



HAL
open science

On the Deployment of Wireless Sensor Networks for Air Quality Mapping: Optimization Models and Algorithms

Ahmed Boubrima, Walid Bechkit, Hervé Rivano

► To cite this version:

Ahmed Boubrima, Walid Bechkit, Hervé Rivano. On the Deployment of Wireless Sensor Networks for Air Quality Mapping: Optimization Models and Algorithms. *IEEE/ACM Transactions on Networking*, 2019, 27 (4), pp.1629-1642. 10.1109/TNET.2019.2923737. hal-02157476

HAL Id: hal-02157476

<https://inria.hal.science/hal-02157476v1>

Submitted on 17 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Deployment of Wireless Sensor Networks for Air Quality Mapping: Optimization Models and Algorithms

Ahmed Boubrima, Walid Bechkit and Hervé Rivano

Abstract— Wireless sensor networks (WSN) are widely used in environmental applications where the aim is to sense a physical phenomenon such as temperature, air pollution, etc. A careful deployment of sensors is necessary in order to get a better knowledge of these physical phenomena while ensuring the minimum deployment cost. In this paper, we focus on using WSN for air pollution mapping and tackle the optimization problem of sensor deployment. Unlike most of the existing deployment approaches, which are either generic or assume that sensors have a given detection range, we define an appropriate coverage formulation based on an interpolation formula that is adapted to the characteristics of air pollution sensing. We derive from this formulation two deployment models for air pollution mapping using integer linear programming while ensuring the connectivity of the network and taking into account the sensing error of nodes. We analyze the theoretical complexity of our models and propose heuristic algorithms based on linear programming relaxation and binary search. We perform extensive simulations on a dataset of the Lyon city, France in order to assess the computational complexity of our proposal and evaluate the impact of the deployment requirements on the obtained results.

Index Terms— Wireless sensor networks (WSN), sensor deployment, air pollution mapping, pollution-aware coverage, heterogeneous connectivity.

I. INTRODUCTION

WIRELESS sensor networks (WSN) are widely used in environmental applications where the aim is to sense a physical phenomenon such as temperature, humidity, air pollution, etc. In this context of application, the use of WSN allows us to understand the variations of the phenomenon over the monitoring region and therefore be able to take adequate decisions regarding the impact of the phenomenon [2]. Air pollution is one of the main physical phenomena that still need to be studied and characterized because it highly depends on other phenomena such as temperature and wind variations. In addition, air pollution is becoming a major threat to human health in urban environments. According to the World Health Organization (WHO), exposure to air pollution is accountable to seven million casualties in 2012. In 2013, the International Agency for Research on Cancer (IARC) classified particulate matter, the main component of outdoor pollution, as carcinogenic for humans. Air pollution is therefore considered as a major issue of modern megalopolis, where the majority of world population lives. As a consequence, the effective

monitoring of pollutant emissions is at the heart of many sustainable development efforts, in particular those of smart cities.

Current air pollution monitoring stations are equipped with multiple lab pollution sensors [3]. These systems are however massive, inflexible and expensive. An alternative – or complementary – solution would be to use wireless sensor networks (WSN) [4] [5]. The progress of electrochemical sensors, that are smaller and cheaper while keeping a reasonable measurement quality, makes the use of WSN for air pollution monitoring viable [6]. The use of WSN for air pollution monitoring may target two objectives [7]: i) the regular air pollution sampling and mapping; and ii) the detection of threshold crossings in order to trigger adequate alerts. Unlike the works presented in [8] [9] where our target was threshold monitoring, we focus in this paper on regular mapping while tackling the deployment issue of sensor nodes.

The deployment optimization is a major challenge in WSN design. The problem consists in determining the optimal positions of sensors and sinks so as to cover the environment and ensure the network connectivity while optimizing an objective function such as the deployment cost or the network lifetime [10]. The network is said connected if each sensor can communicate information to at least one sink node. The coverage issue has been often modeled as a k -coverage problem where at least k sensors should monitor each point of interest. Most research work on coverage uses a simple detection model which assumes that a sensor is able to cover a point in the environment if the distance between them is less than a radius called the detection range [11]. This can be true for some applications like presence sensors but is not suitable for pollution monitoring. Indeed, a pollution sensor needs to take a sample of the air in order to determine the pollution concentration. The notion of detection range is thus irrelevant in this context. Although some WSN-based air pollution monitoring systems are already operating [12] [13] [14], the deployment issue of sensor nodes while taking into account the quality of air pollution maps has not yet been investigated.

