


Article

Optimal Selection of Sampling Points within Sewer Networks for Wastewater-Based Epidemiology Applications

Yao Yao ¹, Yibo Zhu ², Regina Nogueira ³, Frank Klawonn ^{4,*} and Markus Wallner ²

¹ Institute for Information Engineering, Ostfalia University of Applied Sciences, Salzdahlumer Str. 46/48, 38302 Wolfenbüttel, Germany; yao.yao1@ostfalia.de

² Faculty of Civil and Environmental Engineering, Ostfalia University of Applied Sciences, Herbert-Meyer-Str. 7, 29556 Suderburg, Germany; yibo_zhu@web.de (Y.Z.); m.wallner@ostfalia.de (M.W.)

³ Institute of Sanitary Engineering and Waste Management, Leibniz University Hannover, Welfengarten 1, 30167 Hannover, Germany; nogueira@isah.uni-hannover.de

⁴ Biostatistics Research Group, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany

* Correspondence: f.klawonn@ostfalia.de

Abstract: Wastewater-based epidemiology (WBE) has great potential to monitor community public health, especially during pandemics. However, it faces substantial hurdles in pathogen surveillance through WBE, encompassing data representativeness, spatiotemporal variability, population estimates, pathogen decay, and environmental factors. This paper aims to enhance the reliability of WBE data, especially for early outbreak detection and improved sampling strategies within sewer networks. The tool implemented in this paper combines a monitoring model and an optimization model to facilitate the optimal selection of sampling points within sewer networks. The monitoring model utilizes parameters such as feces density and average water consumption to define the detectability of the virus that needs to be monitored. This allows for standardization and simplicity in the process of moving from the analysis of wastewater samples to the identification of infection in the source area. The entropy-based model can select optimal sampling points in a sewer network to obtain the most specific information at a minimum cost. The practicality of our tool is validated using data from Hildesheim, Germany, employing SARS-CoV-2 as a pilot pathogen. It is important to note that the tool's versatility empowers its extension to monitor other pathogens in the future.

Keywords: optimal sampling point; wastewater-based epidemiology; information theory; pathogen surveillance



Citation: Yao, Y.; Zhu, Y.; Nogueira, R.; Klawonn, F.; Wallner, M. Optimal Selection of Sampling Points within Sewer Networks for Wastewater-Based Epidemiology Applications. *Methods Protoc.* **2024**, *7*, 6. <https://doi.org/10.3390/mps7010006>

Academic Editor: Fernando Albericio

Received: 21 November 2023

Revised: 19 December 2023

Accepted: 1 January 2024

Published: 5 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wastewater-based epidemiology (WBE) provides near real-time information on public health status at the community level concerning specific pathogens and can potentially be a powerful tool for fighting pandemics [1–3]. However, studying pathogens based on WBE encompasses several significant challenges related to data representativeness, spatial and temporal variability, accurate population estimation, pathogen decay and dilution, and environmental confounders, strongly influencing virus detection capability. System sensitivity is introduced in this paper based on the mass balance model to standardize the monitoring process, which determines the detectability of sampling points in sewer systems. Utilizing reliable information from wastewater can help us harness the full potential of WBE to better comprehend and treat public health issues.

Due to the high cost and difficulties, only the inflow of wastewater treatment plants (WWTP), the endpoint of a sewer network, is generally used for WBE. However, this has various drawbacks. Signals from different catchments with different wastewater matrices are mixed beside the highly diluted inflow. Environmental conditions such as wastewater composition, temperature, or pH might impact virus decay and detection [4,5]. According to [6–9], there is an apparent information gain in sampling within the sewer network,

rather than only at the endpoints. Thus, this paper aims to use sampling points in the sewer networks besides the inflow of WWTP. The sewer network is a system of pipes and manholes. Generally, the network is distinguished in separate sewer systems, where only wastewater from the population and industry is collected, and combined sewer systems, where additional rainfall is drained within the same pipes.

We develop an entropy-based model to select optimal sampling points to obtain the most specific information using as few sampling points as possible to generate information about pathogens within a community, such as virus distribution, concentration, and developing trends. SARS-CoV-2, as a typical large epidemic, still affects the world today. Our approach is applied to it as a pilot parameter. The general procedure is meant to apply to any other pathogen to facilitate the early detection of outbreaks and the optimization of sampling strategies for profound success in public health interventions.

The structure of our tool combines a monitoring model to detect positive signals from regions with infected populations and an optimization program to select optimal sampling points in the sewer system. Previous studies have focused on the real-time identification of patient zero. Ref. [10] developed two theoretical methods based on binary search algorithms to identify hotspots and patient zero in real-time. However, the presented approach depends strongly on rapid wastewater testing at each target sampling point. Up to now, no rapid wastewater testing is available. Some studies have addressed the general problem of identifying optimal sampling points. In [11], two algorithms based on graph theory combined with greedy optimization were proposed to select sampling points based on approaches by [12]. However, the problem of dilution effects and other parameters, which might impact the detectability of SARS-CoV-2, is only marginally addressed in the studies mentioned above. In [13], a tool that transforms the problem into a min–max problem based on allocating population to a sewer network was designed. The sampling points for this network are minimized by maximizing the covering discharges. However, dilution effects still influence this tool, and its efficiency is only guaranteed in small cities. In [14], the initial concentration and decay rate of SARS-CoV-2 on the detection time and detection likelihood of the virus at downstream nodes were explored. Tools that can identify optimal sampling points were also developed. However, the results and tool only remain applicable to cities of less than 50,000 people. Our approach, in combination with the network topology and the settlement structure, can select the optimal number of sampling points according to the system sensitivity. The system sensitivity is defined as standardizing the virus detectability in the wastewater data, i.e., the minimal number of infected people needed to detect a positive signal. The system's applicability was tested in Hildesheim, Germany, with approximately 104,000 inhabitants.

This paper is structured as follows. Section 2 introduces the study area and data utilized in our research. Moreover, Section 2 outlines the general procedure and the specific parameters related to SARS-CoV-2, along with their uncertainties. It also introduces a mathematical approach grounded in information theory, which establishes connections between settlement structure as represented by residents and sewer topology. Section 3 presents the results, offering a practical perspective on the findings. Finally, in Section 4 we delve into the main discoveries of this study and explore potential avenues for future research.

2. Materials and Methods

2.1. Study Area

Hildesheim is a large city in northern Germany with approximately 104,000 residents. The urban catchment of Hildesheim is divided into 47 sub-catchments in this study based on the sewer network topology (Figure 1). Its main sewer network, which connects the sub-catchments via potential sampling points called candidate nodes, contains approximately 50 km of pipes. Wastewater flows through the sewer network to the northern part of the catchment. Then, it enters the Innerste River, receiving the catchment's water after the WWTP. As the receiving water is not part of our system, it is not discussed further here.

