Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks

Eusebi Calle [a], David Martínez [b], Roser Brugués-i-Pujolràs [a], Miquel Farreras [a], Joan Saló-Grau [a], Josep Pueyo-Ros [b], Lluís Corominas [b,*]

[a] Institute of Informatics and Applications, Universitat de Girona, Girona, Spain
[b] Catalan Institute for Water Research, Emili Grahit 101, 17003 Girona, Spain

ARTICLE INFO

Handling Editor: Frederic Coulon

Keywords:
COVID-19
Wastewater-based epidemiology
Sewer
Sensor placement
Graph theory

ABSTRACT

Selecting sampling points to monitor traces of SARS-CoV-2 in sewage at the intra-urban scale is no trivial task given the complexity of the networks and the multiple technical, economic and socio-environmental constraints involved. This paper proposes two algorithms for the automatic selection of sampling locations in sewage networks. The first algorithm, is for the optimal selection of a predefined number of sampling locations ensuring maximum coverage of inhabitants and minimum overlapping amongst selected sites (static approach). The second is for establishing a strategy of iterations of sample&analysis to identify patient zero and hot spots of COVID-19 infected inhabitants in cities (dynamic approach). The algorithms are based on graph-theory and are coupled to a greedy optimization algorithm. The usefulness of the algorithms is illustrated in the case study of Girona (NE Iberian Peninsula, 148,504 inhabitants). The results show that the algorithms are able to automatically propose locations for a given number of stations. In the case of Girona, always covering more than 60% of the manholes and with less than 3% of them overlapping amongst stations. Deploying 5, 6 or 7 stations results in more than 80% coverage in manholes and more than 85% of the inhabitants. For the dynamic sensor placement, we demonstrate that assigning infection probabilities to each manhole as a function of the number of inhabitants connected reduces the number of iterations required to detect the zero patient and the hot spot areas.

1. Introduction

There is increasing evidence that sewage is a good, unbiased indicator of the prevalence of a virus in a population. The ability to detect SARS-CoV-2 in sewage has been reported by research groups worldwide. Upon confirmation that COVID-19 patients shed SARS-CoV-2 in feces, different studies have provided significant correlation between the concentration of SARS-CoV-2 in sewage and the prevalence of COVID-19 in the corresponding population (Lenzen et al., 2020; Mallapaty, 2020; Medema et al., 2020; Schmidt, 2020). So far, the approach has been successful when monitoring at the wastewater treatment plant (WWTP) level (i.e. integrating all inhabitants from a municipality), but there is limited experience when bringing the approach to a neighborhood level. ‘Upstream’ surveillance for SARS-CoV-2 may facilitate finer spatial detection of the virus in catchments with differing COVID-19 disease burdens, and may help provide information about any mitigation actions implemented at the community level. An example of monitoring at the neighborhood level can be found in Wu et al. (2020) where 11 urban neighborhoods within the wastewater treatment facility’s catchment, representing populations ranging from ~4,000 to ~40,000 individuals, were monitored. GIS (geographic information system) data with catchment outlines was used to aggregate the demographic information for the catchment. Yet, while the selection of the sampling points in Wu et al. (2020) serves the purpose of the study, this might not be optimal from the perspective of a municipality.

Monitoring the traces of SARS-CoV-2 in sewage at the intra-urban scale implies establishing a surveillance network inside the sewage network. Sewer systems are long complex networks of pipes. As an example, the total length of the sewage network across the EU has been estimated at around 3 M kilometers (EurEAU, 2017). A city of about 100,000 inhabitants might have around 300 km of small sewer pipes (building sewer pipes and lateral sewers) and 60 km of bigger main pipes (community sewers collecting sewage from the lateral sewers and transporting it to the WWTPs). It is then not evident where to place

* Corresponding author.
E-mail addresses: eusebi.calle@udg.edu (E. Calle), dmartinez@icra.cat (D. Martínez), u1059680@campus.udg.edu (R. Brugués-i-Pujolràs), miquel.farreras@udg.edu (M. Farreras), u1953621@campus.udg.edu (J. Saló-Grau), jpueyo@icra.cat (J. Pueyo-Ros), lcorominas@icra.cat (L. Corominas).

https://doi.org/10.1016/j.envint.2021.106768
Received 14 April 2021; Received in revised form 25 June 2021; Accepted 6 July 2021
Available online 12 July 2021

autosamplers (or sensors if available in the future) to monitor the concentration of the RNA traces of SARS-CoV-2. Selecting sampling/sensor placement locations can follow several criterion. Given a predefined number of sampling locations to be installed, a municipality might be interested in ensuring maximum coverage of manholes (or of inhabitants) and ensuring a balance in the number of inhabitants covered by each of them (static monitoring site selection from now on). Another approach would be to establish a monitoring procedure to detect the area where the virus is more present (dynamic monitoring site selection from now on) or the zero patient if the analytical method for SARS-CoV-2 would be sufficiently sensitive to detect one virus shedder in an entire community. For either of the two approaches, it is relevant to collect demographic and socioeconomic indicators for the community connected to a specific sampling point. Otherwise, it is not possible to correlate the virus concentration to the number of diagnosed cases, for instance, or to socioeconomic indicators. Furthermore, the cost of autosampler/sensor ownership, maintenance efforts in particular, can still be cost-prohibitive and a balance between the number of sampling locations and costs needs to be guaranteed.

Larson et al. (2020) is the only paper about sampling points selection in sewers related to SARS-CoV-2; the authors propose two strategies to detect the zero patient and to identify zones with high levels of infection. Larson et al. (2020) assume that near-real time SARS-CoV-2 concentrations can be measured, for instance by employing fast tests (Mao et al., 2020b) (Mao et al., 2020a) which are under development but not yet available. Existing approaches to analyze SARS-CoV-2 concentrations imply lab analyzes and deliver results in 24–72 h. Wang et al. (2020) propose adaptive sampling site allocation for the sewage surveillance of the pathogen S. Typhi, by which the locations of sampling sites are dynamically updated to increase the probability of detecting a positive signal of the pathogen. It uses a model to simulate pathogen shedding, pathogen transport and fate in the sewage network, sewage sampling, and detection of the pathogen. Wang et al. (2020) propose stratified sampling for the initial selection of sampling sites, by which the geographic area is divided into a certain number of subareas and one sampling unit is randomly selected from each subarea. Within the scope of wastewater-based epidemiology (not related to SARS-CoV-2) the work from Matus et al. (2019) proposed monitoring sites selection using GIS analysis with city-wide demographic and sewage network information. The approach was semi-automatic, based on the definition of constraints, and did not use optimization.

Monitoring site selection (or sensor placement) has been widely studied for drinking water networks, but only a few studies exist on sewage networks. Kang et al. (2013) determined key sensor locations for non-point pollutant sources management in sewage networks by means of clustering analysis and ANOVA on top of SWMM simulated results. A few examples exist on sensor placement for illicit intrusion detection in sewage networks based on single and multi-objective optimization (Yazdi, 2018) or on Bayesian decision networks (Sambito et al., 2020). Vonach et al. (2018) proposed best sampling locations with the objective of calibrating a hydrodynamic model. Finally, Villez et al. (2016) and Villez et al. (2020) proposed methodologies for sensor placement in WWTPs based on graph theory and mass balances for maximizing the ability to assess and control data quality while minimizing the cost of ownership. Given this background, this paper proposes two algorithms for SARS-CoV-2 monitoring site selection in sewage networks. The algorithms are based on graph-theory and are coupled to a greedy optimization algorithm. To the best of our knowledge, this is the first paper which proposes an algorithm for static SARS-CoV-2 monitoring site selection. Furthermore, this paper enhances the approach proposed in Larson et al. (2020) for dynamic monitoring site selection by i) defining infection probabilities as a function of the number of inhabitants connected to each manhole and ii) evaluating the benefit of combining the static sensor placement outcomes to the dynamic placement algorithms. This paper as well contributes to enhance the selection of the initial selection sites from Wang et al. (2020) by proposing a static sensor placement algorithm which uses optimization. The usefulness of the two types of algorithms is illustrated with a case study in the city of Girona.

2. Materials and methods

This section describes the general methodology followed to obtain optimal monitoring sites for SARS-CoV-2 surveillance and includes a description of the algorithms proposed and the description of the case study used to illustrate their usefulness.

