Detection of SARS-CoV-2 in urban wastewater: An entropy-based multi-objective optimization approach for optimal selection of sampling points

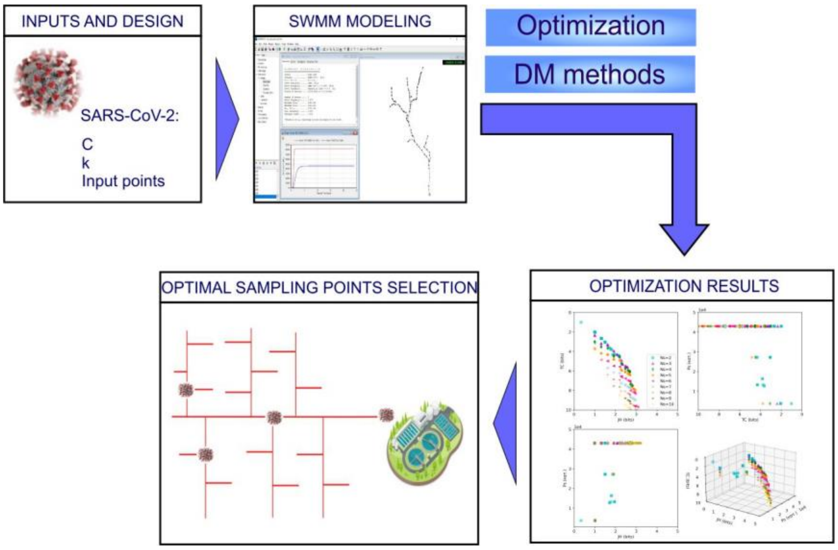
Argyro Gkatzioura[*1], Antigoni Zafeirakou[1]

[1]Department of Civil Engineering, Aristotle University of Thessaloniki, GR- 54124 Thessaloniki, Greece;

gargyro@civil.auth.gr; azafir@civil.auth.gr

[*]Corresponding: Argyro Gkatzioura, Department of Civil Engineering, Aristotle University of Thessaloniki, GR- 54124 Thessaloniki, Greece. Email: gargyro@civil.auth.gr

Graphical abstract



ABSTRACT

Monitoring wastewater for the presence of SARS-CoV-2 proved to be a useful tool during COVID-19 pandemic and it is in place, until today, at many Wastewater Treatment Plants (WWTPs) in Greece and elsewhere. Identification of virus in wastewater can be a valuable indicator for the early detection of the virus's upsurges or the evaluation of the disease's spread. Since this practice might become common in the future, for inspection of SARS-CoV-2 or other viruses, it is important to develop methodologies for identifying the optimum sampling sites, which can provide the most reliable information for the virus's transmission. In the present study, an optimization procedure is proposed for selecting monitoring points, in terms of number and location across the sewer system. Optimization is based on entropy and total correlation, as well as on supervised population. Four sampling points, well distributed across the sewer system, appear to be a well-balanced solution for wastewater surveillance, while more than seven sampling points do not add significant information content.

Keywords: wastewater-based epidemiology; SARS-CoV-2; viruses surveillance; optimization; wastewater; EPA SWMM;

1. Introduction

Wastewater surveillance for the presence of SARS-CoV-2 was widely applied during the COVID-19 pandemic and continues until today. The usefulness of this practice, as a supportive tool to clinical detection of infected population, has been proved valuable during the pandemic. Even though the virus becomes rapidly unviable in wastewater, its genetic fragment remains detectable (Maal-Bared et al., 2021). It is found that there is a strong correlation between the viral concentration detected in wastewater and the reported infections from the clinical tests (Fahrenfeld et al., 2022; Gonzalez et al., 2020; Koureas et al., 2021). Additionally, it is reported that SARS-CoV-2 RNA can be detected in wastewater before the rise of infections and hospitalizations and even before the clinical detection of the first case (D'Aoust et al., 2021; Galani et al., 2022; Hernández-Terrones et al., 2023; Randazzo et al., 2020). Wastewater monitoring for the presence of viruses has multiple intentions. For instance, some of them are the identification of asymptomatic carriers of a disease or those carriers with mild symptoms, an alternative way of monitoring, in case where clinical tests are in deficit or an early-warning tool for predicting diseases' outbreaks. Furthermore, it can offer an insight into disease prevalence within the catchment of the collection system and WWTP.

Wastewater-based surveillance of viruses is not a new tool. It has been in use for decades in order to track presence of viruses of interest in wastewater (Kilaru et al., 2023). Norovirus and hepatitis A virus can be found in wastewater some weeks before an increase in diagnosed cases (Hellmér et al., 2014). Additionally, wastewater is found to be a sensitive indicator for poliovirus circulation in the community (Lago et al., 2003). Also, respiratory viruses such as influenza A virus and respiratory syncytial virus (RSV) can be detected in wastewater, and its concentration presents positive correlation with data from clinical tests (Ahmed et al., 2023). Similarly, SARS-CoV virus was detected in wastewater from hospitals that treat SARS patients, during the virus's outbreak (Wang et al., 2005).

The most common practice among the aforementioned studies and until nowadays amid SARS-CoV-2 surveillance, is to sample untreated wastewater from the influent to the WWTPs, after screening and grit removal. During the following treatment stages, especially after secondary treatment, viruses' genetic markers are very low or undetectable (Carine & Pagilla, 2024), thus they are not appropriate for surveillance. The benefit of sampling at the entrance of the WWTP is that there is usually an already established sampling equipment and that the entire population of the sewershed can be detected for viruses. However, there are certain drawbacks, such as the effect of several in-sewer processes, i.e. dilution, adsorption onto solids and decay of the virus' fragments (Gundy et al., 2009; Kostoglou et al., 2021), which is more intense at the end of the sewer system and this can lead to underestimation of the presence of the virus. In addition, the monitoring of separate regions of the catchment is not possible. To alleviate these drawbacks, more samples should be taken from more upstream points across the sewer system, for example in manholes, interceptors or pumping stations, as proposed in a few studies (Baldovin et al., 2021; Hasan et al., 2021). Word Health Organization (WHO, 2023) general guidance recommends the prioritization of sampling points being representative for larger population, in order to obtain general trend of the viral concentration and release early-warnings, followed by spatially distributed sampling points, such as those located in major sewers or drainage points.