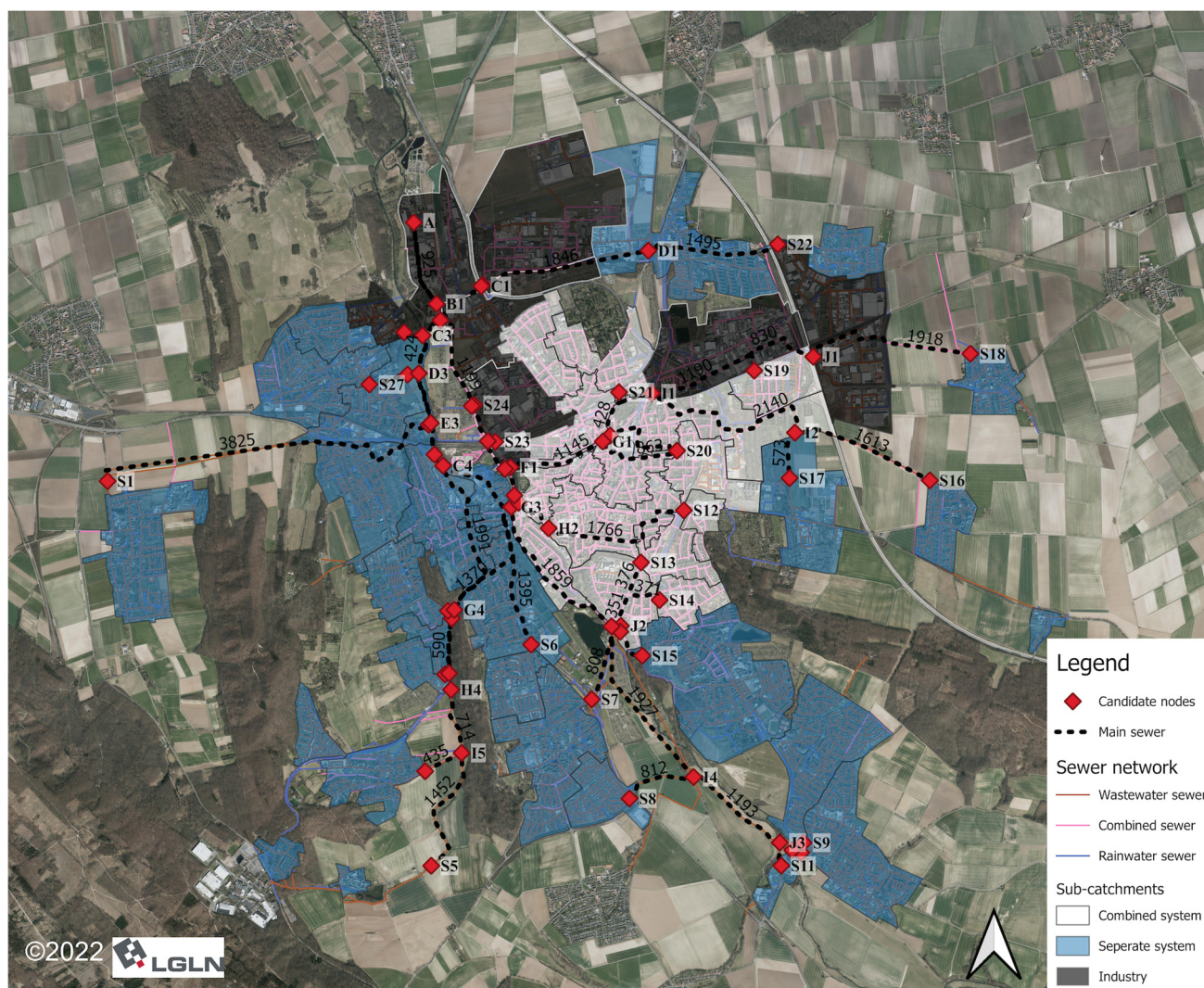


Figure 1. Study area: the main sewer network of Hildesheim with candidate nodes and sub-catchments (source: Landesamt für Geoinformation und Landesvermessung Niedersachsen/LGLN).

The city has two sewer networks: (i) a combined system with combined sewer (~135 km) in the center part of the city and (ii) a separate system with wastewater sewer (~270 km) and rainwater sewer (~300 km) in the outer districts.

2.2. Data

2.2.1. Sampling Data

First, for composite samples, we chose automatic samplers for sample collection. However, automatic samplers are expensive to acquire, maintain, and install, and sampling anywhere in the sewer system is labor- and equipment-intensive (see Figure 2a,b). Therefore, it is critical to select the ideal sampling point possible.

Second, we need selected manholes to be evaluated for overflow before installing the autosampler, as overflow could damage the autosampler. Nevertheless, there is still a risk (Figure 2c,d). Therefore, potential overflow can be used as a further selection criterion in the future, as long as the autosampler is used.

Finally, we also looked at comparing and selecting suitable samples. This is because the quality of the samples (e.g., some sites are prone to toilet paper clogging) and the different wastewater matrices (Figure 2e) can affect the detection of RNA. For calculating the incidences with the RNA concentrations in the water samples, normalization approaches, e.g., by using COD or biomarkers such as CrAssphage, exist to take the impact of the

wastewater matrix into account. In some cases, e.g., heavy rainfall events, a sampling is not recommended.

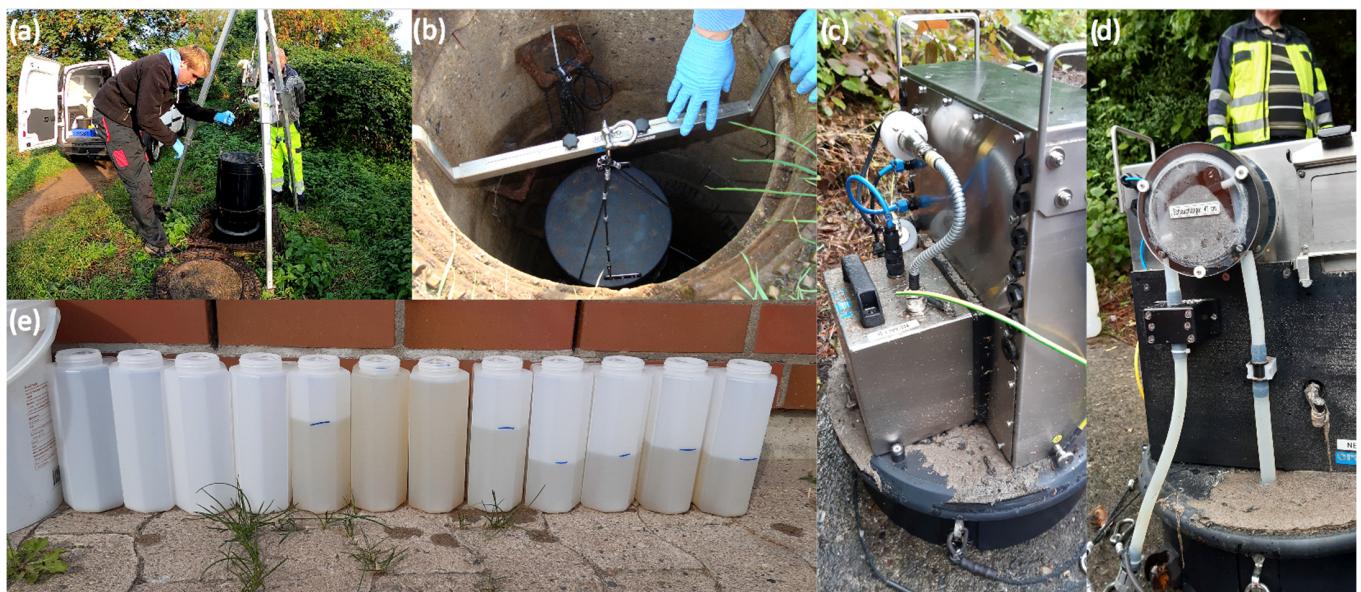


Figure 2. Photos from fieldwork: (a,b) installation of an autosampler in a manhole; (c,d) damaged autosampler due to surcharge; (e) quality of two-hour composite samples over 24 h from one autosampler.

2.2.2. Geographic Data

In this study, the population for each sub-catchment is derived based on the method of estimation of local population density in urban areas introduced by [15], i.e., the number of residents was proportional to the size of living space (the multiplication of building area at the residential area and building height) as the detailed population distribution is only available at the district level (14 districts in total).

Some sub-catchments in the northern part consist mainly of industry. This study does not include these sub-catchments as we focus primarily on domestic wastewater. In total, 63 candidate nodes are defined according to the sewer network's topology. They can be divided into two groups: (i) outlet of sub-catchments (named with capital letter "S" and number) and (ii) intersection nodes in the main sewer (named with capital letters other than "S" and number). Table 1 shows the data used for this study and their sources.

Table 1. Data used, with source.

Data Used	Data Source
Sewer network	Stadtentwässerung Hildesheim (SEHi) (2021) Stadt Hildesheim (2022), https://www.stadt-hildesheim.de/rathaus-verwaltung/buerger-und-ratsinfo/stadtteile/ (accessed on 1 December 2022)
Population statistics	Stadt Hildesheim (2015), https://www.stadt-hildesheim.de/wirtschaft-bauen/stadtplanung-und-stadtentwicklung/stadtentwicklung/flaechennutzungsplan/ (accessed on 1 December 2022)
Land use map	Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) 2022, https://opengeodata.lgln.niedersachsen.de/#dop (accessed on 1 December 2022)
Digital orthophoto (DOP)	LGLN (2022), https://opengeodata.lgln.niedersachsen.de/#lod2 (accessed on 1 December 2022)
3D building model ALKIS-Dataset	LGLN (2021), provided by SEHi

2.3. General Procedure

The general procedure applied to optimize the needed number and location of sampling points in sewer networks for WBE contains two sequential steps. In the first step, the

system sensitivity, i.e., the detectability of the SARS-CoV-2 virus, is defined. The sensitivity refers to a mass balance model utilizing equations from [16], covering the entire process, from virus RNA shedding through transport in the sewer system to wastewater sampling analysis in the lab. The mass balance model facilitates the sewer processes, providing sufficient results for further development. In the second step, an entropy-based mathematical model is formulated to optimize the selection of sampling points (manholes) in the sewer network for WBE applications.

2.4. System Sensitivity

The theoretically minimal number of infected individuals to detect positive signals in a region defines the system sensitivity, calculated based on the mass balance model. The model is presented in Figure 3, including equations from virus shedding to transport in the sewer network and sample analysis, considering the limit of detection of RNA fragments. In this study, RNA concentration in wastewater is calculated only on RNA in stool [17].