2.1. General methodology

The overall approach for obtaining optimal monitoring site selection for SARS-CoV-2 surveillance involves the following steps: i) goal and scope definition, ii) data collection, iii) graph generation, iv) linking demographic and socioeconomic indicators to the graph, v) implementation and execution of the algorithms, and vi) analysis of results. Fig. 1 describes the process flow of this general methodology. Details on the actual application of the general methodology to the specific case-study are provided in Section 2.3.

Goal and scope definition. This first step consists of defining the goal and scope of the study which, in turn, will influence all subsequent steps in terms of data intensity, data quality, algorithm selection, etc. Some examples of ”goal and scope” are given: i) monitoring the spread of the COVID-19 disease in different communities with homogenous socio-economic status of a city during a pandemic; ii) identify in a given city a hotspot area with much higher disease prevalence than others. This step also involves the definition of the monitoring site selection needs (e.g. static vs dynamic), criteria and constraints together with the client profile (municipality or local health authority). Criteria are related to demographic, environmental, health or economic indicators; as an example, population density or socio-economic status of inhabitants are criteria which can be used to make the decision on the sampling sites selection. Constraints can be applied to the criteria, but also can be related to physical constraints in given manholes which do not allow to install equipment for wastewater sampling.

Data collection. The following data are collected: i) sewage network topology; ii) Digital Elevation Model (DEM) of the case under study; iii) cadastral data from the city parcels and iv) demographic and socioeconomic indicators associated to each cadastral parcel.

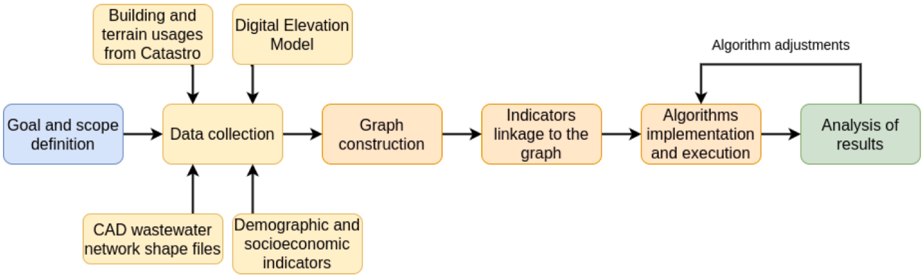

Building and terrain usages from Catastro
Digital Elevation Model
Algorithm adjustments
Goal and scope definition
Data collection
Graph construction
Indicators linkage to the graph
Algorithms implementation and execution
Analysis of results
CAD wastewater network shape files
Demographic and socioeconomic indicators

Fig. 1. General methodology flow.

Graph construction. The sewage network topology is transformed into a graph, where the edges represent the pipes and the nodes represent the manholes. The original sewage network topology must include the pipes, manholes, elevations and direction of the wastewater flow. In most cases, wastewater companies keep this data in GIS databases, which means an additional step is required to transform that data into a graph. The key elements are the coordinates of the nodes and the relationship between the nodes and edges. Data verification and reconciliation is essential at this stage to verify that all the pipes that are supposed to be connected through a manhole are actually connected, and to validate the sewer network slopes.

Indicators linkage to the graph. The physical attributes of the graph (e.g. topological such as node elevation) are assigned. The challenge here is to connect inhabitants (virus shedders) to the nearest manhole. This also provides the possibility to link the manhole with some socioeconomic indicators fed by census or other socioeconomic databases. Hydraulic models can be a source of data (e.g. biomarkers travel time) which might be used as a criteria or constraint by the algorithms. Yet, hydraulic models have not been used in this case-study.

Implementation and execution of the algorithms. Static and dynamic monitoring site selection require computationally expensive algorithms and thus need to be implemented in a relevant computing environment to achieve simulation results at reasonable times. This is of special interest in the case of dynamic monitoring site selection.

Results analysis. The results are analyzed, and if the targets set up in the goal and scope phase are satisfied, the project can be accomplished. Otherwise, a feedback loop to previous phases is needed until the goals are reached, as shown in Fig. 1. Iterations are conducted in the feedback loop to enhance the performance of the algorithms in terms of results and computational time. The final algorithms deployed in this study are transferable to other case-studies with no further upgrade; for transferability it is only necessary to construct the graph of each new case-study and run the algorithms developed in this paper.

2.2. *Static and dynamic algorithms*

The algorithms are based on techniques borrowed from graph theory which allow pipe network configurations to be analyzed (Kesavan and Chandrashekar, 1972). In the case of sewage networks, the pipes correspond to the graph edges and the manholes represent the graph nodes. We provide algorithms for both static and dynamic monitoring site selection. The static challenge implies selecting the locations for permanent monitoring sites which meet the client's (e.g. a municipality or health agency) needs (e.g. maximum coverage at minimal investment). In the case of SARS-CoV-2 sewage surveillance, permanent monitoring sites can be selected where 24 h composite samples are taken and brought to the lab for microbiological analysis that will deliver results in between 24 and 72 h (Ahmed et al., 2020; Rusiñol et al., 2020). The concept of interference is applied in the static monitoring site selection. Interferences appear when two or more monitoring sites overlap in covered manholes. In some cases, interferences might appear as a result of the graph being improperly constructed, and hence a reconciliation exercice is needed before launching the optimization algorithms. The dynamic monitoring site selection challenge implies dynamically and adaptively developing a sequence of manholes to sample and test until it finds the manhole in which the first infected person in a city is connected, or until the group of manholes in a city in which the largest group of infected people are connected is found. For the dynamic approach we build on what was presented in Larson et al. (2020) but also include one enhancement. In Larson et al. (2020), the Bayesian probabilities of infection are equally assigned to all manholes, whereas in our proposal, the Bayesian probabilities are assigned according to the number of inhabitants connected to each manhole; hence we assume that there is a greater chance of finding the origin of infection there. It is assumed in the dynamic monitoring site selection that a portable and fast analytical method (results delivered in few minutes) is available to guarantee several sample-and-test iterations to be executed in a short period of time. The turnaround time for an iteration of sampling, testing and adjustment should ideally be of 24 h; since the concentrations of SARS-CoV-2 can change during the course of a day (and the dynamics are even more pronounced at the community level), it is recommended to analyze 24-h composite samples (Medema et al., 2020), and then the analysis of SARS-CoV-2 concentrations and the launch of the proposed algorithms should take less than an hour. The whole process for detecting the hot spot should ideally be shorter than 1 week.

Table 1 specifies the notation used for the static and dynamic algorithms. In brief, let $\mathscr{G} = (\mathcal{V}, \mathcal{E})$ be the sewage network graph, with a V-element set of nodes $\mathcal{V}$ representing manholes, and an E-element set of links $\mathcal{E} \subset \mathcal{V}^{|2|}$ representing pipes. Additionally, $\mathcal{S}$ (where $\mathcal{S} \subseteqq \mathcal{V}$) denotes an S-element set of nodes with placed monitoring sites.

2.2.1. *Static monitoring site selection algorithm*

A novel algorithm called monitoring site selection (MSS, see Algorithm 1) is proposed. The MSS algorithm presents a greedy approach that optimizes the placement of several K SARS-CoV-2 monitoring sites within the sewage network nodes, dividing the network into K monitoring sites areas or subgraphs. We define the set of nodes that are present in the monitoring site k coverage area (i.e., source nodes) as $\mathcal{C}(k)$, $k \in \mathcal{K}$.

Table 1
Notation concerning the static and dynamic algorithms.

