) x_i x_j}{S_{dist}} + w_{svi} \frac{\sum_{i \in I} SVI(i) x_i}{S_{SVI}} + w_{diss} \frac{\sum_{i \in I} \sum_{j \in I} diss(i, j) x_i x_j}{S_{diss}} \quad (1)$$

where  $\mathbf{x} = (x_1, \dots, x_n)$  and  $x_i$  is a binary variable indicating if plant  $i$  is selected ( $x_i = 1$ ) or not ( $x_i = 0$ ). The terms  $PS_i$ ,  $PD_i$ ,  $dist(i, j)$ ,  $SVI(i)$ , and  $diss(i, j)$  correspond to the population served by plant  $i$ , city's population density associated with plant  $i$ , the distance between plant  $i$  and plant  $j$ , SVI of the county associated with plant  $i$ , and dissimilarity between WWTP signals, respectively.

The variables  $S_{PS} = \sum_{i \in I} PS_i$ ,  $S_{PD} = \sum_{i \in I} PD_i$ ,  $S_{dist} = \sum_{i \in I} \sum_{j \in I} dist(i, j)$ ,  $S_{SVI} = \sum_{i \in I} SVI(i)$ , and  $S_{diss} = \sum_{i \in I} \sum_{j \in I} diss(i, j)$  represent the total values of the corresponding terms over all plants. These sums facilitate normalization for comparison and the introduction of weights. The weights  $w_{PS}$ ,  $w_{PD}$ ,  $w_{dist}$ ,  $w_{SVI}$ , and  $w_{diss}$  assigned to each term in the objective function determine the importance given to the respective factors during the optimization process. To ensure balance, the sum of weights is constrained to be equal to 1:  $w_{PS} + w_{PD} + w_{dist} + w_{SVI} + w_{diss} = 1$ .

The optimization problem is subject to the constraint that the total number of selected plants should be equal to a predefined number  $k$  (i.e.,  $\sum_{i \in I} x_i = k$ ). This constraint plays a critical role in maintaining the efficiency and effectiveness of the WDS system, especially when dealing with resource limitations. The goal of the optimization problem is to select a subset of  $k$  WWTPs that effectively balance multiple factors mentioned above.

## 3.2 Computing a dissimilarity matrix using wavelets

Epidemiological studies often involve analyzing time series data to extract valuable insights and patterns across different scales. However, epidemiological processes are typically not stationary, implying that their statistical properties vary over time. Hence, a global timescale decomposition, where the entire time series is treated uniformly, may not be suitable for studying epidemiological processes [21]. Wavelet analysis is a suitable method for investigating non-stationary time series data, as it performs a local timescale decomposition of the signal [22]. Additionally, it allows us to measure the associations between two or more time series at any frequency band and time window, allowing us to detect synchrony, phase, and coherence patterns [21, 23]. It has been applied to various human infectious diseases, such as measles, influenza, leishmaniasis, and dengue [23, 24, 25, 26].

### 3.2.1 Dissimilarity matrix

We apply the discrete wavelet transform (DWT) to the SARS-CoV-2 concentration time series data using the Daubechies 4 wavelet family. Decomposing the data into four levels ( $L = 4$ ), this process yields coefficients at different scales and positions. At each level ( $l$ ), we obtain approximation coefficients ( $cA_l$ ) and detail coefficients ( $cD_l$ ), both of length  $N/(2^l)$ , where  $N$

is the length of the wastewater time series.  $cA_l$  contains information about the overall signal trend, while  $cD_l$  contains high-frequency information or details,  $l = 1, \dots, L$ . These wavelet coefficients capture the characteristics of the signal at different scales and positions in both the time and frequency domains. We use these coefficients, excluding detail coefficients at the first two levels, to compute a dissimilarity matrix that measures how different the time series are based on their wavelet coefficients; see details in SM.

We used the correlation distance as the dissimilarity measure. The correlation distance between  $u$  and  $v$  is defined as  $1 - \frac{(u-\bar{u}) \cdot (v-\bar{v})}{\|(u-\bar{u})\|_2 \|(v-\bar{v})\|_2}$ , where  $\bar{u}$  is the mean of the elements of  $u$  and  $x \cdot y$  is the dot product of  $x$  and  $y$ . A low correlation distance (high positive correlation) indicates a low dissimilarity, and vice versa. Other distance measures, such as the Euclidean distance, may not be suitable for time series with different scales or magnitudes since it only considers the absolute difference between two vectors. Time series with similar patterns but distinct scales could yield high dissimilarity with Euclidean distance, while correlation distance would give a low dissimilarity. Therefore, correlation distance may be better for comparing time series with different scales or magnitudes.

We use the PyWavelets library in Python to perform the DWT, and subsequently compute the correlation distance matrix using the scikit-learn library. This matrix is denoted as  $diss(\cdot, \cdot)$  in our optimization function. Refer to Figure 3 for a visual representation of the computed correlation distance matrix.

The primary concept behind using wavelet coefficients to compare signals is to emphasize the patterns and relative changes in the data across different scales. Wavelet coefficients are relative measures that describe the variations in the data at different scales, regardless of the absolute magnitude of the data. These coefficients are normalized in the sense that they quantify the relative importance of different frequency components in the data. When comparing wavelet coefficients from two-time series, we are essentially comparing how their frequency content and patterns vary across scales, which makes it possible to identify similarities or differences in their underlying structures, even if the original magnitudes of the series are different.

### 3.2.2 Simulated annealing

Optimal location problems are solved through exact methods or heuristic approaches, depending on the problem's size and complexity [27]. Exact methods find the global optimum but are only feasible for small problems. Heuristic approaches aim to find good-quality solutions within a limited amount of computation time or space [28]. We use the SA algorithm, a heuristic approach that has proven effective in solving complex combinatorial optimization problems [29].

SA is inspired by the annealing process in metallurgy, where a metal is heated quickly to a high “temperature” (a parameter) and then gradually cooled to a more ordered state. By gradually cooling down, SA facilitates effective exploration of the solution space and finds an optimal solution that meets our objectives. The “temperature” parameter in SA controls the balance between exploration (searching for new solutions) and exploitation (improving existing solutions). This algorithm reduces the risk of falling into local minima common to iterative improvement methods, as unlike the latter, it accepts solutions that worsen the objective function [28].

Let  $I = \{1, \dots, n\}$  denote the set of WWTPs to be optimized. Algorithm S1 in SM delineates the steps for maximizing the objective function  $f(\mathbf{x})$ , given in Equation(1), using SA, while ensuring that the total number of selected plants equals  $k < n$ . To initialize the candidate solution  $\mathbf{x}_0$ , we randomly select  $k$  distinct numbers from  $I$ , forming a set of indices denoted as  $I_k$ . The entries of  $\mathbf{x}_0$  are set to 1 if their corresponding indices are in  $I_k$ , and 0 otherwise. For generating a new candidate solution  $\mathbf{x}_{\text{new}}$ , we replace a randomly chosen number from  $I_k$  with a new random number from  $I - I_k$  (i.e., the set of elements in  $I$  but not in  $I_k$ ). Following the guidelines outlined in [30], we adapt the “temperature” based on whether we have accepted  $10 \times n$  solutions or generated  $100 \times n$  candidate solutions. In the reported results, we fix  $T_{\text{max}} = 10^5$  (maximum temperature),  $T_{\text{min}} = 10^{-7}$  (minimum temperature), and  $\alpha = 0.1$  (cooling schedule parameter). Then, we compute the energy in both the current and the new candidate solutions:  $E = f(\mathbf{x}_0)$  and  $E_{\text{new}} = f(\mathbf{x}_{\text{new}})$ , respectively. Then, we calculate the change in energy,  $\Delta E = E - E_{\text{new}}$ . If  $\Delta E < 0$ , indicating an improvement, we accept the new solution  $\mathbf{x}_{\text{new}}$ . Otherwise, we generate a random number  $r$  from a uniform distribution  $U(0, 1)$ . If  $r < e^{-\Delta E/T}$ , where  $T$  is the current temperature, we accept  $\mathbf{x}_{\text{new}}$ ; otherwise, we reject it.