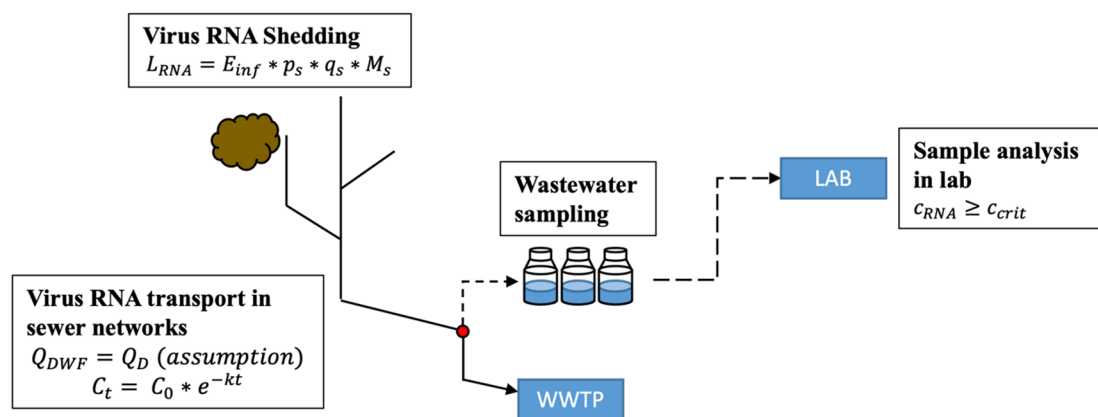


Figure 3. Mass balance model from RNA shedding to sample analysis, where C_t and C_0 are the concentrations of virus RNA in wastewater at time t and time 0, L_{RNA} (copies/day) is the detected RNA load, C_{RNA} (copies/mL) is the detected RNA concentration in the lab, C_{crit} (copies/mL) is the critical detection limit, Q_{DWF} (mL/day) refers to dry weather flow (only domestic wastewater Q_D is considered), p_s (-) is the virus RNA shedding probability in stool, q_s (mL/(person*day)) is the volume of stool produced per individual and day, M_s (copies/mL) is the virus RNA shedding magnitude in stool, k (/day) is the first-order decay value of RNA in wastewater, and t (day) is the flow time of wastewater in the sewer network from RNA input to sampling point.

Merging all variables from the mass balance model and rearranging allows us to calculate E_{inf}^{MIN} to mark a positive signal in a wastewater sample:

$$E_{inf}^{MIN} = \frac{C_{crit} * Q_{DWF}}{p_s * q_s * M_s * e^{-kt}}. \quad (1)$$

The stool volume, q_s , is calculated based on the feces production rate, feces density, and average water consumption (see details in Table 2). The dry weather flow, Q_{DWF} , only considers domestic sewage, assuming the sewers are in good condition without sewer infiltration and exfiltration. Moreover, the samples were taken on Sundays, with only a few industrial activities and impacts on wastewater runoff.

Table 2. Applied values of virus unspecific parameters.

Parameter	Value	Comments	Source
Feces production rate	128 g/(person*day)	Wet mass	[18]
Feces density	1.06 g/mL		[19]
Average water consumption	128 L/(person*day)	The value from 2019	[20]

Some other variables are virus-specific, namely c_{crit} , p_s , M_s , and k , and their values were obtained via a literature review (Table 3). A summary of their values in different literature references can be found in Appendix A Table A1.

Table 3. Statistical information on the virus-dependent parameters from the literature review ¹.

Factors	Labels	Unit	Min	25%	50%	75%	Max
RNA shedding magnitude	M_s	(\log_{10} copies /mL)	2.90	4.15	4.70	6.04	7.10
RNA shedding probability	p_s	(%)	10.1	29.0	48.1	54.5	83.3
RNA decay in wastewater	k	(/day)	0.06	0.09	0.14	0.26	0.67
RNA critical detection limit	c_{crit}	(copies/mL)	3.70	62.66	88.75	177.28	533.78

¹ Numbers in bold indicate the best value in each row.

2.5. Optimization of Sampling Point Location

Information theory can evaluate the degree of dependence or redundancy between monitors [21] and is widely applied to the design and/or evaluation of monitoring networks for hydrological applications [22]. This section will elaborate on its mathematical background for selecting the optimal location of sampling points in sewer networks.

2.5.1. Information Theory

Information is always viewed as a reduction in uncertainty. The Shannon entropy developed in information theory serves as a measure of information [23]. The information of event c with probability $p(c)$ is expressed as $-\log_2 p(c)$ [24].

In this study, the entropy is understood as the information capacity of signals and can be calculated for any random variable with a finite domain [25]. Different entropies and entropy-related measures are used to quantify the information content, namely entropy, joint entropy, and total correlation, as shown in Figure 4. Suppose circles A and B in Figure 4 are two sampling points. The size of each circle represents the gained information content.

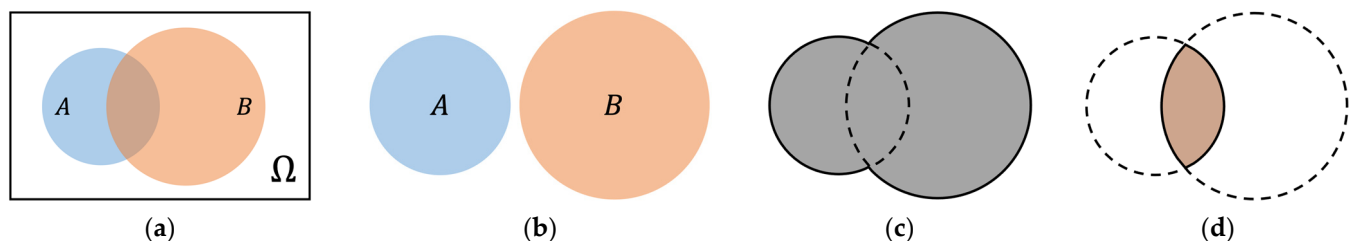


Figure 4. Venn diagrams describing (a) the total system, Ω , and two events (A and B), (b) individual entropy, (c) joint entropy, and (d) total correlation [26].

Figure 4b illustrates the entropy for each sampling point, C_i . In this example, the entropy of B is larger than A's, which means that B provides more information (reduces the uncertainty more substantially) than A. Formally, the entropy, $H(C_i)$, of a random variable, C_i , can be calculated by

$$H(C_i) = -\sum_{j=1}^{n_i} p(c_i^j) \log_2 p(c_i^j), \quad \sum_{j=1}^{n_i} p(c_i^j) = 1, i \in \{1, N\}, \quad (2)$$

where N is the number of random variables (in our case, the number of sampling points) and n_i is the number of all expected elementary events of random variable C_i with values c_i^j and their related probability distribution, $p(c_i^j)$. According to [24], the base of 2 in the logarithm is justified by the expected answers considering the monitoring location design, which is either "select" or "do not select" a sampling point.

Figure 4c represents the joint entropy, which shows the information content covered by both sampling points. If N random variables (C_1, C_2, \dots, C_N) are considered, the total

information content can be calculated by the joint entropy $H(C_1, C_2, \dots, C_N)$, which is defined as

$$H(C_1, C_2, \dots, C_N) = - \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \dots \sum_{j_N=1}^{n_N} p(c_1^{j_1}, c_2^{j_2}, \dots, c_N^{j_N}) \log_2 p(c_1^{j_1}, c_2^{j_2}, \dots, c_N^{j_N}), \quad (3)$$

where $c_i^{j_i}$ is the j_i -th elementary event of random variable C_i , (n_1, n_2, \dots, n_N) are the numbers of elementary events of corresponding variables (C_1, C_2, \dots, C_N) , and $p(c_1^{j_1}, c_2^{j_2}, \dots, c_N^{j_N})$ is the joint probability of events $(c_1^{j_1}, c_2^{j_2}, \dots, c_N^{j_N})$.

Another interesting measure in information theory is the total correlation $TC(C_1, C_2, \dots, C_N)$, which describes the shared information amount of N random variables. The total correlation can be expressed by the difference between the individual entropies and the joint entropy,

$$TC(C_1, C_2, \dots, C_N) = \sum_{i=1}^N H(C_i) - H(C_1, C_2, \dots, C_N), \quad (4)$$

where $\sum_{i=1}^N H(C_i)$ describes the sum of entropy of N random variables, (C_1, C_2, \dots, C_N) , and $H(C_1, C_2, \dots, C_N)$ stands for the joint entropy, calculated by Equation (3). Figure 4d illustrates the redundant information (total correlation) of two sampling points (the overlapped area of two circles).

2.5.2. Probability Distribution

As mentioned in the previous section, the key factor in information theory is the underlying probability distribution, which must be determined for individual problem domains. In most previous research works [27–30], it was derived from analyzing information through hydrodynamic simulations of the sewer network, including mass transport. In this study, a pragmatic approach combining information from simulations with network topology and settlement structure is developed to define the probability distribution. Figure 5 illustrates the probability distribution of two sub-catchments. For simplicity, it depends only on the number of residents in each sub-catchment. Nonetheless, additional factors from the epidemiological point of view, such as demography, socioeconomics, and population density, can be easily included in formulating the probability distribution on demand.

The tools developed are intended to control outbreaks at an early stage or even prevent them altogether through early detection. In the simplest situation, we search for “patient zero”. The probability distribution is simply a step function (Figure 5a). The probability, $p(X_i)$, that the positive (infected) signal comes from the sub-catchment, X_i , is expressed by

$$p(X_i) = \frac{N_i}{\sum_{i=1}^{n_x} N_i}, \quad (5)$$

where N_i is the population of the sub-catchment, X_i , and n_x is the total number of sub-catchments in the city. The infected probability for each sub-catchment is calculated based on the binomial theorem. However, due to the dilution effect, decay of the virus RNA in wastewater, and the detection limit of the analytical method, etc., the virus RNA load from one single patient may not be detected at the sampling point. Therefore, the input must be seen as a hotspot with more individuals infected. In this situation, Figure 5b shows a more rational definition of hotspots via probability distributions across sub-catchments. In this study, we assume that all potentially infected individuals belong to one sub-catchment, so a hotspot can be regarded as “patient zero”. The influences of the dilution effect, etc., are quantified by Equation (1) to define the detectability of each candidate node. Thus, Equation (5) can be used in this study, which will be justified with the results later.