K	number of monitoring sites to place; $1 \leqslant K \leqslant V$ (fixed parameter)
$\mathcal{K} = \{1, 2, \ldots, K\}$	set of (indices) of the monitoring sites
$\mathcal{N}(v), v \in \mathcal{V}$	set of neighbor nodes of the node v; $\mathcal{N}(v) \subseteqq \mathcal{V} \setminus \{v\}$
$\mathcal{C}(k), k \in \mathcal{K}$	set of nodes that are present in the monitoring site k coverage area (i.e., source nodes); $\mathcal{C}(k) \subseteqq \mathcal{V}$
$\mathcal{C}$	set of nodes that are present in at least one monitoring site coverage area; $\mathcal{C} := \bigcup_{k \in \mathcal{K}} \mathcal{C}(k)$; The size (number of nodes) of this set is called **coverage** $C = \|\mathcal{C}\|$
$\mathcal{I}(k), k \in \mathcal{K}$	set of nodes in the monitoring site k coverage area that are present also in at least another monitoring site j coverage area; $\mathcal{I}(k) \subseteqq \mathcal{C}(k)$; $\mathcal{I}(k) := \{v \in \mathcal{C}(k) \cap \mathcal{C}(j) : j \in \mathcal{K}, j \neq k\}$
$\mathcal{I}$	set of nodes that are present in at least two monitoring site coverage areas; $\mathcal{I} := \bigcup_{k \in \mathcal{K}} \mathcal{I}(k)$; The size (number of nodes) of this set is called **interference** $I = \|\mathcal{I}\|$
$\mathcal{U}(k), k \in \mathcal{K}$	set of nodes that are present only and exclusively in the monitoring site k coverage areas; $\mathcal{U}(k) := \{v \in \mathcal{C}(k) \setminus \mathcal{I}(k)\}$
$\mathcal{U}$	set of nodes that are present in one and only one monitoring site coverage area; $\mathcal{U} := \bigcup_{k \in \mathcal{K}} \mathcal{U}(k)$; The size (number of nodes) of this set is called **unique coverage** $U = \|\mathcal{U}\|$
C_{max}	number of nodes of the largest monitoring site coverage area; $C_{max} := \max(\|\mathcal{C}(k)\|, k \in \mathcal{K})$
C_{min}	number of nodes of the smallest monitoring site coverage area; $C_{min} := \min(\|\mathcal{C}(k)\|, k \in \mathcal{K})$
D	number of nodes **difference** between the largest and the smallest monitoring site coverage areas; $D := C_{max} - C_{min}$
$\mathcal{A}$	set of **artificial** source nodes in the area covered by a sensor with $deg^{-} := 0$ and population associated to it.
$\mathcal{P}$	set of nodes in the area covered by a sensor after **normalising** and **simplifying** $\mathcal{C}$.
$\mathcal{T}$	set of nodes of Hot Spot neighborhood/node of Patient Zero in the area covered by a sensor. For PZ, $\|\mathcal{T}\| := 1$.

The MSS algorithm starts from a random combination of nodes with placed monitoring sites $\mathcal{S}$, and then iterates to find better combinations by moving each monitoring site $s \in \mathcal{S}$ through its neighbouring nodes. We propose an evaluation function called monitoring site selection evaluation (MSE, see Function 1) which allows an optimization metric after each execution of the MSS algorithm to be estimated. The MSS algorithm halts when it is not possible to find an $\mathcal{S}' \neq \mathcal{S}$ monitoring site set by which for all neighboring nodes of each $s \in \mathcal{S}$ the optimization metric does not improve as compared to the $\mathcal{S}$ set.

Algorithm 1. Monitoring site selection (MSS) algorithm.

```
Input: $K$: number of monitoring sites to place.
        $\mathcal{G} \leftarrow \{\mathcal{V}, \mathcal{E}\}$: wastewater network with node set $\mathcal{V}$ and link set $\mathcal{E}$.
        $\mathcal{C}(k), k \in \mathcal{K}$: set of nodes that are present in each monitoring site area $k$.
        $\mathcal{N}(v), v \in \mathcal{V}$: set of neighbor nodes of each node $v$.
Output: $\mathcal{S}, \forall_{s \in \mathcal{S}}\ s \in \mathcal{V}, S = K$: set of nodes with collocated monitoring sites ($S = |\mathcal{S}|$).
$O$: optimization value of the monitoring sites placed on the node set $\mathcal{S}$.
1. Obtain a random sample of nodes with collocated monitoring sites $\mathcal{S}$ with $K$
   elements of $\mathcal{V}$.
2. Compute the optimization value $O$ for the sample $\mathcal{S}$.
3. $\forall s \in \mathcal{S}$ picked randomly:
   (a) $\forall n \in \mathcal{N}(v), n \notin \mathcal{S}$ picked randomly:
       i. Obtain a new sample $\mathcal{X}$ removing $s$ and adding $n$ ($\mathcal{X} \leftarrow \mathcal{S} \bigcup \{n\} \setminus \{s\}$).
       ii. Compute the optimization value $P$ for sample $\mathcal{X}$.
       iii. If $P > O$, set $\mathcal{S} \leftarrow \mathcal{X}, O \leftarrow P$, and go to step 3).
4. $\mathcal{S}$ contains the resulting set of nodes with collocated monitoring sites, and $O$
   contains the optimization value of the placed monitoring sites.
```

The MSS searches for large, non-interference, equal-sized coverage areas. Starting from $\mathcal{S}$ monitoring sites, the MSE maximizes the unique coverage U and minimizes the difference D between the maximum C_{max} and minimum C_{min} sizes of the resulting network coverage areas. These measures are normalized taking into account the total number of network nodes V. The U measure is proposed in order to take into account the interference I between the coverage areas of each monitoring site that we want to minimize. In that way, the maximization of the unique coverage U also minimizes the interference I between coverage areas.

Function 1. Monitoring sites evaluation (MSE) function.

```
Input: $\mathcal{G} \leftarrow \{\mathcal{V}, \mathcal{E}\}$: wastewater network with node set $\mathcal{V}$ and link set $\mathcal{E}, V = |\mathcal{V}|$.
        $\mathcal{K}$: set of (indices) of the monitoring sites to test, $K = |\mathcal{K}|$.
        $\mathcal{C}(k), k \in \mathcal{K}$: set of nodes that are present in the monitoring site $k$ coverage
  area; $\mathcal{C}(k) \in \mathcal{V}$.
Output: $O$: optimization value of the provided monitoring sites sample $\mathcal{K}$.
1. Initialize the set of nodes covered by a unique monitoring site $\mathcal{U}$. $\mathcal{U} \leftarrow \varnothing$
2. Initialize the set of interference nodes, covered by multiple monitoring sites $\mathcal{I}$.
   $\mathcal{I} \leftarrow \varnothing$
3. Obtain the maximum $C_{max}$ and minimum $C_{min}$ values of nodes covered by each
   single monitoring site $k, \forall k \in \mathcal{K}$. $C_{max} = \max(|\mathcal{C}(k)|, k \in \mathcal{K}), C_{min} = \min(|\mathcal{C}(k)|, k \in \mathcal{K})$
4. $\forall k \in \mathcal{K}$:
   (a) $\forall v \in \mathcal{C}(k)$:
       i. If the node $v \notin \mathcal{I}$ and $v \notin \mathcal{U}$, then add $v$ to $\mathcal{U}$. $\mathcal{U} \leftarrow \mathcal{U} \bigcup \{v\}$
       ii. Otherwise if the node $v \notin \mathcal{I}$ and $v \in \mathcal{U}$, then remove $v$ from $\mathcal{U}$ and add it to
           $\mathcal{I}$. $\mathcal{U} \leftarrow \mathcal{U} \setminus \{v\}$ $\mathcal{I} \leftarrow \mathcal{I} \bigcup \{v\}$
5. Compute and return the optimization value that is the difference between the
   unique coverage $U = |\mathcal{U}|$ and the difference between $C_{max}$ and $C_{min}$. This is also
   normalized with $V$. There are also two weight variables $w, y$ (1 by default) that
   could be modified in order to prioritize one measure over the other. $o \leftarrow$
   $\frac{(w \times U - y \times (C_{max} - C_{min}))}{V}$
```

The MSS needs to be computed several times (i.e., iterations) to find the best K node combination to place the monitoring sites on $\mathcal{S}$ according to the MSE. The optimal number of required iterations may vary depending on sewage network size and topology. It is up to the network administrator to define the number of iterations as an stop criteria upfront, that may be input manually by the user or an automated decision based on the accumulated O value improvement in the MSS algorithm iteration results.

2.2.2. Dynamic monitoring site selection algorithm

After running the MSE Algorithm 1 and obtaining a subgraph of the sewage network with a monitoring site (sensor) s as output (i.e., $\mathcal{G}(s)$), we then use the dynamic monitoring approach proposed in Larson et al. (2020) to home in on either a possible patient zero or the hot spot neighborhood in that subgraph when its sensor s detects SARS-CoV-2 RNA traces $\mathcal{F}$. Larson et al.'s approach consists of assigning Bayesian probabilities of infection to all possible *source nodes* based on professional beliefs and applying a “binary search” using these probabilities. Our implementation assigns the Bayesian probabilities according to inhabitants connected to each node as we believe that the higher the population in the area is, the greater the chances are of finding infected people.

Two algorithms have been implemented: i) the Patient Zero (PZ) algorithm and ii) the Hot Spot (HS) algorithm. The PZ algorithm assumes there exists only one case of COVID-19 in a community (Patient Zero) and tries to find the minimum sequence of manholes to test in order to locate that first source of infection. The HS algorithm, on the other hand, assumes that many individuals are already infected and seeks to find the cluster in the sewage network with the largest SARS-CoV-2 RNA load; in other words, locate the hot spot.

