# UCLA

## UCLA Previously Published Works

**Title**

Optimal allocation and operation of sewer monitoring sites for wastewater-based disease surveillance: A methodological proposal

**Permalink**

**Authors**

Kim, Keugtae
Ban, Min Jeong
Kim, Sungpyo
et al.

**Publication Date**

**DOI**

Peer reviewed

Research article

# Optimal allocation and operation of sewer monitoring sites for wastewater-based disease surveillance: A methodological proposal

Keugtae Kim [a], Min Jeong Ban [b], Sungpyo Kim [c], Mi-Hyun Park [d], Michael K. Stenstrom [e], Joo-Hyon Kang [b,*]

[a] Department of Environmental and Energy Engineering, The University of Suwon, 17 Wauan-gil, Bongdam-eup, Hwaseong-si, Gyeonggi-do, 18323, Republic of Korea
[b] Department of Civil and Environmental Engineering, Dongguk University-Seoul, 30, Pildong-ro 1gil, Jung-gu, Seoul, 04620, Republic of Korea
[c] Department of Environmental Engineering, Korea University-Sejong, 2 511, Sejong-ro, Sejong City, 30019, Republic of Korea
[d] Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
[e] Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, 90096, USA

A B S T R A C T

Wastewater-based epidemiology (WBE) is drawing increasing attention as a promising tool for an early warning of emerging infectious diseases such as COVID-19. This study demonstrated the utility of a spatial bisection method (SBM) and a global optimization algorithm (i.e., genetic algorithm, GA), to support better designing and operating a WBE program for disease surveillance and source identification. The performances of SBM and GA were compared in determining the optimal locations of sewer monitoring manholes to minimize the difference among the effective spatial monitoring scales of the selected manholes. While GA was more flexible in determining the spatial resolution of the monitoring areas, SBM allows stepwise selection of optimal sampling manholes with equiareal subcatchments and lowers computational cost. Upon detecting disease outbreaks at a regular sewer monitoring site, additional manholes within the catchment can be selected and monitored to identify source areas with a required spatial resolution. SBM offered an efficient method for rapidly searching for the optimal locations of additional sampling manholes to identify the source areas. This study provides strategic and technical elements of WBE including sampling site selection with required spatial resolution and a source identification method.

## 1. Introduction

In the 21st century, the world has been increasingly under threats posed by emerging infectious diseases such as severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and the coronavirus disease 2019 (COVID-19). In particular, the ongoing global pandemic of COVID-19 declared in March 2020 has caused more than 264 million cases with more than 5.23 million confirmed deaths as of December 2021 (World Health Organization, 2003). The COVID-19 pandemic also has a significant impact on the global economy, causing a decrease of 3.5% in the global gross domestic product (GDP) in 2020 (World Bank Group, 2021).

To avoid disastrous impacts of infectious diseases, early detection of emerging disease outbreaks in communities is critical as it enables rapid responses such as an early warning to the public and proactive

transmission control activities. Accordingly, early detection of disease outbreaks is becoming increasingly important in recent decades due to increased frequencies of emerging and re-emerging infectious diseases including novel coronaviruses, tuberculosis, cholera, and poliovirus (Hui, 2006; Jones et al., 2008; World Health Organization, 2018). Response measures to infectious diseases generally rely on clinical surveillance oriented to symptomatic patients, which might be ineffective for early detection of the spread of infectious diseases and revealing its true scale including subclinical infection, thus prohibiting preemptive quarantine activities.

Environmental surveillance has been suggested as an alternative to clinical surveillance for monitoring the spread of infectious diseases (Brouwer et al., 2018; Andrews et al., 2020; Medema et al., 2020). In environmental surveillance, samples are collected from environmental matrices involved in pathogen transmissions such as drinking water,

sewage, air, and other media and tested for evidence of microbial/viral contamination (Asghar et al., 2014; Andrews et al., 2020). Proactive environmental surveillance can detect infection earlier than the recording of clinical cases. As such, it can provide an early warning of infectious disease outbreaks (Moran-Gilad et al., 2016). Alternatively, environmental surveillance can be reactively performed to assess the magnitude of infection after the identification of outbreaks (Hellmér et al., 2014; Brouwer et al., 2018; Ahmed et al., 2020).

Environmental surveillance that systematically collects and analyzes sewage samples for pathogenic reagents or chemical compounds, called wastewater-based epidemiology (WBE), is recently drawing significant attention due to the global pandemic of COVID-19 (Sims and Kasprzyk-Hordern, 2020). WBE is considered an effective approach to monitoring community health and risk for infectious disease because sewage contains pathogens or causative agents (e.g. viral RNA of SARS-CoV-2 for COVID-19) of diseases excreted from both symptomatic and asymptomatic individuals in a catchment (Kim et al., 2015; Ahmed et al., 2020). Especially, WBE is regarded as a promising tool for the early warning of disease outbreaks and for assessing the efficacy of public health interventions. Previous studies have demonstrated the detection of influenza and enteric viruses such as influenza A virus, norovirus, hepatitis A virus, and poliovirus (Heijnen and Medema, 2011; Hellmér et al., 2014; Asghar et al., 2014). More recently, the global pandemic of COVID-19 has motivated active WBE applications for monitoring SARS-CoV-2 at different spatial settings across the world (Table S1).

Despite the potential utility of WBE revealed by previous studies, strategic and technical elements for implementing a WBE program merit further research, including sampling sites, sample storage and transport, quantification methods, and data interpretation and its uncertainty management (Barceló, 2020; Medema et al., 2020; Yao et al., 2021; Wade et al., 2022). In addition, sufficient spatial and temporal resolution of a WEB program is required for timely and effective detection and warning of disease outbreaks (World Health Organization, 2003). Typical sampling sites are inlet points of wastewater treatment plants (WWTP) or sewer manholes. However, sampling sites need to be determined based on the size of the source population to be monitored (i.e., spatial resolution), high-risk areas, and costs including laboratory workloads.

The spatial resolution of a WBE program can vary depending on the program purpose and target analytes. World Health Organization (2003) has recommended a size of 100,000–300,000 source population for regular monitoring of poliovirus circulation in a community. However, the spatial resolution of existing WBE studies varies including a building level (150–200 populations) (Betancourt et al., 2021; Davó et al., 2021; Gibas et al., 2021; Harris-Lovett et al., 2021), a zip code level (Barrios et al., 2021), and a catchment level (up to 1,500,000 populations) (Hellmér et al., 2014; Fumian et al., 2019; Ahmed et al., 2020) (Table S1). Too low spatial resolution would fail to collect representative samples (Medema et al., 2020), whereas too high spatial resolution increases the dilution of target analytes (Foladori et al., 2020), presumably decreasing the sensitivity of detection.

A few studies have recently proposed optimal site selection algorithms for the sewer monitoring of an infectious disease or hotspot identification at a catchment level. Wang et al. (2020) developed a simulation model for optimal sewer monitoring site allocation to maximize the detection sensitivity of Salmonella Typhoid in a catchment considering shedding, loss, decay, and transport of the pathogen, demonstrating the proposed method in a hypothetical sewer network. Larson et al. (2020) proposed a Bayesian probability-based, binary searching algorithm for identifying the zero-patient location and the most infected spot neighborhood of SARS-CoV-2 using hypothetical hotspots and link probability flows in a sewer network of a small city. Calle et al. (2021) expanded the work of Larson et al. (2020) by defining the probability flows by the number of inhabitants, and proposed site searching methods for regular sewer monitoring and hotspot

identification of SARS-CoV-2 based on the graph theory coupled with a greedy optimization algorithm; their proposed site selection method, however, may not completely cover the entire catchment area depending on the number of monitoring sites and site-specific sewer network structure.