In this paper, we propose novel optimization models and algorithms for WSN deployment in the context of air pollution mapping based on a preliminary work and few discussions that we presented in [1]. Unlike most of the existing deployment approaches, which are either generic or assume that sensors have a given detection range, we base on pollution-aware interpolation to define an appropriate mathematical formulation

This work is an extended version of a IEEE LCN 2016 paper [1]. Ahmed Boubrima, Walid bechkit and Hervé Rivano are with Université de Lyon, INRIA, INSA-Lyon, CITI-INRIA, F-69621, Villeurbanne, France. Emails: {ahmed.boubrima,walid.bechkit}@insa-lyon.fr, herve.rivano@inria.fr.

of coverage quality in the context of air pollution mapping. We formulate the quality of air pollution mapping of a given sensor network depending on the sensing error of the deployed nodes and the estimation error of pollution concentration at locations where no sensor is deployed. The estimation error is defined as the difference between the ground truth (or real) value of pollution concentration and the concentration obtained by applying an adequate interpolation method on the measurements of sensor nodes. We derive from our formulation of coverage quality two deployment models using integer programming modeling. The first deployment model allows us to minimize the network deployment cost while ensuring a required coverage quality. In the second model, we propose instead to optimize the coverage quality without exceeding a given deployment budget. Both optimization models ensure the connectivity of the network based on the flow concept, which guarantees that the deployed sensors are able to send pollution data to at least one sink node. We also take into account the sensing error of sensor nodes, the impact of weather conditions on the variations of pollution concentrations and the heterogeneity of sensing and communication characteristics of nodes. We analyze the theoretical complexity of our models and propose heuristic algorithms based on linear programming relaxation and the concept of binary search.

We perform extensive simulations on a dataset of the Lyon city, France in order to assess the computational complexity of the proposed models and algorithms. We show that the less the pollution variability in the deployment region, the higher is the time of solving the optimization models and therefore the higher is the need of using near-optimal heuristics. We also evaluate the performance of our proposal in terms of coverage and connectivity in order to derive engineering insights on the optimal WSN deployment. We show that when considering heterogeneity, accurate sensors are usually more used than low accuracy sensors despite being more expensive. We also show that the choice of the communication technology depends on the requested precision of pollution monitoring.

The remaining of this paper is organized as follows. We first review the related works on the deployment issue of WSN in section II. Then, we present in details our mathematical formulation of air pollution mapping quality in section III. Next, we present our optimization models and heuristic algorithms in sections IV and V. After that, we present the simulation data set and analyze the obtained results in section VI. Finally, we discuss the main characteristics and open issues of our proposal in section VII and conclude the paper in section VIII.

II. RELATED WORKS: WSN DEPLOYMENT

The deployment issue of wireless sensor networks has been addressed extensively in the literature where several mathematical models, optimal algorithms and near-optimal heuristics have been proposed [15]. The problem has been defined in multiple ways depending on the context of the deployment. The main issues targeted in the literature are coverage, connectivity, network lifetime and the network deployment cost. In this section, we identify what lacks in the literature and motivates the need of an application-aware

deployment approach for air pollution mapping where the aim is to place sensors in order to optimize their number and the quality of pollution maps that result from the data gathered by sensors once deployed. We present the related works based on their coverage definition while identifying their formulation of connectivity and network lifetime.

Existing deployment approaches are either event-aware [11], [16]–[20] or correlation-aware [21]–[25]. In the first case, a sensor is assumed to have a detection range, usually circular, within which the sensor is capable of detecting any event that may happen. The second class of deployment approaches is based on the correlation that sensor measurements may present in order to select the minimum number of sensing nodes.

A. Event-aware deployment methods

Chakrabarty et al. [11] represent the deployment region as a grid of points and propose a nonlinear formulation for minimizing the deployment cost of sensors while ensuring complete coverage of the deployment region. Then, they apply some transformations to linearize the first model and obtain an ILP formulation. The authors formulate coverage based on the distance between the different points of the deployment field. Each sensor has a circular detection area, which defines the points that the sensor can cover. Unfortunately, this measure of coverage is inadequate to the air pollution monitoring since a sensor positioned at a point A cannot cover a neighboring point B if there is a difference between pollution concentrations at these two points.

Altinel et al. [16] proposed another formulation based on the Set Cover Problem, which is equivalent to the aforesaid model but less complex. They also extend their formulation to take into account the probabilistic sensing of sensor nodes while assuming that a node is able to cover a given point with a certain predefined probability. Despite that, this new formulation is still generic since the dependency between the errors of the deployed sensors is not considered. However, this has to be taken into account when doing air pollution estimation.

Chang et al. [17] proposed to use data fusion in the definition of coverage in order to take into account the collaborative detection of targets. They based in their work on a probabilistic sensing model to define the probability of target detection and the false alarm rate. Then, they formulated a nonconvex optimization problem minimizing the number of nodes under coverage constraints. They presented resolution algorithms and showed that the obtained solutions are near-optimal and hence very close to the optimal ones. Still, this work considers the existence of a detection range.

Recent works have targeted the connectivity and multi-objective deployment issues. The authors of [18] formulate connectivity based on the flow problem while assuming that sensors generate flow units in the network and verify if sinks are able to recover them. Another connectivity formulation has been introduced in [19] where authors base on an assignment approach. They introduce in their ILP formulation new variables to define the communication paths between sensors and sinks. However, this model involves more variables than

the one based on the flow problem and is therefore more complex. In another work [20], authors study the trade-off between coverage, connectivity and energy consumption. They formulate the problem as an ILP model and then propose a multi-objective approach to optimize coverage, the network lifetime and the deployment cost while maintaining the network connectivity.

B. Correlation-aware deployment methods

In [22], Roy et al. tackled the problem of finding the most informative locations of sensors for monitoring environmental applications. They assume the existence of a set of data snapshots characterizing the phenomenon to monitor. Then, they formulate the problem to find the best locations of sensors in order to reconstruct the data of the whole phenomenon with a required precision. Two optimization models are proposed to handle both stationary and non stationary-fields. An iterative resolution algorithm is proposed to solve the two deployment problems. Unfortunately, this work is based on a strong assumption; that is input data is perfect, which is not the case of air pollution where simulated data may present some errors.