The HS algorithm works as follows. At each iteration it seeks for the manhole whose Bayesian probability of infection is the highest. After testing it, if the viral load $\mathcal{F}'$ is high compared to the previously tested manhole, we know that the infected area is upstream from this point and we can discard all the network nodes downstream. Otherwise, the upstream nodes are discarded. Hence, at each iteration the population associated to the remaining nodes is approximately the same as those associated to the eliminated ones. In order to simplify the simulation, for each iteration we assume that the viral load $\mathcal{F}'$ is boolean. The algorithm halts when the *stopping rule*, which is defined by the user, is reached. The PZ algorithm is a special instance of the HS algorithm, where the *stopping rule* is reached when there is only one source node left.

Both PZ and HS algorithms share a prior three-step process which has been defined below for the sake of clarity.

Given $\mathcal{G}(s) \leftarrow \{\mathcal{V}(s), \mathcal{E}(s)\}, c_v$ is the population associated to the node $v \in \mathcal{V}(s)$:

i **Create artificial nodes:** All *source nodes* must have $deg^{-}(v) = 0$. To achieve this, for each inner node in the graph (so-called “original node”) with the associated population, we create a new node (so-called ”artificial node”) with the same associated population and connected to that original node. From any artificial node we can easily obtain its original node. Let $\mathcal{A}$ be the set with all artificial nodes from $\mathcal{G}(s)$.

(a) Create empty set $\mathcal{A}$

(b) $\forall v \in \mathcal{V}(s)$ such that $deg^{-}(v) > 0$ and $c_v > 0$, add new node v' and edge (v', v) to $\mathcal{G}(s)$, make the citizens $c_{v'} \leftarrow c_v$, define $original(v') \leftarrow v$, add v' to $\mathcal{A}$.

(c) Return $\mathcal{A}$

ii **Simplify and normalise:** Dispense with useless nodes for the calculus in order to simplify the graph and assign a Bayesian probability to the *source nodes* based on its associated population.

(a) $\forall v \in \mathcal{V}(s)$, if $deg^{-}(v) = 0$ and $c_v = 0$, remove it.

(b) $\forall v \in \mathcal{V}(s)$ with $deg^{-}(v) = 1, deg^{+}(v) = 1$ and $c_v = 0$, add edge to $\mathcal{G}(s)$ going from predecessor of v to the successor of v. Remove vertex v

(c) Let $\mathcal{P} \subset \mathcal{V}(s)$ be the set of all $v \in \mathcal{V}(s)$ such that $deg^{-}(v) = 0$.

(d) Normalise $\mathcal{P}$: $\forall p \in \mathcal{P}, probability(p) \leftarrow \frac{c_p}{\sum_{p \in \mathcal{P}} c_p}$

(e) Return $\mathcal{P}$

iii **Propagate probabilities:** Propagate probabilities from *source nodes* to all other nodes, such that each inner node's probability is the sum of probabilities of upstream nodes. Let $\mathcal{P}$ be the set obtained after **simplifying and normalising** $\mathcal{G}(s)$.

```
    (a) $\forall v \in (\mathcal{V}(s)\backslash\mathcal{P})$, and $\mathcal{W}$ is all the predecessors of
        $v$ : $probability(v) \leftarrow \sum_{w \in \mathcal{W}} \frac{probability(w)}{deg^{+}(w)}$
```

Algorithm 2. Hot Spot detection algorithm.

```
Input: $\mathcal{G}(s) \leftarrow \{\mathcal{V}(s), \mathcal{E}(s)\}$: Directed graph such that it only has one node $s \in \mathcal{S} | deg^{+}(s) = 0$ corresponding to sensor node.
        $c_v \mid \forall v \in \mathcal{S}, c_v$ are the citizens of $v$.
        $\mathcal{F}$: Viral load detected in sensor node $s$.
Output: $\mathcal{T} | \mathcal{T} \subset \mathcal{V}(s)$: Nodes of Hot Spot neighborhood.
1. $\mathcal{A} \leftarrow$ Create artificial nodes
2. $\mathcal{P} \leftarrow$ Simplify and normalise graph
3. Define stopping rule.
4. While not stopping rule:
    (a) Propagate probabilities
    (b) Find node $t \in (\mathcal{V}(s)\backslash\mathcal{A})$ such that $t \leftarrow \min_{v \in (\mathcal{V}(s)\backslash\mathcal{A})} \left| probability(v) - \mathcal{F}/2 \right|$. In case of a tie, choose the node with the larger probability.
    (c) Let $\mathcal{G}'$ be a subgraph of $\mathcal{G}(s)$ having $t$ and all the ancestors of $t$.
    (d) Test node $t$.
        i. $\mathcal{F}' \leftarrow$ detected viral load
        ii. If $\mathcal{F}' \geqslant \mathcal{F}/2$ then $\mathcal{G}(s) \leftarrow \mathcal{G}', F \leftarrow F'$
        iii. Else, $\mathcal{G}(s) \leftarrow \mathcal{G}(s)\backslash\mathcal{G}', \mathcal{F} \leftarrow \mathcal{F} - \mathcal{F}'$
    (e) $\mathcal{P} \leftarrow$ Simplify and normalise graph
5. $\mathcal{T} \leftarrow \mathcal{V}(s)\backslash\mathcal{A}$
6. Return $\mathcal{T}$
```

2.3. Case study

The usefulness of the algorithms was illustrated with the sewage network of Girona (Girona, northern Catalonia, Spain). Girona is a city of 101,852 inhabitants, with a metropolitan area shaped by seven municipalities that together have a total of 148,504 inhabitants (Source: 2019 electoral roll). Girona is a typical compacted western Mediterranean city, with mixed uses and clearly divided between the old town and the modern peripheral. It extends 39.1 km^2 at the confluence of the Ter, Onyar, Galligants, and Güell rivers and has a population density of 2,605 inhab./km^2. The Girona sewage network consists of 9,718 manholes with a total number of 148,504 inhabitants connected on them, resulting in a large network of 13.79 km in diameter and with a total of 338 kms of pipes. The basic topological characteristics of the network are: 9,718 nodes (V); 10,185 edges (E); an average node degree of 2.1 ($\overline{D}$); a diameter of 9,718 ($\varnothing$); and an average shortest path length of 9,580 ($\overline{d}$).

The topological data from the community sewer network was provided by the municipality of Girona through GIS that included feature geometry, attributes, etc. First, these files were combined to generate a GraphML file format which is compatible with the Network Robustness Simulator (NRS) (BCDS, 2021) used for graph analysis. The output format is a unique file in GraphML format which contains both nodes and edges, including their attributes. GraphML is an XML based format (GraphML, 2001). Next, a data verification and reconciliation approach was followed. The obtained graph was then checked for inconsistencies in disconnected nodes and/or edges. It was also important to check the additional data for outliers and discuss possible errors with the water company to ensure greater precision.

The citizens living in a household are estimated to be 2.7 citizens per household. This assumption is taken from the ratio of inhabitants in 2019 in Girona (101,852 citizens) and the surrounding villages connected to the sewer system, including Salt (31,362 citizens), Vilablareix (2,897 citizens), Sarrià de Ter (5,170 citizens), Aiguaviva (756 citizens), Fornells de la Selva (2,650 citizens) and Sant Gregori (3,817 citizens), which gives a final total of 148,504 citizens. These data were obtained from the Catalan Statistics Institute (idescat, 2019) and are divided by the total number of households (53,466 households in 2019).

The cadastral parcels from the city were associated to each manhole. The cadastral database for Girona was downloaded from the official Spanish Spatial Data Infrastructure, which is based on the European INSPIRE Directive (2007/2/EC), and transformed into a geojson file that contained the geometries as well as the alphanumeric information linked to the cadaster parcels. Next, each cadastral parcel was linked to the nearest manhole following a negative slope, adhering to the assumption that water is transported by gravity. The official DEM from the Catalan Cartographic and Geological Institute at a 2×2 resolution (ICGC, 2020) was used to estimate the z coordinate of the centroid of each cadastral parcel and manhole; hence, each cadastral parcel was connected to the nearest sewer origin with an equal or lower elevation. However, to overcome EDM and other inaccuracies, when the distance between the parcel and the manhole exceeded a defined threshold (100 m in this case), the algorithm searched for the closest higher manholes in a progressive way (1 m added in each new search) until the distance was lower than the threshold or the maximum z tolerance was reached (3 m in this case). To run these calculations, the scripts were developed on a PyQGIS console on QGIS v. 3.10 (QGIS, 2008). The number of inhabitants connected to each manhole was used as an input to the dynamic sensor placement algorithm to assign the Bayesian probabilities of infection to the source nodes. It was not used as an input to the static sensor placement algorithm.