## 4 Results

We developed a flexible framework that can be used to evaluate resource planning scenarios, tailored to address specific challenges and priorities within a WDS system. We consider two overarching scenarios: 1) a disease-agnostic approach that considers only geographic and population attributes in resource allocation, and 2) a disease-informed approach that additionally considers dissimilarities in the dynamics of wastewater-derived COVID-19 data. The first scenario facilitates the prioritization of vulnerable populations using the SVI, while the second further facilitates the prioritization of unique wastewater signal contributions to the WDS dataset. In each scenario, we reduced the number of WWTPs from baseline by 40%. We explore three configurations within each scenario that align with potential system

priorities, achieved by assigning varying weights to the parameters of the objective function. Sustainable, efficient, and equitable allocation of WDS resources is essential for the continued effectiveness of surveillance efforts in the long term. Our framework offers a strategy and tool for public health officials and policymakers involved in wastewater-based disease surveillance and resource allocation.

### Scenario 1: Disease-agnostic optimization approach

In Scenario 1, we integrated data on the location and population served by 84 WWTPs in California as the baseline condition, with city-level population density and the county-level SVI associated with each facility. We assigned a weight of zero to the normalized wastewater concentration dissimilarity term in the optimization function, representing scenarios for which disease dynamics are not considered in site selection. Reducing the number of plants by 40% resulted in a total of 50 remaining WWTPs in each solution. We explored three configurations of weight combinations to gain an understanding of how weights influence the solution:

- **Scenario 1A: Prioritizing SVI** ( $w_{SVI} = 1, w_{PS} = w_{PD} = w_{dist} = w_{diss} = 0$ ). This scenario prioritizes plants with high SVI, aiming to improve representation in areas with increased social vulnerabilities. The optimization solution (see Figure 1B) prioritizes the selection of plants situated in areas with higher SVI scores, which tend to correspond to communities that experienced disproportionate impacts from COVID-19.
- **Scenario 1B: Prioritizing population served** ( $w_{PS} = 1, w_{PD} = w_{dist} = w_{SVI} = w_{diss} = 0$ ). Prioritizing population served directs attention to ensuring that surveillance efforts are concentrated in areas where a larger number of people reside ( $w_{PS} = 1$ ) (see Figure 1C).
- **Scenario 1C: A balanced approach** ( $w_{PS} = w_{PD} = w_{dist} = w_{SVI} = w_{diss} = 0.25, w_{diss} = 0$ ). This scenario seeks an optimal solution that balances multiple factors equally (see Figure 1D).

Findings in Figure 1 underscore key insights essential for determining the allocation of disease surveillance resources. In the scenario prioritizing SVI, numerous plants in California’s Central Coast region are omitted (Figure 1B). However, plants in the Central Valley and South Interior regions are retained, which aligns with the high SVI scores observed in those areas. This highlights a potential trade-off: while prioritizing SVI helps address vulnerable communities, it might lead to gaps in geographic coverage. In the scenario prioritizing population served (Figure 1C), two plants in the North region, one in the South Coast, and several in the Central region of California are eliminated. Communities represented by some of these plants correspond to regions in California that experienced a significant burden during the COVID-19 pandemic. The removal of these plants could have important implications for public health and environmental equity, particularly in communities that already have

limited access to healthcare and socioeconomic disparities. The allocation of WWTPs guided by a balanced set of priorities, as described in Figure 1D, seems to provide reasonable results despite the significant downsizing of plants.

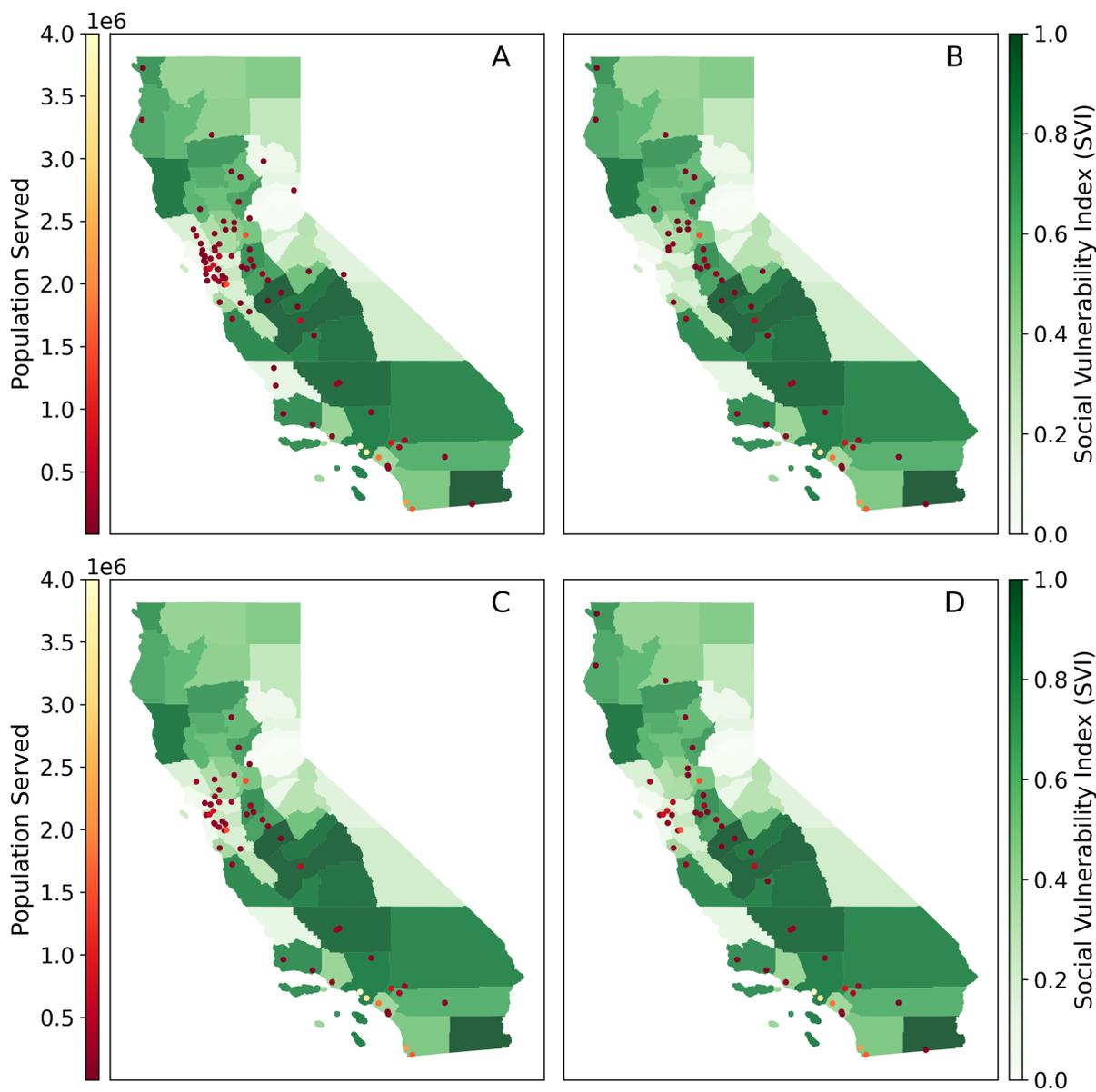


Figure 1: **A.** Geographic distribution of WWTPs across California monitoring SARS-CoV-2, with emphasis on the population served and SVI at the county level. Optimal solutions assuming a 40% reduction of active plants and different scenarios: **B.** Prioritizing SVI, **C.** Prioritizing Population Served, and **D.** Balanced approach.

Figure 2 describes the distribution of SVI, population served, and density among all 84 WWTPs included in this scenario. As expected, scenarios prioritizing specific factors such

as SVI and population served display higher values and less variability in these variables compared to all plants. Interestingly, a balanced approach consistently results in the selection of WWTPs with similar distributions for the population served and density while also representing communities with higher SVI scores. Notably, the two plants that ceased data collection consistently appear as optimal solutions in multiple scenarios. One plant, located in the Central Valley, emerged as an optimal solution across all scenarios, while the other, located in a Southern County of CA, was optimal in the SVI and balanced priority scenarios.