In this study, the concept of an effective monitoring area (EMA; i.e., the area exclusively monitored by a designated monitoring manhole) was applied to determine the spatial monitoring scale and the optimal sewer monitoring site allocation to support a WBE program. Based on the concept of EMA, two methods were proposed for optimal selection of sewer monitoring manholes for infectious disease surveillance that covers the entire catchment area—a spatial bisection method (SBM) and a global optimization algorithm (i.e., genetic algorithm, GA). The performances of the two methods with a real-scale sewer network of a large catchment were illustrated and compared. Furthermore, SBM was proposed as a follow-up practice for scaling down the infected area upon detecting disease outbreaks at a regular monitoring manhole, and its performance in a real catchment was demonstrated. The SBM-based site searching demonstration is an extension of the works of Larson et al. (2020) and Calle et al. (2021) by providing a computational method for selecting the best bisecting sites. That is, to realize SBM for site selection in this study, a recursive calculation algorithm with a binary matrix of manhole topology and local service areas of individual manholes of the sewer network was used for computing EMA (or spatial monitoring scale).

## 2. Methods

### 2.1. Study area and data preparation

The study catchment was a sewershed section (Seoho sewershed) in Suwon City, Republic of Korea. The catchment has a total building area of 2,705,586 $m^2$, serving a total population of 220,993. The catchment has a combined sewer system with a total of 5,424 manholes, interconnecting storm-drain pipes and sewer pipes with many closed-loops. A public WWTP with a treatment capacity of 47,000 $m^2$/d is located at the outlet of the catchment.

Spatial data of sewer pipes and manholes (2020) were obtained in ESRI Shapefile format from Suwon City. A land use feature layer (2018) was obtained from the Ministry of the Environment, Korea. We modified the existing sewer structure such that features of pipe networks that only carry stormwater and any closed-loops that are uninvolved in the dry weather sewer flow were removed; In this way, only gravity-induced unidirectional flow of sewage water during dry days was considered to characterize the sewer network structure for WBE. Features for the buildings from the land use layer were extracted and corresponding local service areas for individual manholes were calculated. All area calculations were based on the footprint area of buildings. All spatial processing and mapping were performed using Quantum GIS 3.10 (QGIS development team).

The numbers of daily confirmed cases of COVID-19 in the study area were obtained from Suwon City (https://www.suwon.go.kr/web/safesuwon/corona/PD_monitor.do, accessed on Sep. 1, 2021). The past data on COVID-19 prevalence from the date of the first case report to recent time (Feb. 2, 2020–Aug. 31, 2021) allow us to develop a strategy in a retrospective way for operating a WBE program. To demonstrate the searching algorithm for the source area of infection, a nursing home in the study area that has experienced multiple outbreak events since the beginning of the COVID-19 pandemic was considered a known hotspot to be identified through WBE.

### 2.2. Calculation of the effective monitoring area

A single monitoring manhole can be operated at the outlet point of a catchment, or inlet point of a public WWTP to monitor the entire catchment area. However multiple monitoring manholes may be
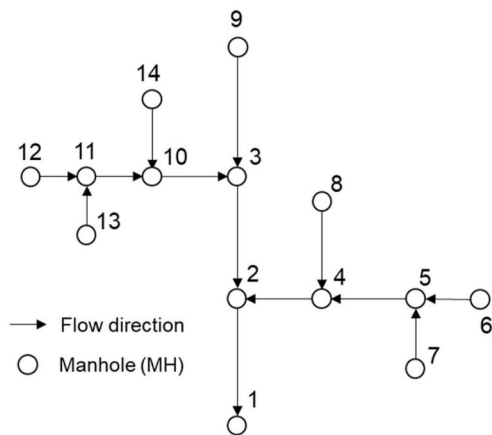
required in a large catchment depending on the maximum possible spatial resolution (monitoring area or population) associated with the detection sensitivity of the target analytes. In this case, it is ideal to evenly distribute the monitoring area or population served by each monitoring manhole. This is an optimization problem for determining the best locations of the monitoring manholes as a different set of monitoring manholes would result in different monitoring areas (or populations) for each manhole depending on the structure of the sewer pipe network.

Calculating the monitoring area of a manhole requires information on manhole connectivity and the local service area (LSA) of each manhole—an area that drains into a manhole via no other upstream manholes. Fig. 1a shows a simple sewer network example, illustrating the topology of the manholes. An ideal sewer network has a dendric structure without a closed-loop so that sewage can flow through a pipe unidirectionally, thus uniquely defining the catchment area served by each

manhole. It encompasses the entire upstream area that drains into the manhole (called "cumulative service area (CSA)" hereafter, and illustrated in Fig. 1b). The connectivity information between adjacent pairs of the upstream and downstream manholes is stored in a squared binary matrix ("topology matrix of the manholes") whose elements are 1's in case of direct connection and 0's otherwise. Based on the topology matrix of manholes, the CSA of each manhole is computed by searching for connected manholes in the upstream direction and their LSAs using a recursive calculation in the upstream direction as follows (see Fig. 1b for illustration):

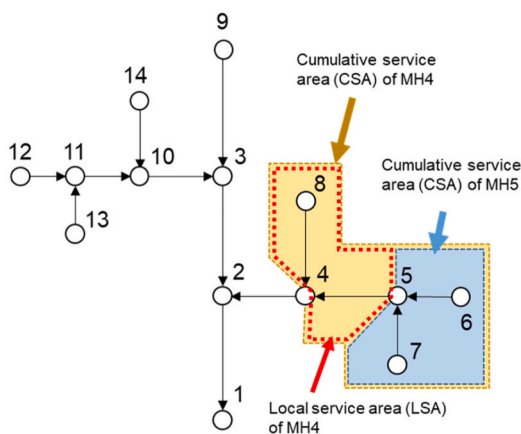$$CSA_i = LSA_i + \sum_{i^+=1}^{n_{i^+}} CSA_{i^+} \qquad (1)$$

where $CSA_i$ and $LSA_i$ are CSA and LSA of the manhole $i$, respectively, $CSA_{i^+}$ is CSA of the adjacent upstream manhole at the $i^+$ th upstream



(a)



CSA of MH2 = LSA of MH4 + CSA of MH5

(b)
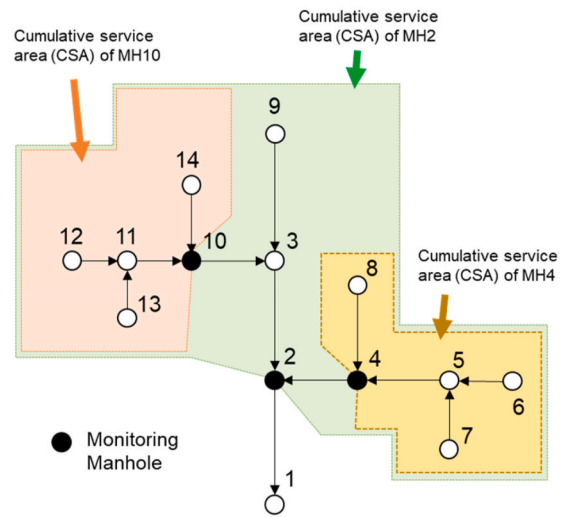


EMA of MH2 = CSA of MH2 − (CSA of MH4 + CSA of MH10)

(c)

**Fig. 1.** Concepts of manhole topology database, the cumulative service area, and the effective monitoring area for a hypothetical sewer network: (a) Construction of a binary matrix representing the manhole topology. Direct connection and no direct connection (or no connection) between two manholes are indicated by 1 and 0, respectively; (b) Illustration of the cumulative service area (CSA); (c) illustration of the effective monitoring area (EMA).

branch of the manhole $i$, and $n_{i+}$ is the total number of the adjacent manholes in the upstream branches of the manhole $i$.