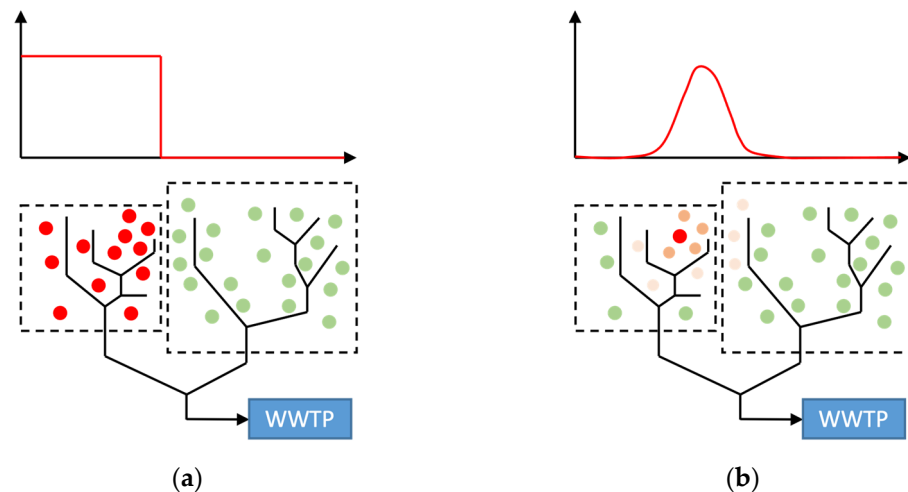


Figure 5. Different definitions of the probability distribution using network topology and settlement structure: (a) sub-catchment-dependent probability distribution, (b) realistic probability distribution. Dots indicate potentially infected individuals (green dot: the probability that the individual is infected is zero, red dot: the lighter the red, the less likely the individual is to be infected).

2.5.3. Signal Matrix and Entropy

As mentioned previously, system sensitivity represents the influence factors of the virus RNA load, such as the dilution effect, virus RNA decay in wastewater, and the detection limit of the analytical method. The RNA load determines whether a sampling point can detect the infected (positive) signal. At a specific system sensitivity, the potential signals detected by each sampling point form a signal matrix. Moreover, the signal matrix combines the information from the network topology and the settlement structure. We use a hypothetical sewer network with 2500 residents and six sub-catchments, as shown in Figure 6, to explain our approach. For simplicity, only main sewers are considered in this study.

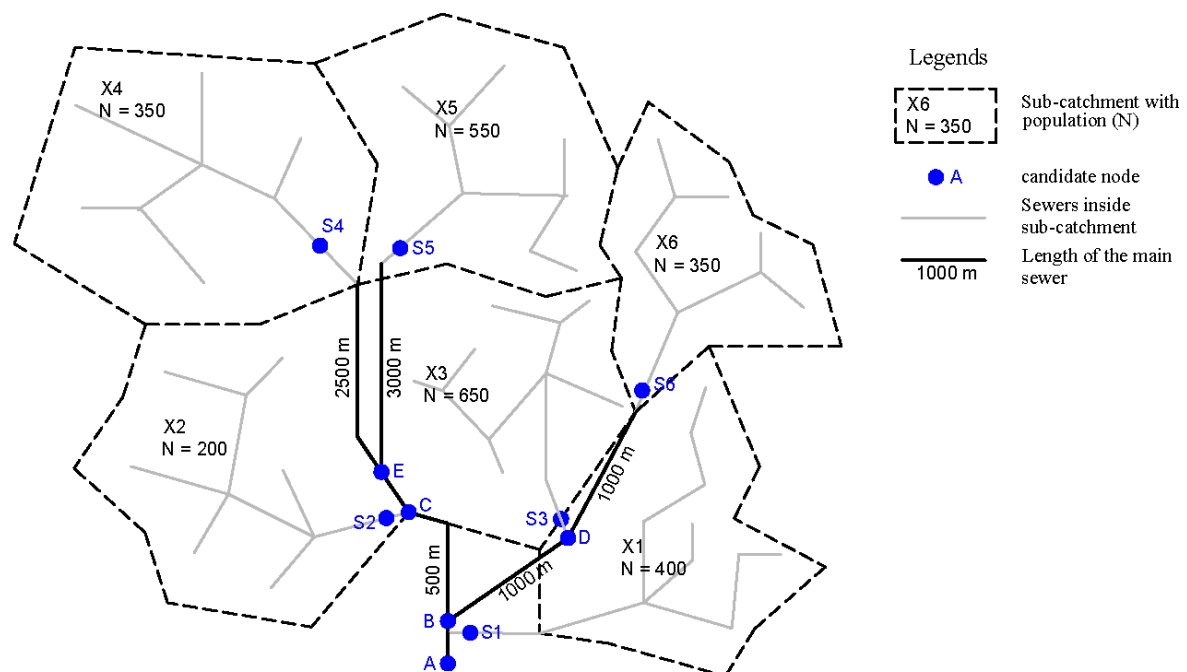


Figure 6. Hypothetical sewer network with 6 sub-catchments and 11 candidate nodes.

The candidate nodes S1 to S6 for the sampling are pre-selected according to the sewer network topology, including the outlets of sub-catchments X1 to X6. Further candidate nodes A to E exist at the main sewer's intersections.

As mentioned in Section 2.5.2, we assume all patients belong to only one sub-catchment because we aim to prevent or detect an outbreak early. The system sensitivity determines whether a positive signal can be detected in the candidate nodes for this sub-catchment, not only based on network topology. Besides potential signals, the probability of one sub-catchment being infected is shown in Table 4. The probability is calculated by residential density (See Equation (5)). For example, the “infected individual” belonging to sub-catchment X3 occurs with a probability of 0.26 and leads to a signal from candidate nodes S3 (the sampling point of this sub-catchment), D, and B (according to network topology) at this specific system sensitivity, 1:2200. Although candidate node A must theoretically show a positive signal based on network topology, considering other influence parameters (such as dilution effect) quantified by system sensitivity, it does not show a signal. A signal matrix can be built (see Table 4) to represent the potential signals (positive) of each sub-catchment when it is the source of the outbreak and no signals (negative) when not detected to be infected.

After constructing the signal matrix, the entropy calculation is formulated to find optimal sampling points covering the highest information content. For this purpose, we quantify their importance according to Equation (1),

$$H(C_i) = - \sum_{j=1}^{n_i} p(c_{i+}^j) \log_2 p(c_{i+}^j) - \left(1 - \sum_{j=1}^{n_i} p(c_{i+}^j)\right) \log_2 \left(1 - \sum_{j=1}^{n_i} p(c_{i+}^j)\right), \quad (6)$$

where $p(c_{i+}^j)$ is the probability of candidate node C_i showing a potential positive signal because of an infected sub-catchment X_j , n_i is the total number of sub-catchments (same by all candidate node C_i in this study), and $(1 - \sum_{j=1}^{n_i} p(c_{i+}^j))$ is the probability of this candidate node C_i detecting no signals. Note that if no signal is detected in all sub-catchments, the probabilities of all combinations without a signal are summarized for the entropy calculation.

Table 4. An example of a potential signal matrix using the system sensitivity of 1:2200 based on the hypothetical model with the probability of the infected individual from a specific sub-catchment (+ positive: signal detected on candidate node, − negative: no signal detected).

Source X	Candidate Nodes C											Probability
	A	B	C	D	E	S1	S2	S3	S4	S5	S6	$p(X_i)$
X1	−	−	−	−	−	+	−	−	−	−	−	0.16
X2	−	+	+	−	−	−	+	−	−	−	−	0.08
X3	−	+	−	+	−	−	−	+	−	−	−	0.26
X4	−	+	+	−	+	−	−	−	+	−	−	0.14
X5	−	+	+	−	+	−	−	−	−	+	−	0.22
X6	−	+	−	+	−	−	−	−	−	−	+	0.14

2.5.4. Objective Function and Optimization Algorithm

An objective function called maximum information and minimum redundancy (MIMR) as an entropy measure for the optimization process is used by several studies [21,27,29]. In this study, a simplified form of the MIMR criterion based on [30] is applied,

$$MIMR_C = \lambda_1 H(C_1, C_2, \dots, C_N) - \lambda_2 TC(C_1, C_2, \dots, C_N) \rightarrow MAX, \text{ where } C = \{C_1, \dots, C_N\}, \quad (7)$$

where $H(C_1, C_2, \dots, C_N)$ is the joint entropy of N sampling points (see Equation (3)), $TC(C_1, C_2, \dots, C_N)$ is the total correlation (see Equation (4)), and λ_1 and λ_2 are the in-

formation redundancy weights to provide weights for joint entropy and total correlation ($\lambda_1 + \lambda_2 = 1$).

$$\operatorname{argmax}_{S \subseteq C} \text{MIMR}_S, |S| \leq z, \quad (8)$$

where z is the pre-defined maximum number of sampling points and S is the set of selected sampling points.