An optimization procedure can be employed for the determination of the optimum number and combination of sampling points. Optimization approaches are widely applied in water and hydraulic engineering problems (Bakanos & Katsifarakis, 2019) (Panagoulia et al., 2017). However, only a few studies apply optimization methods in wastewater-based epidemiology field in order to select monitoring points. Domokos et al. (2021), aimed to find the best combination of monitoring points in a way that each of them can monitor a certain part of the sewer network, engaging a GIS-based application. Larson et al. (2020) and Calle et al. (2021) proposed methodologies for consecutive measurements to a series of nodes, in order to identify the source of the virus or the hot spot of the infections, using the optimization coupled with graph-theory. Calle et al. (2021) proposed an additional optimization algorithm for maximizing observed population and avoiding overlapping of observed areas, in order to select an optimum number of sampling points for surveillance of SARS-

CoV-2. Similarly, Kim et al. (2022) developed a methodology for selecting monitoring points for wastewater based epidemiology, that each of them can monitor a different part of the sewer network with similar area sizes, using a genetic optimization algorithm, as well as a special bisection method. Nevertheless, all the aforementioned studies used a graphical representation of the sewer system and did not take into consideration the fate of the viral particles. The only two studies that account, to some extent, for the viral RNA loss across the sewer system are the one by Yao et al. (2024) and Wang et al. (2020). Both these studies proposed optimization algorithms to determine optimal sampling points based on detection likelihood of each node. Specifically for SARS-CoV-2 the first study and for Salmonella Typhi the second one. Yao et al. assumed a constant flow velocity and constant travel time in the sewer system, for the computation of viral RNA loss. While Wang et al. adjusted parameters, such as travel time and percentage loss of viral RNA, based on probability distributions. However, neither of them used a hydraulic simulation of the sewer system or a simulation of the viral particles fate as a step before optimization procedure. Thus, the structure of the sewer system was not considered nor the hydraulic conditions and in-sewer processes that possibly affect the fate and transport of viral particles.

This study considers sampling from two or more points across the sewer system. A four-step methodology is proposed in order to find the best number and combination of sampling points. It involves the hydraulic modeling of the sewer system and the simulation of the fate and transport of the viral particles, followed by an optimization procedure, the application of decision making tools and the validation of the procedure. The aim of the study is to select an optimum number along with the optimum locations of monitoring points, for detecting viral traces in the wastewater. The finally selected set of sampling points should be adequately informative for the presence or not of viral RNA. It is important the wastewater-based surveillance of viruses to be able to serve as a useful tool for early detection of viruses' upsurges, allowing the prompt employment of measures to prevent the wider spread of diseases. In addition, the proposed method aims to aid the community in the assessment of viruses' circulation and better control of a pandemic situation.

2. Materials and methods

The proposed four-step optimization procedure is depicted in Figure 1. Initially, the hydraulic modelling of the sewer system, with the simulation of fate and transport of SARS-CoV-2 RNA in the sewer system is employed. Data extracted from the hydraulic simulation, is utilized in the calculation of objective functions' values required for the optimization algorithm. The optimization algorithm is then executed and a pareto-front containing the non-dominated solutions is generated. Following that, a decision making process is carried out for picking certain well-balanced solutions among the non-dominated solutions. Finally, a validation of the results is carried out.

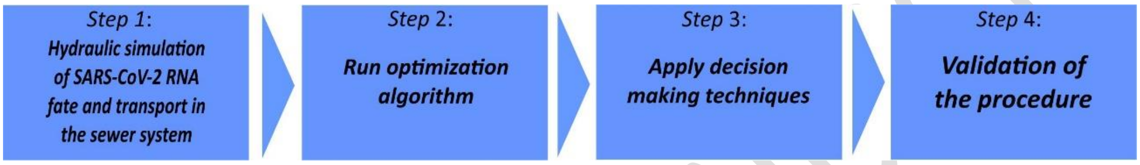


Figure 1: Schematic presentation of the proposed four-step optimization methodology

2.1. Hydraulic model for the sewer system

The United States Environmental Protection Agency's Storm Water Management Model (EPA's SWMM) is used for the hydraulic simulation of the sewer system and for the virus' fate and transport. SWMM is a dynamic rainfall runoff simulation model; it can simulate both wastewater quantity and quality, of the runoff and routing through the sewer system (Rossman, 2015).

The collection system of the city of Kozani, in Greece, is used as a study area, for the application of the proposed methodology. The collection system of Kozani is a combined system, which carries stormwater and wastewater to the nearby Wastewater Treatment Plant (WWTP). Kozani's WWTP treats urban wastewater of an equivalent population of 48,000 (43,000 from the city of Kozani and 5,000 from the village of Krokos). Local water supply and collection system services company (D.E.Y.A.K.) provided all the data of the collection system and wastewater inflows to WWTP. Due to lack of some data about the collection system, only the main pipeline system is used for further analysis (Figure 2). WWTP's inflow data is used for calibration of the model, in terms of wastewater inflows at each node. Dry weather conditions are taken into account and no sub-catchments are included in the model. Simulation of 48-hours is applied, with kinematic wave as the flow rooting method. Discharge of the virus into the collection system starts with the start of the simulation.

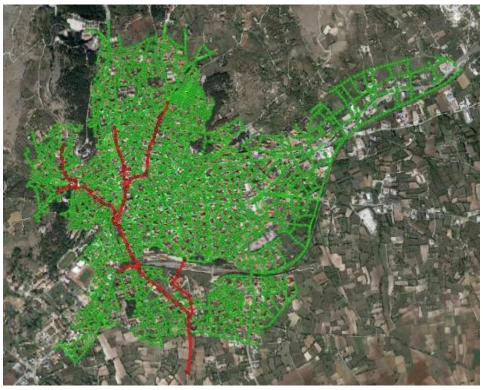


Figure 2: Sewer system of the Kozani city and the main part of the system (with red line) that is considered

First order decay is considered for SARS-CoV-2 RNA, as proposed by (Bivins et al., 2020) and (Hokajärvi et al., 2021). Mean first-order decay rate constant (k) of SARS-CoV-2 found to be in the range of 0.08–3.4 day^{-1}, based on the literature review (Ahmed et al., 2020; Bivins et al., 2020; Hokajärvi et al., 2021; Weidhaas et al., 2021; Yang et al., 2022). For the current study, the higher reported value of 3.4 day^{-1} is adopted. This value is considered as the least favourable scenario, since it causes rapid reduction of viral particles concentration. On the other hand, there is no available data of initial concentration, since it depends on the number of infected persons, and there are large uncertainties in estimation of the exact extracted rate of SARS-CoV-2 RNA per infected person (Li et al., 2021). As a result, for the present study, the relative low value of 3×10^4 gene copies/L (GC/L) is assumed for initial concentration. Moreover, it is presumed that initial concentration is inserted into the collection system from one node at a time, in order to consider an early stage of the disease spread with low number of infected persons. More detailed description of the hydraulic model, the results as well as a sensitivity analysis for different scenarios of initial concentration and decay rates can be found in Gkatzioura & Zafeirakou (2023). Sixteen different scenarios are run, with a different entrance point of the viral concentration for each one of them and twenty-one possible sampling points are examined. The nodes with the higher number of connected population are chosen as candidate entrance points, since these points have higher possibility of including the first infected persons. Central nodes and junctions of more than two pipes are chosen as candidate sampling points, also taking into consideration their accessibility according to our communication with the responsible company (D.E.Y.A.K.). Furthermore, 24h composite samples of 100 mL of wastewater, collected every hour are considered, and data is extracted from the second day of the simulation, when the concentration values are stabilized. Simulation data is extracted from SWMM model by a code written using the pyswmm Python package (McDonnell et al., 2020).

