

Article

Information Theory Solution Approach to the Air Pollution Sensor Location–Allocation Problem

Ziv Mano ¹, Shai Kendler ^{1,2} and Barak Fishbain ^{1,*}

¹ Faculty of Civil & Environmental Engineering, Technion—Israeli Institute of Technology, Haifa 3200003, Israel; ziv.mano@campus.technion.ac.il (Z.M.); skendler@technion.ac.il (S.K.)

² Environmental Physics Department, Israel Institute for Biological Research, 24 Lerer St., Ness Ziona 7410001, Israel

* Correspondence: fishbain@technion.ac.il; Tel.: +972-4-8293177

Abstract: Air pollution is one of the prime adverse environmental outcomes of urbanization and industrialization. The first step toward air pollution mitigation is monitoring and identifying its source(s). The deployment of a sensor array always involves a tradeoff between cost and performance. The performance of the network heavily depends on optimal deployment of the sensors. The latter is known as the location–allocation problem. Here, a new approach drawing on information theory is presented, in which air pollution levels at different locations are computed using a Lagrangian atmospheric dispersion model under various meteorological conditions. The sensors are then placed in those locations identified as the most informative. Specifically, entropy is used to quantify the locations' informativity. This entropy method is compared to two commonly used heuristics for solving the location–allocation problem. In the first, sensors are randomly deployed; in the second, the sensors are placed according to maximal cumulative pollution levels (i.e., hot spots). Two simulated scenarios were evaluated: one containing point sources and buildings and the other containing line sources (i.e., roads). The entropy method resulted in superior sensor deployment in terms of source apportionment and dense pollution field reconstruction from the sparse sensors' network measurements.



Citation: Mano, Z.; Kendler, S.; Fishbain, B. Information Theory Solution Approach to the Air Pollution Sensor Location–Allocation Problem. *Sensors* **2022**, *22*, 3808. <https://doi.org/10.3390/s22103808>

Academic Editor: Lorena Parra

Received: 21 March 2022

Accepted: 6 May 2022

Published: 17 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Keywords: air pollution; environmental monitoring networks; location–allocation models; sensors' array; information theory

1. Introduction

In recent years, the negative impact of air pollution on health and climate change has become a major environmental issue. According to the World Health Organization, air pollution has emerged as the deadliest form of pollution and the fourth leading risk factor for premature deaths worldwide, accounting for about seven million deaths in 2012 [1] with a toll of about US \$225 billion in lost labor income in 2013 [2,3]. Hence, controlling and monitoring air pollution is crucial. Routine monitoring is typically done by standardized air quality monitoring (AQM) stations spread thinly due to their size and cost [4]. Therefore, the effective deployment of AQM is crucial.

Advances in sensory and communication technologies have made the deployment of portable and relatively low-cost Micro-Sensing air pollution Units (MSUs) feasible. These sensors can be spread more densely and provide higher spatial resolution data. Recent studies evaluating these sensors in laboratory and field trials have shown that these units are less accurate than standard laboratory equipment or AQM stations; however, their sheer number makes it possible to effectively capture air pollution spatiotemporal variability [5–8]. While these sensors are becoming increasingly available compared to AQM, procuring, maintaining, and operating many MSUs is still a demanding task. Hence, sensor networks remain limited in size, so even for MSU networks, an optimal deployment strategy is critical.

Placing sensors in optimal locations so that the sensory network provides valuable environmental information is known as the *location–allocation* problem and has attracted considerable attention for many years [9–12]. The sensor location–allocation problem has also been studied for water resource management [13–15], structural health monitoring [16], soil contamination [17], and many other domains. While different systems pose different challenges, the sensor location–allocation problem can be viewed as a case of choosing the best subset of sensor locations from a set of candidates that results in a desirable outcome under budget constraints, which usually dictate the number of sensors and their properties. Formulating the location–allocation problem in this fashion serves to cast it as the well-known knapsack problem [11]. Thus, the location–allocation problem is NP-hard. Since there is no computationally efficient solution, heuristic approaches are often applied.

There are many examples of such heuristics. Zou et al. [18] utilized sensor deployment spatial proximity models and the theoretical reliability of Gaussian dispersion processes of air pollutants to build a Gaussian weighting function-aided proximity model (GWFP). Li et al. [19] used inverse distance weight (IDW) interpolation coupled with a geographic information system (GIS) to assess the particle matter (PM) dense pollution field for the placement of sensors in locations with the highest pollution (i.e., hot spots). Another method that capitalizes on GIS capabilities was presented by Alsahli and Harbi [20], where land use was inferred from GIS data, and sensors were deployed based on a greedy algorithm that traded off highly polluted with highly populated areas. However, the use of GIS systems requires detailed information on the target region. Often, these data are unavailable or grossly inaccurate [21]. Furthermore, placing the sensors in locations where the substance recorded by the sensors is the highest or near populated areas does not guarantee optimality in terms of pollution field reconstruction and source apportionment.

Optimization-based methods have also been used to solve the location–allocation problem. Two main problems have been addressed: optimizing network operations through connectivity and coverage [22] and optimizing air-quality sensing. For the latter, Boubrima et al. cast the optimization problem as a minimum cost problem that finds optimal sensors and sink locations, ensuring air pollution coverage and network connectivity [23–25]. Zoroufchi-Benis et al. [26] defined the optimization problem as a minimum fitness problem with multi-objective functions with the aim of ensuring maximum coverage, continuity of the coverage area, the least overlap among coverage areas, maximum detection of violations over ambient air standards, and sensitivity of monitoring stations to emission sources. Al-Adwani et al. [27] formulated the monitoring cost minimization problem as a minimum set cover [28], where the maximum number of overlapping points in space was correlated with the maximum number of peaks. In this work, the pollution dispersion model consisted of a Gaussian plume model to describe the dispersion of continuous emissions in steady-state conditions and a Gaussian puff model that simulated instantaneous emissions. The findings showed that the cost of monitoring could be reduced without a concomitant loss of information by minimizing the number of stations. Kumar et al. [29] presented a deterministic spatial sampling design to capture intra-city variability in air pollution. Their objective was to draw a sample of households that best represented the spatial distribution of ambient air pollution while maximizing the variance in the preliminary estimates of air pollution with the minimum number of sample sites. The algorithm ensured that the sample sites were informative for addressing inferences by emphasizing certain population or environmental characteristics. Kanaroglou et al. [10] developed a methodology for selecting monitoring sites based on spatial variations in air pollution and the distribution of addresses over the target area. The network density increased with concentration variability and population. The method specified a continuous demand surface (for monitoring) over the area. Lerner et al. [11] cast the air pollution location–allocation problem as the knapsack problem, where a given sensor’s utility in a given location was inferred from the sensor’s physicochemical characteristics and land-use analysis.

These works all dealt with the three main factors that affect the solution of the location–allocation problem: land use, meteorological conditions, and pollution signal characteristics.