The optimization of Equation (7) leads to a monitoring network with a minimal number of sampling points (set S) from original sampling points (set C), where the information content of these sampling points has been maximized, while redundant information among these sampling points is minimized. To find the optimal solution to Equation (7), the MIMR-based greedy selection algorithm derived from [30] is adapted (see the pseudo codes Algorithm 1) to select sampling points iteratively (see Equation (8)). The first sampling point, S_1 , is chosen based only on the entropy, $H(C_i)$, where no total correlation is considered. The MIMR criterion is applied from selecting the second sampling point until the theoretical maximum joint entropy (JE^{MAX}) is reached, or a defined number of sampling points (z) is achieved.

Algorithm 1. MIMR-based greedy selection algorithm to find optimal sampling points.

Procedure FIND_OPTIMAL_S (C, N, X, S, z)

Input: candidate set C including all N candidate nodes, catchment set X including all sub-catchments, and desired number of candidate nodes z .

Output: selection set S including optimally selected candidate nodes.

- 1 Initialize maximum joint entropy $JE^{MAX} \leftarrow H(X)$ using Equation (3), selection set $S \leftarrow \emptyset$, temporary joint entropy $JE^{TEMP} \leftarrow NULL$.
- 2 **for** $c \in C$ **do**
- 3 Calculate entropy $H(c)$ for each candidate node using Equation (6)
- 4 **end for**
- 5 Assign optimal candidate node $s \leftarrow \operatorname{argmax}_c [H(c)]$
- 6 Update $C \leftarrow C \setminus \{s\}$
- 7 Update $S \leftarrow S \cup \{s\}$
- 8 **while** $JE^{TEMP} \neq JE^{MAX}$ and $|S| \leq z$ **do**
- 9 **for** $c \in C$ **do**
- 10 Calculate $\text{MIMR}_{S \cup \{c\}}$ using Equation (7)
- 11 **end for**
- 12 Find local optimal candidate node $s \leftarrow \operatorname{argmax}_c [\text{MIMR}_{S \cup \{c\}}]$ using Equation (8)
- 13 Update $C \leftarrow C \setminus \{s\}$
- 14 Update $S \leftarrow S \cup \{s\}$
- 15 Assign temporary joint entropy $JE^{TEMP} \leftarrow H(S)$
- 16 **end while**
- 17 **return** selection set S

3. Experimental Results and Discussion

We determine the minimum number of infected individuals required to trigger a positive signal at specific candidate nodes within the sewer network in Section 3.1. This forms the foundational premise for the subsequent discussion of the optimization results of sampling points within the catchment in Hildesheim in Section 3.2.

3.1. Determination of System Sensibility

The theoretical minimal number of infected people for each candidate node to obtain a positive signal, E_{inf}^{MIN} , must be determined before the optimization of sampling points by actual infectious number, which can be calculated using Equation (1). However, the used values of virus-dependent parameters for this equation vary between studies, as mentioned in Table 3. Thus, different combinations of virus-dependent parameter values and their impact on detecting SARS-CoV-2 RNA in wastewater were analyzed in this experiment based on a longitudinal section of the sewer network with nine candidate nodes, as shown

in Figure 7. An average flow velocity of 0.5 m/s and the longest flow path from the main sewer was used to calculate the transport time.

The impact of the variability of the different virus-dependent parameter values was analyzed by two methods. The first was changing only one parameter between its min and max values by holding all other parameters with their median values from Table 3. The resulting minimal required number of infections are shown in Figure 7a. The decay value (k) influenced the results marginally, as the virus typically has only a short residential time until the sample is taken. The parameter with the most significant uncertainty is the shedding magnitude (M_s), as values from the literature vary exponentially. Focusing on the M_s in Figure 7a, at candidate node “A” the number of infected individuals needed to detect a positive signal exceeded the total population. This means no positive signal would be detected even if the entire population was infected. Compared to the median values, their optimal values from Table 3 were utilized in the second method. The results are illustrated in Figure 7b. With the optimal parameter combination, circa 40 infected individuals are needed in the catchment (104,231 residents) to detect a positive signal. In other words, one infected individual out of 2641 noninfected individuals could be detected under the combination of optimal parameter values. This value is close to our experience with real-world data, where SARS-CoV-2 RNA was detected in wastewater samples from the WWTP for incidences of approximately 40 infected individuals per 100,000 residents (1 infected individual in 2500). Thus, the optimal parameter values were used for the following analysis in Section 3.2.

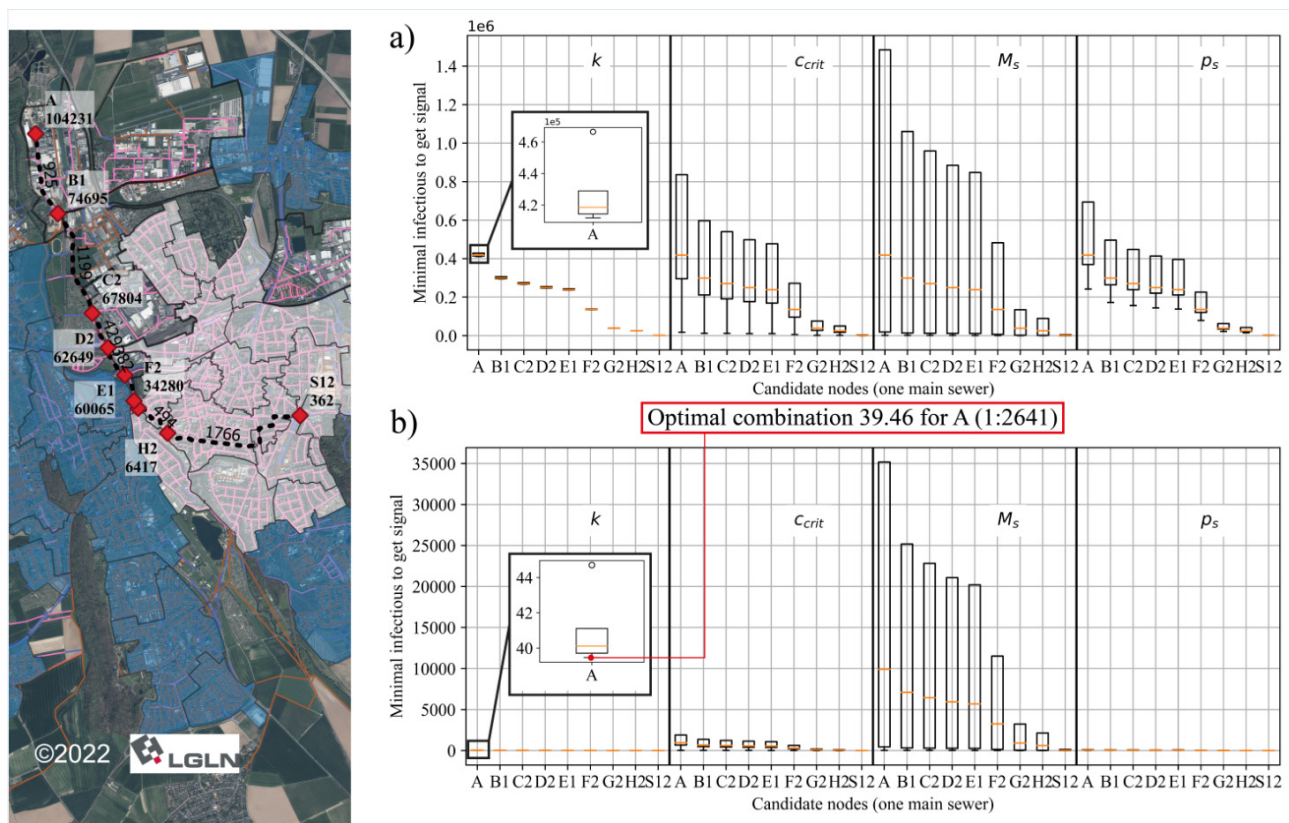


Figure 7. Longitudinal section of candidate nodes (capital letter with number) and covered residents (number below nodes' identifier). The minimal number of infected individuals to detect a positive signal by changing one parameter and holding (a) all other parameters with their median values and (b) all other parameters with their optimal values. Virus-dependent parameters: k —Virus RNA decay in wastewater, c_{crit} —Critical detection limit, M_s —Virus RNA shedding magnitude, p_s —Virus RNA shedding probability.

Regardless of the virus-dependent parameter values, it can be observed in Figure 7 that more infected people are required to detect a positive signal in the downstream nodes of the sewer network compared to the upstream nodes. This can be attributed to the dilution effect. The downstream nodes are connected to more sub-catchments and residents, so more domestic water will be produced.

3.2. Optimization of Sampling Points

For the optimization of the MIMR (Equation (8)), the underlying optimal set of virus-specific parameters determined by the previous section is highlighted in Table 3. Furthermore, regarding the information redundancy weights, we gave more weight to the joint entropy than to the total correlation. Thus, λ_1 is chosen to be 0.8 and λ_2 as 0.2.

As introduced in Section 2.5.3, the signal matrix demonstrates the detected signals to calculate entropy for selecting optimal sampling points, which depends on the number of infected individuals in one sub-catchment and the E_{inf}^{MIN} . The system sensitivity in this study is 1:2641, as determined in Section 3.1. Table 5 shows the optimal number of sampling points depending on the number of infected individuals. The first column indicates the number of infected individuals. The second column depicts the number of minimal selected candidate nodes. The third and fourth columns show the maximum joint entropy and its relation to the theoretical maximum joint entropy of the covered area. The fifth and sixth columns demonstrate the covered population with the sampling points and the percentage of covered people in the entire population of the catchment.