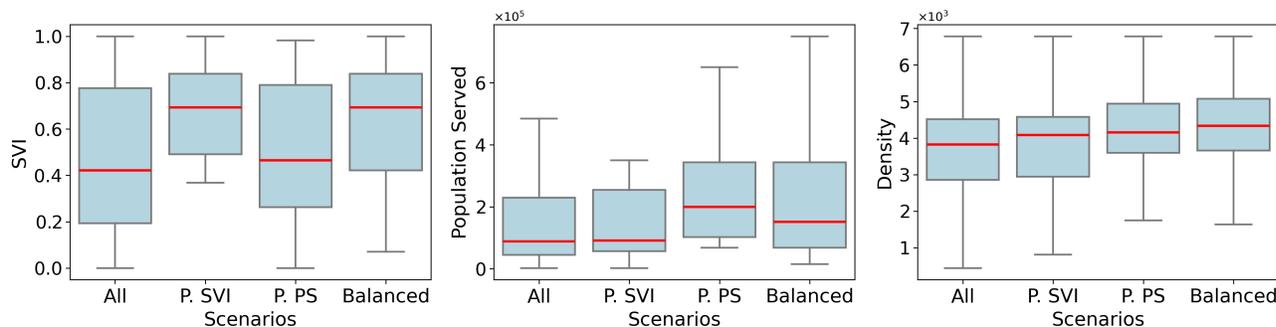


Figure 2: Distribution of the SVI, population served, and population density for all WWTPs that have WDS for SARS-CoV-2 in California (All), and for the optimal WWTPs found in each scenario: prioritizing SVI (P.SVI), prioritizing population served (P.PS), and a balanced approach (Balanced). Outliers were excluded from the boxplot for improved visualization; see Figure S.5 in SM for a plot including outliers.

## Scenario 2: Disease-informed optimization approach

Scenario 2 represents a planning scenario in which the dynamics of the community disease trends are considered alongside geographic and population factors in the selection of sampling sites. We include a dissimilarity term that takes into account the differences between normalized wastewater concentrations, giving preference to those with distinct signal patterns. We consider 19 WWTPs as a baseline in the scenario with a wastewater signal dissimilarity term in the optimization problem in Equation (1). The dissimilarity matrix, which represents the correlation distance between discrete wavelet transforms of normalized SARS-CoV-2 wastewater concentrations, is illustrated in Figure 3. By incorporating dissimilarity term in the objective function, our goal is to present the results of the optimization problem and demonstrate the impact of this addition.

- **Scenario 2A: Prioritizing disease signal dissimilarity** ( $w_{diss} = 1, w_{PS} = w_{PD} = w_{dist} = w_{SVI} = 0$ ). This scenario emphasizes dissimilarity between WDS signals, aiming to optimize the representation of unique signal patterns across WWTPs (see Figure 4A).
- **Scenario 2B: Equal factors** ( $w_{PS} = w_{PD} = w_{dist} = w_{SVI} = w_{diss} = 0.2$ ). This scenario

seeks an optimal solution that weights multiple factors equally (see Figure 4B).

– **Scenario 2C: A balanced approach for SVI and disease signal dissimilarity** ( $w_{SVI} = w_{diss} = 0.35, w_{PS} = w_{PD} = w_{dist} = 0.1$ ). In this scenario, we emphasize both the SVI and dissimilarity metrics, while also taking into account population density, spatial distances, and population served (see Figure 4C).

Prioritizing dissimilarities in our optimization process can improve our capacity for early detection, response, and tailored interventions. Figure 4 illustrates the WWTPs selected in the optimization problem prioritizing WW signal dissimilarity (Figure 4B), initially with 19 plants (Figure 4A). Subsequently, it showcases the plants that remained after adopting an equal factors approach (Figure 4C) and a balanced approach for SVI and disease signal dissimilarity (Figure 4D).

Figure S.4 in SM displays the dissimilarity matrix for each optimal solution and scenario. In this particular scenario, prioritizing the dissimilarity term means that we are placing emphasis on highlighting the distinctions between the selected WWTPs with regard to their wastewater signal characteristics.

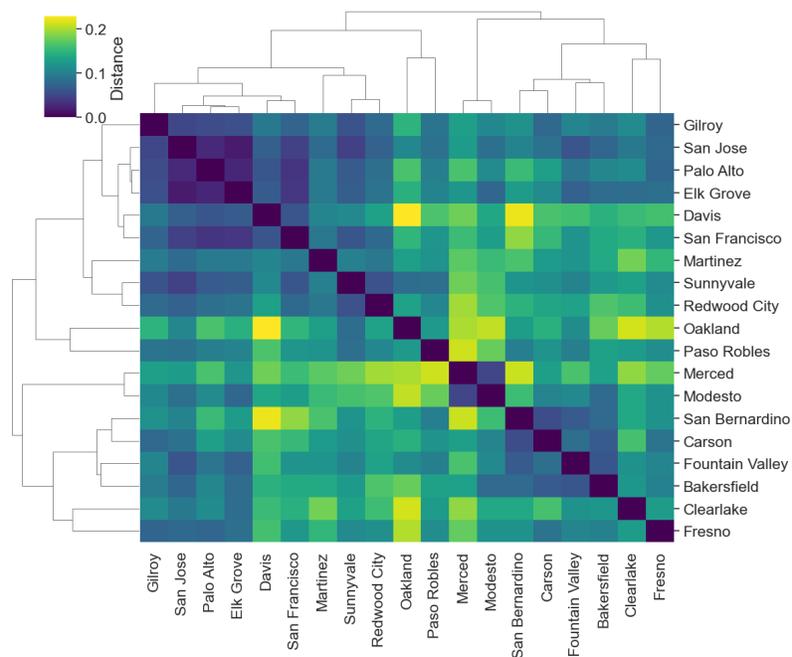


Figure 3: Dissimilarity matrix depicting wavelet coefficients for 19 WWTPs. Alongside the matrix, dendrograms illustrate hierarchical clustering, showcasing how WWTPs group together based on similarities in wavelet coefficient.

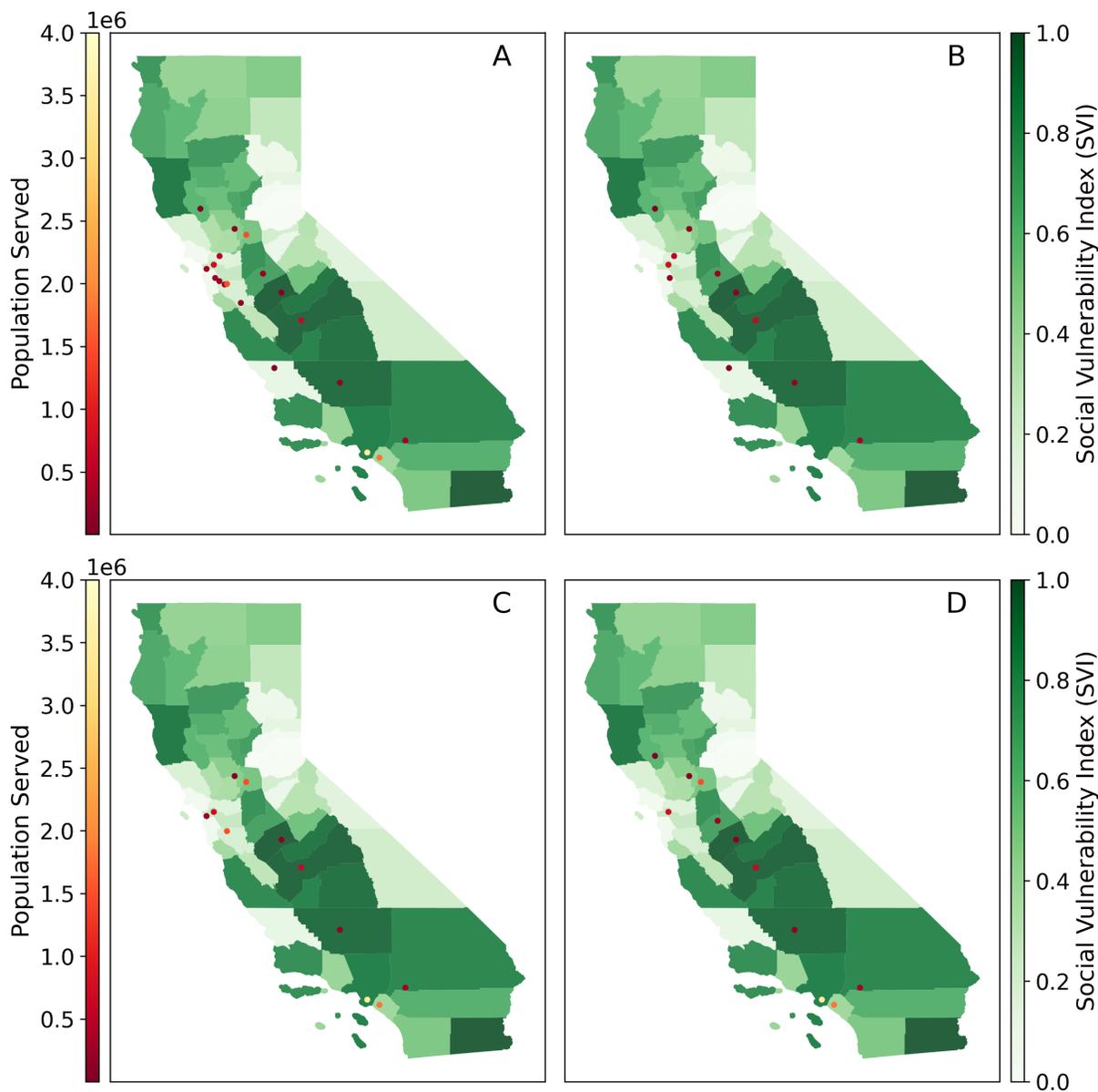


Figure 4: Optimal solutions obtained assuming signal dissimilarity and 19 initial WWTPs. Each priority scenario assumes a goal of a 40% reduction in the number of plants included in the optimization problem. **A.** Initial Plants. **B.** Prioritizing disease signal dissimilarity. **C.** Equal factors. **D.** A balanced approach for SVI and disease signal dissimilarity.