In [24], Krause et al. tackle the same problem based on the assumption that the variations of the phenomenon are Gaussian. They also assume a pre-deployment phase allowing to gather data that can be used to characterize the phenomenon. In order to select the best positions of sensors, they use the concept of mutual information in order to define the quality of a given topology. After the formulation of the problem, they use the sub-modularity of mutual information to define a polynomial algorithm. This work considers only coverage and is extended in [25] to take into account the cost of connectivity where the links qualities are assumed to be Gaussian. Since air pollution is not necessarily Gaussian, this work does not fit our application case.

The mathematical characteristics of the correlation-aware deployment problem has been studied by Ranieri et al. in [21] while considering a generic form. A greedy heuristic is proposed to solve the problem. They perform extensive simulations to show that their algorithm is capable of solving the problem in a short time compared to the existing heuristics while providing a near optimal solution.

In [23], authors consider an already deployed sensor network and propose an algorithm to define a sensing topology to select active sensors and turn off the others. They estimate the variations of the phenomenon in an online way to decide whether a sensor is to keep active or not. In contrary to this work, in our case, the sensing locations have to be chosen in an offline way since the selection of sensing points is performed before the network deployment.

C. Discussion

Even if the recent works take into account network constraints like connectivity and energy consumption, all coverage formulations either assume that sensors have a given detection range, which is the case of event-aware methods, or the assumption is instead made on the distribution of sensor measurements, which is the case of correlation-aware

methods. Novel application-aware deployment methods have been recently proposed to consider the characteristics of the application case in the design of the deployment approach; examples include the work of [26] on wind monitoring and the work of [27] for hot server detection in data centers. Following the same direction, we propose in the next section to consider the context of air pollution mapping in order to define an appropriate formulation of coverage quality and then we derive optimization models and heuristic algorithms in the following sections.

III. MATHEMATICAL FORMULATION OF AIR POLLUTION MAPPING QUALITY

A. Characterization of the deployment region

We consider as input the map of a given urban area that we call the deployment region. Let \mathcal{P} be a set of discrete points approximating the deployment region at a high-scale ($|\mathcal{P}| = \mathcal{N}$). The set \mathcal{P} can be obtained using a 2D or 3D discretization. Our goal in this paper is to be able to determine with a high precision the concentration value at each point $p \in \mathcal{P}$. We ensure that for each point $p \in \mathcal{P}$, either a sensor is deployed or the pollution concentration can be estimated with a high precision based on the data gathered by the neighboring deployed sensors. We also ensure that all the deployed sensors can send their data to at least one sink node while optimizing the positions of sink nodes. The main notations used in this section are presented in TABLE I.

In general case, the set \mathcal{P} is considered as the set of potential positions of WSN nodes. However, in smart cities applications, some restrictions on node positions may apply because of authorization or practical issues. For instance, in order to alleviate the energy constraints, we may place sensors on only lampposts and traffic lights as experimented in [28]. When this is the case, we do not consider as potential positions the points $p \in \mathcal{P}$ where sensors cannot be deployed.

We use decision variables x_p (respectively y_p) to specify if a sensor (respectively a sink) is deployed at point p or not. For sink nodes, we consider as potential positions the set $\hat{\mathcal{P}} = \mathcal{P} \cup \mathcal{P}_S$ where \mathcal{P}_S is a set of additional potential positions where only sinks can be deployed; i.e. $x_p = 0 \forall p \in \mathcal{P}_S$. In addition, we assume that sensors and sinks cannot be deployed at the same location; i.e. $x_p + y_p \leq 1$.

B. Interpolation formulation

As claimed before, our idea is to base on interpolation in order to formulate coverage quality based on the sensing error of the deployed nodes and the estimation error of pollution concentration at locations where no sensor is deployed. Interpolation formulate the estimated concentration \hat{Z}_p at a given location $p \in \mathcal{P}$ as a weighted combination of the measured concentrations Z_q , $q \in \mathcal{P}$ [29]. The weights of the measured concentrations W_{pq} are called correlation coefficients and can be evaluated in a deterministic way based on the distance between the location of the measured concentration and the location of the estimated concentration. In this case, which is called the Inverse Distance Weighting interpolation, \hat{Z}_p is

evaluated using formula 1. The correlation coefficients can be also evaluated in a stochastic way, the most used method doing so is called kriging. Without loss of generality, we focus in this paper on the case of deterministic interpolation.

$$\widehat{\mathcal{Z}}_p = \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} * \mathcal{Z}_q}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq}} \quad (1)$$