2.2. Optimization procedure

For the optimization procedure, concentration data is extracted from SWMM model at each node of the sewer system and are utilized for the calculation of parameters that are required for the objective functions estimation. Two different optimization concepts, O1 and O2, are examined in order to choose the best number and combination of sampling points across the sewer system. The objective functions under evaluation are joint entropy and total correlation of each combination of sampling points in order to find the most informative solutions, in addition to population that can be inspected from each group of sampling points. Information Theory is developed by Shannon (1948) and introduced *Entropy* as a measure of information content of an independent variable, *Joint Entropy* as a measure of the amount of information content between two or more random variables and *Total Correlation* as the amount of information shared by two or more variables. These parameters have been used in the past for optimal design of water quality and quantity monitoring systems (Alfonso et al., 2014; Banik et al., 2017; Brentan et al., 2021; Chen et al., 2022; Jafari et al., 2022). *Joint Entropy* is used as a measure of the information content of a combination of sampling points, and it is calculated based on the viral particles' concentration at each node of the collection system. *Total Correlation* is a metric regarding the information shared between the nodes or dependency of a combination of sampling points. In addition, it serves as a limiter for the redundant information. Observed population is proposed from WHO as a criterion for choosing sampling points.

Entropy (H), Joint Entropy (JH) and Total Correlation (TC) are calculated with the formulas (1), (2) and (3) as shown below:

$H(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log p(x_i)$ (1)

where, H(X): the entropy of the discrete random variable X, $x_1, x_2, \ldots x_n$: discrete values of variable X and $p(x_i)$: probabilities of values $x_1, x_2, \ldots x_n$ of variable X.

$JH(X_1, X_2, \ldots X_n) = -\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \ldots \sum_{i_n=1}^{n_n} p(x_{i1}, x_{i2}, \ldots x_{in}) \cdot \log p(x_{i1}, x_{i2}, \ldots x_{in})$ (2)

where, $JH(X_1,X_2,\ldots X_n)$: joint entropy of the independent random variables $X_1, X_2, \ldots X_n$, $p(x_{i1}, x_{i2},\ldots x_{in})$: joint probability of $X_1,X_2,\ldots X_n$ variables.

Entropy and joint entropy measurement units depend on logarithm's base. If base *e* is used, then the measurement units are nats. If base 2 is used, then the measurement units are bits. For this study, logarithm with base 2 is used and thus, H, JH and TC values are presented in bits.

$TC(X_1, X_2, \ldots X_n) = \sum_{i=1}^{n} H(X_i) - JH(X_1, X_2, \ldots X_n)$ (3)

where, $TC(X_1,X_2,\ldots X_n)$: total correlation of the variables $X_1, X_2, \ldots X_n$.

In this research, variables X_i are the concentration values at each candidate sampling point.

Population (P_s) is the upstream population that can be observed from a particular node. Maximum population, that can be observed from a certain combination of sampling points is considered:

$P_s = \max P_s(X_i), \quad i = 1 \ldots . N$ (4)

where $P_s(X_i)$ is the population that can be observed from the sampling point (node) X_i.

The two optimization concepts (O1 and O2) that are applied and the objective functions f_i for each one of them, are described herein:

O1: f_1 = maxJH, f_2 = minTC

O2: f_1 = maxJH, f_2 = minTC, f_3 = maxP_s

The aforementioned optimization procedure is executed, with respect to the samples acquired from a set of 2 to 10 sampling points each time (N_s = [2,3,…10]).

For the realization of the optimization procedures, the Nondominated Sorting Genetic Algorithm (NSGA-II), an elitist multi-objective evolutionary algorithm proposed by Deb et al. (2002), is selected. The NSGA-II algorithm introduces a fast nondominated sorting procedure that reduces the computational complexity. It also includes elitism, since the mating pool is created by combining the parent and offspring populations. Furthermore, the main functions of the algorithm are the calculation of non-domination and crowding-distance, as well as the sorting of solutions based on the fitness and spread. Thus, the selection of diverse solutions is reassured. The optimization algorithm is ran with the Python's package, pymoo 0.5.0 (Blank & Deb, 2020).

2.3. Decision making techniques

Multi-objective optimization procedure concludes to a pareto front, which frequently is comprised of a large number of non-dominated solutions. These solutions are the best trade-offs between the objective functions. Decision makers have to pick one or more of these solutions and the final decision should also be based on criteria such as accessibility of nodes and cost. However, due to numerous pareto optimal solutions, decision making tools can be applied supportively. Herein, Pseudo-Weight Vector proposed by Deb (2001) and Simple Additive Weighting (SAW) (Zionts & Wallenius, 1983) methods are applied for the decision-making step. For applying the Pseudo-Weight Vector method (PW), the relative distance of each solution from the worst value in each objective function is calculated. This value is called pseudo-weight vector, which is a metric of the solutions' location across the pareto-front. For example, the boundary solutions that favour one objective function over the rest, will have values of pseudo-weight vector equal to 1 for the favoured objective and equal to 0 for all the others. Then, the desired location of the solution or pseudo-weights are set by the user and the solutions with the greatest proximity to these values are chosen. For the application of SAW method a "score" for each solution is calculated. The "score" is the sum of the products of each objective function's normalized values and the weights that are assigned, by the user, to each of them. The solutions are then ranked based on the highest score. In all cases, equal weighs are considered for all objective functions. Furthermore, a decision based on "other criteria" (OC) is explored. More specifically, OC refers to selecting solutions that include WWTP as one of the sampling points and at the same time being ranked high at one or at both the applied decision making methods.

2.4 Validation of the proposed procedure

For validation, 4 extra scenarios are ran with the SWMM model. For each of these scenarios, a different node, different from the nodes considered during the application of the optimization methodology, is considered as the point where the initial concentration of SARS-CoV-2 enters the sewer system. This modelling results are used in order to evaluate the ability of the proposed solutions to detect the virus.