Given the information on CSAs of the monitoring manholes, the effective monitoring area (EMA) of a monitoring manhole is defined as an area monitored by that manhole excluding the CSAs of the nearest monitoring manholes at its upstream branches as follows (see Fig. 1c for illustration):

$$EMA_j = CSA_j - \sum_{j^+=1}^{n_{j^+}} CSA_{j^+} \qquad (2)$$

where $EMA_j$ and $CSA_j$ are EMA and CSA of the monitoring manhole, $j$, respectively, $CSA_{j^+}$ is CSA of the nearest upstream monitoring manhole at the $j^+$ th upstream branch of the monitoring manhole $j$, $n_{j^+}$ is total number of the upstream branches of the monitoring manhole $j$ that have upstream monitoring manholes.

### 2.3. Selection of the best monitoring manhole locations using a global optimization algorithm

The best strategy for locating a given number of monitoring manholes across the catchment is to assign equal EMAs to all monitoring manholes, and the approximate solution is achieved by minimizing the maximum value among the EMAs of the manholes selected as the monitoring stations. Therefore, determining the optimal locations of a given number of the monitoring manholes can be formulated as an integer optimization problem as follows:

$$minimize : \max_{\substack{j \in U \\ U \subset S, \ n(U)=N}} (EMA_j) \qquad (3)$$

where S is a set of all monitoring manholes in the catchment as elements, U is a subset of S with N number of required monitoring manholes. N is determined by dividing the entire catchment area (or total population) by the ideal EMA for each monitoring manhole (or the maximum possible monitoring area).

Although any global optimization algorithm can be used to solve the optimization problem in Eq. (3), a genetic algorithm (GA) was applied to compare its performance with that of SBM. A GA is a heuristic search reflecting the natural selection process such as mutation, crossover, and selection (Mitchell, 1996). GAs have been used to solve nonlinear optimization problems in many environmental applications including pipe network optimization (Haghighi and Bakhshipour, 2012; Gupta et al., 1999), optimization of pollutant control strategies (Srivastava et al., 2002; Arabi et al., 2006) and parameter estimation of environmental models (Fazal et al., 2005; Massoudieh et al., 2008).

Different numbers of monitoring manholes ranging from 2 to 16 were selected using the GA. For each case, the population size (i.e., number of candidate solutions) of the GA was 200, where each individual contained a vector of input variables (chromosome) consisting of integer values (genes) that represent selected monitoring manholes. For each generation, 5% elite individuals were selected and subject to crossover (with 0.8 of crossover fraction) and mutation to produce the next generation. The GA was terminated when the average change in the penalty fitness function value (Kalyanmoy, 2000) became less than $10^{-6}$ or the number of generations reached 100 times the number of input variables (i.e., the number of monitoring manholes). Computer codes for implementing the GA and SBM were written in MATLAB R2020b (Mathworks, Massachusetts, USA).

### 2.4. Stepwise selection of sewer monitoring manholes using spatial bisection

As an alternative to the global optimization algorithm for monitoring manhole placement, we employed a bisection method that can be applied to spatial pipe network data. The bisection method is a mathematical approximation method to find the root of any continuous

function by repeatedly bisecting the interval of independent variables and then selecting the subinterval that contains the root (Burden and Faires, 1985). We adopted the core idea of the mathematical bisection for the spatial data of the sewer network, called SBM. SBM repeatedly bisects the catchment until a spatial resolution required by a WBE program is obtained or the infected area is found, similar to the root searching procedure in the mathematical bisection method.

To provide a spatial resolution required for regular sewage monitoring, a spatial bisection-based stepwise selection strategy for monitoring manholes is proposed as shown in Fig. 2. In a given catchment with an outlet monitoring manhole, the addition of one monitoring manhole at any location inside the catchment results in two subcatchments (the added manhole is called "inside monitoring manhole" hereafter): an upstream one that drains to the inside monitoring manhole and a downstream one that drains to the outlet monitoring manhole without passing through the inside monitoring manhole (Fig. 2a). These two subcatchment areas correspond to EMAs of the inside monitoring manhole and the outlet monitoring manhole ($EMA_{up}$ and $EMA_{down}$). The location of the inside monitoring is determined to minimize the difference between EMAs of the two subcatchments, thus providing a consistent spatial resolution of a WBE program. This optimal selection of the monitoring manhole is achieved using the Brute Forth approach to scan through all inside manholes within a given catchment (Fig. 2b). The manhole selection procedure is repeated for each subcatchment resulting from the previous manhole selection step until the required number of monitoring manholes or the spatial resolution of the monitoring is obtained (Fig. 2b). The spatial resolution is estimated as follows:

$$A_{des} = \frac{A_{tot}}{N} = \frac{A_{tot}}{2^p} \qquad (4)$$

where $A_{des}$ and $A_{tot}$ are the spatial resolution of the WBE (or monitoring population) and total catchment area (or total population), respectively, and $p$ is the number of spatial bisections. Note that in this study, the quantification of the spatial resolutions including $A_{des}$, $A_{tot}$, CAS, LSA, and EMA is based on the total footprint area of buildings, although different spatial units such as populations can be used if available; If the population census data with sufficiently high spatial resolution (such as building level) are available, EMA can be calculated based on the population rather than the building footprint area. For example, areas and buildings with large population density or visitors such as nursing homes and community centers might be given priority when selecting monitoring manholes. Designers can choose different numbers of monitoring manholes depending on target analytes, their quantification sensitivity, and cost.

SBM was applied to select different numbers of the monitoring manholes in the study area ranging from 2 to $2^4$. For the case of 4 and 8 monitoring manholes, the percent difference between the maximum and minimum EMAs among the selected manholes was compared with the those of a global optimization algorithm (genetic algorithm, GA) to evaluate the relative performance of SBM in selecting the optimal places of the monitoring manholes.

### 2.5. Identification of infected areas

When regular monitoring at multiple manholes is infeasible in many cases due to limited budget, a single monitoring location typically at the outlet point of the entire catchment (e.g., inlet point of a WWTP) might be a more practical situation. In this case, SBM or GA can be used to further identify the source areas of an infectious disease in the catchment upon detecting disease outbreaks at the outlet point. That is, additional samplings can be simultaneously conducted at the inside manholes selected by SBM or GA to identify hotspots upon detection at the outlet of the entire catchment; a stepwise binary search (Larson et al., 2020) may be impractical due to the time requirements for the
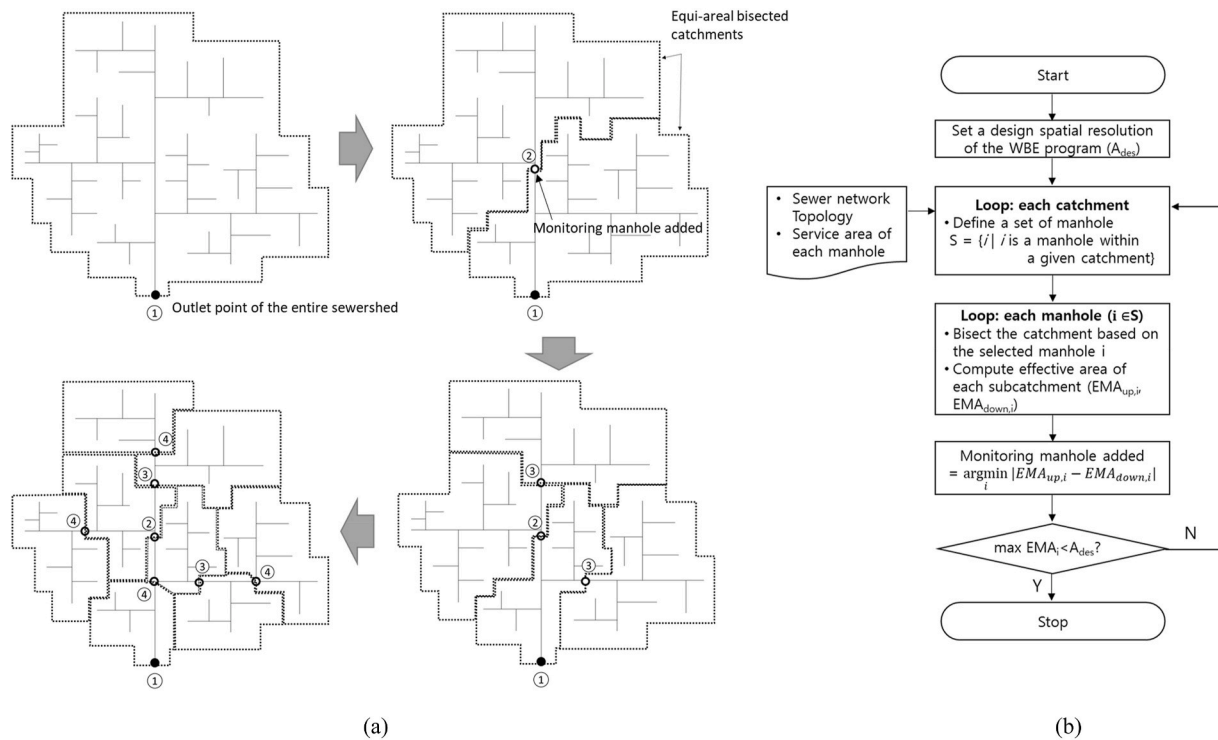
**Fig. 2.** Selection procedure of monitoring manholes based on the spatial bisection method: (a) Conceptual example; (b) Flow chart.