Sets and parameters	
\mathcal{P}	Set of points (the deployment region)
\mathcal{N}	Number of points
\mathcal{K}	Set of sensor types
\mathcal{T}	Set of time snapshots
\mathcal{G}_p	Ground truth pollution concentrations
M_p^t	Simulated pollution concentrations ⁽¹⁾ (using atmospheric dispersion simulation)
\mathcal{Z}_p	Measured pollution concentrations
$\widehat{\mathcal{Z}}_p$	Estimated pollution concentrations (using interpolation)
m_p^t	Simulation errors ⁽¹⁾ (using atmospheric dispersion simulation)
s_p^k	Sensing errors ⁽²⁾
\mathcal{W}_{pq}	Correlation coefficients
\mathcal{B}_{pq}	Correlation-existence coefficients
\mathcal{D}	The correlation distance function
d	Maximum correlation distance
α	Attenuation coefficient of the correlation distance
β	Wind coefficient of correlation
$\Gamma^k(p)$	Communication neighborhoods ⁽²⁾
\mathcal{R}^k	Communication ranges ⁽²⁾
A_{pq}^k	Adjacency matrix ⁽²⁾
E	Tolerated estimation error
I	The maximum number of sinks
δ_p^k	The cost of sensors ⁽²⁾
ψ_p	The cost of sinks
\mathcal{F}	Deployment cost of the whole network
J	Deployment budget
Decision variables of optimization models	
x_p^k	Define whether a sensor of type k is deployed at point p or not; ⁽²⁾ $x_p^k \in \{0, 1\}$, $p \in \mathcal{P}$
y_p	Define whether a sink is deployed at point p or not; $y_p \in \{0, 1\}$, $p \in \mathcal{P}$
Auxiliary variables of optimization models	
e_p	Estimation error of pollution concentration at point p ; $e_p \geq 0$, $p \in \mathcal{P}$
g_{pq}	Flow quantity transmitted from node p to node q $g_{pq} \in \{0, 1, \dots\}$, $p \in \mathcal{P}$, $q \in \Gamma(p)$

⁽¹⁾ When considering only one time snapshot, t is omitted.

⁽²⁾ When considering only one type of nodes, k is omitted.

TABLE I: Main notations used in our proposal.

Formula 1 can be generalized in order to take into account that: 1) $\widehat{\mathcal{Z}}_p = \mathcal{Z}_p$ when $x_p = 1$; and 2) \mathcal{Z}_q values should be considered in the fraction only when $x_q = 1$. Hence, we get formula 2 in addition to constraint 3 that ensures that the denominator is never equal to zero. \mathcal{B}_{pq} parameters define whether there is a correlation between points p and q or not; that is, $\mathcal{B}_{pq} = 1$ when $\mathcal{W}_{pq} > 0$.

$$\widehat{\mathcal{Z}}_p = x_p \cdot \mathcal{Z}_p + (1 - x_p) \cdot \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot \mathcal{Z}_q}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \quad (2)$$

$$\sum_{q \in \mathcal{P}} \mathcal{B}_{pq} \cdot x_q \geq 1 \quad (3)$$

C. Correlation coefficients

In the context of air pollution estimation, a variant of Inverse Distance Weighting interpolation can be used in order to calculate the \mathcal{W}_{pq} parameters based on the wind velocity and direction in addition to the distance between the two points p and q . For instance, authors in [30] use formula 4 where $\mathcal{D}(p, q)$ is the distance function and α is the attenuation coefficient of the correlation distance (i.e. for greater values of α , very low correlation coefficients are assigned to far points). \mathcal{B}_{pq} is calculated based on d , which is called the maximum correlation distance and defines the range of the correlated neighboring points of a given point. The last parameter of formula 4 is β , which allows us to increase the correlation when points p and q are in the direction of wind; that is, $\beta = 1$ if the angle between wind direction and $\vec{p}\vec{q}$ or $\vec{q}\vec{p}$ is less than a threshold $\mathcal{T}\mathcal{H}_\beta$, otherwise $\beta = 0$.

$$\mathcal{W}_{pq} = \mathcal{B}_{pq} \cdot \left(\frac{1}{\mathcal{D}(p, q)^\alpha} + \beta \right) \quad (4)$$

$$\mathcal{B}_{pq} = \begin{cases} 1 & \text{if } q \in \text{Disc}(p, d) \\ 0 & \text{if } q \notin \text{Disc}(p, d) \end{cases} \quad (5)$$

Note that the formulation of \mathcal{W}_{pq} in formula 4 does not take into account the impact of wind velocity. In addition, wind direction from p to q or q to p is assumed to have the same impact. To overcome those limitations, formula 6 can be used instead to calculate \mathcal{W}_{pq} parameters. Here, \vec{u} denotes the wind vector, $|\vec{u}|$ the normalized wind velocity in the range $[0, 1]$ and $\cos(\vec{p}\vec{q}, \vec{u})$ is the cosine of the angle between vectors $\vec{p}\vec{q}$ and \vec{u} . Thanks to the use of the cosine function, correlation values are reduced when the direction of wind is opposite to $\vec{p}\vec{q}$.

$$\mathcal{W}_{pq} = \mathcal{B}_{pq} \cdot \frac{1}{\mathcal{D}(p, q)} \cdot |\vec{u}| \cdot \frac{(1 + \cos(\vec{p}\vec{q}, \vec{u}))}{2} \quad (6)$$

In order to take into account the impact of the urban topography on the dispersion of pollutants, \mathcal{D} can be defined as the shortest distance along the roads network. This allows us to assign small correlation values to points that are separated by buildings, even if they are close with respect to the euclidean distance [31].