3. Results and discussion

3.1 Optimization results

Results of the optimization procedure are presented and discussed hereafter. Figure 3 depicts the pareto optimal solutions of optimization concept O1 and for Ns = 2 – 10 number of sampling points. The optimal solutions are those that are located at the upper, right part of the chart. Total correlation (TC), as expected, is increased (deteriorated) with the increased number of sampling points, because of the increase of redundant information between selected sampling points. On the other hand, joint entropy (JH) is increased (improved) while the number of sampling points is increased, since the information content is also increased. However, after 4 sampling points the ascent rate is limited. The maximum, the mean and the 3rd quartile values of JH rise, equal to 30%, 41% and 49% from 2 to 4 sampling points respectively. Similarly, they equal to 13%, 26%, 11% from 4 to 7 sampling points respectively. And finally, they equal to 0%, 6 %, 3% from 7 to 10 sampling points accordingly. It appears that a set of 4 to 7 sampling points can perform well with respect to the information content (maximizing joint entropy). Considering that the total correlation should be as low as possible and that not too many samples can be analyzed simultaneously, since the analysis is conducted in the lab, which is time consuming and expensive, it can be concluded that 4 sampling points is a well-balanced and feasible solution.

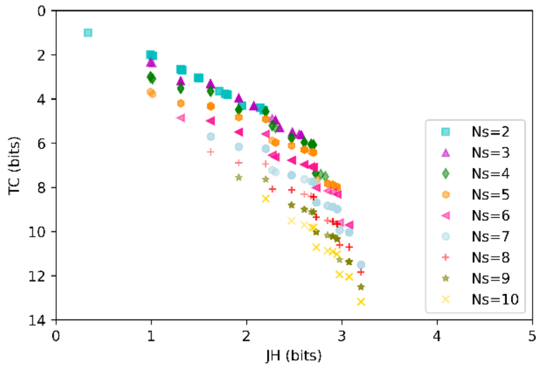


Figure 3: Pareto-optimal solutions derived from optimization concept O1

Similar outcomes can be drawn from optimization concept O2, which consists of three optimization objectives. Figure 4 depicts resulting pareto fronts for this optimization procedure. JH and TC are increased with larger number of sampling points. The inspection of the whole population can be achieved as well with less sampling points, especially for more than 3. Thus, JH appears to be the objective function that the final selection should be based on. The maximum, the mean and the 3rd quartile values of JH rise, equal to 30%, 42%, 44% from 2 to 4 sampling points respectively. Similarly, they equal to 13%, 23% and 9% from 4 to 7 sampling points respectively. And finally, they equal to 0%, 10%, 4% from 7 to 10 sampling points accordingly. It appears that 4 to 7 sampling points are capable of achieving balanced values from all objectives. Similarly with optimization concept O1, 4 sampling points appeared to be the most appropriate alternative, in O2 concept.

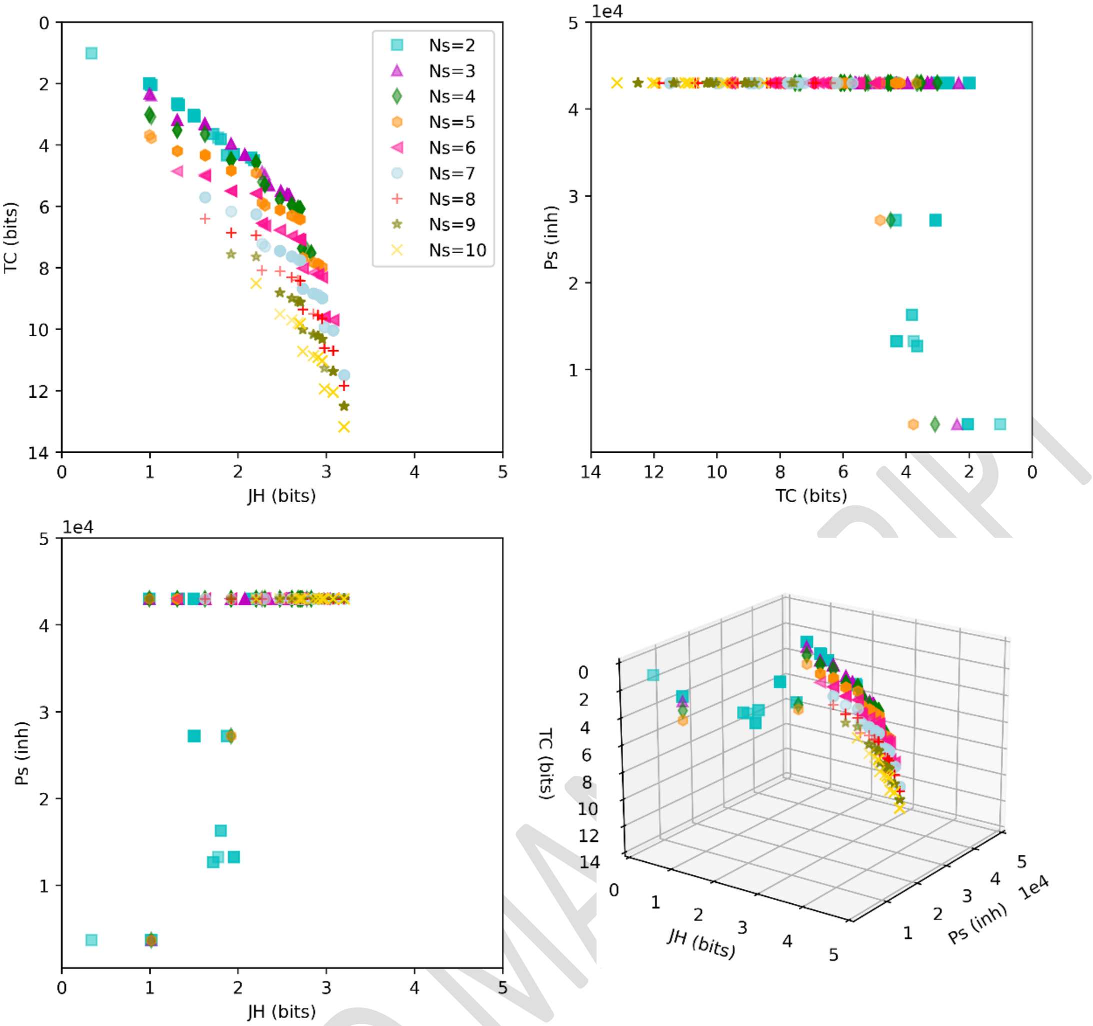


Figure 4: Pareto-optimal solutions derived from optimization concept O2

The Entropy and Joint Entropy values are case-dependant, they depend on the x_i values of X_i variables, x_i values' discretization method and on the number of bins used for probabilities' calculations; similar is the case for the TC values. The observed population is also network-dependant. However, a comparison with other studies can be made only based on the number of selected inspection points. Available studies also propose similar number of sampling points for reliable wastewater quality parameters monitoring. Domokos et al. (2021) chose 6 nodes for SARS-CoV-2 inspection and Calle et al. (2021) also selected 5-7 points for SARS-CoV-2 inspection, whereas Lee (2013) concluded to 7 nodes for water quality monitoring.