Table 5. Relationship between the number of infected individuals in one sub-catchment and entropies ¹. Numbers in brackets indicate that one more sampling point is needed to cover the catchment. E_{inf} is the number of infected individuals in the sub-catchment, N_S is the smallest number of sampling points needed to reach the maximum joint entropy, JE^{MAX} , $theo.JE_{CA}^{MAX}$ is the theoretical maximum joint entropy of the covered areas, $N_{covered}$ is the maximum number of covered populations, and N_{total} is the total number of populations.

E_{inf}	N_S	JE^{MAX}	$\frac{JE^{MAX}}{theo.JE_{CA}^{MAX}}$	$N_{covered}$	$\frac{N_{covered}}{N_{total}}$
(-)	(-)	(bits)	(%)	(-)	(%)
1	17	1.86	37.4	26,000	24.9
2	20	3.55	71.3	58,046	55.7
3	20	4.43	89.1	80,925	77.6
4	16	4.85	97.4	94,018	90.2
6	15	4.85	97.4	94,018	90.2
8	11	4.95	99.4	99,834	95.8
9	8	4.95	99.4	99,834	95.8
10	7 (8)	4.98	100.0	99,076 (104,231)	95.1 (100.0)
12	6 (7)	4.98	100.0	99,076 (104,231)	95.1 (100.0)
13	5 (6)	4.98	100.0	99,076 (104,231)	95.1 (100.0)
23	4 (5)	4.98	100.0	99,076 (104,231)	95.1 (100.0)
24	3 (4)	4.98	100.0	99,076 (104,231)	95.1 (100.0)
26	3 (4)	4.98	100.0	98,750 (104,231)	94.7 (100.0)
29	2	4.98	100.0	104,231	100.0
40	1	4.98	100.0	104,231	100.0

¹ Bold and underlined numbers indicate the reached values for the optimal set of sampling points.

According to Equation (1), the bigger the E_{inf} , the higher the load and the concentration of RNA. Therefore, reducing N_S generally comes with a larger E_{inf} . The values in brackets in the column $N_{covered}$ resulted from probability distribution and entropy.

The entropy for each candidate node and different numbers of infected individuals ($E_{inf} = 3, 10, 24, 40$) are shown in Figure 8 to demonstrate the impact of the regional distribution of candidate nodes. For low RNA concentrations (small E_{inf}), the highest entropy is reached close to the outlets of the sub-catchments as further downstream dilution

effects would reduce the detectability of SARS-CoV-2 in the wastewater. Therefore, if an early warning system is built to detect a small number of infected individuals within a catchment, more sampling points close to the sub-catchment outlets would be needed. Increasing RNA concentration (higher E_{inf}) allows fewer sampling points to cover more citizens and areas. For the case of 40 infected individuals, the problem of redundant information can be seen for the left main sewer close to the outlet. While the entropy indicated a high information content of the nodes at the end of this main sewer, it was evident that the information gain would only be minimal when adding two sampling points in a row (see red circles in Figure 8). This underlines the importance of including the total correlation as an additional measure in the objective function $MIMR_C$ (see Equation (7)).

The case of ten infected individuals and eight selected optimal candidate nodes is discussed in more detail. It was seen that the JE^{MAX} is reached, and the whole population could be observed for this combination (see bold values in Table 5). The selected candidate nodes with corresponding sub-catchments are shown in Figure 9.

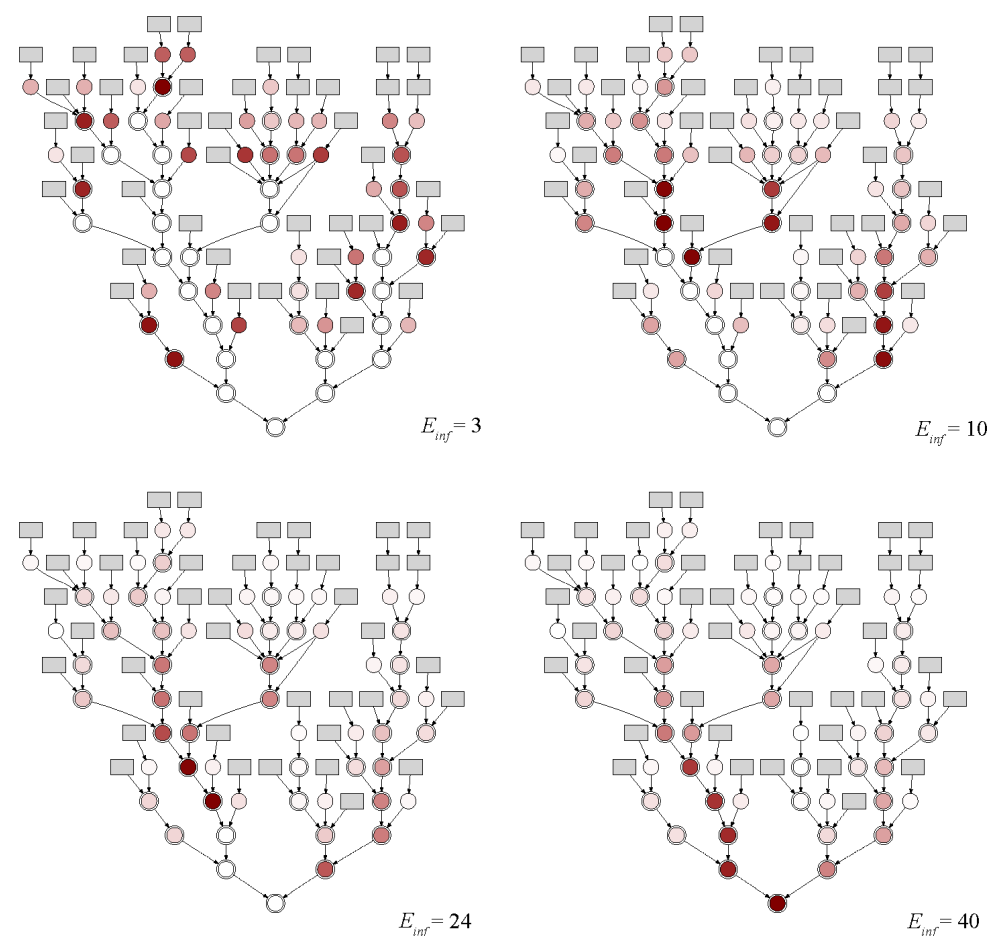
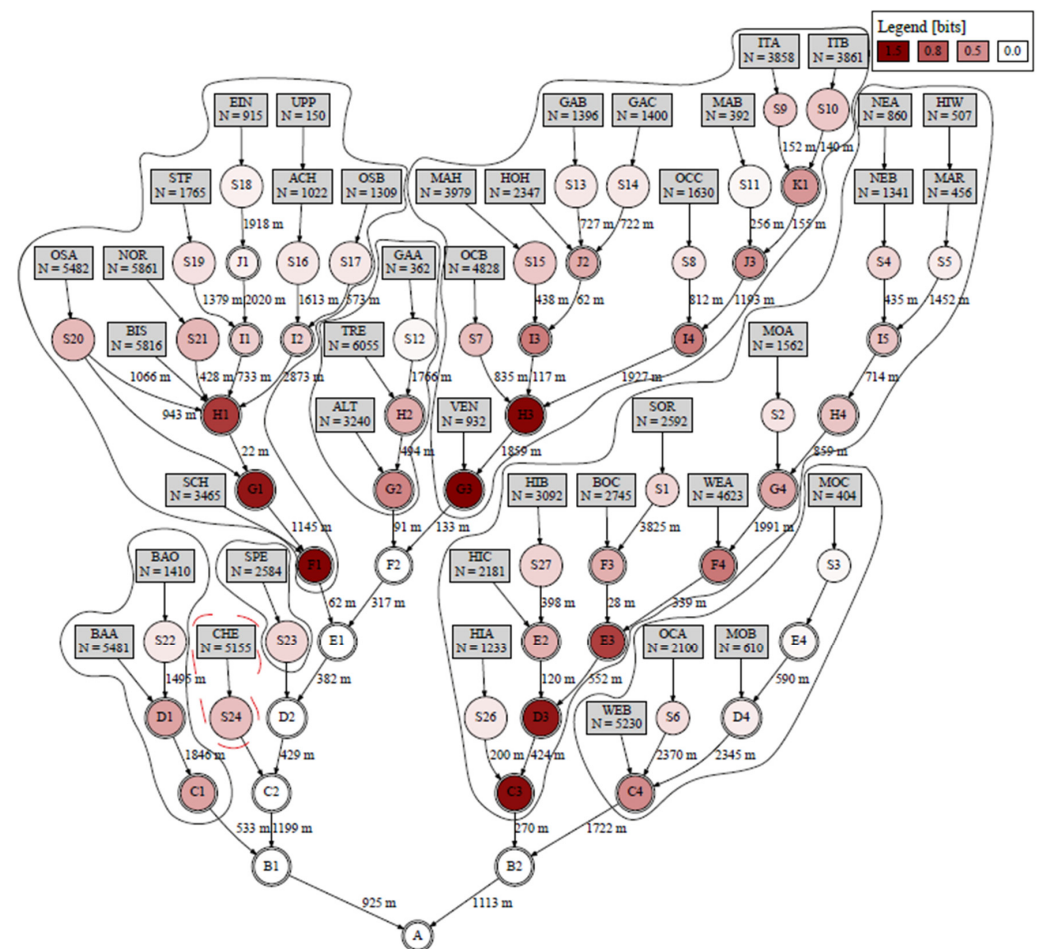


Figure 8. Color maps of entropy based on different E_{inf} within the detection capability 1:2641. The darker the color of the node, the higher the entropy.