The proposed formulation of correlation coefficients \mathcal{W}_{pq} corresponds to the application of the Inverse Distance Weighting interpolation as already claimed. This is why the correlation coefficients are calculated based on the distance between sensing locations and estimated locations. However, our formulation is general enough to fit other deterministic interpolation methods mainly natural neighbor interpolation but also others like nearest neighbor interpolation. Natural

neighbor interpolation can be obtained by setting the α parameter of formula 4 to a sufficiently high value so that only nearest neighbors are taken into account in the interpolation formula. In this case, the correlation coefficients can be also estimated using Voronoi diagrams [32]. Indeed, we first determine the Voronoi diagram corresponding to the sensors located in the neighborhood of point p where the interpolation will be calculated. Then, we update the Voronoi diagram by considering point p in addition to its neighboring sensors. As a result, the size of the Voronoi cells corresponding to each neighboring sensor location is reduced due to the adding of the new Voronoi cell corresponding to point p . Finally, we get the correlation coefficient of each neighboring sensor q by evaluating the proportion of the Voronoi cell of sensor q which has been lost after having updated the Voronoi diagram. As in formula 4, we can also take into account the effect of wind direction by adding the β parameter to the final result.

D. Approximation of sensor measurements

In order to be able to use the interpolation formula 2, we need to know the concentrations that can be measured at every point $p \in \mathcal{P}$; i.e. \mathcal{Z}_p . However, since the aim of our work is to determine the best positions of sensors, the network is not yet deployed and hence \mathcal{Z}_p values cannot be considered as input to our formulation. Fortunately, using atmospheric dispersion simulators [33], we can obtain simulated pollution concentrations that can be used to approximate measured pollution concentrations \mathcal{Z}_p . These simulated concentrations depend on weather conditions and pollution emissions. Let \mathcal{G}_p denote the ground truth (or real) value of pollution concentration at point p and \mathcal{M}_p denote the simulated pollution concentration obtained by means of the atmospheric dispersion simulator. The relationship between \mathcal{G}_p and \mathcal{M}_p (respectively \mathcal{G}_p and \mathcal{Z}_p) is presented in formula 7 (respectively in formula 8) where m_p is the simulation error given by the dispersion simulator at point p and s_p is the sensing error in the measurements of a node deployed at point p . We now derive formula 9 that characterizes the error of approximating the measured concentrations \mathcal{Z}_p by the simulated concentrations \mathcal{M}_p .

$$\mathcal{M}_p - \mathcal{G}_p \in [-m_p, m_p] \quad (7)$$

$$\mathcal{Z}_p - \mathcal{G}_p \in [-s_p, s_p] \quad (8)$$

$$\mathcal{Z}_p - \mathcal{M}_p \in [-s_p - m_p, s_p + m_p] \quad (9)$$

E. Basic coverage quality formulation

We define the quality of pollution coverage at point p as the estimation error e_p of the interpolation method; i.e. the absolute difference between the estimated concentration $\widehat{\mathcal{Z}}_p$ and the ground truth concentration \mathcal{G}_p . Hence, we get formula 10 where UB is the least upper bound (or maximum) function, which is used because of the uncertainty in $\widehat{\mathcal{Z}}_p$ and \mathcal{G}_p .

$$e_p = UB(|\widehat{\mathcal{Z}}_p - \mathcal{G}_p|) \quad (10)$$

In order to be able to use this coverage quality measure in the optimization models, we need to express it depending only on the known parameters, this means that we should not use parameters \mathcal{Z} and \mathcal{G} in the estimation error formulation. First we simplify formula 10 to obtain formula 11 and then formula 12 based on the fact that either $x_p = 1$ or $1 - x_p = 1$. Then, we get formula 13 based on the relationship between \mathcal{Z}_p and \mathcal{G}_p presented in formula 8.

$$e_p = UB(|x_p \cdot (\mathcal{Z}_p - \mathcal{G}_p) + (1 - x_p) \cdot \frac{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q}|) \quad (11)$$

$$e_p = x_p \cdot UB(|\mathcal{Z}_p - \mathcal{G}_p|) + (1 - x_p) \cdot \frac{UB(|\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)|)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \quad (12)$$

$$e_p = \frac{x_p \cdot s_p + (1 - x_p) \cdot UB(|\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)|)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \quad (13)$$

In order to eliminate \mathcal{Z}_q and \mathcal{G}_p from 13, notice that the numerator of the fraction can be simplified to get formula 14. Also, notice that the members of the summations in formula 14 can be simplified as shown in formulas 15 and 16.

$$\begin{aligned} & UB(|\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)|) \\ &= Max \{ UB(\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)), \\ & \quad UB(-\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{Z}_q - \mathcal{G}_p)) \} \\ &= Max \{ \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot UB(\mathcal{Z}_q - \mathcal{G}_p), \\ & \quad \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot UB(\mathcal{G}_p - \mathcal{Z}_q) \} \end{aligned} \quad (14)$$

$$\begin{aligned} UB(\mathcal{Z}_q - \mathcal{G}_p) &= UB(\mathcal{M}_q - \mathcal{M}_p + \mathcal{Z}_q - \mathcal{M}_q + \mathcal{M}_p - \mathcal{G}_p) \\ &= \mathcal{M}_q - \mathcal{M}_p + UB(\mathcal{Z}_q - \mathcal{M}_q) + UB(\mathcal{M}_p - \mathcal{G}_p) \\ &= \mathcal{M}_q - \mathcal{M}_p + s_q + m_q + m_p \end{aligned} \quad (15)$$

$$UB(\mathcal{G}_p - \mathcal{Z}_q) = \mathcal{M}_p - \mathcal{M}_q + s_q + m_q + m_p \quad (16)$$