3.2 Decision making results

After carefull examination of the results obtained from the optimization procedure, it is concluded that 4 sampling points are adequate to monitor SARS-CoV-2 in wastewater. Sampling from more than

4 points is costly for the utility companies that operate collection systems, thus the proposed 4-sampling points solutions may be considered as a realistic scenario. Still, a large number of non-dominated solutions of 4 sampling points exists and it is necessary to choose some particular solutions among them. The number of solutions that are found by the NSGA-II algorithm is equal to the initial population size, unless a special technique is applied during the algorithm to reduce them, such as a threshold or a filtering technique. Alternatively a multiple criteria decision making technique can be applied after the completion of the optimization algorithm (Deb, 2001; Wang & Rangaiah, 2017). The latter is applied in this research. The solutions selected after application of decision making tools and the respective objective functions' values, for each optimization concept and decision making method are displayed in Table 1. Results from Pseudo-weight vector (PW) and Simple additive weighting (SAW) decision making methods are compared with Other Criteria (OC) in terms of JH, TC and Ps. Furthermore, the objective functions' values, when samples are taken only at the entrance of the WWTP, are also shown for comparison. The selected solutions are depicted in Figure 5.

Table 1: Solutions derived from decision-making (DM) techniques (Pseudo-weight vector (PW), Simple additive weigh (SAW), other criteria (OC)), for O1 and O2 optimization concepts and WWTP as the only sampling point.

Optimization	DM	Ns	Solutions	JH	TC	P_S
O1	PW (0.5,0.5)	4	1263, 1937, 2185, WWTP	2.2	4.6	*(43,000)*
	PW (0.8,0.2) & SAW	4	1290, 1714, 1835, 2185	2.7	6.1	*(43,000)*
	OC	4	1263, 1937, 2185, WWTP	2.2	4.6	*(43,000)*
O2	PW (0.333,0.333,0.333)	4	128, 1263, 1872, 2185	1,9	4,5	27,200
	PW (0.6,0.2,0.2) & SAW	4	1290, 1714, 1835, 2185	2.7	6.1	43,000
	OC	4	1263, 1835, 2128, WWTP	2.2	4.6	43,000
WWTP		1	WWTP	**H** = 0.7	-	43,000

The two methods, PW and SAW, select solutions in a different way. PW method finds the location of a solution in the pareto-front, while SAW method ranks the solutions based on a score emphasizing solutions closer to the best solutions for each objective function. It is worth mentioning that for each application of the decision making methods, not single, but groups of 5-8 solutions are selected. This happens due to the fact that a number of different solutions have similar values for certain objectives, and thus it is more intricate for the decision making tools to pick a distinct solution. Only certain of these solutions are shown in Table 1. Here, for the PW method, we choose the solutions located in the middle, with weightages w_1=0.5 and w_2=0.5 for O1 and w_1=0.333, w_2=0.333, w_3=0.333 for O2. However, for the O2 optimization concept, the solution set of sampling points is not representative for the whole population, thus it is not eligible. The two methods conclude to the same solutions if pseudo-weights values equal to w_1=0.80 and w_2=0.20 for O1 and w_1=0.60, w_2=0.20, w_3=0.20 for O2 respectively are targeted by PW method in order to pick a solution closer to JH's best value. SAW method appears to emphasize maximization objectives and has better performance for these objectives (JH, PS), while chosen solutions have worse values for the single minimization objective (TC). Nonetheless, selected sets of sampling points are well allocated across the sewer system for all decision making tools. The decision based on OC considers only solutions that include WWTP as a sampling point and it picks solutions which are ranked high simultaneously for the other two decision making methods. These solutions under both O1 and O2 optimization concepts have good overall performance. However, they exhibit lower values for JH, compared to the other selected solutions from either PW or SAW. Therefore, they can be representative of the whole population (P_S) and they achieve average values of JH, TC. The inclusion of the entrance to the WWTP as a sampling point is an efficient alternative, since regular sampling points are usually there and the required equipment and capacity are already in place.

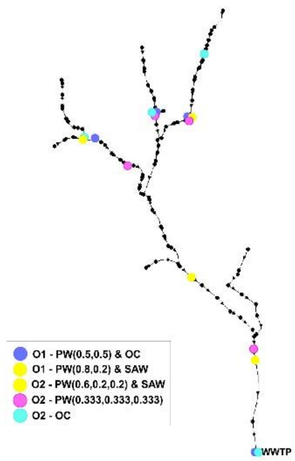


Figure 5: Solutions's location after the decision making step

If the samples are taken only from the entrance to the WWTP, all the population can be monitored, though entropy is low, lower than the ordinary joint entropy's values achieved from the solutions derived from the proposed optimization methodology. Additionally, the probability of detecting virus traces at the entrance to the WWTP is equal to 25%, while the maximum detection probability of the whole system is equal to 56%. This means that there is a higher possibility to fail to detect the viral traces, especially at the begging of a virus surge where the infected persons are fewer, viral load is low and the effect of in-sewer reduction of viral concentration is more intense. This is a critical point, because the early detection can help in taking measures on time and prevent the spread of the virus more effectively. Therefore, this research provides evidence that more upstream sampling points have to be included in SARS-CoV-2 monitoring programmes.

3.3 Validation results

Validation is performed for the solutions included in Table 1. They are evaluated for their ability to detect the virus under validation scenarios. The evaluation results are presented in Table 2, for the 2 optimization concepts, 3 decision making tools and 4 entrance points. Detection is achieved in 59% of the whole examined cases, for the rest of them the concentration are very low and maybe not detectable. This percentage is in agreement with the maximum detection probability of the system, which equals to 56%. Overall successfulness rate of both optimization concepts are similar, detection is achieved at 8 out of 16 cases. Similarly, solutions identified from the PW method with equal weights, as well as with the PW with unequal weights the SAW method and other criteria, are able to detect the virus at 6 out of 8 cases. Under all validation scenarios the WWTP as the only sampling point has very low concentration of viral particles and thus probably undetectable.