sampling and measurement. The number of inside monitoring manholes (and spatial resolution of the monitoring) for identifying the infected areas depends on the program budget and resources. The infected catchments can be identified based on the measurement outcome at the inside monitoring manholes and selected for further scaling down the infected area by adding more samplings in the upstream subcatchments; here we presumed that the level of infection in a catchment can be estimated by the pathogen/virus detection intensity at the outlet (Medema et al., 2020; McMahan et al., 2021). In this study, a nursing home in the study area was considered a known source location to demonstrate the performance of SBM in identifying the source area of an infectious disease.

## 3. Results and discussion

### 3.1. Optimal manhole selection using GA and SBM

Considering the ideal unidirectional flow of sewage water, the sewer network structure of the study area was idealized for WBE, with a total of 669 representative manholes extracted out of 5,424 manholes. The resulting sewer network and building footprints are shown in Fig. 3. The green circle is the location of the public WWTP and the inlet point of the WWTP was always selected as the monitoring site for the entire study area. The red star symbol is the location of the nursing home as a known hotspot of the COVID-19. The black dots are the extracted manholes for the unidirectional flow of sewage.

Fig. 4 illustrates the convergence of the fitted objective function (Eq. (3)) as the GA progressed in selecting optimal places of 3 and 8 the monitoring manholes (including the inlet point of the WWTP). In this case, the GA terminated after 98 and 280 generations for 3 and 8 monitoring manholes, respectively. As the number of monitoring manholes increased, the required generation and computation time for the convergence proportionally increased. Although the GA may provide acceptable manhole selection particularly for a few monitoring manholes, the global convergence was hardly achieved as indicated by a large gap between the mean and the best values of the penalty function.

Fig. 5 is the box plots showing the distributions of % differences between maximum and mean EMA values from 10 repeated runs of GA for each of the different numbers of the monitoring manholes. As shown in Fig. 5, a large variability among different runs for a given number of the monitoring manholes and the relative difference between the maximum and the mean values of EMA increased as the number of the monitoring manholes increased.

Fig. 6 visualizes the locations of selected monitoring manholes and their monitoring areas in different colors by the SBM and the GA for regular monitoring with different spatial resolutions. For the SBM, one to four repetitions of spatial bisection were conducted to select 2, 4, 8, and 16 manholes including the inlet point of the WWTP, respectively (Fig. 6a to d). As shown in Fig. 6a–d, subcatchments with selected manholes monitored were divided with approximately equal spatial resolutions. From the first selection step (Fig. 6a), EMAs of the two monitoring manholes were 1,363,313 m$^2$ and 1,342,273 m$^2$, respectively, revealing a successful division of the entire catchment into equal areas with the SBM. Similarly, the other three cases of additional manhole selection (n = 4, 8, and 16) divided the EMA of the catchment (Fig. 6b–d and Table S2) almost equally. Fig. 6e and f shows the locations of 3 and 8 manholes selected by the GA, respectively. Except for the case of two monitoring manholes (n = 2, Fig. 6a), the GA and the SBM selected different locations for a given number of monitoring manholes. In general, the SBM was superior to the GA to obtain consistent spatial resolution for the monitoring (compare Fig. 6d and f).

The GA was more flexible in selecting varying numbers of the monitoring manholes compared to the SBM, but it suffered from convergence to a local minimum, resulting in large variabilities in the site selection options for individual runs of the algorithm with an inconstant spatial resolution (i.e., unequal EMAs). This shortcoming of global optimization is because the penalty function of the optimization problem in this study is highly discontinuous with many local minima in the searching space. To illustrate the existence of a significant number of local minima in the search space, the degree of equality among the EMAs (i.e., the ratio between mean and the maximum values of EMA) as a function of different choices of two additional monitoring manholes
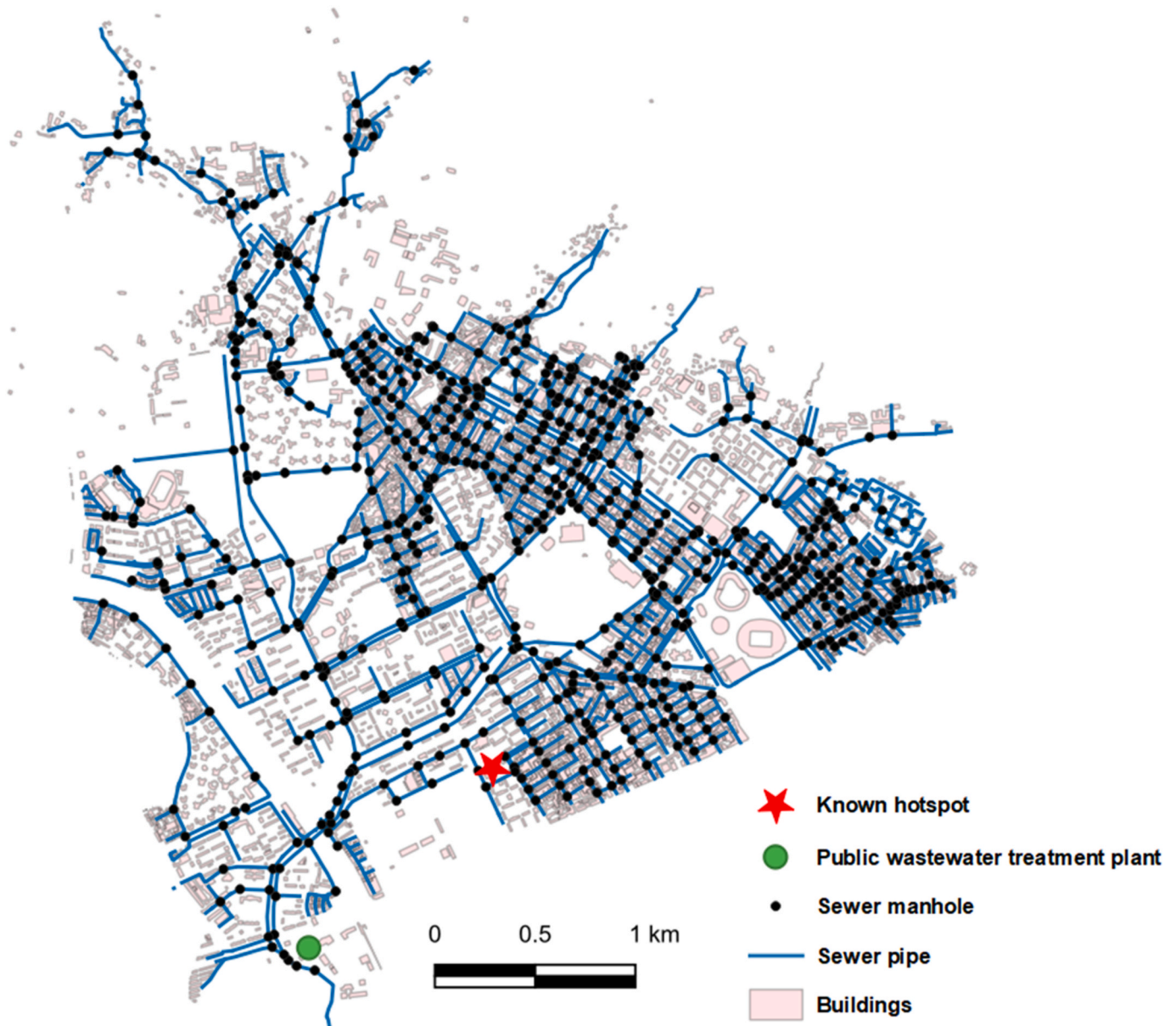
**Fig. 3.** Sewer network and building footprints of the Seoho sewershed.