3. Results and discussion

Below, we discuss the numerical results that illustrate the considerations of this paper. For that purpose, we used the Girona sewage network instance, i.e., *girona-wastewater*.

3.1. Static monitoring site selection

The results obtained for the case study of Girona confirm that the MSS algorithm performed well. The solution obtained for each of the eight tests (each of them fixing the number of monitoring sites from one to eight, $K = \{1,2,\ldots,8\}$), result in a coverage (in manholes) larger than 60%, an interference smaller than 3% and a maximum difference of 25% amongst monitoring sites coverage (Table 2).

The solution obtained when fixing one monitoring site is arbitrary, as it corresponds to the selection of the sampling point at the end of the sewage network (the entrance of the WWTP) with 100% coverage and 0% interference. For the remaining tests, an optimal solution was found which balances coverage, interference and manholes' coverage equity amongst sites. The results show that when fixing two and three monitoring sites, the coverage reduces down to 65%; after fixing four or more monitoring sites the total coverage is always larger than 75% (Fig. 2a).

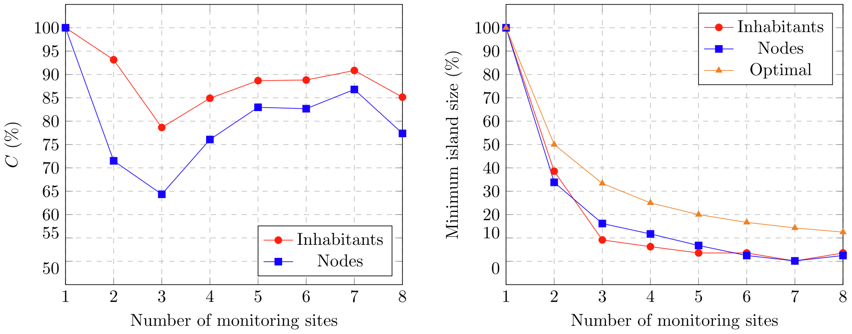
The algorithm uses the number of manholes as an input (not the inhabitants connected to each manhole). As the MSS algorithm minimizes the difference of the number of nodes between the largest and the smallest monitoring site coverage areas (D), our results show adequate minimum island sizes from one to five monitoring site placements (Fig. 2b) (up to a maximum of five monitoring site placements is recommended in the *girona-wastewater* network as a larger number of placements result on small-sized islands, which should be avoided). The minimum island sizes for nodes and inhabitants are compared with the theoretical optimal solution, which considers that all of the obtained monitoring site coverage areas are equally sized. The solutions provided show a good correlation between the number of manholes and inhabitants covered for each monitoring site coverage area for all tests, this is the particular case in the Girona catchment with a population density range of 17.7–759.9 inh/km^2 amongst neighborhoods.

The results show that the larger the number of monitoring sites the smaller the distance from the furthest node to the respective sampling points (ID values in the table). Given the potential attenuation of RNA signal along the sewage network transport Hart and Halden (2020) a constraint might be added in the selection of monitoring sites related to the maximum distance between the points of discharge of SARS-CoV-2 RNA traces and each monitoring site.

Table 2
MSS results data for *girona-wastewater* network, from one to eight monitoring sites.

K	*C* (%)	*I* (%)	*D* (%)	IN	WD (m)	ID (m)	IH	HC (%)
1	100	0	0	9718	0	13785	151248	100
2	71.51	2.45	6.4	3909, 3287	4561, 4457	6513, 9328	58331, 82561	93.15
3	64.35	0.01	8.3	2378, 2306, 1571	6247, 4452, 4618	7538, 6618, 4232	61007, 44102, 13845	78.65
4	76.07	0.01	12.76	2378, 2306, 1571, 1138	6247, 4452, 4618, 4182	7538, 6618, 4232, 3655	61007, 44102, 13845, 9468	84.91
5	82.97	0.01	17.83	658, 1571, 2306, 2391, 1138	812, 4618, 4452, 6021, 4182	3825, 4232, 6618, 7764, 3655	5404, 13845, 44102, 61313, 9468	88.68
6	82.67	0.1	21.24	1138, 240, 2143, 658, 1571, 2294	4182, 6255, 6253, 812, 4618, 4756	3655, 1465, 7531, 3825, 4232, 6319	9468, 7920, 53589, 5404, 13845, 44102	88.81
7	86.79	0.44	24.31	2306, 658, 408, 2379, 17, 1571, 1138	4452, 812, 4447, 6239, 5085, 4618, 4182	6618, 3825, 2778, 7546, 214, 4232, 3655	44102, 5404, 3401, 61007, 186, 13845, 9468	90.85
8	77.38	0.99	15.15	240, 1712, 1512, 1130, 596, 658, 1138, 630	6255, 6371, 6263, 4736, 5317, 812, 4182, 6301	1465, 4699, 4607, 4034, 2376, 3825, 3655, 7484	7920, 25075, 47133, 11067, 16254, 5404, 9468, 6456	85.14

K – number of placed monitoring sites, *C* (%) – normalized network nodes coverage (*C*, in %), *I* (%) – normalized network nodes interference (*I*, in %), D (%) – normalized difference between C_{max} and C_{min} (*D*, in %), IN – islands number of nodes, WD – distance between the WWTP and each node where the monitoring sites are placed (in meters), ID – islands diameter (i.e., monitoring sites coverage areas diameter, in meters), IH – island number of inhabitants (inhabitants size), HC – inhabitants coverage (in %).



(a) Coverage.

(b) Minimum island sizes.

Fig. 2. Coverage and minimum island sizes (one to eight monitoring sites, *girona-wastewater*).

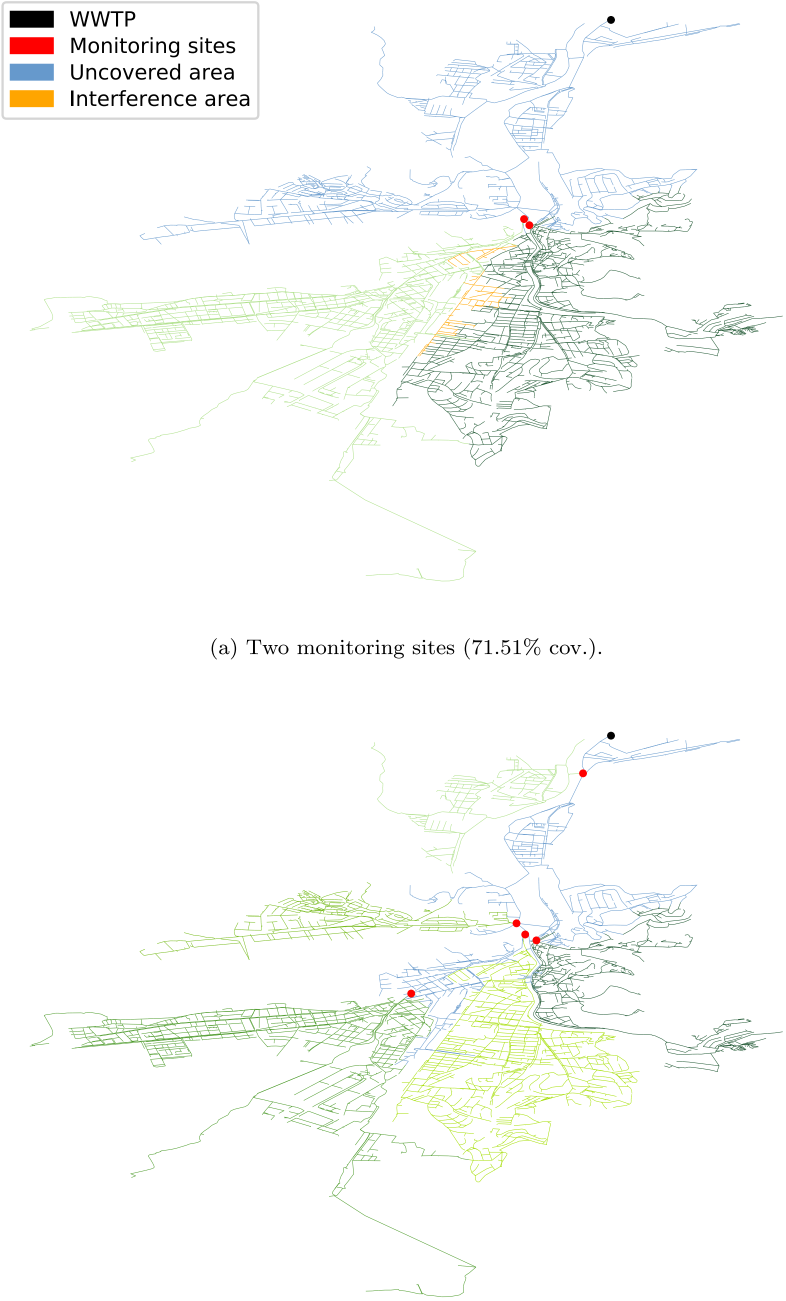
Fig. 3 shows the results for the tests that fixed two and five monitoring sites. When fixing two monitoring sites, the two covered areas (in green) represent 71% of the manholes. The uncovered areas (in blue) are the ones located further downstream in the sewage network (closest to the WWTP). When fixing five monitoring sites, the coverage increases (up to 83%) and smaller residential areas (as compared to the ones covered by the fixing two sites test) are included. Again, the manholes in the areas close to the WWTP cannot be captured in the final solutions because of their small number of manholes and their potential to generate interference as they are located downstream.