At then end, the expression of the estimation error depending only on known parameters is given in formula 17.

$$\begin{aligned} e_p &= x_p \cdot s_p + \frac{(1 - x_p)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \cdot Max \{ \\ & \quad \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{M}_q - \mathcal{M}_p + s_q + m_q + m_p), \\ & \quad \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{M}_p - \mathcal{M}_q + s_q + m_q + m_p) \} \end{aligned} \quad (17)$$

F. Multi-scenario coverage quality formulation

In the expression of the quality of pollution coverage at point p given in formula 17, the values of \mathcal{M}_p and m_p are considered as constants. We recall that \mathcal{M}_p and m_p are obtained by simulation depending on pollution emissions and weather conditions. This means that when the sensor network is operating at a given moment t , the estimation error e_p corresponds to the expression presented in 17 only if \mathcal{M}_p and m_p values are obtained while considering the weather conditions and pollution emissions that correspond to the moment t . In order to cope with that, \mathcal{M}_p and m_p should be defined as random variables and not constants. By doing so, e_p as defined in formula 17 also becomes a random variable. In order to derive a deterministic formula from the expression presented in 17, we first propose to consider a set of time snapshots \mathcal{T} where each snapshot $t \in \mathcal{T}$ corresponds to a potential scenario of weather conditions and pollution emissions. Then, we define \mathcal{M}_p^t and m_p^t as, respectively, simulated pollution concentration and simulation error at point p while considering weather conditions and pollution emissions corresponding to the snapshot t . Finally, we define in formula 18 the new formulation of pollution coverage quality as the maximum error among all the time snapshots \mathcal{T} . Our formulation remains general and can also accommodate the mean error instead of the maximum error if desired by the application. Note that the number of the time snapshots is a key factor in this new definition of coverage quality. Indeed, the more the number of snapshots, the better is the approximation of the random variables \mathcal{M}_p and m_p . Another key factor is the granularity of the snapshots which can be as low as hourly or as high as yearly.

$$e_p = \text{Max}_{t \in \mathcal{T}} \left\{ x_p \cdot s_p + \frac{(1 - x_p)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q} \cdot \text{Max} \left\{ \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{M}_q^t - \mathcal{M}_p^t + s_q + m_q^t + m_p^t), \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot x_q \cdot (\mathcal{M}_p^t - \mathcal{M}_q^t + s_q + m_q^t + m_p^t) \right\} \right\} \quad (18)$$

G. Taking into account sensing heterogeneity

So far, we have been considering that all the sensor nodes have the same sensing error at a given point p . The coverage quality formulation can be more general by considering a set of sensor types \mathcal{K} where the sensing error depends on the type of the sensor; i.e the sensing error of a sensor of type $k \in \mathcal{K}$ at point $p \in \mathcal{P}$ is denoted s_p^k . In this case, the index k is also added to variables x_p in order to denote the fact that a sensor of type k is placed at point p , hence we get x_p^k variables. Note that only one type can be chosen to be deployed at point p as formulated in formula 19.

$$\sum_{k \in \mathcal{K}} x_p^k \leq 1, \quad p \in \mathcal{P} \quad (19)$$

In order to take into account the heterogeneity of sensor nodes in the formulation of coverage quality, we transform formula 18 to formula 20 by adding the index k to the sensing errors and replacing x_p variables by $\sum_{k \in \mathcal{K}} x_p^k$. Since the denominator has been changed, we need to transform formula 3 to formula 21 in order to ensure that the denominator is always greater than 0.

$$e_p = \text{Max}_{t \in \mathcal{T}} \left\{ \sum_{k \in \mathcal{K}} x_p^k \cdot s_p^k + \frac{(1 - \sum_{k \in \mathcal{K}} x_p^k)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k} \cdot \text{Max} \left\{ \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_q^t - \mathcal{M}_p^t + s_q^k + m_q^t + m_p^t), \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_p^t - \mathcal{M}_q^t + s_q^k + m_q^t + m_p^t) \right\} \right\} \quad (20)$$

$$\sum_{q \in \mathcal{P}} \mathcal{B}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \geq 1, \quad p \in \mathcal{P} \quad (21)$$

H. Taking into account the reliability of sensor measurements

The measurements of sensor nodes may be different compared to the ground truth values because of environmental conditions like high temperatures for instance [34]. Moreover, the way sensors react to environmental conditions may be different from a sensor to another even if they measure the same pollutant because the sensing quality highly depends on the brand of the sensors (also denoted in our paper as the type of the sensors). Indeed, during high temperatures for instance, the drift of sensor measurements is not the same depending on their brand.

We recall that the difference between sensor measurements and ground truth values (a.k.a. the sensing errors) is denoted by s_p^k in our definition of coverage quality in formula 20. Here, p is the point where the sensor is deployed and k is the type (or the brand) of the sensor. In order to take into account the impact of weather conditions on sensing errors, we add the index t to s_p^k to get s_p^{kt} where $t \in \mathcal{T}$ and \mathcal{T} is the set of snapshots characterizing weather conditions as defined in section III-F.