with an outlet monitoring manhole is shown in Fig. 7. As illustrated in Fig. 7, due to this discontinuous nature of the penalty function, the global minimum is hardly obtained particularly for a large number of the monitoring manholes using the GA. On the other hand, the SBM requires less computational cost (e.g., computation times of the SBM and the GA for selecting 8 manholes using a 3.3 GHz CPU computer were 0.74 s and 1251 s, respectively) and provides a better selection strategy for the monitoring manholes than the GA, as illustrated in Fig. 5, comparing the % different between maximum and minimum EMAs from the GA and the SBM. Therefore, the SBM can be an alternative to the GA when only a few monitoring manholes are needed to obtain the required spatial resolution of the sewer monitoring.

### 3.2. Scaling down to the source area based on SBM

One advantage of SBM over GA is its ability to rapidly search for additional sampling manholes to scale down to the source area of infection. Fig. 8 illustrates how additional samplings aided by SBM scale down to the known source area of infection in the study area. In this

example, upon detection at the outlet point of the catchment (which is the inlet point of the WWTP), 15 additional manholes were determined by four stepwise bisections of SBM to scale down to the infected area until a required spatial resolution is achieved; the selected source area of infection was colored in red in Fig. 8. In this case, the source area was identified with a spatial resolution of around 0.16 km$^2$ (or 13,800 people) with additional 15 samplings upon detection at the outlet manhole.

### 3.3. Considerations and limitations for practical applications

The proposed SBM or GA can help with flexible selection of the sewer monitoring manholes in accordance with the prevalence dynamics. Fig. 9 illustrates the numbers of daily confirmed cases of COVID-19 in the Seoho sewershed. In the Seoho sewershed, after the first case was reported in February 2020, the daily cases were maintained under five for around 10 months until December 2020 (defined as 'low prevalence stage'). The daily cases significantly increased thereafter maintaining relatively high numbers of daily cases with the continued increase of the
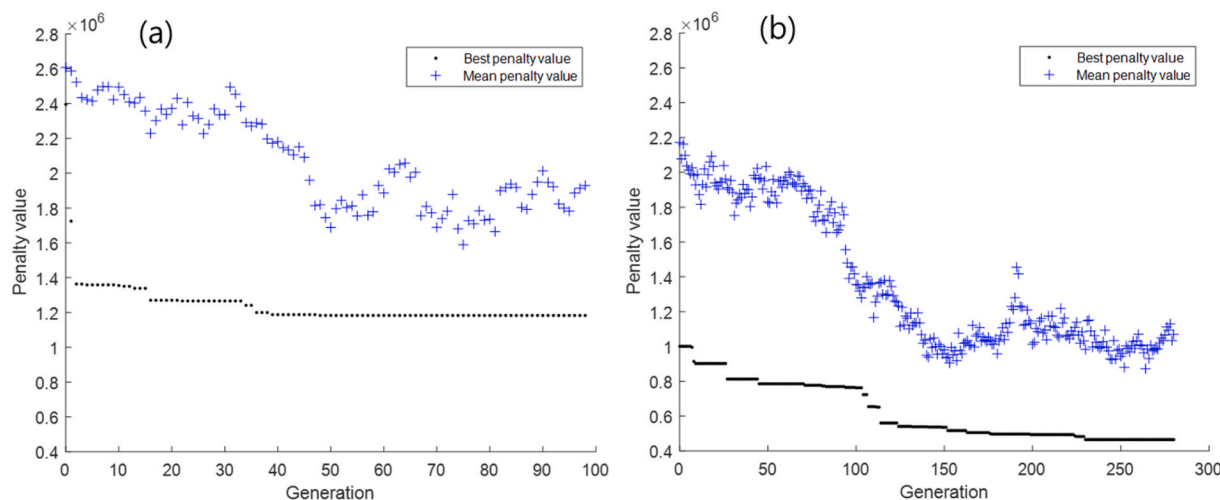
**Fig. 4.** Convergence of the penalty function with the generation in the genetic algorithm. (a) N = 3; (b) N = 8.
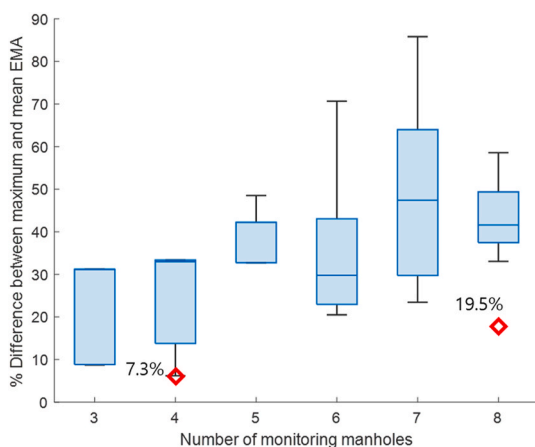


**Fig. 5.** Percent difference between maximum and mean values of the effective monitoring area for different numbers of monitoring manholes. Boxplots are from 10 repeated runs of the genetic algorithm and diamonds indicate the values obtained from the spatial bisection method.

cases and larger daily variabilities until recently (defined as 'high prevalence stage').

In the low prevalence stage, detection of infection occurrences for an early warning and identification of hotspots would be the primary goal of a WBE program. In normal times, a single manhole at the catchment outlet can be regularly monitored for an early warning, and upon detection at the outlet, additional sampling can be conducted in the upstream manholes to identify the infected area. SBM or GA can help select the required numbers of regular or additional sampling manholes to provide a required spatial resolution for the disease surveillance or source area identification upon detection at the outlet. The detection sensitivity of a monitoring manhole can be site-specific depending on the catchment area served by the manhole (i.e., CSA), sampling method (e.g., grab or composite), sewage characteristics, the recovery efficiency of the quantification method, and detection limit of the instrument (Zhu et al., 2021). Currently, a few studies reported detection sensitivity of the viral RNA of SARS-CoV-2: Hart and Halden (2020) reported the theoretically possible ranges of detection sensitivity as $5 \times 10^{-5}\%$–1% (1 individual among 100 to 2,000,000 persons) based on a computational analysis; and Ahmed et al. (2020) reported a detection sensitivity ranging from 0.028% to 0.18% (170–1090 individuals in 600,000 persons) based on the Monte Carlo simulation and detection limit of RT-qPCR technique. In the Seoho sewershed, the outlet manhole may

successfully detect the low prevalence present in the entire catchment area if the detection sensitivity is greater than $4.5 \times 10^{-4}\%$, which is in the possible detection range by Hart and Halden (2020). Due to smaller CSA with less dilution effect, the detection sensitivity of additional sampling manholes in the upstream is typically greater than that of the outlet manhole, enabling promising outcomes for source area identification by additional samplings upon detection at the outlet.