## 5 Discussion

In this paper, we present an optimization framework to evaluate scenarios for the selection of WWTPs for wastewater-based disease surveillance. Our analysis in California revealed that prioritizing social vulnerability in defining the optimization problem enables more equitable

integration of communities with a history of significant COVID-19 burden when planning the allocation of resources for WDS. These findings can inform public health surveillance efforts, particularly as WDS is adapted to help manage infectious diseases beyond COVID-19, including for influenza and respiratory syncytial virus. By focusing on areas with higher vulnerability, which are often characterized by limited resources and greater health disparities, resources for surveillance efforts can be allocated more effectively. This strategy helps ensure that communities with elevated risks of infection and severe symptoms receive greater attention and support in disease monitoring and response.

Our findings make a significant contribution to the current literature by bringing to light the complexities of optimizing WDS resource allocation. We demonstrate how prioritizing key factors in the optimization process could influence planning decisions for resource distributions. Our results also highlight how different communities offer unique contributions to a WDS program. For instance, we observed that two WWTPs within the sampling frame ceased WDS data collection during the study period, yet these two facilities were consistently retained in most of our optimization scenarios. Our framework offers a mechanism to evaluate the significance of WDS program changes in the context of seeking equitable representation and cost efficiencies. We recommend ongoing evaluation of site distributions to maintain fair and inclusive coverage. Greater investments in community engagement and partnership-building programs can support marginalized communities and guarantee they have a say in decision-making processes related to WDS.

Actively addressing equity in WDS planning and resource allocation is crucial to strengthening pandemic preparedness. One solution is to adopt a more comprehensive strategy that gives priority to vulnerable communities in the distribution of monitoring resources. This could entail re-evaluating the standards used to choose monitoring sites to ensure sufficient coverage of regions with elevated social vulnerability. Our approach is optimized from the 84 WWTPs that have had WDS due to data availability. The scenarios examined in this study serve as valuable entry points, offering initial insights into potential WWTP configurations, but there are over 200 WWTPs in California that could be further evaluated in the future. The framework can also be extended to other states and countries to guide resource allocation for WDS. It can be customized and adjusted to ensure that surveillance efforts are fair, efficient, and in line with the changing requirements of diverse communities, ultimately enhancing our collective capacity to detect and mitigate public health risks on a global level.

While this framework provides valuable insights and serves as a basis for decision-making, we recognize several limitations and needs for further improvement. First, achieving more practical and context-specific outcomes will require accounting for decision-maker preferences and constraints in future refinements. For instance, budget limitations, workforce availabil-

ity, and infrastructure capacity can be incorporated to inform proposed solutions. Second, reliability and accuracy of the optimization model are reliant on the quality and availability of data pertaining to population distribution, social vulnerability, historical disease burden, and other related variables. Enhancing data transparency and accessibility can empower communities to participate actively in WDS efforts and advocate for their requirements.

Finally, integrating spatial and temporal dynamics into upcoming optimization models could provide significant improvements to the modeling framework. The distribution of population and disease burden may vary spatially and temporally, requiring dynamic modeling approaches to account for these fluctuations accurately. Furthermore, ethical considerations and other equity concerns should also be factored into the optimization process to guarantee a fair and unbiased allocation of disease surveillance resources [31, 32]. To this end, it is essential to establish clear guidelines and criteria that prioritize the needs of underserved and marginalized communities. This may involve conducting thorough assessments of socio-economic factors, healthcare access, and vulnerable indices to identify areas with the greatest need for surveillance resources. Moreover, stakeholder engagement and community involvement are crucial components of ensuring equity in resource allocation. By actively seeking input from affected communities, local leaders, and advocacy groups, decision-makers can gain valuable insights into the specific challenges and priorities of different population groups.

## Acknowledgments

This research was conducted as part of Healthy Central Valley Together (HCVT), a collaborative research project to enable public health action from wastewater-based disease surveillance. Research support for HCVT was provided by a philanthropic gift to the University of California, Davis. We extend our gratitude to the California Department of Public Health (CDPH) and WastewaterSCAN for their collaboration and for generously sharing wastewater data. We would like to express our gratitude to Alexandria B. Boehm, Marlene K. Wolfe, and Jesus De Loera for their valuable comments. We extend our sincere gratitude to the dedicated teams at the wastewater treatment plants, as well as the invaluable support from public health entities and all other collaborators involved in the collection and curation of the data used in this study.

## Declaration of competing interest

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

## References

- [1] John R Paul, James D Trask, and Sven Gard. II. Poliomyelitic virus in urban sewage. *The Journal of Experimental Medicine*, 71(6):765–77, 1940.
- [2] James D Trask, AJ Vignec, and John R Paul. Poliomyelitis virus in human stools. *Journal of the American Medical Association*, 111(1):6–11, 1938.
- [3] Colleen C Naughton, Jr Roman, Fernando A, Ana Grace F Alvarado, Arianna Q Tariqi, Matthew A Deeming, Krystin F Kadonsky, Kyle Bibby, Aaron Bivins, Gertjan Medema, Warish Ahmed, Panagis Katsivelis, Vajra Allan, Ryan Sinclair, and Joan B Rose. Show us the data: global COVID-19 wastewater monitoring efforts, equity, and gaps. *FEMS Microbes*, 4:xtad003, 01 2023.
- [4] Olga E. Hart and Rolf U. Halden. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of The Total Environment*, 730:138875, 2020.
- [5] CDC. National wastewater surveillance system (nwss). CDC National Wastewater Surveillance System, 2020.
- [6] CDC. Wastewater surveillance: Progress in the united states. <https://www.cdc.gov/nwss/progress.html>, March 2023. Centers for Disease Control and Prevention (CDC).
- [7] Teresa Kumblathan, Yanming Liu, Gursharan K. Uppal, Steve E. Hrudey, and Xing-Fang Li. Wastewater-based epidemiology for community monitoring of sars-cov-2: Progress and challenges. *ACS Environmental Au*, 1(1):18–31, 2021.
- [8] Damià Barceló. Wastewater-Based Epidemiology to monitor COVID-19 outbreak: Present and future diagnostic methods to be in your radar. *Case Studies in Chemical and Environmental Engineering*, 2:100042, 2020.
- [9] Guangming Jiang, Yanchen Liu, Song Tang, Masaaki Kitajima, Eiji Haramoto, Sudipti Arora, Phil M. Choi, Greg Jackson, Patrick M. D’Aoust, Robert Delatolla, Shuxin Zhang, Ying Guo, Jiangping Wu, Yan Chen, Elipsha Sharma, Tanjila Alam Prosun,