Table 2: Validation's results, for 2 optimization concepts (O1 &O2), 3 decision making tools (PW, SAW, OC) and 4 entrance nodes (1713, 1391, 1849, 1260).

	Entrance nodes	**1713**	**1391**	**1849**	**1260**
Optimization	**DM**				
O1	PW (0.5,0.5)	Low Conc.	Detected	Detected	Detected
	PW (0.8,0.2) & SAW	Detected	Detected	Detected	Low Conc.
	OC	Low Conc.	Detected	Detected	Detected

O2	PW (0.333,0.333,0.333)	Low Conc.	Detected	Detected	Detected
	PW (0.6,0.2,0.2) & SAW	Detected	Detected	Detected	Low Conc.
	OC	Low Conc.	Detected	Detected	Detected
WWTP		Low Conc.	Low Conc.	Low Conc.	Low Conc.

3.4 Limitations of the study

Some limitations of the study are the assumption that has to be made for the initial concentration and the fate of the virus in the sewer system due to the lack of an increased amount of data for robust calibration of the model. Another limitation is the consideration only of the main part of the sewer system, due to the lack of reliable data for the rest of the system. If the whole data is available and a bigger part of the sewer system is simulated, more possible entrance points can be considered and a part of them can be used for validation of the optimization's results. Also, the travel time of the viral load will be larger and this will affect the viral concentration at nodes. In the present study, 16 nodes are considered for the discharge of the initial concentration and 4 nodes for validation. Additionally, only dry weather conditions are considered in the present study. Under wet weather conditions, dilution effect may be more intense and can cause reduction of SARS-CoV-2 RNA concentration at nodes. On the other hand increase flow rates may cause the detachment of viral particles from suspended solids or biofilm and its resuspension in the wastewater phase (Maere et al, 2021).

Despite the limitations, the proposed methodology was implemented successfully and yielded acceptable results. The lack of data for calibration is alleviated by comparison with other studies and the similar order of magnitude results for viral concentration at the entrance to the WWTP are noticed. In the scenario with initial concentration equal to 3×10^4 GC/L and decay rate equal to 3.4 day^{-1}, maximum concentration values at the entrance of the WWTP are found between 1,300–2,000 GC/L (capacity of the WWTP: 16,500 m^3/day). The simulated data of the study are in the low side of the reported data from other studies which are in the range of 750–34,000 GC/L (capacity 5,200–241,000 m^3/day) (Hasan et al., 2021), 100–100,000 GC/L (capacity 57,000–200,000 m^3/day) (Gonzalez et al., 2020), 12-22,000 GC/L (Medema et al., 2020) and 3,000–20,000 GC/L (capacity 13,400–1,100,000

m^3/day) (Westhaus et al., 2021). The WWTP of the Kozani city is a small plant and has lower capacity than other reported WWTPs. The initial concentration is considered to enter the sewer system from one node at a time and the decay rate is assumed equal to the highest from the reported values. Thus, the low values of concentration found herein are reasonable and comparable with other studies. Furthermore, the results in terms of the optimal number of monitoring points are in accordance to other studies as already have been mentioned and validation results are satisfactory.

4. Conclusions

The present study proposes an optimization methodology for selecting the optimum sampling points for wastewater surveillance for SARS-CoV-2. The proposed methodology comprises a four-step procedure; hydraulic simulation - optimization - decision making - validation. In similar studies, the selection of the optimum sapling points for wastewater based epidemiology is based on graphical representation of the collection system. In the present study, hydraulic simulation of the sewer system with fate and transport simulation of viral particles are implemented. The simulation resulted viral concentrations are used as the input to calculate the optimization parameters. Subsequently, a decision making step is applied as a supportive tool for selecting the most suitable set of sampling points among the pareto-optimal solutions.

Summarizing, the main conclusions of the study are the following:

- It was determined that 4-7 sampling points, well distributed across the collection system, are sufficient for monitoring the SARS-CoV-2 in wastewater. In addition, it was found that more sampling points do not improve the monitoring efficiency.
- 4 sampling points, including the entrance to the WWTP, are proposed herein as a well-balanced solution. The proposed solution is based on the results from the optimization methodology and other criteria as well, such as time, cost and convenience.
- The entrance to WWTP as the only sampling point, show higher possibility of failing to detect the viral traces. This fact is based on the in-sewer reduction of the viral particles concentration, that may render them undetectable at the end of the collection system.

- The two optimization concepts end up in non-dominated sets that have many solutions in common.
- For optimization O1 with the two objectives, all the high ranked solutions from decision making tools also achieve the observation of all the population. This means that JH and TC may be sufficient objectives and no more objectives, that increase computation complexity, need to be included.
- It is important to mention that the user's preferences are significantly affecting the results of the decision making methods. For the application of the SAW method, the normalization methodology and the appointed weightage are the critical parameters. Whereas for the PW method, is the desirable place (or desirable pseudo-weights) of the solution across the pareto front. In the present study, normalization of the values using the best values of each objective and equal weights for the SAW method concludes to the same best solutions with PW, with desirable placement the one that favors JH.

Finally, there is the need for more data for better calibration and validation of the model. Initial concentration of virus depends on number of infected persons and the virus's excretion rate per person. The data about these parameters is sparce and with high uncertainty. Similarly, sparse is the data about the fate and transport of the viral particles across the sewer system. More data of the geometry and flow parameters of the sewer system is also necessary, but unfortunately is not always available in Greece, especially in smaller cities. The aforementioned limitations constitute the basis for future work on the research area of virus fate and transport in sewer systems. The effect of flow regimes, temperature and adsorption of viral particles on solids, organic compounds and biofilms are aspects that should be studied in depth.

AUTHOR CONTRIBUTIONS:

Conceptualization: Argyro Gkatzioura and Antigoni Zafeirakou; Methodology: Argyro Gkatzioura; Software: Argyro Gkatzioura; Validation: Argyro Gkatzioura; Formal analysis: Argyro Gkatzioura;

Investigation: Argyro Gkatzioura; Resources: Argyro Gkatzioura and Antigoni Zafeirakou; Data curation: Argyro Gkatzioura; Writing-original draft preparation: Argyro Gkatzioura; Writing-review and editing: Argyro Gkatzioura and Antigoni Zafeirakou; Visualization: Argyro Gkatzioura; Supervision: Antigoni Zafeirakou; Project administration: Argyro Gkatzioura and Antigoni Zafeirakou.;

REFERENCES

Ahmed, W., Bertsch, P. M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., … Bivins, A. (2020). Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environmental Research*, *191*(August), 110092. https://doi.org/10.1016/j.envres.2020.110092