The latter has mostly been considered in terms of the signal's extreme points, i.e., hotspots. However, no previous work has attempted to associate these three factors in one framework to provide a more comprehensive solution.

This paper presents a new approach to the location–allocation problem, which takes the topography of the observed area and its meteorology, as well as its expected pollution signal characteristics, into account. This is done by utilizing Lagrangian atmospheric dispersion models and information theory to solve the location–allocation problem. Sensors are regarded as information sources that vary due to the nature of pollution dispersion and variations in climatic conditions. The sensors are placed in a configuration that maximizes the amount of information, i.e., the joint information according to Shannon entropy, and minimizes redundancy, i.e., the mutual information.

2. Materials and Methods

2.1. Simulation Study

A simulation was carried out to evaluate sensor deployment strategies. A GRAZ Lagrangian (GRAL) dispersion model, combined with a Prognostic Wind Field Model GRAMM [30,31] was used to facilitate the examination of a wide range of scenarios differing in source characteristics and environmental conditions.

GRAL is an open-source air pollution simulator developed by the Graz University of Technology (TUG) and the Government of Styria, Austria. The GRAL can model a wide range of spatial scales, from street-level through whole cities to a state-wide scale. The model takes as inputs topography, including buildings and infrastructure, sources with their emission profiles, and wind fields. The output is a spatiotemporal dense pollution map over the study region [32]. To compute this map and given the topography, the model takes into consideration building's downwash effects through microscale modeling. The sources may be of different types, including surface road networks and point sources, such as tunnel ventilation outlets and industrial stacks, tunnel portals, and area sources. GRAL has been used extensively in regulatory assessments and scientific studies, such as calculating the impacts of road traffic or industry on air pollution, and it has been extensively validated in several different countries and contexts [30].

It is important to note that, with respect to meteorology, the GRAL model takes as input solely the wind field and does not regard any other chemical reactions and transformations between substances [33]. To this end, the wind field is computed by the GRAMM meteorological model, which is based on the Reynolds-averaged Navier–Stokes equations (RANS equations) and the law of mass conservation [34]. Thus, the term meteorology is limited to the wind field and its stability class.

The region of interest, Ω , was modeled as a flat 500×500 m area divided into a grid of 5×5 m cells, $\{\omega\} \in \Omega$. In this area, two different scenarios were examined to present the capability of the method for a wide range of substances (gaseous and particulate matter) and topographical complexities:

1. A Small NeighBorHood, SNBH, consists of six buildings of different sizes and five-point sources emitting PM10 at different rates.
2. A Central Business District, CBD, consists of 35 buildings, three-line sources (i.e., roads), and five-point sources. NO_x was emitted from the point sources and the roads at different rates. Figure 1 depicts the two computer-generated scenarios.

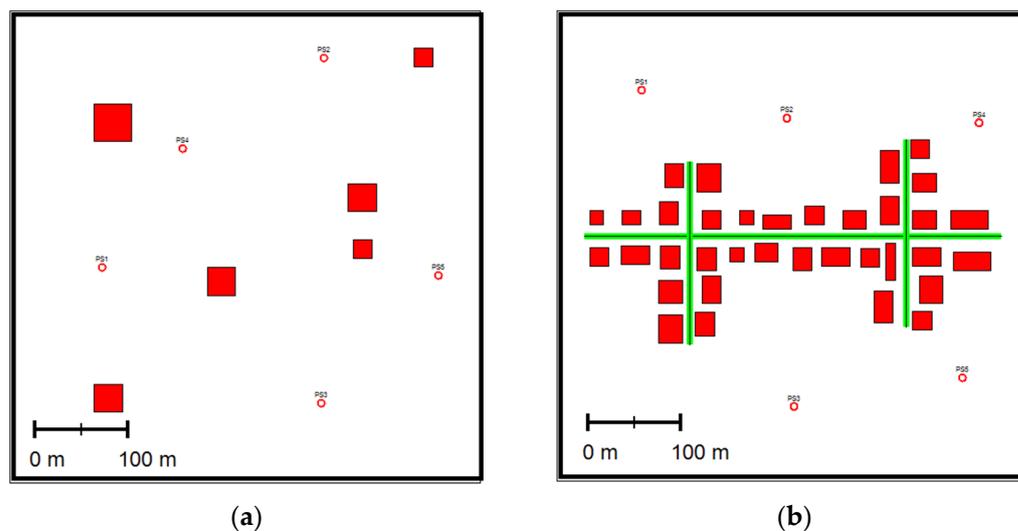


Figure 1. Schemas of the two scenarios: (a) small neighborhood, SNBH; (b) CBD, where the red squares indicate buildings; the green lines indicate line sources; and the circles indicate point sources (PS).

For each meteorological condition (a given wind speed and direction), a dense pollution map was computed, where each grid point was represented by several pollution values, each of which corresponded to a specific meteorological condition.

The meteorology input for the GRAL simulation was obtained from real-life data collected by the Israel Meteorology Service at the Hadera Port station (lon: lat, 34.8815: 32.4732) on 1 May 2020. For T , the total time in hours of the analysis, the temporal resolution of the meteorological data *in minutes* is $\Delta\Phi$; thus, for a $T = 24$ h period, the number of samples was: $|q| = T \cdot \frac{60}{\Delta\Phi}$. Here, $T = 24$, $\Delta\Phi$ was a 10-min interval, so that $|q| = 144$. The meteorological input for the GRAL simulation also contained the atmospheric stability classes (A–G), which were computed based on an atmospheric stability classification scheme [35].

The GRAL building prognostic approach was used with the default parameters. The maximum number of iterations for the internal flow field solver was 500 iterations, and concentration levels were measured three meters above ground level. A complete list of the parameter sets used for the GRAL computations is provided in the Supplementary Materials.

2.2. System Overview

When addressing the location–allocation problem in a given region for the first time, there is likely to be little information on pollution behavior in that specific region. On the other hand, information on the *static* attributes of the land, such as topography, land use, meteorological, and the locations of buildings, and potential pollution sources are more readily available. Thus, relying on static factors constitutes a more feasible approach. While we assume that potential sources' locations are known a priori, the method suggested here is still applicable when sources' locations are not known. This is discussed in the discussion section. Using *static* factors facilitated problem formulation, as described below.

In the initial stage, it is assumed that all sources emitted at a constant and equal rate (zero approximation). In this work, a constant rate of 100 kg/h was chosen arbitrarily. Then, a simulation generates a dense pollution map describing the pollution level, C_{ω}^q , for meteorological condition q at location $\omega \in \Omega$. This process resulted in $|q|$ different maps. Based on these $|q|$ dense pollution maps, the locations that maximized a decision criterion were selected, resulting in a set of several possible deployment configurations. These possible configurations were evaluated for their ability to locate and quantify the source term under the simplistic zero approximation. Then, the optimal configuration was tested on other, more realistic source terms.