In Figure 9, the entropy for nodes close to the WWTP (node A) was zero, as here no positive signal was detected because of dilution effects. The entropy was also relatively small for the outlet nodes of single sub-catchments (node identifiers starting with S) because these sampling points covered only a few residents. Table 6 shows the stepwise optimization of sampling points with entropy definitions from Section 2.3 and the covered population for each node. The joint entropy reached the maximum value with seven sampling points.

Table 6. Optimization process of sampling points (system sensitivity 1:2641, $E_{inf} = 10$).

N_S	S	JE	TC	$MIMR_S$	$N_{covered}$
1	G3	1.51	0.00	1.21	24,623
2	F1	2.88	0.11	2.28	25,785
3	C3	4.02	0.34	3.15	21,192
4	G2	4.40	0.51	3.42	9657
5	C4	4.72	0.71	3.64	8344
6	C1	4.91	0.92	3.74	6891
7	S23	4.98	1.02	3.78	2584
8	S24	4.98	-	-	5155

**Figure 9.** Entropy for each node and optimal sampling point covering the corresponding sub-catchments with eight sampling points (system sensitivity 1:2641, $E_{inf} = 10$). The numbers of the lines indicate the length of the specific sewer.

With each added sampling point ($s \in S$), the joint entropy (JE) and the value of the objective function ($MIMR_S$) increase until seven sampling points are reached. The selected candidate nodes as sampling points have the highest information content with minor redundancy for this scenario (detectability of one infected individual out of 2641 noninfected and 10 infected individuals). However, only one sub-catchment was not observed with the defined sampling points, when the entropy reached maximum for this combination of sampling points. One further sampling point must be implemented for this sub-catchment (see candidate node S24 in Figure 9). Subsequently adding the remaining sub-catchment CHE with sampling point (candidate node) S24, the entire population of Hildesheim can be surveilled. Some of the chosen sampling points will provide positive

signals with less than ten infected individuals, e.g., for node S23 with 2584 residents, only one infected individual would be sufficient to detect a positive signal.

4. Conclusions

This paper develops a pandemic early warning system using samples from a sewer network. This system consists of two sequential parts: a signal detection program based on the mass balance model and an optimal sampling point selection program based on information theory. The signal detection program calculates the theoretical system sensitivity, i.e., the minimum number of infections for which a positive signal can be obtained at a sampling point, standardizing the detectability of sampling points in the WBE, such as the dilution effect. Following this, the actual number of infected persons and the minimum number of sampling points are used to mathematically optimize the location of sampling points in the sewer network, considering the network topology and the settlement structure.

The key findings of this study are as follows:

- Virus specific parameter values of SARS-CoV-2 from the literature are currently not sufficient for parametrizing our model.
- Number and locations for the sampling points depends on the expected sensitivity of the system.
- Increasing the number of sampling points does not necessarily improve the information content.
- Virus-related uncertainties have an impact on the placement and number of sampling points, but this impact is offset by the expected sensitivity.
- For the case study of Hildesheim, only 8 sampling points and less than 10 infected individuals per sub-catchment were required to identify potentially infected sub-catchments.

However, some limitations remain in this paper:

- The probability distribution function is simply based on the assumption that all infected people come from the same sub-catchment. For a better representation, epidemiological data could be used to estimate real infection distributions, as shown in Figure 5b.
- The flow time used to calculate the system sensitivity simply uses a constant. For further studies, 1D sewer models can be applied to better estimate the flow time and also simulate RNA loss, which is another limitation of the current approach.

However, since our practical detection limits agree with the theoretical calculations, this can be ignored for demonstrating the developed method. However, it is necessary for a more realistic or complex sewer network.

Author Contributions: Conceptualization, M.W., R.N. and F.K.; methodology, M.W. and F.K.; software, Y.Y. and Y.Z.; validation, Y.Y. and Y.Z.; formal analysis, Y.Y. and Y.Z.; investigation, Y.Y. and Y.Z.; resources, M.W. and F.K.; data curation, Y.Z.; writing—original draft preparation, Y.Y. and Y.Z.; writing—review and editing, M.W., F.K. and R.N.; visualization, Y.Z. and Y.Y.; supervision, M.W., F.K. and R.N.; project administration, M.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund (ERDF) and Lower Saxony, grant numbers ZW 7-85094959 and ZW 7-85125149.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank Stadtentwässerung Hildesheim (SEHi) and the Screening Team for their support, and the European Regional Development Fund (ERDF) for funding. This research is part of the project SCREENING from Ostfalia University of Applied Sciences in Germany.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Virus-dependent parameter values from literature review.

Parameter: Virus RNA Shedding Magnitude		
Values (\log_{10} copies/mL)	Literature	Comments
2.9	[31]	
3.75	[32]	Units adjusted
4.55	[16]	Units adjusted
4.7	[33]	
5.8	[32]	
6.28	[34]	Units adjusted
7.1	[31]	
Parameter: virus RNA Shedding probability		
Values (%)	Literature	Comments
10.1	[35]	
15.3	[33]	
29	[36]	
47.7	[37]	
48.1	[33]	
53.4	[38]	
54.5	[16]	
55	[39]	
83.3	[32]	
Parameter: virus RNA decay in wastewater		
Values (-)	Literature	Comments
0.06	[40]	
0.084	[41]	
0.09	[42]	
0.183	[43]	
0.286	[41]	
0.67	[42]	
Parameter: critical detection limit		
Values (copies/mL)	Literature	Comments
3.7	[44]	
9.2	[44]	
39.04	[45]	
59.4	[45]	
72.42	[45]	
78.96	[45]	
79.08	[45]	
98.42	[45]	
133.02	[45]	
159.08	[45]	
183.34	[45]	
301.22	[45]	
374.86	[45]	
533.78	[45]	