- Jiawei Zhao, Manish Kumar, Ryo Honda, Warish Ahmed, and Jon Meiman. Moving forward with covid-19: Future research prospects of wastewater-based epidemiology methodologies and applications. *Current Opinion in Environmental Science & Health*, 33:100458, 2023.
- [10] Jill S McClary-Gutierrez, Mia C Mattioli, Perrine Marcenac, Andrea I Silverman, Alexandria B Boehm, Kyle Bibby, Michael Balliet, Daniel Gerrity, John F Griffith, Patricia A Holden, et al. SARS-CoV-2 wastewater surveillance for public health action. *Emerging infectious diseases*, 27(9), 2021.
- [11] Jessica R. Mosher, Jim E. Banta, Rhonda Spencer-Hwang, Colleen C. Naughton, Krystin F. Kadonsky, Thomas Hile, and Ryan G. Sinclair. An environmental equity assessment using a social vulnerability index during the sars-cov-2 pandemic for siting of wastewater-based epidemiology locations in the united states. *Geographies*, 4(1):141–151, 2024.
- [12] National Academies of Sciences, Engineering, and Medicine. *Wastewater-based Disease Surveillance for Public Health Action*. The National Academies Press, Washington, DC, 2023.
- [13] Keugtae Kim, Min Jeong Ban, Sungpyo Kim, Mi-Hyun Park, Michael K Stenstrom, and Joo-Hyon Kang. Optimal allocation and operation of sewer monitoring sites for wastewater-based disease surveillance: A methodological proposal. *Journal of Environmental Management*, 320:115806, 2022.
- [14] Yuke Wang, Christine L Moe, Shanta Dutta, Ashutosh Wadhwa, Suman Kanungo, Wolfgang Mairinger, Yichuan Zhao, Yi Jiang, and Peter FM Teunis. Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation. *Epidemics*, 31:100391, 2020.
- [15] Richard C Larson, Oded Berman, and Mehdi Nourinejad. Sampling manholes to home in on SARS-CoV-2 infections. *PloS ONE*, 15(10):e0240007, 2020.
- [16] Eusebi Calle, David Martínez, Roser Brugués-i Pujolràs, Miquel Farreras, Joan Saló-Grau, Josep Pueyo-Ros, and Lluís Corominas. Optimal selection of monitoring sites in cities for sars-cov-2 surveillance in sewage networks. *Environment International*, 157:106768, 2021.
- [17] GovOps. Covid-19 wastewater surveillance data. california. California Open Data, 2024.
- [18] US Cities Database. SimpleMaps. Accessed on 10/15/2023.

- [19] CDC/ATSDR. CDC/ATSDR SVI Data and Documentation Download. CDC/ATSDR SVI Data and Documentation Download, 2022. Accessed on 10/15/2023.
- [20] Alexandria B Boehm, Marlene K Wolfe, Krista R Wigginton, Amanda Bidwell, Bradley J White, Bridgette Hughes, Dorothea Duong, Vikram Chan-Herur, Heather N Bischel, and Colleen C Naughton. Human viral nucleic acids concentrations in wastewater solids from central and coastal california usa. *Scientific Data*, 10(1):396, 2023.
- [21] Bernard Cazelles, Mario Chavez, Guillaume Constantin de Magny, Jean-Francois Guégan, and Simon Hales. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*, 4(15):625–636, 2007.
- [22] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.
- [23] Hélele Broutin, Jean-François Guégan, Eric Elguero, François Simondon, and Bernard Cazelles. Large-Scale Comparative Analysis of Pertussis Population Dynamics: Periodicity, Synchrony, and Impact of Vaccination. *American Journal of Epidemiology*, 161(12):1159–1167, 2005.
- [24] Khoa TD Thai, Bernard Cazelles, Nam Van Nguyen, Long Thi Vo, Maciej F Boni, Jeremy Farrar, Cameron P Simmons, H Rogier van Doorn, and Peter J de Vries. Dengue dynamics in binh thuan province, southern vietnam: periodicity, synchronicity and climate variability. *PLoS neglected tropical diseases*, 4(7):1–8, 2010.
- [25] Bryan T Grenfell, Ottar N Bjørnstad, and Jens Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, 2001.
- [26] Yury E Garcia, Shu-Wei Chou-Chen, Luis A Barboza, Maria L Daza-Torres, J Cricelio Montesinos-Lopez, Paola Vasquez, Juan G Calvo, Miriam Nuno, and Fabio Sanchez. Common patterns between dengue cases, climate, and local environmental variables in costa rica: A wavelet approach. *PLOS Global Public Health*, 3(10):e0002417, 2023.
- [27] Richard L Church and Alan T Murray. *Business site selection, location analysis, and GIS*. John Wiley & Sons Hoboken, NJ, 2009.
- [28] Ana MT Amorim, Alexandre B Gonçalves, Luís Miguel Nunes, and António Jorge Sousa. Optimizing the location of weather monitoring stations using estimation uncertainty. *International Journal of Climatology*, 32(6):941–952, 2012.

- [29] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
- [30] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007.
- [31] Sarah Hawkes and Kent Buse. Socially Constructed Determinants of Health: The Case for Synergies to Arrive at Gendered Global Health Law. *Public Health Ethics*, 13(1):16–28, 07 2020.
- [32] Amy L Fairchild, Ali Akbar Haghdoost, Ronald Bayer, Michael J Selgelid, Angus Dawson, Abha Saxena, and Andreas Reis. Ethics of public health surveillance: new guidelines. *The Lancet Public Health*, 2(8):e348–e349, 2017.
- [33] Krystin F Kadonsky, Colleen C Naughton, Mirjana Susa, Rachel Olson, Adam Gushgari, John Cosgrove, Alexandria B Boehm, and Heather N Bischel. Expansion of wastewater-based disease surveillance to improve health equity in california’s central valley: sequential shifts in case-to-wastewater and hospitalization-to-wastewater ratios. *Frontiers in Public Health*, 11:1141097, 2023.
- [34] Maria L Daza-Torres, J Cricelio Montesinos-López, Heather N Bischel, Colleen C Naughton, Angel N Desai, Marlene K Wolfe, Alexandria B Boehm, and Miriam Nuño. A mixed-effects model to predict covid-19 hospitalizations using wastewater surveillance. *medRxiv*, pages 2023–08, 2023.
- [35] Revvity. *chemagic Viral DNA/RNA 300 Kit H96*, 2024.
- [36] E.M. Symonds, Karena H. Nguyen, V.J. Harwood, and M. Breitbart. Pepper mild mottle virus: A plant pathogen with a greater purpose in (waste)water treatment development and public health management. *Water Research*, 144:1–12, 2018.
- [37] Aaron Topol, Marlene Wolfe, Brad White, Krista Wigginton, and Alexandria B Boehm. High throughput sars-cov-2, pmmov, and bcov quantification in settled solids using digital rt-pcr. *Protocols.io*, 2021.
- [38] Marlene K. Wolfe, Aaron Topol, Alisha Knudson, Adrian Simpson, Bradley White, Duc J. Vugia, Alexander T. Yu, Linlin Li, Michael Balliet, Pamela Stoddard, George S. Han, Krista R. Wigginton, and Alexandria B. Boehm. High-frequency, high-throughput quantification of sars-cov-2 rna in wastewater settled solids at eight publicly owned

treatment works in northern california shows strong association with covid-19 incidence. *mSystems*, 6(5):10.1128/msystems.00829–21, 2021.

- [39] Aaron Topol, Marlene Wolfe, Brad White, Krista Wigginton, and Alexandria B. Boehm. High throughput sars-cov-2, pmmov, and bcov quantification in settled solids using digital rt-pcr. *protocols.io*, 5, 2022.
- [40] MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit. Product Information. Accessed on 10/15/2023.
- [41] Bio-Rad Corporation. *Droplet Digital PCR Applications Guide*, 2024.
- [42] Shuchen Feng, Adelaide Roguet, Jill S McClary-Gutierrez, Ryan J Newton, Nathan Kloczko, Jonathan G Meiman, and Sandra L McLellan. Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. *Acs Es&T Water*, 1(8):1955–1965, 2021.
- [43] Aaron Topol, Marlene Wolfe, Brad White, Krista Wigginton, and Alexandria B Boehm. High Throughput SARS-COV-2, PMMOV, and BCoV quantification in settled solids using digital RT-PCR V.2. *protocols.io*, 2021.