2.3. Algorithmic Approach

Based on the $|q|$ dense pollution maps generated by the simulation for the different meteorological conditions, under the zero-approximation assumption, for each location $\omega \in \Omega$, we obtained a set of measurements $\{C_\omega^q\}$. The measuring units of $\{C_\omega^q\}$ for gaseous matter can be parts per million (ppm), parts per billion (ppb) or $\mu\text{gr}/\text{m}^3$; for particulate matter it can be particle number or $\mu\text{gr}/\text{m}^3$. Regardless, it is important to note that the method and entropy measure are invariant to the measuring units. The analysis of these sets constituted the decision criterion, $DM_\omega \forall \omega \in \Omega$, to allocate the sensors. For $P(C_\omega^q)$ the empirical probability function of C_ω^q , two different metrics were compared—Entropy (Equation (1)) and Hot Spot (Equation (2)):

$$DM_\omega = - \sum_q P(C_\omega^q) \log P(C_\omega^q) \quad (1)$$

$$DM_{\omega \in \Omega} = \sum_q C_\omega^q \quad (2)$$

The method is based on finding a set of sensor positions that maximize the entropy or hot spot score and have the lowest correlations between the sensor readings over time. The correlation between the two sensors is defined as the Pearson correlation index between the pollution concentration sets of the different meteorological conditions. This is accomplished using an iterative algorithm allocated to one sensor in each iteration. The following notations are used to describe the algorithm: let d_{max} and d be the maximum available sensors for deployment and the number of sensors already deployed by the algorithm, respectively, and $\Omega^d \subseteq \Omega$ is the set of locations with sensors. For each location $\omega \in \Omega$, the set of neighboring locations is B^ω . The union of all the cells' neighboring sensors is then denoted by B^d . B^d allows for avoiding placing sensors too close to each other, which would have represented the same information. Note that B^d is automatically updated as Ω^d is updated.

Using the notation above, at each iteration, the candidate locations for placing a sensor are $\Omega^N = \Omega \setminus (\Omega^d \cup B^d)$. For N , the maximum candidate locations for placing a sensor in each iteration, the locations are selected by taking the minimum between $|\Omega^N|$ and N locations in Ω^N with the highest DM score. Using small N values leads to sensor allocations mainly influenced by the DM score, whereas using large N values leads to sensor allocations mainly influenced by the correlation. A specific location is then selected from this set of candidate locations by computing the correlation between each candidate location and previously selected sensing locations and choosing the location with the lowest cumulative correlation. Here for each $d \leq d_{max}$, the N producing the best results is used. Typically, the N values were between 30 and 50. The algorithm appears in the box below (Algorithm 1):

Algorithm 1: Placing d_{max} sensors with the highest DM score and the lowest correlation

1. Set: $\Omega^d = \phi$
 2. Compute $DM_\omega \forall \omega \in \Omega$
 3. Place the first sensor in the location with the highest DM score,

$$\omega^* = \max_{\omega \in \Omega} DM_\omega$$
 4. Set:
 - a. $\Omega^d \leftarrow \omega^*$
 - b. $d \leftarrow d_{max} - 1$
 5. While $d > 0$ do:
 - a. Find $\left\{ \min(N, |\Omega \setminus (\Omega^d \cup B^d)|) \right\}$ candidate locations in all optional locations, $\Omega \setminus (\Omega^d \cup B^d)$, with the highest DM score:

$$\Omega^N = \max_{\omega \in \Omega \setminus (\Omega^d \cup B^d)} DM_\omega$$

$$|\Omega^N| = \min(N, |\Omega \setminus (\Omega^d \cup B^d)|)$$
 - b. For each location in $\omega^n \in \Omega^N$, compute the correlation of C_ω^q with all locations where sensors have already been placed; i.e., $\omega^d \in \Omega^d$.
 - c. Place the next sensor in location ω^* within Ω^N , which presents the lowest summation of correlations with all locations in Ω^d :

$$\omega^* = \min_{\omega^n \in \Omega^N} \sum_{\omega^d \in \Omega^d} corr(C_{\omega^d}^q, C_{\omega^n}^q)$$
 - d. Set:
 - i. $\Omega^d \leftarrow \Omega^d \cup \omega^*$
 - ii. $d \leftarrow d_{max} - 1$
-

Here d_{max} was arbitrarily set to 30 and B^{ω^d} , the restricted area in which sensors could not be placed around already deployed sensors, was a rectangular area of 4900 square meters (70 m × 70 m), centered at ω^d .

2.4. Deployment Evaluation

2.4.1. Formulation

The deployment evaluation was conducted using the notation in Nebenzal et al. [36]. Recall that $\{D\}$ is the set of sensors and $\{S\}$, a set of sources. $d \in \{D\}$ is then a sensor located at $\omega^d \in \Omega$. For atmospheric conditions q , d records a pollution level of $C_{\omega^d}^q$. The source $s \in S$, located at $\omega^s \in \Omega$, emits at a rate of r^s . m_{ds}^q is the pollution transfer function, which associates sensor d 's readings, located at $\omega^d \in \Omega$, with the emission of source s at atmospheric condition q :

$$C_{\omega^d}^q = m_{ds}^q \cdot r^s \tag{3}$$

For a multiple source scenario, each sensor's reading consists of the contribution of all sources, i.e.,:

$$C_{\omega^d}^q = \sum_{s \in \{S\}} m_{ds}^q \cdot r^s \tag{4}$$

For the set $\{D\}$, the sources' contributions for each of the sensors can be written in a matrix form:

$$\vec{C}^q = M^q \vec{r}^s \tag{5}$$

where \vec{C}^q is the row measurement vector, M^q is the transfer matrix consisting of m_{ds}^q , and \vec{r}^s is a column vector of the pollution emission rates. M^q can be inferred either through empirical measurements or through a dispersion model.

Sensor deployment is then evaluated based on its ability to predict the source term; i.e., the real emission vector, r^s . This is achieved by formulating an optimization problem for a given meteorological condition, q :

$$\begin{aligned} \min_{\vec{\hat{r}}} \quad & \|M^q \vec{\hat{r}} - \vec{C}^q\| \\ \text{s.t.} \quad & \text{all elements in } \vec{\hat{r}} \geq 0 \end{aligned} \quad (6)$$

The solution to Equation (6) provides $\vec{\hat{r}}$, which represents the estimation of the source term, r^s , and thus its quality, is given by the normalized difference between the estimation of the source term and the true emission vector:

$$E = \frac{\|\vec{\hat{r}} - \vec{r}^s\|}{\|\vec{r}^s\|} \quad (7)$$

The computation of $\vec{\hat{r}}$ is repeated for every condition, q . In this case, 144 meteorological conditions are used, resulting in 144 source-term vector estimates. The element-wise median is taken as the emission rate of each source in the vector.