In the high prevalence stage (Fig. 9), forecasting patient trends can be more important than an early warning or hotspot identification. Because of a larger number of infected individuals within a time window of active viral shedding, a smaller number of monitoring manholes with more frequent samplings will be needed. Here, we assumed that the viral shedding rate of an infected individual stays constant for a week (although it is not true) and the prevalence level is defined by cumulative cases during the preceding seven days. In the case of the Seoho sewershed (Fig. 9), the daily prevalence level is mostly greater than 10. Thus, the outlet monitoring manhole can efficiently detect the level of disease prevalence at a detection sensitivity of $4.5 \times 10^{-3}\%$ during the high prevalence stage in the study area.

The suggested design and operation method of sewer monitoring stations based on SBM or GA can be optimized by offsetting cost, considering the required resolution of the detection (spatial monitoring scale and sampling frequency), the prevalence stage, the predetermined threshold of prevalence level (i.e., detection sensitivity), sampling and experimental method, and total budget for the WBE program. In addition, more weights can be given to the areas of high infection risk such as hospitals, nursing homes, or districts with vulnerable classes during the determination of the spatial scale in the spatial bisection procedure.

Despite the potential utility of the proposed site allocation methods in a WBE program, successful implementation of these methods might be limited by uncertainties in measurement quality and data interpretation originating from various site-specific factors such as leakage or dilution of the sewer, sewer characteristics, temperature and time driven decay of the analytes, temporally and individually varying shedding rates, and population mobility (Wade et al., 2022). Therefore, a WBE program needs techniques for handling these uncertainties and correct prevalence estimation based on the measured data to ensure practical feasibility of the proposed site allocation methods, which is beyond the scope of the current study.

## 4. Conclusions

This study aimed to provide an efficient tool to better design a WBE program. Performances of a global optimization algorithm, GA, and SBM were compared in determining the optimal locations of the monitoring
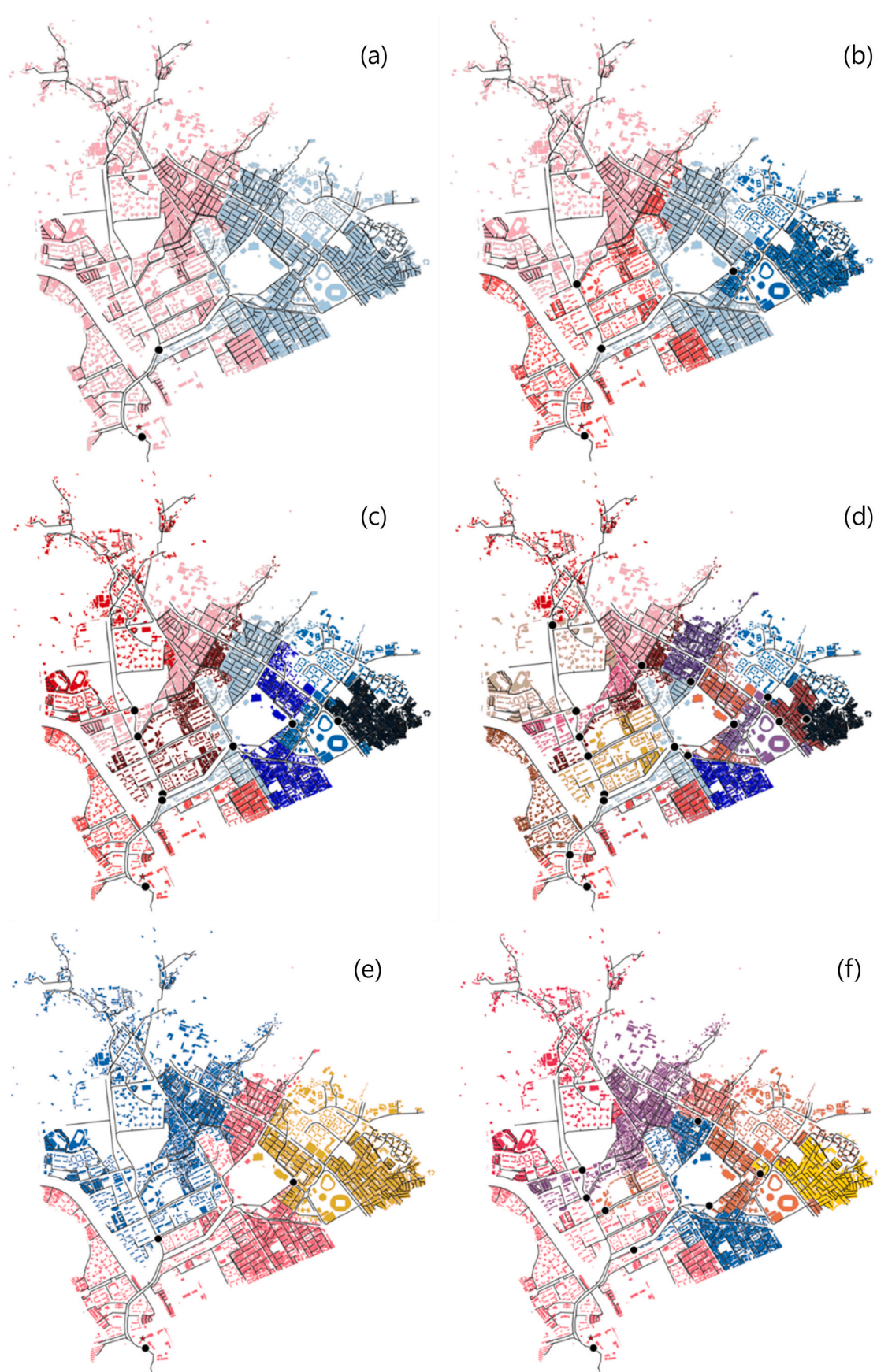
**Fig. 6.** Selection of optimal places of the monitoring sites up to $2^4$ manholes in the Seoho sewershed for regular monitoring: (a) Fist step selection using SBM (n = 2); (b) Second step selection using SBM (n = 4); (c) Third step selection using SBM (n = 8); (d) Fourth step selection using SBM (n = 16); (e) Selection of three manholes by GA; (f) Selection of eight manholes by GA. Black solid circles indicate selected manholes. Subcatchments (effective monitoring areas) are in different colors.
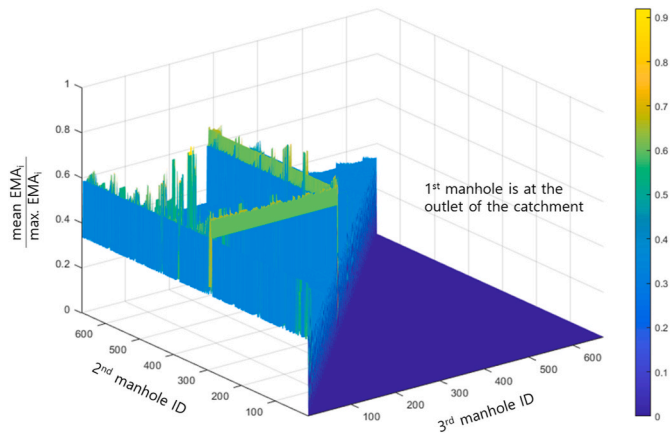
**Fig. 7.** Ratio between mean and maximum values of the effective monitoring area (EMA) with all different selections of two manholes given an outlet monitoring manhole. The candidate monitoring manholes in the catchment were numbered from 1 to 669, with 1 being the default monitoring manhole at the catchment outlet (the 1st monitoring manhole).