# Supplementary Material: Optimizing Spatial Distribution of Wastewater-based Disease Surveillance to Advance Health Equity

J. Cricelio Montesinos-López<sup>1,+</sup>, Maria L. Daza-Torres<sup>1,+</sup>, César Herrera<sup>2</sup>, Yury E. García<sup>1</sup>, Colleen C. Naughton<sup>3</sup>, Heather N. Bischel<sup>4</sup>, Miriam Nuño<sup>1</sup>

**1** Department of Public Health Sciences, University of California Davis, CA, USA

**2** Department of Mathematics, Purdue University, Indiana 47907, United States

**3** Department of Civil and Environmental Engineering, University of California Merced, CA, USA

**4** Department of Civil and Environmental Engineering, University of California Davis, CA, USA

+ co-first author

## List of supplementary materials

- 1 S1 Wavelet decomposition.
- 2 S2 Comparative analysis of wastewater data using discrete wavelet transform.
- 3 S3 Algorithm performance.
- 4 S4 Supplementary figures.
- 5 S5 Simulated Annealing algorithm.
- 6 S6 Wastewater methods description.
- 7

## 8 **S1 Comparative analysis of wastewater data using discrete wavelet transform**

### 10 **S1.1 Wavelet decomposition**

11 The wavelet transform decomposes non-stationary time series into a time-frequency domain,  
12 allowing interpretation of temporal variability. It analyzes signals at multiple resolutions  
13 and scales, using basis functions called wavelets. Wavelets are mathematical functions used  
14 to decompose data into different frequencies and scales. The continuous wavelet transform  
15 (CWT) is given by the following equation:

$$F(x) = \frac{1}{\sqrt{\xi}} \int x(t) \cdot \psi^* \left( \frac{t - \tau}{\xi} \right) \quad (2)$$

16 where  $F(x)$  is the wavelet transform for the signal  $x(t)$  as a function of time,  $\xi$  is the scale  
17 parameter,  $\tau$  is the time parameter, and  $\psi$  is the mother wavelet or the basis function with  
18  $*$  denoting the complex conjugate.

19 The CWT is computationally expensive, but the discrete wavelet transform (DWT) offers  
20 efficiency. DWT uses sub-band coding, which is a technique of dividing the signal into  
21 frequency bands or sub-bands, and then sampling each sub-band at a lower rate. The DWT  
22 achieves this by passing the signal through a series of high-pass and low-pass filters, which  
23 produce detail and approximation coefficients, respectively. The approximation coefficients  
24 estimate the rough features of the original data, while the detail coefficients capture frequent  
25 movements, such as sudden changes or spikes, that might be hidden in the original data due  
26 to trends or seasonal fluctuations. The DWT facilitates multi-resolution representations of  
27 signals, aiding pattern detection algorithms by revealing structures at various time scales.

## 28 **S1.2 Comparative analysis**

29 In this paper, we used data from two sources, the California Department of Public Health  
30 (CPDH) and WastewaterSCAN (WWSCAN) [17, 20], each of which implemented different  
31 sampling and laboratory analytical protocols to measure the concentration of SARS-CoV-2  
32 RNA (N gene) and PMMoV (Tables S.2-S.3). The CDPH laboratory analyzed raw wastew-  
33 ater from composite influent samples, with concentrations of each gene target reported as  
34 genome copies per volume of wastewater. The WWSCAN laboratory analyzed dewatered  
35 settled solids from composite influent samples or, more often, dewatered solids from primary  
36 clarifier sludge, with concentrations of each gene target reported as genomic copies per mass  
37 of dry sludge. Trends in normalized wastewater concentrations (N/PMMoV, a dimensionless  
38 metric) obtained by the two methods have been extensively validated against trends COVID-  
39 19 clinical data. However, the magnitudes of the normalized wastewater concentrations differ  
40 between the procedures. Distinct laboratory procedures can also introduce differing noise,  
41 biases, or variations into the data, which may manifest as rapid fluctuations or high-frequency  
42 components.

43 A primary challenge thus lies in directly comparing wastewater signals acquired by dif-  
44 ferent methods due to the inherent biases introduced by different laboratory procedures. To  
45 help mitigate these biases, we employed a wavelet transform, which extracts patterns and  
46 features from the data to compare and analyze the wastewater signals more robustly. Specif-  
47 ically, we decompose the data into multiple low-frequency and high-frequency components,  
48 allowing us to focus independently on high-frequency details. We hypothesize that variations  
49 in high-frequency components may arise from differences in laboratory methodologies rather  
50 than fluctuations in disease dynamics. We conduct an exploratory analysis to investigate the

51 success of a wavelet transformation in distinguishing patterns related to disease dynamics  
52 rather than features specific to the laboratory methods.

53 We analyzed normalized wastewater concentration data collected from the city of Davis  
54 between May 7, 2022, and September 28, 2022, by two different laboratories applying highly  
55 similar methods (referred to as Lab 1 and Lab 2). For a detailed description of this data,  
56 please refer to [33]. We used the normalized wastewater concentration of SARS-CoV-2 RNA  
57 detected in each wastewater sample for the two laboratories. We employed the discrete  
58 wavelet transformation utilizing the Daubechies 4 wavelet family to decompose the data into  
59 multiple levels. During this process, coefficients at different scales and positions are generated.  
60 At each level ( $l$ ), we obtain two sets of coefficients: approximation coefficients ( $cA_l$ ), repre-  
61 senting low-frequency components) and detail coefficients ( $cD_l$ ), representing high-frequency  
62 components). Both  $cA_l$  and  $cD_l$  have a length of  $N/2^l$ .  $cA_l$  captures information about the  
63 overall trend of the signal, while  $cD_l$  contains high-frequency details,  $l = 1, \dots, L$ .

64 We utilized the PyWavelets library for the wavelet decomposition, specifically the `pywt.wavedec`  
65 function. We employed the `pywt.dwt_max_level` to determine the appropriate decomposi-  
66 tion level function. The `pywt.wavedec` function returns the coefficients ( $cA4, cD1, cD2, cD3, cD4$ ).  
67 Figure S.1 shows the wavelet coefficients for the two-time series. We observed higher differ-  
68 ences between these two series in the detail coefficients  $cD_2, cD_3$ .

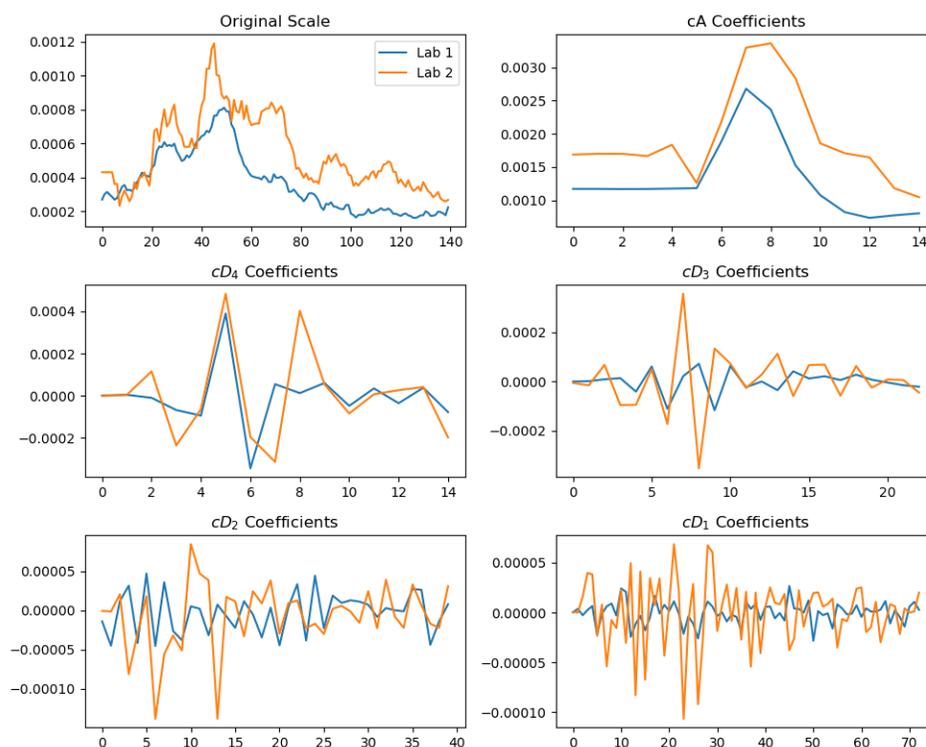


Figure S.1: Wavelet Coefficients Comparison between Lab 1 and Lab 2 time series.

69 We calculated the distance correlation between the data from the two labs using different  
 70 data representations: original scale, wavelet coefficients (including all levels), and wavelet  
 71 coefficients excluding detail coefficients at different levels. The results, as shown in Table  
 72 S.1, indicate that the distance correlation is lower (indicating higher similarity) when we use  
 73 their wavelet coefficients instead of the series in the original scale. We also observed that  
 74 excluding detail coefficients at the first two levels further decreased the distance correlation,  
 75 as these coefficients primarily capture changes in high frequencies within the data and may  
 76 be associated with laboratory-specific methods.