2.4.2. Source Term Estimation

The zero approximation is a single value source term vector; i.e., $c \cdot \vec{1}$, with c arbitrarily set to 100 kg/h. Hence, sensor placement is based on a simplified source term vector. However, the evaluation was carried out using more realistic emission profiles that serve to evaluate the capability of the network to estimate complex source term vectors. Four configurations were used, one for the SNBH and three for the CBD scenarios, as listed in Table 1. Note that the 5th point source (PS5) in SNBH and CBD.1 is zero. Thus, not all potential sources need to be active.

Table 1. Emission rates from different sources in different scenarios.

		PS1	PS2	PS3	PS4	PS5	Line 1	Line 2	Line 3
SNBH		100	50	200	100	0	N/A	N/A	N/A
CBD	CBD.1	150	100	50	200	0	10	15	20
	CBD.2	5	3	9	7	10	1000	1000	1000
	CBD.3	50	30	90	70	10	100	200	150

2.4.3. Comparison of Deployment Methods

Three deployment methods are evaluated in this study: random deployment and the application of Algorithm 1, using either the hotspot approach or the entropy to compute the DM score. The random deployment is justified because sensor installation is often governed by the availability of infrastructures, such as public facilities, power sources, utility poles, and communication towers. Random deployment is evaluated for each atmospheric condition, q , for $1 \leq d_{max} \leq 30$. In these cases, for a given d the only two limitations on sensor placement are to avoid non-vacant grid points and to satisfy the “Box-Out” criterion; i.e., the possible candidate locations have to satisfy $\omega \in \Omega \setminus (\Omega^d \cup B^d)$. Once the d_{max} sensors are placed, deployment is evaluated using the method detailed above. Since this is a random process, the average error is computed for several different deployments (in this work, 50).

3. Results and Discussion

3.1. Dense Pollution Maps

The dense maps were computed for all 144 meteorological conditions extracted from a real dataset in Hadera, Israel, using the GRAL/GRAMM Lagrangian atmospheric dispersion model. Figure 2 shows four dense pollution maps in $\mu\text{gr}/\text{m}^3$, for the two scenarios under a zero-approximation emission vector, where (a) and (b) depict the pollution over SNBH in two different meteorological conditions, and (c) and (d) over CBD in two other meteorological configurations. Note that the color scale is different for the SNBH and CBD scenarios. Note that these values, as defined Equation (2), are cumulative rather than instantaneous values.

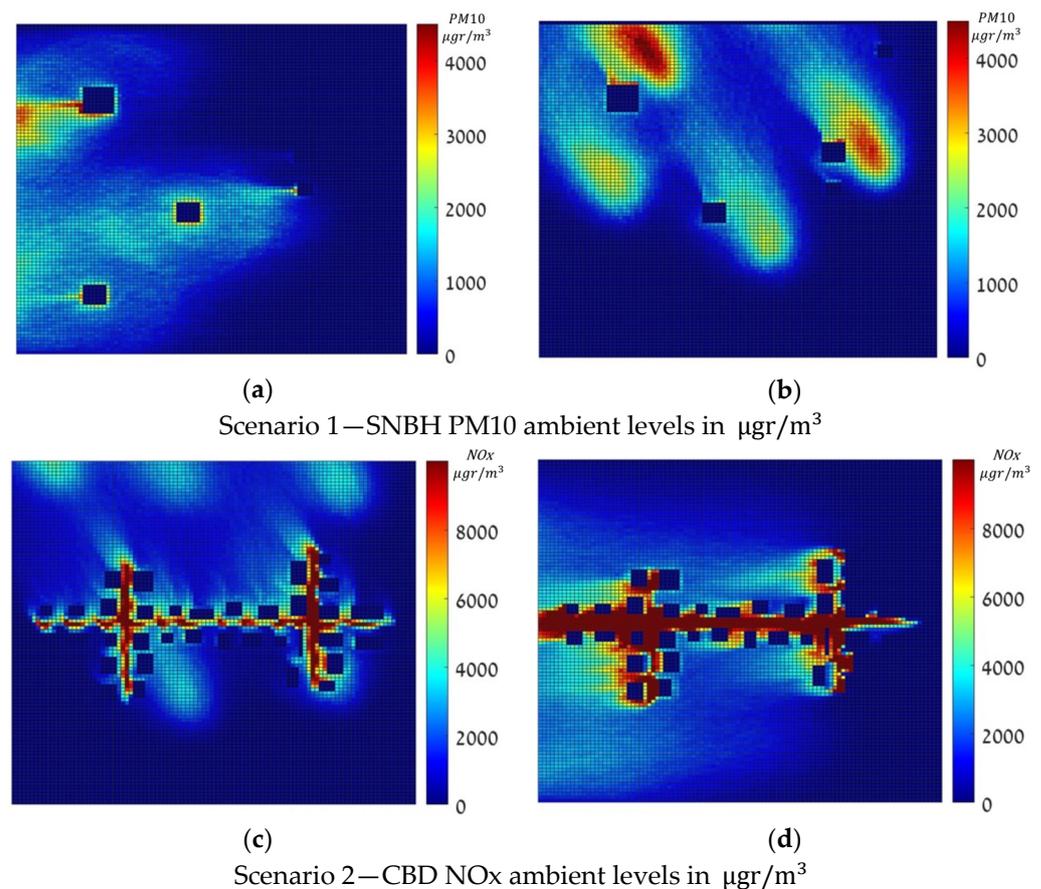


Figure 2. Dense Pollution Maps, in $\mu\text{gr}/\text{m}^3$, under a zero-approximation emission vector.

3.2. Decision Matrix

Based on the 144 dense pollution maps, two types of decision matrices, Entropy (Equation (1)) and hotspot (Equation (2)), were computed. The entropy and hotspot decision matrices are presented in Figures 3 and 4, respectively, where the right-hand side (a) in both figures represents the decision matrix for the SNBH, and the left-hand side shows the same matrix for CBD. It shows that for the SNBH scenario, similar patterns are obtained for both methods, whereas for the CBD scenario, each metric resulted in a different pattern. Hence, the data gain did not necessarily coalign with the locations with the highest pollution.