manholes. Although GA allowed more flexible selection of the total number of the monitoring manholes and spatial resolution of the monitoring, it was less efficient in evenly assigning the monitoring areas to the monitoring manholes and required greater computational cost compared to SBM. SBM was demonstrated as an efficient approach to rapidly scale down to source areas of infectious disease. The proposed methods provide options for WBE designers to determine the most appropriate locations of sewage surveillance with a required source population or spatial resolution in a large catchment, or identifying hotspots with the least number of additional samplings upon detection from regular sewage surveillance. Designers can offset between sample representativeness (i.e., monitored source population) and quantification accuracy by adjusting spatial resolutions with different numbers of monitoring manholes. The proposed methods can support WBE for early detection and warning of the (re)emerging infectious diseases outbreaks; identifying hotspots to be designated for quarantine and

forecasting the dynamics of the prevalence level in the community prior to the medical diagnostic reporting; and thereby allowing preparedness of the healthcare facilities against disease outbreaks. A WBE program can have more than one objective among the objectives described above, or focus more on one objective over another depending on the progress status or level of the disease prevalence.

### Credit author statement

**Keug Tae Kim**: Conceptualization, Writing – review & editing. **Min Jeong Ban**: Data curation, Formal analysis. Sungpyo Kim: Data curation, Funding acquisition. **Mi-Hyun Park**: Writing – review & editing, Validation. **Michael K. Stenstrom**: Investigation, Validation. **Joo-Hyon Kang**: Conceptualization, Methodology, Investigation, Validation, Writing – original draft.

### Declaration of competing interest

The authors declare that they have no known competing financial
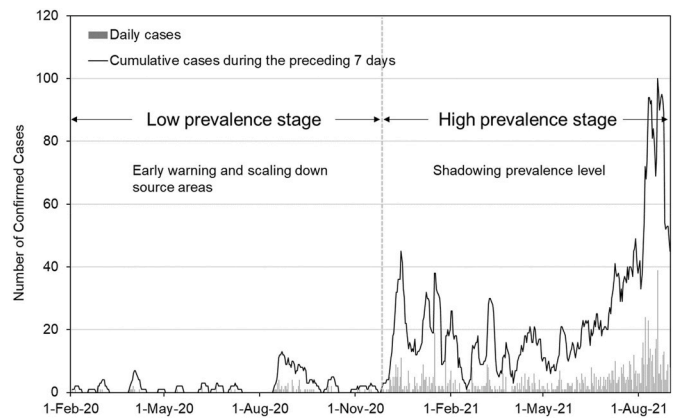


**Fig. 9.** Temporal variation and stages of COVID-19 prevalence in the Seoho sewershed, with a proposed operational strategy of sewer monitoring stations.
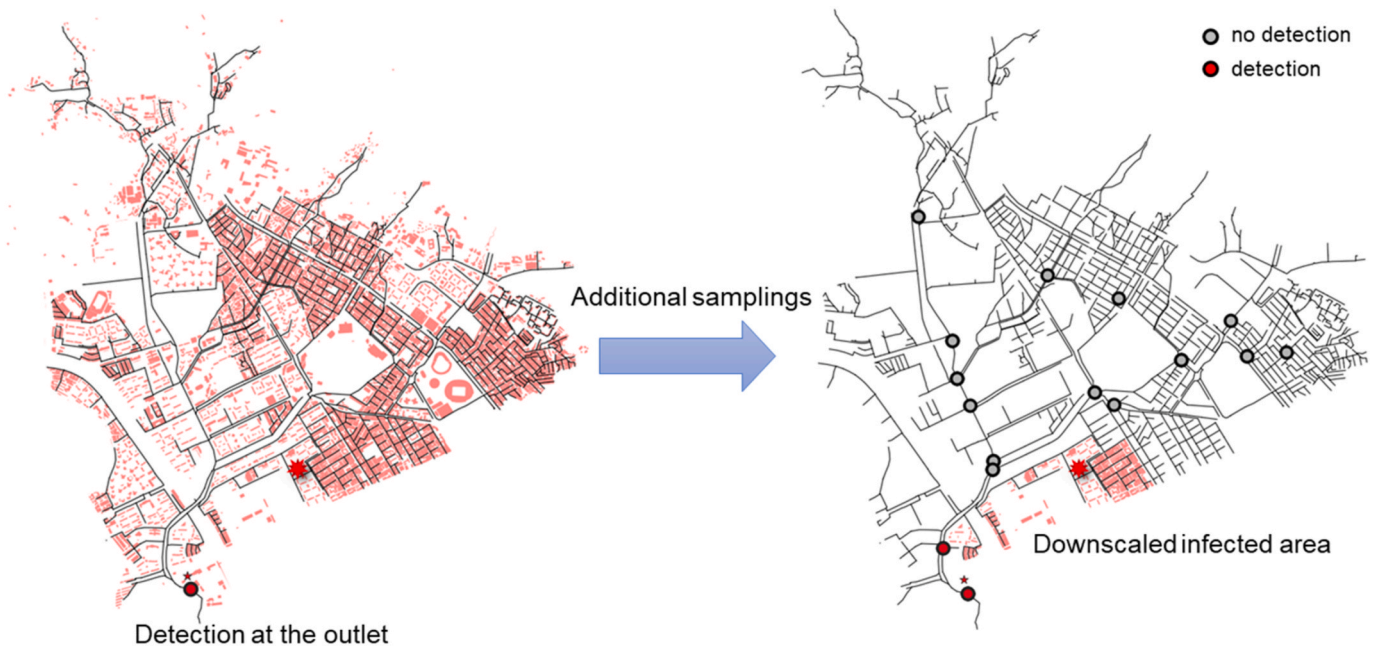


**Fig. 8.** Downscaling of an infected area using additional samplings at the upstream manholes selected by SBM upon detection at the outlet. Red-colored areas indicate the areas that contain an infection source. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2022.115806.

## References

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tscharke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K.V., Mueller, J., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Sci. Total Environ. 728, 138764.

Andrews, J.R., Yu, A.T., Saha, S., Shakya, J., Aiemjoy, K., Horng, L., Qamar, F., Garrett, D., Baker, S., Saha, S., Luby, S.P., 2020. Environmental surveillance as a tool for identifying high-risk settings for typhoid transmission. Clin. Infect. Dis. 71 (Suppl. 2), S71–S78.

Arabi, M., Govindaraju, R.S., Hantush, M.M., 2006. Cost-effective allocation of watershed management practices using a genetic algorithm. Water Resour. Res. 42, W10429.

Asghar, H., Diop, O.M., Weldegebriel, G.W., Malik, F., Shetty, S., Bassioni, L.E., Akande, A.O., Maamoun, E.A., Zaidi, S., Adeniji, A.J., Burns, C.C., Deshpande, J., Oberste, M.S., Lowther, S.A., 2014. Environmental surveillance for polioviruses in the global polio eradication initiative. JID (J. Infect. Dis.) 210 (Suppl. 1), S294–S303.

Bank Group, World, 2021. Global Economic Prospects. June 2021, Washington, DC.

Barceló, D., 2020. 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.

Barrios, R.E., Lim, C., Kelley, M.S., Li, X., 2021. SARS-CoV-2 concentrations in a wastewater collection system indicated potential COVID-19 hotspots at the zip code level. Sci. Total Environ. 880, 149480.