Ahmed, W., Bivins, A., Stephens, M., Metcalfe, S., Smith, W. J. M., Sirikanchana, K., … Simpson, S. L. (2023). Occurrence of multiple respiratory viruses in wastewater in Queensland, Australia: Potential for community disease surveillance. *Science of the Total Environment*, *864*(November 2022), 161023. https://doi.org/10.1016/j.scitotenv.2022.161023

Alfonso, L., Ridolfi, E., Gaytan-Aguilar, S., Napolitano, F., & Russo, F. (2014). Ensemble entropy for monitoring network design. *Entropy*, *16*(3), 1365–1375. https://doi.org/10.3390/e16031365

Bakanos, P. I., & Katsifarakis, K. L. (2019). Optimizing operation of a large-scale pumped storage hydropower system coordinated with wind farm by means of genetic algorithms. *Global Nest Journal*, *21*(4), 495–504. https://doi.org/10.30955/gnj.002978

Baldovin, T., Amoruso, I., Fonzo, M., Buja, A., Baldo, V., Cocchio, S., & Bertoncello, C. (2021). SARS-CoV-2 RNA detection and persistence in wastewater samples: An experimental network for COVID-19 environmental surveillance in Padua, Veneto Region (NE Italy). *Science of the Total Environment*, *760*, 143329. https://doi.org/10.1016/j.scitotenv.2020.143329

Banik, B. K., Ph, D., Alfonso, L., Ph, D., Cristo, C. Di, Ph, D., … Sc, D. (2017). Evaluation of

Different Formulations to Optimally Locate Sensors in Sewer Systems. *J. Water Resour. Plann. Manage*, *143*(7), 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000778.

Bivins, A., Greaves, J., Fischer, R., Yinda, K. C., Ahmed, W., Kitajima, M., … Bibby, K. (2020). Persistence of SARS-CoV-2 in Water and Wastewater. *Environmental Science and Technology Letters*, *7*(12), 937–942. https://doi.org/10.1021/acs.estlett.0c00730

Blank, J., & Deb, K. (2020). Pymoo: Multi-Objective Optimization in Python. *IEEE Access*, *8*, 89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567

Brentan, B., Carpitella, S., Barros, D., Meirelles, G., Certa, A., & Izquierdo, J. (2021). Water Quality Sensor Placement: A Multi-Objective and Multi-Criteria Approach. *Water Resources Management*, *35*(1), 225–241. https://doi.org/10.1007/s11269-020-02720-3

Calle, E., Martínez, D., Brugués-i-Pujolràs, R., Farreras, M., Saló-Grau, J., Pueyo-Ros, J., & Corominas, L. (2021). Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks. *Environment International*, *157*. https://doi.org/10.1016/j.envint.2021.106768

Carine, M. R., & Pagilla, K. R. (2024). A mass balance approach for quantifying the role of natural decay and fate mechanisms on SARS-CoV-2 genetic marker removal during water reclamation. *Water Environment Research*, *96*(4), 1–17. https://doi.org/10.1002/wer.11015

Chen, J., Dai, Z., Dong, S., Zhang, X., Sun, G., Wu, J., … Soltanian, M. R. (2022). Integration of Deep Learning and Information Theory for Designing Monitoring Networks in Heterogeneous Aquifer Systems. *Water Resources Research*, *58*(10). https://doi.org/10.1029/2022WR032429

D'Aoust, P. M., Graber, T. E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., … Delatolla, R. (2021). Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Science of the Total Environment*, *770*. https://doi.org/10.1016/j.scitotenv.2021.145319

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. https://doi.org/10.1109/4235.996017

Domokos, E., Sebestyén, V., Somogyi, V., Trájer, A. J., Gerencsér-Berta, R., Horváth, B. O., … Abonyi, J. (2021). Identification of sampling points for the detection of SARS-CoV-2 in the sewage system. *Sustainable Cities and Society*, *76*(June 2021), 103422. https://doi.org/10.1016/j.scs.2021.103422

Fahrenfeld, N. L., Morales Medina, W. R., D'Elia, S., Deshpande, A. S., & Ehasz, G. (2022). Year-long wastewater monitoring for SARS-CoV-2 signals in combined and separate sanitary sewers. *Water Environment Research*, *94*(8), 1–13. https://doi.org/10.1002/wer.10768

Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., … Thomaidis, N. S. (2022). SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Science of the Total Environment*, *804*, 150151. https://doi.org/10.1016/j.scitotenv.2021.150151

Gkatzioura, A., & Zafeirakou, A. (2023). Optimal Selection of Sampling Points for Detecting SARS-CoV-2 RNA in Sewer System Using NSGA-II Algorithm. *Water (Switzerland)*, *15*(23). https://doi.org/10.3390/w15234076

Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M. H., Yetka, K., … Gonzalez, D. (2020). COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Research*, *186*, 116296. https://doi.org/10.1016/j.watres.2020.116296

Gundy, P. M., Gerba, C. P., & Pepper, I. L. (2009). Survival of Coronaviruses in Water and Wastewater. *Food and Environmental Virology*, *1*(1), 10–14. https://doi.org/10.1007/s12560-008-9001-6

Hasan, S. W., Ibrahim, Y., Daou, M., Kannout, H., Jan, N., Lopes, A., … Yousef, A. F. (2021). Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: Surveillance of COVID-19 epidemic in the United Arab Emirates. *Science of the Total Environment*, *764*, 142929. https://doi.org/10.1016/j.scitotenv.2020.142929

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., … Norder, H. (2014). Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Applied and Environmental Microbiology*, *80*(21), 6771–6781. https://doi.org/10.1128/AEM.01981-14

Hernández-Terrones, L. M., Carrillo-Reyes, J., Ayala-Godoy, J. A., Guerrero-Ruiz, E., Vargas, L. M. G., Prado-Guevara, B. A., … Buitrón, G. (2023). Monitoring of SARS-CoV-2 RNA in wastewater: A surveillance tool to foresee infection's evolution in the Mexican Caribbean. *Water Environment Research*, *95*(4), 1–9. https://doi.org/10.1002/wer.10859

Hokajärvi, A. M., Rytkönen, A., Tiwari, A., Kauppinen, A., Oikarinen, S., Lehto, K. M., … Pitkänen, T. (2021). The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland. *Science of the Total Environment*, *770*, 145274. https://doi.org/10.1016/j.scitotenv.2021.145274

Jafari, H., Nazif, S., & Rajaee, T. (2022). A multi-objective optimization method based on NSGA-III for water quality sensor placement with the aim of reducing potential contamination of important nodes. *Water Supply*, *22*(1), 928–944. https://doi.org/10.2166/ws.2021.222