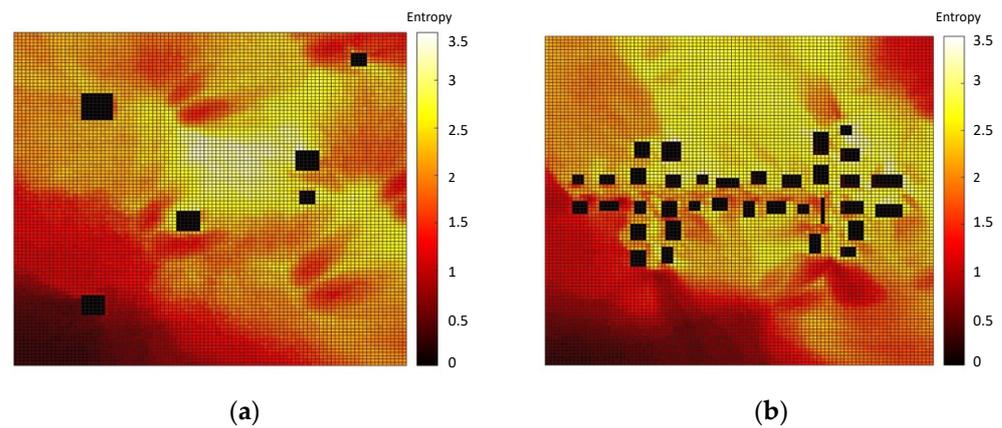


Figure 3. Entropy decision measures for the SNBH (a) and CBD (b) scenarios. Entropy has no units.

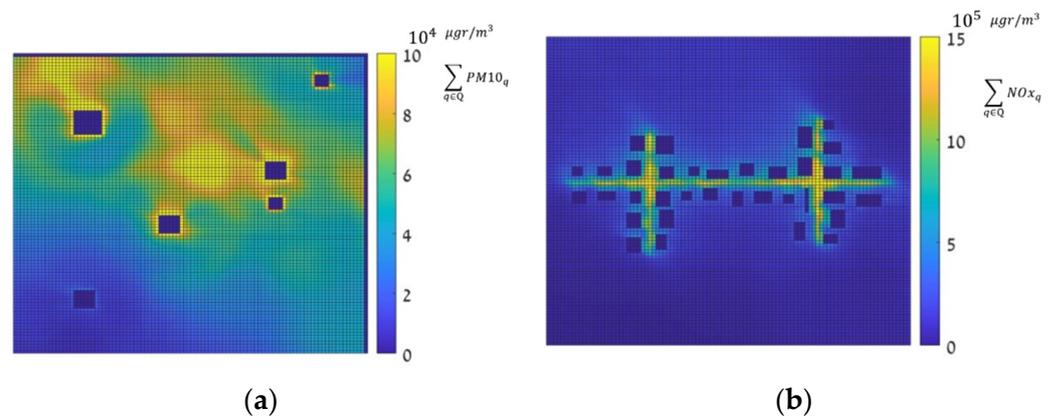


Figure 4. Hot spot decision measure, in $\mu\text{gr}/\text{m}^3$, for the SNBH (a) and CBD (b) scenarios. Note that the color scale is different for figure (a,b).

3.3. Sensor Placement

The optimal deployments of 30 sensors, as dictated by Algorithm 1 for the two DMs, entropy and hotspot, are presented Figure 5. The locations identified by the entropy DM are marked in blue, while the locations identified by the hotspot DM are marked in red. Comparing the two deployments, while considering Figures 3 and 4 show that the main difference between the two deployments is that while the hotspot assigns the locations in the epicenter of the pollution field, the entropy allocates locations with larger pollution field gradients. This conclusion co-aligns with the conclusion reached by Kendler and Fishbain [37].

The optimal sensor deployment was assessed in terms of source-term estimation accuracy (Equation (7)). Figures 6–8 show the error of the source term estimation for each of the deployment methods (entropy, hot spot, and random) as a function of d_{max} , for each SNBH, CBD.1, CBD.2, and CBD.3 scenario, as detailed in Table 1. The graphs show the error (left) and the cumulative error (right) as a function of the number of sensors. The graphs clearly indicate that, in all cases, the error did not reach stagnation and decreased steadily with each sensor added to the array in a similar way for all three metrics. The cumulative error for the entropy metric deployment was slightly better for each additional sensor deployed.

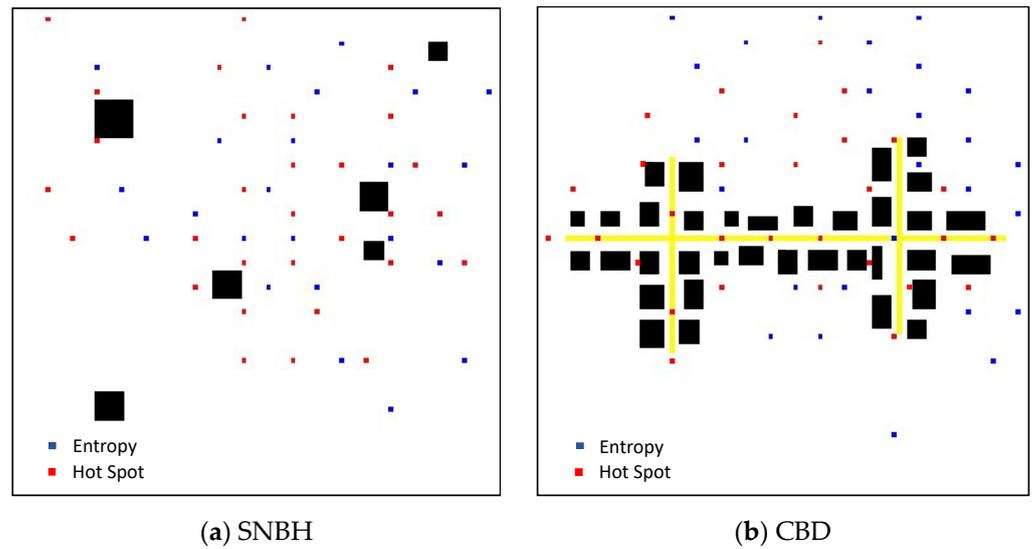


Figure 5. Optimal deployment as computed by Algorithm 1 for the SNBH (a) and CBD (b) configurations. The locations identified by the Entropy DM are marked in blue, while the locations identified by the hotspot DM are marked in red.

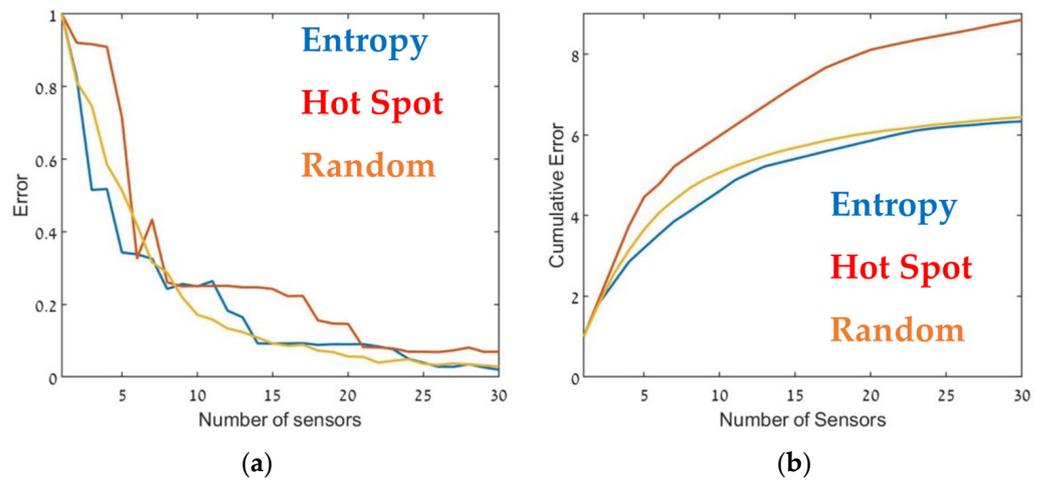


Figure 6. Error (Equation (7)) as a function of the number of sensors, d_{max} (a) and the cumulative error (b) for SNBH. Entropy in blue, hot spot in red, and random in orange.