Table S.1: Distance correlation between Lab 1 and Lab 2.

Data	Distance Correlation
Original Scale	0.1774
Wavelet Coefficients	0.0261
Wavelet Coefficients (Excluding $cD_1$ )	0.0271
Wavelet Coefficients (Excluding $cD_1, cD_2$ )	0.0298

77 Furthermore, we reconstructed the time series from the wavelet coefficients, excluding  
 78 coefficients at level 1, levels 1 and 2, and levels 1, 2, and 3. When we exclude detail coefficients  
 79 at level 1 and at levels 1 and 2, the reconstructed signal closely aligns with the original series,  
 80 with only small variations removed, Figure S.2. However, excluding detail coefficients at the  
 81 third level results in a desynchronization of the reconstructed signal compared to the original  
 82 scales, leading to a loss of information about the trend. This highlights the importance  
 83 of preserving essential features while eliminating high-frequency components during signal  
 84 reconstruction.

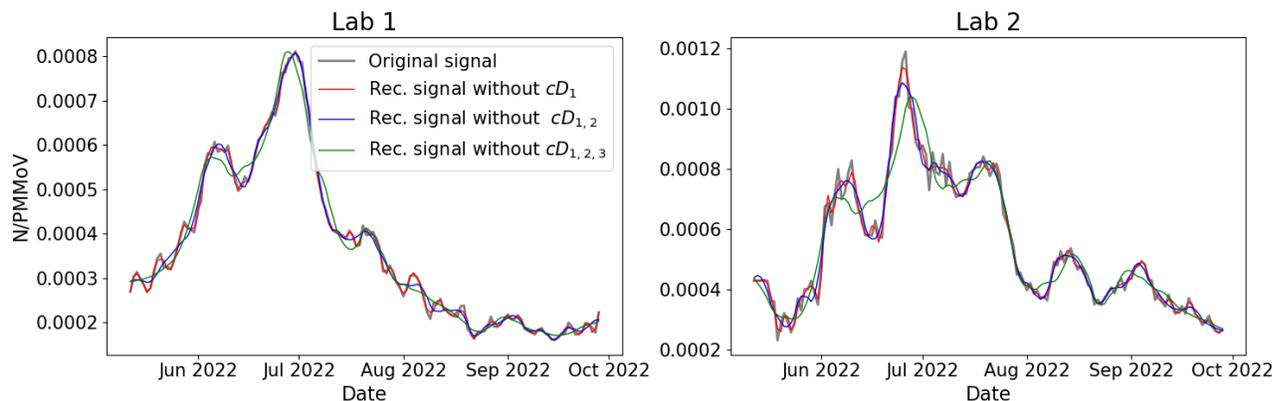


Figure S.2: Reconstructed time series from wavelet coefficients excluding different levels.

85 Based on our findings, we chose to use wavelet coefficients, excluding detail coefficients at

86 levels 1 and 2, to construct the dissimilarity matrix with correlation distance. This approach  
87 helped us identify similarities between signals from different laboratories while minimizing  
88 the biases introduced by varying measurement methods. We will explore this hypothesis  
89 further in the future analysis to assess the variability in different frequencies and scales and  
90 its relation with different labs, procedures, temperature, and other variables.

## 91 **S2 Algorithm performance**

92 To validate the robustness of the algorithm, we executed it 100 times, consistently converging  
93 to the same solution in each run. This demonstrates the stability and reliability of the  
94 algorithm across multiple iterations. Figure S.3 illustrates the energy trajectories for these  
95 runs, providing a visual representation of the algorithm's convergence consistency.

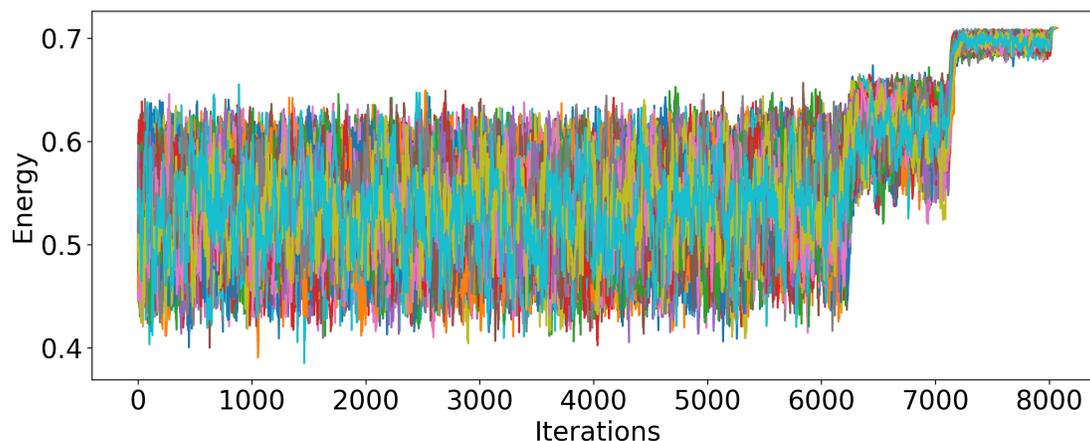


Figure S.3: Evaluation of the optimization function is performed in each iteration of simulated annealing. The traces correspond to 100 trajectories that were run to demonstrate their performance.

96 **S3** Supplementary figures

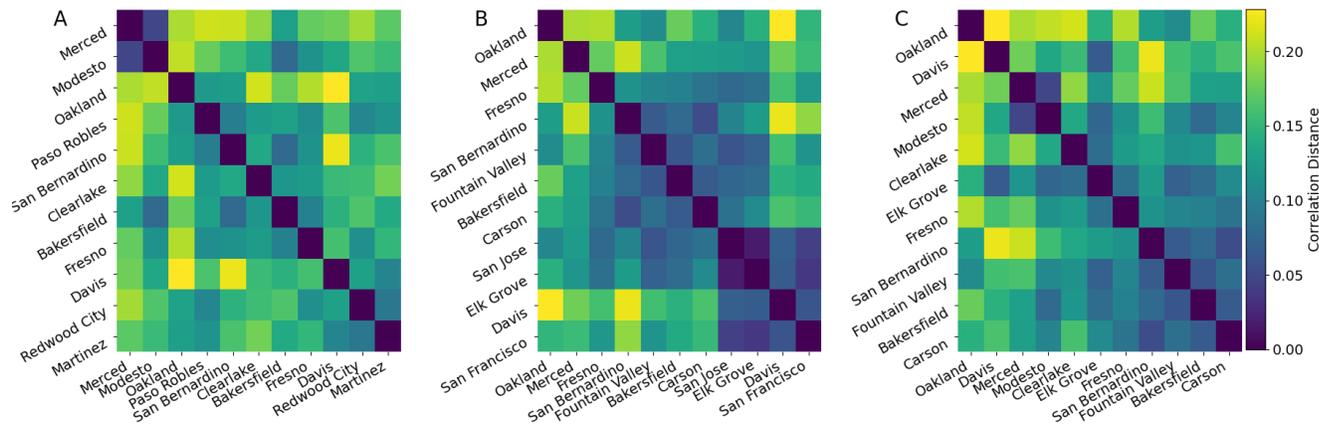


Figure S.4: Dissimilarity matrix for the optimal solution in each scenario. **A.** Prioritizing disease signal dissimilarity. **B.** Equal factors. **C.** A balanced approach for SVI and disease signal dissimilarity.

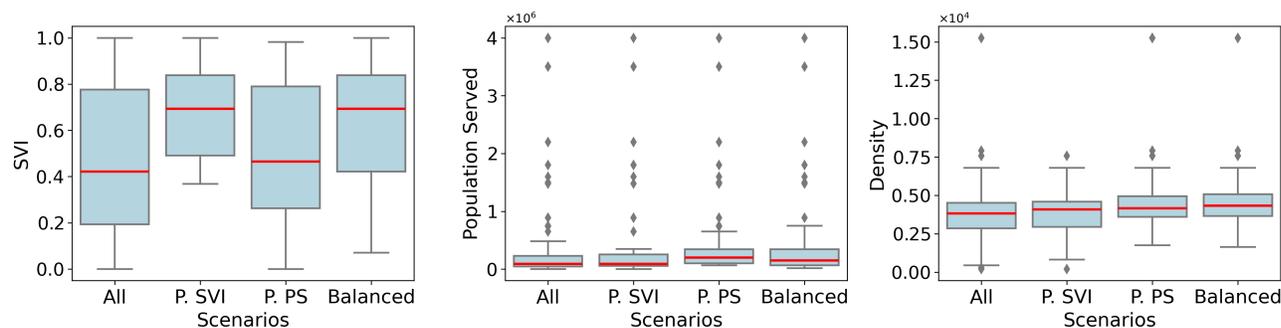


Figure S.5: Distribution of the SVI, population served, and population density for all WWTPs that have WDS for SARS-CoV-2 in California (All), and for the optimal WWTPs found in each scenario: prioritizing SVI (P.SVI), prioritizing population served (P.PS), and a balanced approach (Balanced).