The identification of s_p^{kt} values is necessary before the deployment process. Fortunately, sensing errors can be characterized during the calibration process of sensor nodes [35]. The idea is to expose the sensors to different weather conditions and then compare their measurements to those provided by reference stations. For each weather condition (i.e. for each $t \in \mathcal{T}$), several measurements are performed and the s_p^{kt} value corresponds to the maximum error of the sensor node in those measurements.

During the interpolation process, measurements with high sensing errors should be removed from formula 20. In order to ignore the bad quality measurements in the coverage quality formulation, we first add the indices k and t to the correlation coefficients \mathcal{W}_{pq} defined in section III-C; i.e. \mathcal{W}_{pq}^{kt}

corresponds then to the weight of the measurement of sensor k deployed at point q during weather condition t . Then, we propose to set \mathcal{W}_{pq}^{kt} to 0 when the sensing error during weather condition t , i.e. s_q^{kt} , exceeds a given threshold defined as input to our deployment approach.

In the following section, we design optimization models to find the optimal locations of sensor nodes while using our pollution-aware definition of coverage quality. Sensors are deployed in order to reduce the estimation error of interpolation at locations where no sensor is deployed. Thanks to the use of s_p^{kt} parameters and \mathcal{W}_{pq}^{kt} correlation coefficients in our coverage formulation, the optimization models presented in the following section will ensure in the deployment results that sensors having good precision will be preferred over those having bad precision as follows:

- If for some weather condition t , a sensor node k has a sensing error exceeding the threshold value, the measurements of the node in question occurring during weather condition t will not be taken into account in the interpolation formulation.
- Sensor nodes with s_p^{kt} exceeding the threshold of the sensing error for all the weather conditions are considered as defective and will not be deployed at all.

IV. OPTIMIZATION MODELS

A. *MIN_COST: Deployment cost minimization under coverage quality requirements*

1) *Objective function:* We first denote by δ_p^k (respectively ψ_p) the deployment cost of a sensor of type k (respectively a sink) at point p . The network deployment cost to minimize is thus given as follows:

$$\mathcal{F} = \sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{K}} \delta_p^k * x_p^k + \sum_{p \in \mathcal{P}} \psi_p * y_p \quad (22)$$

2) *Air pollution mapping constraints:* First, let the function $f(\{e_p, p \in \mathcal{P}\})$ denote the estimation error of a given sensor network topology. f is calculated based on the estimation error at all the points that define the deployment region. Without loss of generality, we consider in this paper that f is either the maximum or the mean function. The main constraint of air pollution mapping is defined in formula 23 and ensures that the estimation error in the deployment region does not exceed the required precision E that we call the tolerated estimation error. The linear definition of function f is given in constraint 24 (respectively 25) in the case of the max function (respectively the mean function) where $N = |\mathcal{P}|$.

$$f \leq E \quad (23)$$

$$f \geq e_p, \quad p \in \mathcal{P} \quad (24)$$

$$f = \sum_{p \in \mathcal{P}} e_p / N \quad (25)$$

Constraint 23 together with constraint 20 defined in the previous section ensure that the pollution estimation error of the resulting network is bounded by the tolerated error E . However, constraint 20 is too complex and should be linearized in order to be solved efficiently by a mathematical solver and this by eliminating the *Max* and division functions in addition to the product between variables x_p^k , x_q^k and e_p . Note that since

we are constraining e_p variables to minimum values, the equality in constraint 20 can be transformed to an inequality as in constraint 26. Therefore, the *Max* function can be easily eliminated and thus we get constraints 27 and 28.

$$e_p \geq \text{Max}_{t \in \mathcal{T}} \left\{ \sum_{k \in \mathcal{K}} x_p^k \cdot s_p^k + \frac{(1 - \sum_{k \in \mathcal{K}} x_p^k)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k} \cdot \text{Max} \left\{ \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_q^t - \mathcal{M}_p^t + s_q^k + m_q^t + m_p^t), \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_p^t - \mathcal{M}_q^t + s_q^k + m_q^t + m_p^t) \right\} \right\} \quad (26)$$

$$e_p \geq \sum_{k \in \mathcal{K}} x_p^k \cdot s_p^k + \frac{(1 - \sum_{k \in \mathcal{K}} x_p^k)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k} \cdot \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_q^t - \mathcal{M}_p^t + s_q^k + m_q^t + m_p^t), \quad p \in \mathcal{P}, t \in \mathcal{T} \quad (27)$$

$$e_p \geq \sum_{k \in \mathcal{K}} x_p^k \cdot s_p^k + \frac{(1 - \sum_{k \in \mathcal{K}} x_p^k)}{\sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k} \cdot \sum_{q \in \mathcal{P}} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} x_q^k \cdot (\mathcal{M}_p^t - \mathcal{M}_q^t + s_q^k + m_q^t + m_p^t), \quad p \in \mathcal{P}, t \in \mathcal{T} \quad (28)$$

We simplify constraints 27 and 28 by multiplying their both sides by the denominator; thus we get, respectively, constraints 29 and 30 where h_{pq}^{kt} and l_{pq}^{kt} variables are defined in formulas 31 and 32.