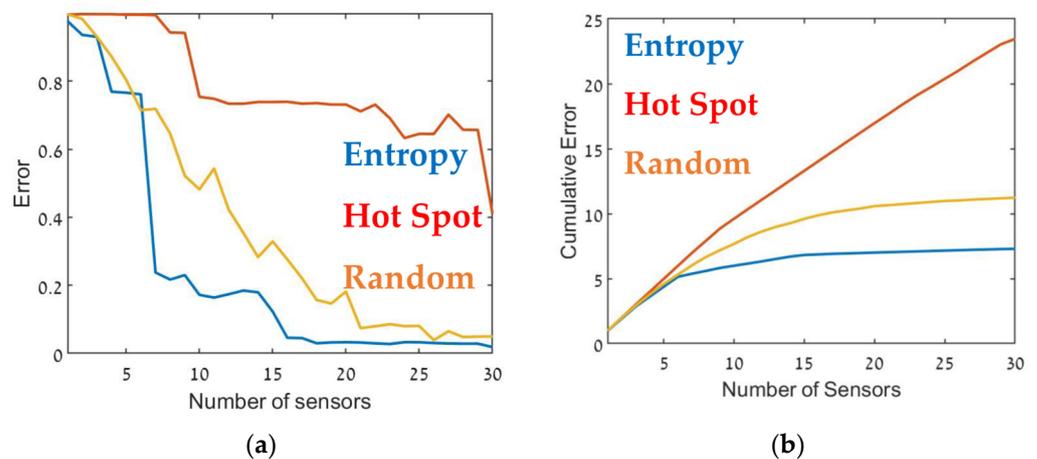


Figure 7. Error as a function of the number of sensors, d_{max} (a) and the cumulative error (b) for the CBD.1 configuration (see Table 1). Entropy in blue, hot spot in red, and random in orange.

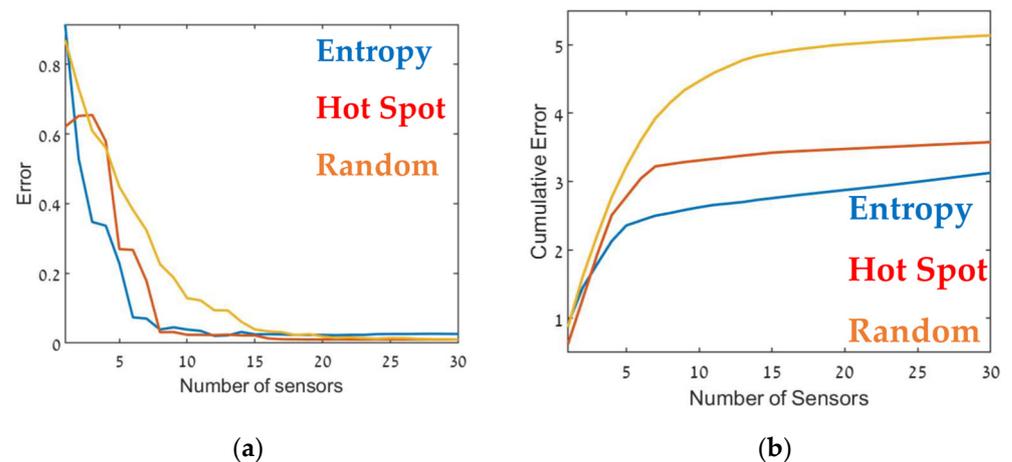


Figure 8. Error as a function of the number of sensors, d_{max} (a) and the cumulative error (b) for the CBD.2 configuration (see Table 1). Entropy in blue, hot spot in red, and random in orange.

Although, in general, similar results were obtained in all cases, there were some interesting differences. In the case of CBD.1, where the point sources were dominant, with larger emission rates by one order of magnitude than the line sources, the Entropy metric significantly outperformed the hotspot metric. When the line sources were two orders of magnitude higher than the point sources (CBD.2), the difference between the entropy and hotspot metrics was less pronounced. In the case of the CBD.3 scenario in which the line source emission rate was only one order of magnitude higher than the point source, the entropy metric was marginally superior for sensor arrays comprised of fewer than 10–12 sensors. Increasing the number of sensors beyond 15 decreased the error when using the entropy metric but only led to a minor improvement in the hotspot metric. This difference was most noticeable for large sensor arrays, where the error for the entropy metric was five times lower than the hotspot metric. These findings suggest that the entropy metric was superior to the other two methods since it provided lower error and evidenced greater stability to changes in the site.

Table 2 presents the cumulative error obtained from each case for the maximum number of sensors, i.e., the overall error. The table suggests that in all cases, using the entropy metric for deploying the sensors resulted in an overall lower error compared to the hot spot or random deployment metrics. In certain specific cases, for example, Figure 9, for 12–13 sensors, the random or hot spot metric emerged as slightly better than entropy, but the overall trend was clear. Further, using the hot spot metric resulted in inconsistent performance, where in some cases the results for this metric were similar to those obtained using entropy (for example, Figure 8), but in others was considerably worse (Figure 7).

Table 2. The cumulative error in each scenario for each deployment metric. Best result, for each scenario (SNBH, CBD.1, CBD.2 and CBD.3), are highlighted in bold.