Betancourt, W.Q., Schmitz, B.W., Innes, G.K., Prasek, S.M., Pogreba Brown, K.M., Stark, E.R., Foster, A.R., Sprissler, R.S., Harris, D.T., Sherchan, S.P., Gerba, C.P., Pepper, I.L., 2021. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. Sci. Total Environ. 779, 146408.

Brouwer, A., Eisenberg, J.N.S., Pomeroy, C.D., Shulman, L.M., Hindiyeh, M., Manor, Y., Grotto, I., Koopman, J.S., Eisenberg, M., 2018. Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. Proc. Natl. Acad. Sci. USA 115 (45), E10625–E10633.

Burden, R.L., Faires, J.D., 1985. Numerical Analysis, third ed. PWS Publishers.

Calle, E., Martinez, D., Brugues-i-Pujolras, R., Farreras, M., Salo-Grau, J., Pueyo-Ros, J., Corominas, L., 2021. Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks. Environ. Int. 157, 106768.

Davó, L., Seguí, R., Botija, P., Beltrán, M.J., Álbert, E., Torres, I., López-Fernández, P.A., Ortí, R., Maestre, J.F., Sánchez, G., Navarro, D., 2021. Early detection of SARS-CoV-2 infection cases or outbreaks at nursing homes by targeted wastewater tracking. Clin. Microbiol. Infect. 27 (7), 1061–1063.

Fazal, M.A., Imaizumi, M., Ishida, S., Kawachi, T., Tsuchinara, T., 2005. Estimating groundwater recharge using the SMAR conceptual model calibrated by genetic algorithm. J. Hydrol. 303 (4), 56–78.

Foladori, P., Cutrupi, F., Segata, N., Manara, S., Pinto, F., Malpei, F., Bruni, L., La Rosa, G., 2020. SARS-CoV-2 from faeces to wastewater treatment: what do we know? A review. Sci. Total Environ. 743, 140444.

Fumian, T.M., Fioretti, J.M., Lun, J.H., dos Santos, I.A.L., White, P.A., Miagostovich, M. P., 2019. Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. Environ. Int. 123, 282–291.

Gibas, C., Lambirth, K., Mittal, N., Juel, M.A.I., Barua, V.B., Brazell, L.R., Hinton, K., Lontai, J., Stark, N., Young, I., Quach, C., Russ, M., Kauer, J., Nicolosi, B., Chen, D., Akella, S., Tang, W., Schlueter, J., Munir, M., 2021. Implementing building-level

SARS-CoV-2 wastewater surveillance on a university campus. Sci. Total Environ. 782, 146749.

Gupta, I., Gupta, A., Khanna, P., 1999. Genetic algorithm, for optimization of water distribution systems. Environ. Model. Software 14 (5), 437–446.

Haghighi, A., Bakhshipour, A.E., 2012. Optimization of sewer networks using an adaptive genetic algorithm. Water Resour. Manag. 26, 3441–3456.

Harris-Lovett, S., Nelson, K.L., Beamer, P., Bischel, H.N., Bivins, A., Bruder, A., Butler, C., Camenisch, T.D., De Long, S.K., Karthikeyan, S., Larsen, D.A., Meierdiercks, K., Mouser, P.J., Pagsuyoin, S., Prasek, S.M., Radniecki, T.S., Ram, J.L., Roper, D.K., Safford, H., Sherchan, S.P., Shuster, W., Stalder, T., Wheeler, R.T., Korfmacher, K.S., 2021. Wastewater surveillance for SARS-CoV-2 on college campuses: initial efforts, lessons learned, and research needs. Int. Environ. Res. Public Health 18, 4455.

Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges. Sci. Total Environ. 730, 138875.

Heijnen, L., Medema, G., 2011. Surveillance of influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in The Netherlands. J. Water Health 9 (3), 434–442.

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. Appl. Environ. Microbiol. 80 (21), 6771–6781.

Hui, E.K.W., 2006. Reasons for the increase in emerging and re-emerging viral infectious disease. Microb. Infect. 8, 905–916.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451, 990–994.

Kalyanmoy, D., 2000. An efficient constraint handling method for genetic algorithms. Comput. Methods Appl. Mech. Eng. 186 (2–4), 211–338.

Kim, K.Y., Lai, F.Y., Kim, H.Y., Thai, P.K., Mueller, J.F., Oh, J.E., 2015. The first application of wastewater-based drug epidemiology in five South Korean cities. Sci. Total Environ. 524–525, 440–443.

Larson, R.C., Berman, O., Nourinejad, M., 2020. Sampling manholes to home in on SARS-CoV-2 infections. PLoS One 15 (10), e0240007.

Massoudieh, A., Abrishamchi, A., Kayhanian, M., 2008. Mathematical modeling of first flush in highway storm runoff using genetic algorithm. Sci. Total Environ. 398, 107–121.

McMahan, C.S., Self, S., REnnert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J.A., Popat, S.C., Karanfil, T., Freedman, D.L., 2021. COVID-19 wastewater epidemiology: a model to estimate infected populations. Lancet Planet. Health 5, e874–881.

Medema, G., Been, F., Heijnen, L., Petterson, S., 2020. Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: opportunities and challenges. Current Opinion in Environmental Science & Health 17, 49–71.

Mitchell, M., 1996. An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA.

Moran-Gilad, J., Kaliner, E., Gdalevich, M., Grotto, I., 2016. Public health response to the silent reintroduction of wild poliovirus to Israel, 2013-2014. Clin. Microbiol. Infect. 22, S140–S145.

Sims, N., Kasprzyk-Hordern, B., 2020. Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level. Environ. Int. 139, 105689.

Srivastava, P., Hamlett, J.M., Robillard, P.D., Day, R.L., 2002. Watershed optimization of best management practices using AnnAGNPS and a genetic algorithm. Water Resour. Res. 38 (3), 3–14, 3-1.

Wade, M.J., Jacomo, A.L., Armenise, E., Brown, M.R., Bunce, J.T., Cameron, G.J., Fang, Z., Farkas, K., Gilpin, D.F., Grahan, D.W., Grimsley, J.M.S., Hart, A., Hoffman, T., Jackson, K.J., Jones, D.L., Lilley, C.J., McGrath, J.W., McKinely, J.M., McSparron, C., Nejad, B.F., Moran, M., Quintela-Baluja, M., Roberts, A.M.I., Singer, A.C., Souque, C., Speight, V.L., Sweetapple, C., Walker, D., Watts, G., Weightman, A., Kasprzyk-Hordern, B., 2022. Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: lessons learned from the United Kingdom national COVID-19 surveillance programmes. J. Hazard Mater. 424, 127456.

Wang, Y., Moe, C.L., Dutta, S., Wadhwa, A., Kanungo, S., Mairinger, W., Zhao, Y., Jiang, Y., TEunis, P.F.M., 2020. Designing a typhoid environmental surveillance study: a simulation model for optimum sampling site allocation. Epidemics 31, 100391.

World Health Organization (WHO), WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/(Accessed on 3 December 2021).

World Health Organization (WHO), 2003. Guidelines for Environmental Surveillance of Poliovirus Circulation. Department of Vaccines and Biologicals, Geneva, Switzerland.

World Health Organization (WHO), 2018. Managing Epidemics: Key Facts about Major Deadly Diseases. World Health Organization.

Yao, L., Zhu, W., Shi, J., Xu, T., Qu, G., Zhou, W., Yu, X.F., Zhang, X., Jiang, G., 2021. Detection of coronavirus in environmental surveillance and risk monitoring for pandemic control. Chem. Soc. Rev. 50, 3656–3676.

Zhu, Y., Oishi, W., Maruo, C., Saito, M., Chen, R., Kitajima, M., Sano, D., 2021. Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. Sci. Total Environ. 767, 145124.