## 97 S4 Simulated Annealing algorithm

---

**Algorithm 1** Simulated Annealing Algorithm

---

**Input:** Number of plants  $n$ , objective function  $f(\mathbf{x})$ , positive integer  $k < n$ , maximum temperature  $T_{\max}$ , minimum temperature  $T_{\min}$ , and cooling schedule parameter  $\alpha$

**Output:** Solution  $\mathbf{x} \in \{0, 1\}^n$

- 1: Enumerate each plant from 1 to  $n$  and set  $I = \{1, \dots, n\}$
- 2: Initialize  $\mathbf{x} = (x_1, \dots, x_n)$  by randomly selecting  $k$  distinct numbers from  $I$ , forming a set of indices denoted as  $I_k$ . Set

$$x_i = \begin{cases} 1, & i \in I_k \\ 0, & i \notin I_k \end{cases}$$

- 3: Compute  $E = f(\mathbf{x})$
- 4:  $T \leftarrow T_{\max}$
- 5: **while**  $T > T_{\min}$  **do**
- 6:      $A \leftarrow 0$  ▷ Number of accepted solutions
- 7:      $G \leftarrow 0$  ▷ Number of generated solutions
- 8:     **while**  $A < 10 \times n$  and  $G < 100 \times n$  **do**
- 9:         Randomly choose a element  $l$  from  $I_k$  ▷ Index (plant) to be replaced
- 10:         Randomly choose a element  $m$  from  $I - I_k$  ▷ Index to be added
- 11:         Set  $\mathbf{x}_{\text{new}} = \mathbf{x}$  with  $x_l = 0$  and  $x_m = 1$
- 12:         Compute  $E_{\text{new}} = f(\mathbf{x}_{\text{new}})$
- 13:         Calculate  $\Delta E = E - E_{\text{new}}$
- 14:         **if**  $\Delta E < 0$  **then**
- 15:              $\mathbf{x} \leftarrow \mathbf{x}_{\text{new}}$
- 16:              $E \leftarrow E_{\text{new}}$
- 17:         **else**
- 18:             Generate  $r \sim \text{Uniform}(0, 1)$
- 19:             **if**  $r < e^{-\Delta E/T}$  **then**
- 20:                  $\mathbf{x} \leftarrow \mathbf{x}_{\text{new}}$
- 21:                  $E \leftarrow E_{\text{new}}$
- 22:             **end if**
- 23:         **end if**
- 24:     **end while**
- 25:      $T \leftarrow \alpha \cdot T$
- 26: **end while**
- 27: **return**  $\mathbf{x}$

## 98 S5 Wastewater methods description

99 Wastewater data from March 21, 2022, to May 21, 2023, for WWTPs located in Oakland  
 100 (Alameda), Martinez (Contra Costa), Merced (Merced), Elk Grove (Sacramento), Paso Rob-  
 101 les (San Luis Obispo), Palo Alto (San Mateo), San Jose (Santa Clara), Modesto (Stanislaus),  
 102 and Davis (Yolo) have been previously published in [34]. For the remaining WWTPs, we  
 103 present an overview of the methods employed as reported by CDPH. Specifically, we refer to  
 104 the lab identifiers listed in Table S.3 for the WWTPs described in Table S.2.

County	City	POTWs	Population Served	Lab Id	Data Source
Alameda*	Oakland	East Bay Municipal Utility District Main Wastewater Treatment Plant	740,000	VLT	WWSCAN
Contra Costa*	Martinez	Central Contra Costa Sanitary District Wastewater Treatment Plant	487,300	VLT	WWSCAN
Fresno	Fresno	Fresno/Clovis RWRP	650,000	DWRL	CDPH
Kern	Bakersfield	City of Bakersfield Plant #2	168,750	DWRL	CDPH
Lake	Clearlake	City of Clearlake	13,200	DWRL	CDPH
Los Angeles	Carson	Joint Water Pollution Control Plant	3,500,000	VLT	WWSCAN
Merced *	Merced	Merced Wastewater Treatment Plant	91,000	VLT, VLT1	WWSCAN
Orange	Fountain Valley	OC San (Orange County Sanitation District) Reclamation Plant No. 1	1,800,000	DWRL	CDPH
Sacramento*	Elk Grove	Sacramento Regional County Sanitation District	1,480,000	VLT	WWSCAN
San Bernardino	San Bernardino	Margaret H Chandler WWRF, San Bernardino	325,000	DWRL	CDPH
San Francisco	San Francisco	Oceanside Water Pollution Control Plant	250,000	VLT	WWSCAN
San Luis Obispo*	Paso Robles	City of Paso Robles Wastewater Treatment Plant	31,037	VLT1	WWSCAN
San Mateo*	Palo Alto	Palo Alto Regional Water Quality Control Plant	236,000	VLT	WWSCAN
San Mateo	Redwood City	Silicon Valley Clean Water	1992,000	VLT	WWSCAN
Santa Clara	Sunnyvale	City of Sunnyvale Water Pollution Control Plant	153000	VLT	WWSCAN
Santa Clara*	San Jose	San Jose Santa Clara	1,500,000	VLT	WWSCAN
Santa Clara	Gilroy	South County Regional Wastewater Authority	110,338	VLT2	WWSCAN
Stanislaus*	Modesto	Modesto Wastewater Primary Treatment Facility	230000	VLT, VLT1	WWSCAN
Yolo*	Davis	City of Davis Wastewater Treatment Plant	68,000	VLT, VLT1	WWSCAN

Table S.2: Description of publicly owned treatment works (POTWs) included in this study. Plants located in cities marked with an asterisk (\*) indicate those for which wastewater data during the study period have been published here [34]. Lab Id (DWLR; VLT, VLT1, VLT2) corresponds to methods summarized in Table S.3

Collection Method		Processing Method			Quantification Method				Sample					Quantification Results		Lab Id
Sample type	Sample matrix	Extr. method	Rec eff target	Rec eff spike	PCR type	LOD ref	Hum frac target	Hum frac target ref	Quant stan type	Stan ref	Inh method	Num no target control	Major lab method desc	gene target	PCR gene target ref	
grab	primary sludge	[35]	bcov vaccine	dewatered solids	ddpcr	(i)	PPMoV	[36]	DNA	(ii)	(iii)	more than 3		N	[37]	VLT
grab	primary sludge	[35]	bcov vaccine	dewatered solids	ddpcr	[38]	PPMoV	[36]	RNA	(ii)	none	more than 3	Verily solids method	N	[39]	VLT1
24-hr flow-weighted composite	raw wastewater	[35]	bcov vaccine	dewatered solids	ddpcr	(i)	PPMoV	[36]	DNA	(ii)	(iii)	more than 3		N	[37]	VLT2
24-hr time-weighted composite	raw wastewater	[40]	bcov vaccine	raw sample	ddpcr	[41]	PMMoV	[42]	RNA	[43]	none	3	(iv)	N1	[43]	DWRL

Table S.3: Summary of different methods employed in the collection, processing, quantification, and analysis of samples for the detection of viral pathogens in wastewater as reported by CDPH [17]. Abbreviations: Extr. method - Extraction method; Rec eff - Recovery efficiency; PCR - Polymerase Chain Reaction; LOD ref - Limit of detection reference; Hum frac - Human fecal fraction; Quant stan type - Quantification standard type; Stan ref - Standard reference; Inh method - Inhibition method; Lab - Laboratory; Major lab method desc - Major laboratory method description. References are provided where applicable. (i) 3 droplets as a minimum for a positive detection. (ii) Synthetic DNA oligo purchased from IDT. (iii) Solutions titrated with varying concentrations of solids to identify a concentration at which inhibition of the SARS-CoV-2 assays was minimized. (iv) This lab uses a lab method distinct from other labs in this reporting jurisdiction.