	Entropy	Hot Spot	Random	Max Random
SNBH	6.34	8.85	6.44	13.64
CBD.1	7.29	23.45	11.23	25.14
CBD.2	3.13	3.58	5.14	12.77
CBD.3	6.04	11.25	7.92	23.35

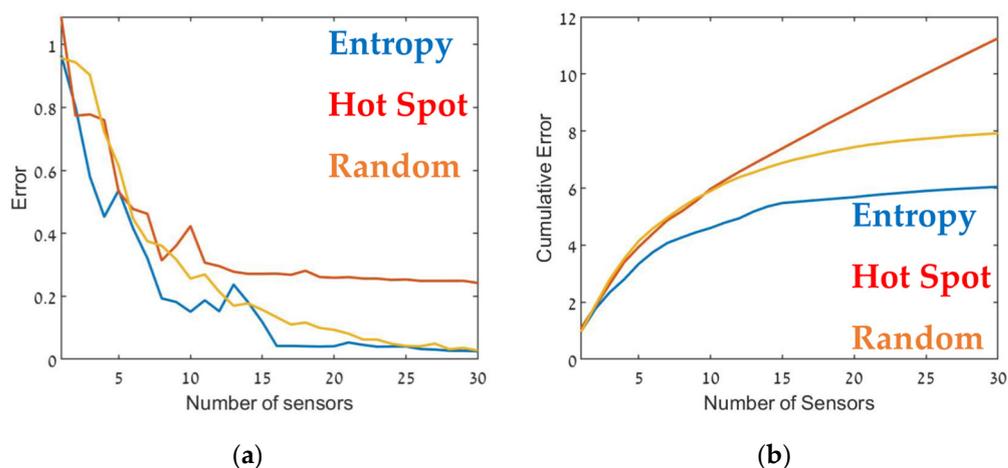


Figure 9. Error as a function of the number of sensors, d_{max} (a) and the cumulative error (b) for the CBD.3 configuration (see Table 1). Entropy in blue, hot spot in red, and random in orange.

Thus, Figure 6 through Figure 9 and Table 2 show that the entropy approach produced superior accuracy and stability over the hot spot or random methods. These findings suggest that sensors should be placed in locations where the information gain is maximal. In the entropy method, informativity is based on a concept taken from information theory, where the entropy of a random variable is the average level of information inherent in the variable's possible outcomes. In this case, the random variable was the pollution level at each grid point, and its possible outcomes were all possible readings of the air pollution sensor.

4. Conclusions

This study presents a new information theory-based approach to the deployment of air pollution sensors. This entropy-based approach was compared to methods based on hot spots and random metrics. The random approach simulated situations in which no previous knowledge was available, and the deployment was mainly dictated by availability and convenience.

The presented method does not require any prior knowledge of the number and source locations, since the optimization problem (Equation (6)) can be solved for sources with a pollution emission rate that equals zero. The ability of the method here to estimate a source with zero emissions (see source PS5 in CBD.1 in Table 1) makes it possible to set the source term vector to be the size of the entire region of interest; thus, $|\vec{\hat{r}}| = |\Omega|$ and a source can be in each $\omega \in \Omega$. In this fashion, locations without a source are estimated to be zero. A similar notion was presented by Nebenzal et al. [36,38]. The study here addresses optimal sensor deployment in a small neighborhood. The algorithm can be easily adapted to a large-scale deployment in a city and even on a national scale.

Future work will include the extension of this work to cases where the number and locations of the sources are unknown, where the simulations will be run with real sensor capabilities (sensitivity and dynamic range), and a comparison to field experiments will be made. The application of the methodology over a large geographical scale will also be sought.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s22103808/s1>, Table S1: List of parameters used for GRAL computation.

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

Funding: This project was funded, in part, by the Israeli Ministry of Environmental Protection grant number 162-7-1 and, in part, by the Israeli ministry of Science & Technology, grant No. #3-15628.

Data Availability Statement: Not applicable.

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

References

1. Smith, K.R.; Bruce, N.; Balakrishnan, K.; Adair-Rohani, H.; Balmes, J.; Chafe, Z.; Dherani, M.; Hosgood, H.D.; Mehta, S.; Pope, D.; et al. Millions Dead: How Do We Know and What Does It Mean? Methods Used in the Comparative Risk Assessment of Household Air Pollution. *Annu. Rev. Public Health* **2014**, *35*, 185–206. [[CrossRef](#)] [[PubMed](#)]
2. World Bank. *The Cost of Air Pollution*; World Bank: Washington, DC, USA, 2016. [[CrossRef](#)]
3. Voegelé, J.; Kemper, K.; Bucknall, J.; Bosquet, B.; Blarel, B.; Shuker, I.; Christian, A.P. *The Global Health Cost of Ambient PM_{2.5} Air Pollution*; World Bank: Washington, DC, USA, 2020.
4. Moltchanov, S.; Levy, I.; Etzion, Y.; Lerner, U.; Broday, D.M.D.M.; Fishbain, B. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Sci. Total Environ.* **2015**, *502*, 537–547. [[CrossRef](#)] [[PubMed](#)]
5. Fishbain, B.; Lerner, U.; Castell, N.; Cole-Hunter, T.; Popoola, O.; Broday, D.; Iñiguez, T.M.; Nieuwenhuijsen, M.; Jovasevic-Stojanovic, M.; Topalovic, D.; et al. An evaluation tool kit of air quality micro-sensing units. *Sci. Total Environ.* **2017**, *575*, 639–648. [[CrossRef](#)]
6. Woo, J.H.; An, S.M.; Hong, K.; Kim, J.J.; Lim, S.B.; Kim, H.S.; Eum, J.H. Integration of CFD-Based Virtual Sensors to A Ubiquitous Sensor Network to Support Micro-Scale Air Quality Management. *J. Environ. Inform.* **2016**, *27*, 85–97. [[CrossRef](#)]
7. Becker, T.; Mühlberger, S.; Braunmühl, C.B.; Müller, G.; Ziemann, T.; Hechtenberg, K.V. Air pollution monitoring using tin-oxide-based microreactor systems. *Sens. Actuators B Chem.* **2000**, *69*, 108–119. [[CrossRef](#)]
8. Barrenetxea, G.; Ingelrest, F.; Schaefer, G.; Vetterli, M. The hitchhiker’s guide to successful wireless sensor network deployments. In Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, Raleigh, NC, USA, 5–7 November 2008; pp. 43–56. [[CrossRef](#)]
9. Cooper, L. Location-Allocation Problems. *Oper. Res.* **1963**, *11*, 331–343. [[CrossRef](#)]
10. Kanaroglou, P.S.; Jerrett, M.; Morrison, J.; Beckerman, B.; Arain, M.A.; Gilbert, N.; Brook, J.R. Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach. *Atmos. Environ.* **2005**, *39*, 2399–2409. [[CrossRef](#)]
11. Lerner, U.; Hirshfeld, O.; Fishbain, B. Optimal deployment of a heterogeneous air quality sensor network. *J. Environ. Inform.* **2019**, *34*, 99–107. [[CrossRef](#)]
12. Kendler, S.; Nebenzal, A.; Gold, D.; Reed, P.M.; Fishbain, B. The effects of air pollution sources/sensor array configurations on the likelihood of obtaining accurate source term estimations. *Atmos. Environ.* **2020**, *246*, 117754. [[CrossRef](#)]
13. Perelman, L.S.; Abbas, W.; Koutsoukos, X.; Amin, S. Sensor placement for fault location identification in water networks: A minimum test cover approach. *Automatica* **2016**, *72*, 166–176. [[CrossRef](#)]
14. Zhao, Y.; Schwartz, R.; Salomons, E.; Ostfeld, A.; Poor, H.V. New formulation and optimization methods for water sensor placement. *Environ. Model. Softw.* **2016**, *76*, 128–136. [[CrossRef](#)]
15. Sankary, N.; Ostfeld, A. Multiobjective Optimization of Inline Mobile and Fixed Wireless Sensor Networks under Conditions of Demand Uncertainty. *J. Water Resour. Plan. Manag.* **2018**, *144*, 04018043. Available online: <https://ascelibrary.org/doi/abs/10.1061/1.4128ASCE%29WR.1943-5452.0000930> (accessed on 29 December 2021). [[CrossRef](#)]
16. Li, D.-S.; Li, H.-N. The state of the art of sensor placement methods in structural health monitoring. *Proc. SPIE* **2006**, *6174*, 1217–1227.
17. Yoshida, I.; Tasaki, Y.; Tomizawa, Y. *Optimal Placement of Sampling Locations for Identification of a Two-Dimensional Space*; Taylor & Francis: Abingdon, UK, 2021.
18. Zou, B.; Zheng, Z.; Wan, N.; Qiu, Y.; Wilson, J. An optimized spatial proximity model for fine particulate matter air pollution exposure assessment in areas of sparse monitoring. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 727–747. [[CrossRef](#)]
19. Feng, Q.; Peng, D.; Gu, Y. Research of regularization techniques for SAR target recognition using deep CNN models. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018), Chengdu, China, 6 May 2019; p. 5. [[CrossRef](#)]
20. Alsahli, M.; Al-Harbi, M. Allocating optimum sites for air quality monitoring stations using GIS suitability analysis. *Urban Clim.* **2018**, *24*, 875–886. [[CrossRef](#)]
21. Walsh, S.J.; Lightfoot, D.R.; Butler, D.R. Recognition and assessment of error in geographic information systems. *Photogramm. Eng. Remote Sens.* **1987**, *53*, 1423–1430. Available online: https://www.asprs.org/wp-content/uploads/pers/1987journal/oct/1987_oct_1423-1430.pdf (accessed on 25 December 2021).
22. Al-Karaki, J.; Gawanmeh, A. The Optimal Deployment, Coverage, and Connectivity Problems in Wireless Sensor Networks: Revisited. *IEEE Access* **2017**, *5*, 18051–18065. [[CrossRef](#)]
23. Boubrima, A.; Matigot, F.; Bechkit, W.; Rivano, H.; Ruas, A. Optimal Deployment of Wireless Sensor Networks for Air Pollution Monitoring. In Proceedings of the 2015 24th International Conference on Computer Communication and Networks (ICCCN), Las Vegas, NV, USA, 3–6 August 2015; IEEE: Piscataway, NJ, USA, 2015.