This is the first time that a static monitoring site selection algorithm has been proposed for SARS-CoV-2 monitoring at fine spatial resolution. The static monitoring site selection algorithm can effectively be applied to municipalities wishing to monitor SARS-CoV-2 RNA traces at a finer scale than an entire municipality. The presence of RNA traces is normally analyzed in tandem with health (number of infected cases, (Medema et al., 2020)), demographic and socioeconomic indicators (Wu et al., 2020). The module developed in this study which links each household to a manhole is essential to be able to aggregate these indicators (from individuals to inhabitants connected to a specific monitoring site). In the particular case of Catalonia, the public COVID-19 prevalence data is only aggregated at the municipality level and at the ABS (primary health area, which are areas defined around primary care health centers in the city) level. As an illustration exercise, we fixed the placement of sensors at the end of each ABS (the downstream manhole of each ABS subcatchment) and estimated the interference. Appendix A shows that locating sensors as a function of ABS results in large interference amongst monitoring sites and hence would not be the preferred option. Therefore, municipalities should make a request to the health authorities to aggregate prevalence data according to the areas covered by the sampling points. The main limitation to bringing the approach into practice is the data quality on the topology of the sewage network; data reconciliation is the most time-consuming step. Launching the optimization for each of the tests took between one and three minutes on a modern laptop (CPU Ryzen 5 4800U, 16 GB RAM). Finally, the approach is equally valid for the placement of monitoring sites for purposes other than tracking the spread of COVID-19, such as estimating the consumption of pharmaceuticals (Escolà Casas et al., 2021) and illicit drugs (González-Mariño et al., 2020) at fine spatial resolution, or detecting illicit discharges of pollutants from industries (Sambito et al., 2020).

3.2. Dynamic monitoring site selection

The dynamic monitoring site selection algorithm is executed over both the entire sewage network (cases 1 and 2) and a reduced network, which includes 1,099 manholes (equivalent to 44,102 citizens), found as an outcome of the static monitoring algorithm (cases 3 and 4). For cases 1 and 3, the Bayesian probabilities of source nodes are assigned according to the connected inhabitants, while for cases 2 and 4 the probabilities are assigned using random numbers from a unit probability distribution. Cases 2 and 4 would be comparable to the methodological proposal from Larson et al. (2020).

Patient Zero. When applying the PZ algorithm to the entire network and with probabilities as a function of population (case 1), the sampling iterations required to identify the first individual discharging SARS-CoV-2 RNA traces in the sewage system range from 8 to 15. When the probabilities are randomly assigned (case 2) the range is smaller, between 10 and 14 iterations. Looking at the median of the distributions one less iteration would be needed when assigning the probabilities as a



(a) Two monitoring sites (71.51% cov.).

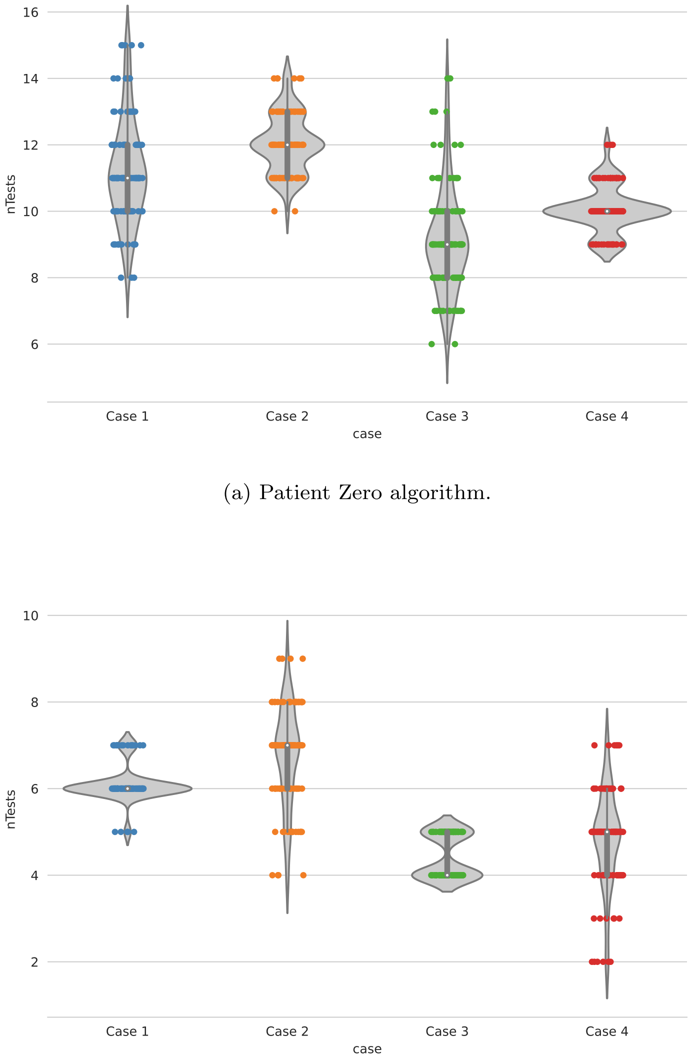
(b) Five monitoring sites (82.97% cov.).

Fig. 3. Coverage areas for two and five monitoring sites placement (*girona-wastewater*).

function of the population. This gain might increase in cities with larger variability in population density amongst neighborhoods (in the case of Girona, the population density across neighborhoods varies between 17.66 and 759.94 inh/km^2).

The application of the PZ algorithm to the reduced network results in a decrease of two iterations to identify the first individual discharging SARS-CoV-2 RNA traces (cases 3 and 4 as compared to 1 and 2). Overall, the strategy resulting in a smaller number of iterations is case 3, which involves departing from a reduced network resulting from the static monitoring site selection algorithm and with the assignment of probabilities as a function of connected inhabitants. The frequency distribution of the number of required samples for the PZ algorithm is shown in Fig. 4a.

Hot Spot. The HS algorithm allows hot spot areas with a large number of potential infected people to be identified (we set 3000 infected inhabitants as the *stopping rule* in this exercise). As compared to the PZ algorithm, the number of iterations reduces by five no matter the case. As shown in Fig. 4b, when assigning the probabilities as a function of the connected inhabitants, one less iteration is needed (looking at the median of the distributions); yet, the spread of the distribution of the resulting number of iterations is much smaller as compared to the assignment of uniform probabilities; this is the opposite of what is shown in Fig. 4a. The most favourable (and conservative) option would be case 3 with large certainty that with four or five iterations the hot spot area would be identified. As compared to case 1, one or two iterations would be saved when departing from a sewage network obtained from the



(a) Patient Zero algorithm.

(b) Hot Spot algorithm.

case 1 – entire wastewater network and Bayesian probabilities as a function of population.
case 2 – entire wastewater network and Bayesian probabilities randomly assigned.
case 3 – island from the MSS algorithm and Bayesian probabilities as case 1.
case 4 – island from the MSS algorithm and Bayesian probabilities as case 2.

Fig. 4. Distribution of sampling points required.

static approach.