References

1. Sims, N.; Kasprzyk-Hordern, B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environ. Int.* **2020**, *139*, 105689. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Wade, M.J.; Jacomo, A.L.; Armenise, E.; Brown, M.R.; Bunce, J.T.; Cameron, G.J.; Fang, Z.; Gilpin, D.F.; Graham, D.W.; Grimsley, J.M.; et al. 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.* **2022**, *424*, 127456. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Mao, K.; Zhang, K.; Du, W.; Ali, W.; Feng, X.; Zhang, H. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Curr. Opin. Environ. Sci. Health* **2020**, *17*, 1–7. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Amoah, I.D.; Kumari, S.; Bux, F. Coronaviruses in wastewater processes: Source, fate and potential risks. *Environ. Int.* **2020**, *143*, 105962. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Praus, P. Information Entropy for Evaluation of Wastewater Composition. *Water* **2020**, *12*, 1095. [\[CrossRef\]](#)
6. Saguti, F.; Magnil, E.; Enache, L.; Churqui, M.P.; Johansson, A.; Lumley, D.; Davidsson, F.; Dotevall, L.; Mattsson, A.; Trybala, E.; et al. Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Res.* **2021**, *189*, 116620. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Prado, T.; Fumian, T.M.; Mannarino, C.F.; Resende, P.C.; Motta, F.C.; Eppinghaus, A.L.F.; do Vale, V.H.C.; Braz, R.M.S.; de Andrade, J.D.S.R.; Maranhão, A.G.; et al. Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. *Water Res.* **2021**, *191*, 116810. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Albastaki, A.; Naji, M.; Lootah, R.; Almeheiri, R.; Almulla, H.; Almarri, I.; Alreyami, A.; Aden, A.; Alghafri, R. First confirmed detection of SARS-CoV-2 in untreated municipal and aircraft wastewater in Dubai, UAE: The use of wastewater based epidemiology as an early warning tool to monitor the prevalence of COVID-19. *Sci. Total Environ.* **2021**, *760*, 143350. [\[CrossRef\]](#)
9. Yaniv, K.; Shagan, M.; Lewis, Y.E.; Kramarsky-Winter, E.; Weil, M.; Indenbaum, V.; Elul, M.; Erster, O.; Brown, A.S.; Mendelson, E.; et al. City-level SARS-CoV-2 sewage surveillance. *Chemosphere* **2021**, *283*, 131194. [\[CrossRef\]](#)
10. Larson, E.R.; Graham, B.M.; Achury, R.; Coon, J.J.; Daniels, M.K.; Gambrell, D.K.; Jonassen, K.L.; King, G.D.; LaRacune, N.; Perrin-Stowe, T.I.; et al. From eDNA to citizen science: Emerging tools for the early detection of invasive species. *Front. Ecol. Environ.* **2020**, *18*, 194–202. [\[CrossRef\]](#)
11. Calle, E.; Martínez, D.; Brugués-i-Pujolràs, R.; Farreras, M.; Saló-Grau, J.; Pueyo-Ros, J.; Corominas, L. Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks. *Environ. Int.* **2021**, *157*, 106768. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Larson, R.C.; Berman, O.; Nourinejad, M. Sampling manholes to home in on SARS-CoV-2 infections. *PLoS ONE* **2020**, *15*, e0240007. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Domokos, E.; Sebestyén, V.; Somogyi, V.; Trájer, A.J.; Gerencsér-Berta, R.; Horváth, B.O.; Tóth, E.G.; Jakab, F.; Kemenesi, G.; Abonyi, J. Identification of sampling points for the detection of SARS-CoV-2 in the sewage system. *Sustain. Cities Soc.* **2022**, *76*, 103422. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Gkatzoura, A.; Zafeirakou, A. Optimal Selection of Sampling Points for Detecting SARS-CoV-2 RNA in Sewer System Using NSGA-II Algorithm. *Water* **2023**, *15*, 4076. [\[CrossRef\]](#)
15. Steinnocher, K.; Younsoo, K.; Köstl, M. *Schätzung der Lokalen Bevölkerungsdichte in Stadtgebieten. Eine Fallstudie aus Daejeon, Korea*; AGIT 2006 Angewandte Geographische Informationsverarbeitung; Wichmann Verlag: Heidelberg, Germany, 2006; pp. 633–638.
16. Li, X.; Zhang, S.; Shi, J.; Luby, S.P.; Jiang, G. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. *Chem. Eng. J.* **2021**, *415*, 129039. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Crank, K.; Chen, W.; Bivins, A.; Lowry, S.; Bibby, K. Contribution of SARS-CoV-2 RNA shedding routes to RNA loads in wastewater. *Sci. Total Environ.* **2022**, *806*, 150376. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Rose, C.; Parker, A.; Jefferson, B.; Cartmell, E. The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Crit. Rev. Environ. Sci. Technol.* **2015**, *45*, 1827–1879. [\[CrossRef\]](#)
19. Brown, D.M.; Butler, D.; Orman, N.R.; Davies, J.W. Gross solids transport in small diameter sewers. *Water Sci. Technol.* **1996**, *33*, 25–30. [\[CrossRef\]](#)
20. Statistisches Bundesamt (Destatis). *Datenreport 2021: Ein Sozialbericht für die Bundesrepublik Deutschland*; Bundeszentrale für Politische Bildung: Bonn, Germany, 2021.
21. Alfonso, L.; Lobbrecht, A.; Price, R. Optimization of water level monitoring network in polder systems using information theory. *Water Resour. Res.* **2010**, *46*, W12553. [\[CrossRef\]](#)
22. Amorocho, J.; Espildora, B. Entropy in the assessment of uncertainty in hydrologic systems and models. *Water Resour. Res.* **1973**, *9*, 1511–1522. [\[CrossRef\]](#)
23. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
24. Keum, J.; Kornelsen, K.; Leach, J.; Coulibaly, P. Entropy Applications to Water Monitoring Network Design: A Review. *Entropy* **2017**, *19*, 613. [\[CrossRef\]](#)
25. Yoo, A.H.; Klyszejko, Z.; Curtis, C.E.; Ma, W.J. Strategic allocation of working memory resource. *Sci. Rep.* **2018**, *8*, 16162. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Keum, J.; Coulibaly, P. Information theory-based decision support system for integrated design of multivariable hydrometric networks. *Water Resour. Res.* **2017**, *53*, 6239–6259. [\[CrossRef\]](#)

27. Alfonso, L.; Lobbrecht, A.; Price, R. Information theory-based approach for location of monitoring water level gauges in polders. *Water Resour. Res.* **2010**, *46*, W03528. [\[CrossRef\]](#)
28. Alfonso, L.; He, L.; Lobbrecht, A.; Price, R. Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *J. Hydroinform.* **2013**, *15*, 211–228. [\[CrossRef\]](#)
29. Banik, A.; Brown, R.E.; Bamburg, J.; Lahiri, D.K.; Khurana, D.; Friedland, R.P.; Chen, W.; Ding, Y.; Mudher, A.; Padjen, A.L.; et al. Translation of Pre-Clinical Studies into Successful Clinical Trials for Alzheimer’s Disease: What Are the Roadblocks and How Can They Be Overcome? *J. Alzheimer’s Dis. JAD* **2015**, *47*, 815–843. [\[CrossRef\]](#)
30. Li, C.; Singh, V.P.; Mishra, A.K. Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy. *Water Resour. Res.* **2012**, *48*, W05521. [\[CrossRef\]](#)
31. Ng, K.; Poon, B.H.; Kiat Puar, T.H.; Shan Quah, J.L.; Loh, W.J.; Wong, Y.J.; Tan, T.Y.; Raghuram, J. COVID-19 and the Risk to Health Care Workers: A Case Report. *Ann. Intern. Med.* **2020**, *172*, 766–767. [\[CrossRef\]](#)
32. Zhang, N.; Gong, Y.; Meng, F.; Bi, Y.; Yang, P.; Wang, F. Comparative study on virus shedding patterns in nasopharyngeal and fecal specimens of COVID-19 patients. *Sci. China Life Sci.* **2021**, *64*, 486–488. [\[CrossRef\]](#)
33. Cheung, K.S.; Hung, I.F.; Chan, P.P.; Lung, K.C.; Tso, E.; Liu, R.; Ng, Y.Y.; Chu, M.Y.; Chung, T.W.; Tam, A.R.; et al. Gastrointestinal Manifestations of SARS-CoV-2 Infection and Virus Load in Fecal Samples from a Hong Kong Cohort: Systematic Review and Meta-analysis. *Gastroenterology* **2020**, *159*, 81–95. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Hoffmann, T.; Alsing, J. Faecal shedding models for SARS-CoV-2 RNA among hospitalised patients and implications for wastewater-based epidemiology. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2023**, *72*, 330–345. [\[CrossRef\]](#)
35. Kim, J.M.; Kim, H.M.; Lee, E.J.; Jo, H.J.; Yoon, Y.; Lee, N.J.; Son, J.; Lee, Y.J.; Kim, M.S.; Lee, Y.P.; et al. Detection and Isolation of SARS-CoV-2 in Serum, Urine, and Stool Specimens of COVID-19 Patients from the Republic of Korea. *Osong Public Health Res. Perspect.* **2020**, *11*, 112–117. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* **2020**, *323*, 1843–1844. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Lin, L.; Jiang, X.; Zhang, Z.; Huang, S.; Zhang, Z.; Fang, Z.; Gu, Z.; Gao, L.; Shi, H.; Mai, L.; et al. Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection. *Gut* **2020**, *69*, 997–1001. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Xiao, F.; Tang, M.; Zheng, X.; Liu, Y.; Li, X.; Shan, H. Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* **2020**, *158*, 1831–1833. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Wu, Y.; Guo, C.; Tang, L.; Hong, Z.; Zhou, J.; Dong, X.; Yin, H.; Xiao, Q.; Tang, Y.; Qu, X.; et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet. Gastroenterol. Hepatol.* **2020**, *5*, 434–435. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Hokajärvi, A.M.; Rytönen, A.; Tiwari, A.; Kauppinen, A.; Oikarinen, S.; Lehto, K.M.; Kankaanpää, A.; Gunnar, T.; Al-Hello, H.; Blomqvist, S.; et al. The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland. *Sci. Total Environ.* **2021**, *770*, 145274. [\[CrossRef\]](#)
41. Ahmed, W.; Bertsch, P.M.; Bibby, K.; Haramoto, E.; Hewitt, J.; Huygens, F.; Gyawali, P.; Korajkic, A.; Riddell, S.; Sherchan, S.P.; et al. Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environ. Res.* **2021**, *770*, 145274. [\[CrossRef\]](#)
42. Bivins, A.; Greaves, J.; Fischer, R.; Yinda, K.C.; Ahmed, W.; Kitajima, M.; Munster, V.J.; Bibby, K. Persistence of SARS-CoV-2 in Water and Wastewater. *Environ. Sci. Technol. Lett.* **2020**, *7*, 937–942. [\[CrossRef\]](#)
43. Mota, C.R.; Bressani-Ribeiro, T.; Araújo, J.C.; Leal, C.D.; Leroy-Freitas, D.; Machado, E.C.; Espinosa, M.F.; Fernandes, L.; Leão, T.L.; Chamhum-Silva, L.; et al. Assessing spatial distribution of COVID-19 prevalence in Brazil using decentralised sewage monitoring. *Water Res.* **2021**, *202*, 117388. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Gerrity, D.; Papp, K.; Stoker, M.; Sims, A.; Frehner, W. Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: Methodology, occurrence, and incidence/prevalence considerations. *Water Res. X* **2021**, *10*, 100086. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Ahmed, W.; Bivins, A.; Metcalfe, S.; Smith, W.J.; Verbyla, M.E.; Symonds, E.M.; Simpson, S.L. Evaluation of process limit of detection and quantification variation of SARS-CoV-2 RT-qPCR and RT-dPCR assays for wastewater surveillance. *Water Res.* **2022**, *213*, 118132. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.