24. Boubrima, A.; Bechkit, W.; Rivano, H. A new WSN deployment approach for air pollution monitoring. In Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 455–460.
25. Boubrima, A.; Bechkit, W.; Rivano, H. Optimal WSN Deployment Models for Air Pollution Monitoring. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 2723–2735. [[CrossRef](#)]
26. Benis, K.Z.; Fatehifar, E.; Shafiei, S.; Nahr, F.K.; Purfarhadi, Y. Design of a sensitive air quality monitoring network using an integrated optimization approach. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 779–793. [[CrossRef](#)]
27. Al-Adwani, S.; Elkamel, A.; Duever, T.; Yetilmesoy, K.; Abdul-Wahab, S. A Surrogate-Based Optimization Methodology for the Optimal Design of an Air Quality Monitoring Network. *Can. J. Chem. Eng.* **2015**, *93*, 1176–1187. [[CrossRef](#)]
28. Hassin, R.; Levin, A. A Better-Than-Greedy Approximation Algorithm for the Minimum Set Cover Problem. *SIAM J. Comput.* **2006**, *35*, 189–200. Available online: <http://www.siam.org/journals/sicomp/35-1/44475.html> (accessed on 29 December 2021). [[CrossRef](#)]
29. Kumar, N.; Nixon, V.; Sinha, K.; Jiang, X.; Ziegenhorn, S.; Peters, T. An Optimal Spatial Configuration of Sample Sites for Air Pollution Monitoring. *J. Air Waste Manag. Assoc.* **2009**, *59*, 1308–1316. [[CrossRef](#)] [[PubMed](#)]
30. Berchet, A.; Zink, K.; Oetli, D.; Brunner, J.; Emmenegger, L.; Brunner, D. Evaluation of high-resolution GRAMM-GRAL (v15.12/v14.8) NOx simulations over the city of Zürich, Switzerland. *Geosci. Model Dev.* **2017**, *10*, 3441–3459. [[CrossRef](#)]
31. Oetli, D.; Sturm, P.; Almbauer, R. Evaluation of GRAL for the pollutant dispersion from a city street tunnel portal at depressed level. *Environ. Model. Softw.* **2005**, *20*, 499–504. [[CrossRef](#)]
32. Oetli, D. Evaluation of the Revised Lagrangian Particle Model GRAL against Wind-Tunnel and Field Observations in the Presence of Obstacles. *Bound.-Layer Meteorol.* **2015**, *155*, 271–287. [[CrossRef](#)]
33. Graz University of Technology. GRAL-Graz Lagrangian Model. 2020. Available online: <http://lampz.tugraz.at/~{}gral/index.php/2-uncategorised/1-description#> (accessed on 24 April 2020).
34. Salim, S.M.; Buccolieri, R.; Chan, A.; di Sabatino, S. Numerical simulation of atmospheric pollutant dispersion in an urban street canyon: Comparison between RANS and LES. *J. Wind Eng. Ind. Aerodyn.* **2011**, *99*, 103–113. [[CrossRef](#)]
35. Woodward, J. Appendix A: Atmospheric Stability Classification Schemes, Estimating the Flammable Mass of a Vapor Cloud. In *Estimating the Flammable Mass of a Vapor Cloud: A CCPS Concept Book*; Wiley: Hoboken, NJ, USA, 2010; pp. 209–212. [[CrossRef](#)]
36. Nebenzal, A.; Fishbain, B.; Kendler, S. Model-based dense air pollution maps from sparse sensing in multi-source scenarios. *Environ. Model. Softw.* **2020**, *128*, 104701. [[CrossRef](#)]
37. Kendler, S.; Fishbain, B. Optimal Wireless Distributed Sensor Network Design and Ad-Hoc Deployment in a Chemical Emergency Situation. *Sensors* **2022**, *22*, 2563. [[CrossRef](#)]
38. Nebenzal, A.; Fishbain, B. Hough-based Interpolation Scheme for Generating Accurate Dense Spatial Maps of Air Pollutants from Sparse Sensing. *Int. Fed. Inf. Process. Adv. Inf. Commun. Technol.* **2018**, *507*, 51–60. [[CrossRef](#)]