The dynamic algorithms are comparable to the ones discussed in Larson et al. (2020), where the Bayesian probabilities are assigned randomly in the same way as we did in both cases 2 and 4. The enhancements in this paper relate i) to the use of the outcome from the static sensor placement algorithm as a starting point which allows the overall number of samples needed to be reduced in both PZ and HS algorithms, and ii) to the addition of information about the number of inhabitants to each manhole which also allows the number of samples on average to be reduced; the latter is relevant in cities like Girona, which show high spatial variability in population density (and hence in the number of inhabitants connected to a manhole). As stated in Larson et al. (2020), the usefulness of the dynamic approach to detect a hot spot within a city is constrained by the availability of devices which offer a fast response in the detection and quantification of SARS-CoV-2 RNA traces. Current methods imply the transport of the samples to a lab where the RNA traces are concentrated (e.g. (Forés et al., 2021)) and then the qPCR is executed, overall with a result being available in between 24 and 48 h. In case 3, iterations are needed to locate the Hot Spot, which means that between three and six days would be needed in total, which is probably too slow to make a decision on an effective mitigation action.

The applicability of the dynamic monitoring site selection algorithms is also constrained by the detection limit of the SARS-CoV-2 analysis in sewage. The lowest incidence resulting in quantifiable SARS-CoV-2 concentration in wastewater differed between community sizes; Rusiñol et al. (2021) found the lowest quantifiable incidence to be 0.11 and 0.82 cases per 1,000 inhabitants for the large and small sized communities respectively and Hata et al. (2021) reported 0.05–0.10 detectable cases per 1,000 inhabitants. Hart and Halden (2020) reported that under a best-case scenario of no in-sewer RNA signal loss, wastewater generation (50–500 L/person/d) and virus shedding (56.6 million-113.2 billion viromes/d) are important variables determining the detectability in community wastewater of a single infected person among one hundred to two million healthy individuals, assuming homogeneous distribution of cases. In the case of SARS-CoV-2 it is really challenging to detect the zero patient; the patient zero might be moving around the city, but also the limit of detection of the analysis of SARS-CoV-2 in wastewater might not allow to detect 1 infected amongst the surveilled community. Results of patient zero are provided in this paper to compare the performance of the algorithm against Larson et al. (2020). Yet, the patient zero algorithm can be used for other purposes than SARS-CoV-2 surveillance, such as the detection of an illegal industrial discharges in sewer systems. Furthermore, there are substantial uncertainties in estimating SARS-CoV-2 loads (Li et al., 2021) which propagate to the calculation of increase or decrease of virus load between two collected samples; high-frequency flow-proportional sampling would reduce uncertainties (yet this is challenging at the intra-city scale) as well as using surrogate viruses as internal or external standards during the analysis, and further improvement on analytical approaches.

Future work will be conducted to connect the algorithm to a mechanistic model that includes SARS-CoV-2 concentration as a variable, using a hydraulic model to estimate the dilution capacity in the sewer network, a SARS-CoV-2 load generation pattern and implementing the in-sewer degradation of SARS-CoV-2. The concentration can then be used by the algorithm as a criteria or a constraint to define best sampling sites. Some attempts in that sense have been published for SARS-CoV-2 (Hart and Halden, 2020) and for other pathogens (Ranta et al., 2001; Wang et al., 2020).

4. Conclusions

This paper demonstrates that it is possible to optimally select sampling points for SARS-CoV-2 sewage surveillance in cities. An algorithm is proposed for the placement of a predefined number of monitoring sites which result in maximum coverage of manholes and minimum interference amongst them (static sensor placement). Two other algorithms are proposed to dynamically sample and analyze to identify patient zero and hot spots in cities (dynamic sensor placement). For the case study of Girona, a static sensor placement of five monitoring sites (or more) results in a coverage greater than 80% of both manholes and inhabitants. The best option for detecting a patient zero and a hotspot area implies assigning probabilities as a function of the number of inhabitants connected to each manhole. Results have demonstrated that when using these probabilities our proposed algorithms enhanced previous proposals in all presented scenarios. As a conclusion for the city of Girona, 11 iterations would be needed to detect the patient zero, and six iterations for identifying a hotspot of about 3,000 infected inhabitants. In the case of combining both algorithms, the number of iterations can be reduced to nine and four, respectively.

CRediT authorship contribution statement

Eusebi Calle: Conceptualization, Funding acquisition, Writing – review & editing, Supervision, Project administration. **David Martínez:** Methodology, Software, Formal analysis, Visualization, Writing – review & editing, Data curation. **Roser Brugués-i-Pujolràs:** Software, Visualization. **Miquel Farreras:** Software, Visualization. **Joan Saló-Grau:** Software, Visualization. **Josep Pueyo-Ros:** Methodology, Data curation, Software, Writing – review & editing. **Lluís Corominas:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

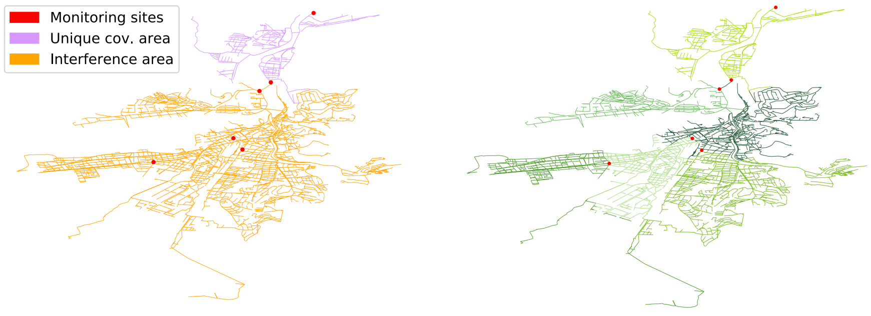
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Lluís Corominas acknowledges the Ministry of Economy and competitiveness for the Ramon and Cajal grant (RYC-2013-14595) and its corresponding I3 consolidation. The authors acknowledge the project VirWASTE (2020PANDE00044) funded by AGAUR. The authors acknowledge the CLEaN-TOUR (CTM2017-85385-C2-1-R) and INVEST (RTI2018-097471-B-C21) projects from the Spanish Ministry of Economy and Competitiveness and thank Generalitat de Catalu-nya through Consolidated Research Group 2017 SGR 1318. ICRA researchers thank funding from the CERCA program. UdG researchers thank funding from Red temática Go2Edge (Ref.: RED2018-102585-T) and Ajut PontUdG2020/23.

Appendix A. ABS monitoring site selection

ABS (Área Básica de Salut) or Basic Health Area, is the clustering method used by the Spanish government to divide a city into different areas. The main criteria is that all of them include at least one primary Health-care facility. A priori it would be interesting to establish the monitoring points considering these areas. However, if we locate the monitoring points considering only the coverage of these areas the 'interference' between them can report a huge error in the expected results. Fig. 5 shows the *girona-wastewater* network placing $k = 6$ monitoring sites, each one monitoring one of the six ABS areas on the network. Each ABS monitoring site is placed in the closest node from an ABS area to the sewage treatment plant. The large level of interference provided by this approach cannot be considered as a feasible solution (88.79% of interference depicted on the orange area (Fig. 5a)). This is produced, as expected, because part of the monitoring points are located closed to the WWTP, covering by themselves the major part of the city.



(a) ABS interference area.

(b) ABS areas.

Fig. 5. ABS monitoring site selection ($k = 6$, 88.79% of interference).

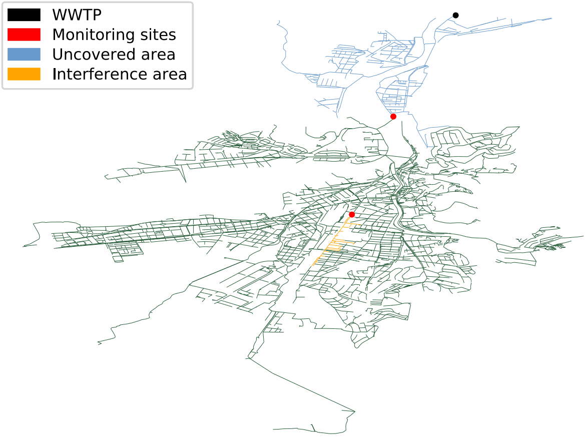


Fig. 6. Interference between two ABS areas.

Consequently, the rest of the monitoring points, located downstream from the WWTP, overlap the covered area. Moreover, in the case of reducing the number of ABS monitored areas, the interference between areas would still persist. This is described in Fig. 6 where the six ABS areas (in green) are also shown. In this case, the interfence area reports a percentage of 1.22. Consequently, it is shown that using ABSs as a clustering method to monitor the city is not possible if the results have to be relevant (without interferences between areas).