Kilaru, P., Hill, D., Anderson, K., Collins, M. B., Green, H., Kmush, B. L., & Larsen, D. A. (2023). Wastewater Surveillance for Infectious Disease: A Systematic Review. *American Journal of Epidemiology*, *192*(2), 305–322. https://doi.org/10.1093/aje/kwac175

Kim, K., Ban, M. J., Kim, S., Park, M. H., Stenstrom, M. K., & Kang, J. H. (2022). Optimal allocation and operation of sewer monitoring sites for wastewater-based disease surveillance: A methodological proposal. *Journal of Environmental Management*, *320*(May), 115806. https://doi.org/10.1016/j.jenvman.2022.115806

Kostoglou, M., Petala, M., Karapantsios, T., Dovas, C., Roilides, E., Metallidis, S., … Papaioannou, N. (2021). SARS-CoV-2 adsorption on suspended solids along a sewerage network: mathematical model formulation, sensitivity analysis, and parametric study. *Environmental Science and Pollution Research*, *29*, 11304–11319.

https://doi.org/10.1007/s11356-021-16528-0

Koureas, M., Amoutzias, G. D., Vontas, A., Kyritsi, M., Pinaka, O., Papakonstantinou, A., … Hadjichristodoulou, C. (2021). Wastewater monitoring as a supplementary surveillance tool for capturing SARS-COV-2 community spread. A case study in two Greek municipalities. *Environmental Research*, *200*(July). https://doi.org/10.1016/j.envres.2021.111749

Lago, P. M., Gary, H. E., Pérez, L. S., Cáceres, V., Olivera, J. B., Puentes, R. P., … Cruz, R. G. (2003). Poliovirus detection in wastewater and stools following an immunization campaign in Havana, Cuba. *International Journal of Epidemiology*, *32*(5), 772–777. https://doi.org/10.1093/ije/dyg185

Larson, R. C., Berman, O., & Nourinejad, M. (2020). Sampling manholes to home in on SARS-CoV-2 infections. *PLoS ONE*, *15*(10 October), 1–20. https://doi.org/10.1371/journal.pone.0240007

Lee, J. H. (2013). Determination of optimal water quality monitoring points in sewer systems using entropy theory. *Entropy*, *15*(9), 3419–3434. https://doi.org/10.3390/e15093419

Li, X., Zhang, S., Shi, J., Luby, S. P., & Jiang, G. (2021). Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. *Chemical Engineering Journal*, *415*(129039). https://doi.org/10.1016/j.cej.2021.129039

Maal-Bared, R., Brisolara, K., Munakata, N., Bibby, K., Gerba, C., Sobsey, M., … Rubin, A. (2021). Implications of SARS-CoV-2 on current and future operation and management of wastewater systems. *Water Environment Research*, *93*(4), 502–515. https://doi.org/10.1002/wer.1446

Maere, T., Haboub, K., & Vanrolleghem, P. A. (2021). Understanding the fate of SARS-CoV-2 in sewer systems: development of a mechanistic model framework. *15th International Conference on Urban Drainage*.

McDonnell, B., Ratliff, K., Tryby, M., Wu, J., & Mullapudi, A. (2020). PySWMM: The Python Interface to Stormwater Management Model (SWMM). *Journal of Open Source Software*, *5*(52), 2292. https://doi.org/10.21105/joss.02292

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in the Netherlands. *Environmental Science and Technology Letters*, *7*(7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Panagoulia, D., Tsekouras, G. J., & Kousiouris, G. (2017). A multi-stage methodology for selecting input variables in ANN forecasting of river flows. *Global Nest Journal*, *19*(1), 49–57. https://doi.org/10.30955/gnj.002067

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*, *181*. https://doi.org/10.1016/j.watres.2020.115942

Rossman, L. A. (2015). *Storm Water Management Model User's Manual Version 5.1.* Cincinnaty, OH: U.S. Environmental Protection Agency (EPA).

Shannon, E. C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *5*(1), 957–1000. https://doi.org/10.1002/j.1538-7305.1968.tb00069.x

Wang, X. W., Li, J., Guo, T., Zhen, B., Kong, Q., Yi, B., … Li, J. (2005). Concentration and detection of SARS coronavirus in sewage from Xiao Tang Shan hospital and the 309th Hospital of the Chinese People's Liberation Army. *Water Science and Technology*, *52*(8), 213–221. https://doi.org/10.2166/wst.2005.0266

Wang, Y., Moe, C. L., Dutta, S., Wadhwa, A., Kanungo, S., Mairinger, W., … Teunis, P. F. (2020). Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation. *Epidemics*, *31*, 100391. https://doi.org/10.1016/j.epidem.2020.100391

Wang, Z., & Rangaiah, G. P. (2017). Application and Analysis of Methods for Selecting an Optimal Solution from the Pareto-Optimal Front obtained by Multiobjective Optimization. *Industrial and Engineering Chemistry Research*, *56*(2), 560–574. https://doi.org/10.1021/acs.iecr.6b03453

Weidhaas, J., Aanderud, Z. T., Roper, D. K., VanDerslice, J., Gaddis, E. B., Ostermiller, J., … LaCross, N. (2021). Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of The Total Environment*, *775*, 145790. https://doi.org/10.1016/j.scitotenv.2021.145790

Westhaus, S., Weber, F. A., Schiwy, S., Linnemann, V., Brinkmann, M., Widera, M., … Ciesek, S. (2021). Detection of SARS-CoV-2 in raw and treated wastewater in Germany – Suitability for COVID-19 surveillance and potential transmission risks. *Science of the Total Environment*, *751*, 141750. https://doi.org/10.1016/j.scitotenv.2020.141750

WHO. (2023). *Environmental surveillance for SARS-CoV-2 to complement other public health surveillance*. Geneva. Retrieved from https://www.who.int/publications/i/item/9789240080638

Yang, S., Dong, Q., Li, S., Cheng, Z., Kang, X., Ren, D., … Qu, J. (2022). Persistence of SARS-CoV-2 RNA in wastewater after the end of the COVID-19 epidemics. *Journal of Hazardous Materials*, *429*(January), 128358. https://doi.org/10.1016/j.jhazmat.2022.128358

Yao, Y., Zhu, Y., Nogueira, R., Klawonn, F., & Wallner, M. (2024). Optimal Selection of Sampling Points within Sewer Networks for Wastewater-Based Epidemiology Applications. *Methods and Protocols*, *7*(6), 1–18.

Zionts, S., & Wallenius, J. (1983). An Interactive Multiple Objective Linear Programming Method for a Class of Underlying Nonlinear Utility Functions. *Management Science*, *29*(5), 519–529. Retrieved from https://doi.org/10.1287/mnsc.29.5.519