References

Ahmed, W., Bertsch, P.M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B.R., Mueller, J.F., Simpson, S.L., Smith, W.J., Symonds, E.M., Thomas, K.V., Verhagen, R., Kitajima, M., 2020. Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. Sci. Total Environ. 739 (June), 139960.

BCDS, 2021. Network Robustness Simulator. Universitat de Girona, BCDS Research Group. http://nrs.udg.edu [Online; accessed 16-mar-2021].

Escolà Casas, M., Schröter, N.S., Zammit, I., Castaño-Trias, M., Rodriguez-Mozaz, S., Gago-Ferrero, P., Corominas, L., 2021. Showcasing the potential of wastewater-based epidemiology to track pharmaceuticals consumption in cities: Comparison against prescription data collected at fine spatial resolution. Environ. Int. 150, 106404.

EurEAU, 2017. Europe's water in figures. An overview of the European drinking water and waste water sectors. Technical report. The European Federation of National Associations of Water Services.

Forés, E., Bofill-Mas, S., Itarte, M., Martínez-Puchol, S., Hundesa, A., Calvo, M., Borrego, C.M., Corominas, L.L., Girones, R., Rusiñol, M., 2021. Evaluation of two rapid ultrafiltration-based methods for SARS-CoV-2 concentration from wastewater. Sci. Total Environ. 768, 144786.

González-Mariño, I., Baz-Lomba, J.A., Alygizakis, N.A., Andrés-Costa, M.J., Bade, R., Bannwarth, A., Barron, L.P., Been, F., Benaglia, L., Berset, J.D., Bijlsma, L., Bodík, I., Brenner, A., Brock, A.L., Burgard, D.A., Castrignanò, E., Celma, A., Christophoridis, C.E., Covaci, A., Delémont, O., Devoogt, P., Devault, D.A., Dias, M. J., Emke, E., Esseiva, P., Fatta-Kassinos, D., Fedorova, G., Fytianos, K., Gerber, C., Grabic, R., Gracia-Lor, E., Grüner, S., Gunnar, T., Hapeshi, E., Heath, E., Helm, B., Hernández, F., Kankaanpaa, A., Karolak, S., Kasprzyk-Hordern, B., Krizman-Matasic, I., Lai, F.Y., Lechowicz, W., Lopes, A., de Alda, M.L., López-García, E., Löve, A.S., Mastroianni, N., McEneff, G.L., Montes, R., Munro, K., Nefau, T., Oberacher, H., O'brien, J.W., Oertel, R., Olafsdottir, K., Picó, Y., Plósz, B.G., Polesel, F., Postigo, C., Quintana, J.B., Ramin, P., Reid, M.J., Rice, J., Rodil, R., Salgueiro-Gonzàlez, N., Schubert, S., Senta, I., Simões, S.M., Sremacki, M.M., Styszko, K., Terzic, S., Thomaidis, N.S., Thomas, K.V., Tscharke, B.J., Udrisard, R., van Nuijs, A.L., Yargeau, V., Zuccato, E., Castiglioni, S., Ort, C., 2020. Spatio-temporal assessment of illicit drug use at large scale: evidence from 7 years of international wastewater monitoring. Addiction 115 (1), 109–120.

GraphML, 2001. The GraphML File Format. http://graphml.graphdrawing.org/ [Online; accessed 21-dec-2020].

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.

Hata, A., Hara-Yamamura, H., Meuchi, Y., Imai, S., Honda, R., 2021. Detection of SARS-CoV-2 in wastewater in Japan during a COVID-19 outbreak. Sci. Total Environ. 758, 143578.

ICGC, 2020. Digital Elevation Model 5x5m. https://www.icgc.cat/Descarregues/Elevacions/Model-d-elevacions-del-terreny-de-5x5-m [Online; accessed 21-dec-2020].

idescat, 2019. El municipi en xifres. https://www.idescat.cat/emex/?id=170792 [Online; accessed 21-dec-2020].

Kang, O.Y., Lee, S.C., Wasewar, K., Kim, M.J., Liu, H., Oh, T.S., Janghorban, E., Yoo, C.K., 2013. Determination of key sensor locations for non-point pollutant sources management in sewer network. Korean J. Chem. Eng. 30 (1), 20–26.

Kesavan, H.K., Chandrashekar, M., 1972. Graph-Theoretic Models for Pipe Network Analysis. J. Hydraulics Div. 98.

Larson, R.C., Berman, O., Nourinejad, M., 2020. Sampling manholes to home in on SARS-CoV-2 infections. PLoS ONE 15.

Lenzen, M., Li, M., Malik, A., Pomponi, F., Sun, Y.Y., Wiedmann, T., Faturay, F., Fry, J., Gallego, B., Geschke, A., Gómez-Paredes, J., Kanemoto, K., Kenway, S., Nansai, K., Prokopenko, M., Wakiyama, T., Wang, Y., Yousefzadeh, M., 2020. Global socio-economic losses and environmental gains from the coronavirus pandemic. PLoS ONE 15 (7 July), 1–13.

Li, X., Zhang, S., Shi, J., Luby, S.P., Jiang, G., 2021. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chem. Eng. J. 415, 129039.

Mallapaty, S., 2020. How sewage could reveal true scale of coronavirus outbreak. Nature 580 (9), 176–177.

Mao, K., Zhang, H., Yang, Z., 2020a. An integrated biosensor system with mobile health and wastewater-based epidemiology (iBMW) for COVID-19 pandemic. Biosens. Bioelectron. 169 (January).

Mao, K., Zhang, H., Yang, Z., 2020b. Can a Paper-Based Device Trace COVID-19 Sources with Wastewater-Based Epidemiology? Environ. Sci. Technol. 54 (7), 3733–3735.

Matus, M., Duvallet, C., Soule, M.K., Sean M. Kearney, S., Endo, N., Ghaeli, N., Brito, I., Ratti, C., Kujawinski, E.B., Alm, E.J., 2019. 24-hour multi-omics analysis of residential sewage reflects human activity and informs public health. bioRxiv preprint.

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. Curr. Opin. Environ. Sci. Health 17, 49–71.

QGIS, 2008. A Free and Open Source Geographic Information System. https://www.qgis.org/en/site/ [Online; accessed 21-dec-2020].

Ranta, J., Hovi, T., Arjas, E., 2001. Poliovirus surveillance by examining sewage water specimens: Studies on detection probability using simulation models. Risk Anal. 21 (6), 1087–1096.

Rusiñol, M., Martínez-Puchol, S., Forés, E., Itarte, M., Girones, R., Bofill-Mas, S., 2020. Concentration methods for the quantification of coronavirus and other potentially pandemic enveloped virus from wastewater. Curr. Opin. Environ. Sci. Health 17, 21–28.

Rusiñol, M., Zammit, I., Itarte, M., Forés, E., Martínez-Puchol, S., Girones, R., Borrego, C., Corominas, L., Bofill-Mas, S., 2021. Monitoring waves of the COVID-19 pandemic: Inferences from WWTPs of different sizes. Sci. Total Environ. 787, 147463.

Sambito, M., Di Cristo, C., Freni, G., Leopardi, A., 2020. Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification. J. Hydroinformat. 22 (1), 46–60.

Schmidt, C., 2020. Watcher in the wastewater. Nat. Biotechnol. 38 (8), 917–920.

Villez, K., Vanrolleghem, P.A., Corominas, L., 2016. Optimal flow sensor placement on wastewater treatment plants. Water Res. 101 (1), 75–83.

Villez, K., Vanrolleghem, P.A., Corominas, L., 2020. A general-purpose method for Pareto optimal placement of flow rate and concentration sensors in networked systems – With application to wastewater treatment plants. Comput. Chem. Eng. 139, 106880.

Vonach, T., Tscheikner-Gratl, F., Rauch, W., Kleidorfer, M., 2018. A heuristic method for measurement site selection in sewer systems. Water (Switzerland) 10 (2), 1–16.

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

Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M.A., Bushman, M., Chai, P.R., Duvallet, C., Erickson, T.B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W.P., Huang, K.H., Lee, W.L., Matus, M., McElroy, K.A., Nagler, J., Rhode, S.F., Santillana, M., Tucker, J.A., Wuertz, S., Zhao, S., Thompson, J., Alm, E.J., 2020. SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. medRxiv.

Yazdi, J., 2018. Water quality monitoring network design for urban drainage systems, an entropy method. Urban Water J. 15 (3), 227–233.