$$\sum_{q \in \mathcal{P}: \mathcal{B}_{pq}=1} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} h_{pq}^{kt} \geq 0, \quad p \in \mathcal{P}, t \in \mathcal{T} \quad (29)$$

$$\sum_{q \in \mathcal{P}: \mathcal{B}_{pq}=1} \mathcal{W}_{pq} \cdot \sum_{k \in \mathcal{K}} l_{pq}^{kt} \geq 0, \quad p \in \mathcal{P}, t \in \mathcal{T} \quad (30)$$

$$h_{pq}^{kt} = x_q^k \cdot (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - (1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (\mathcal{M}_q^t - \mathcal{M}_p^t + s_q^k + m_q^t + m_p^t)), \quad (p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (31)$$

$$l_{pq}^{kt} = x_q^k \cdot (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - (1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (\mathcal{M}_p^t - \mathcal{M}_q^t + s_q^k + m_q^t + m_p^t)), \quad (p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (32)$$

We now linearize the definition of variables h_{pq}^{kt} in constraints 33 and 34 where H is a big number defined to relax constraint 33 when $x_q^k = 1$ and constraint 34 when $x_q^k = 0$. The same thing is performed on l_{pq}^{kt} variables to get constraints 35 and 36 where L is a big number defined to relax constraint

35 when $x_q^k = 1$ and constraint 36 when $x_q^k = 0$. At the end, the linear formulation of air pollution mapping is ensured by constraints 21, 23, 29, 30, 33, 34, 35 and 36.

$$\begin{aligned}
-H * x_q^k &\leq h_{pq}^{kt} \leq H * x_q^k, (p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (33) \\
-H * (1 - x_q^k) + (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - \\
(1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (M_q^t - M_p^t + s_q^k + m_q^t + m_p^t)) \\
&\leq h_{pq}^{kt} \leq \\
H * (1 - x_q^k) + (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - \\
(1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (M_q^t - M_p^t + s_q^k + m_q^t + m_p^t)), \\
(p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (34)
\end{aligned}$$

$$\begin{aligned}
-L * x_q^k &\leq l_{pq}^{kt} \leq L * x_q^k, (p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (35) \\
-L * (1 - x_q^k) + (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - \\
(1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (M_q^t - M_p^t + s_q^k + m_q^t + m_p^t)) \\
&\leq l_{pq}^{kt} \leq \\
L * (1 - x_q^k) + (e_p - \sum_{k' \in \mathcal{K}} x_p^{k'} \cdot s_p^{k'} - \\
(1 - \sum_{k' \in \mathcal{K}} x_p^{k'}) \cdot (M_q^t - M_p^t + s_q^k + m_q^t + m_p^t)), \\
(p, q) \in \mathcal{P} : \mathcal{B}_{pq} = 1, k \in \mathcal{K}, t \in \mathcal{T} \quad (36)
\end{aligned}$$

a) *Case of the mean error of pollution snapshots*:: As stated in the previous section, our coverage formulation can be adapted in order to consider the mean error of pollution snapshots rather than the maximum error. First, let e_p^t be a decision variable denoting the estimation error at point p corresponding to time snapshot $t \in \mathcal{T}$. Therefore, $e_p = \frac{1}{|\mathcal{T}|} \cdot \sum e_p^t$. Note that for each $t \in \mathcal{T}$, e_p^t is defined as in formulas 27 and 28. Hence, the linearized form of coverage formulation is obtained as in formulas 29, 30, 33, 34, 35 and 36.

3) *Network connectivity constraints*: We formulate the connectivity constraint as a network flow problem. We consider the same potential positions set \mathcal{P} for sensors and sinks. We first denote by $\Gamma^k(p)$ where $p \in \mathcal{P}$ and $k \in \mathcal{K}$, the set of neighbors of a sensor node of type k deployed at point p . This set can be determined using sophisticated propagation models. It can be also determined using the binary disc model, in which case $\Gamma^k(p) = \{q \in \mathcal{P} \text{ where } q \in \text{Disc}(p, \mathcal{R}^k)\}$ where \mathcal{R}^k is the communication range of of a sensor of type k . Let A_{pq}^k be equal to 1 if $q \in \Gamma^k(p)$ and equal to 0 otherwise. Now, we define the flow variables g_{pq} as the flow quantity transmitted from a node located at point p to another node located at point q . The idea is to suppose that each sensor of the resulting WSN generates a flow unit in the network, and then verify if these units can be recovered by sink nodes. The following constraints ensure that the deployed sensors and sinks form a connected wireless sensor network; i.e. each sensor can communicate with at least one sink.

$$g_{pq} \leq N * (y_p + \sum_{k \in \mathcal{K}} x_p^k * A_{pq}^k), (p, q) \in \widehat{\mathcal{P}} \quad (37)$$

$$g_{pq} \leq N * (y_q + \sum_{k \in \mathcal{K}} x_q^k * A_{qp}^k), (p, q) \in \widehat{\mathcal{P}} \quad (38)$$

$$\begin{aligned}
&\sum_{k \in \mathcal{K}} x_p^k - N * y_p \\
&\leq \sum_{q \in \widehat{\mathcal{P}}} g_{pq} - \sum_{q \in \widehat{\mathcal{P}}} g_{qp} \leq \\
&\sum_{k \in \mathcal{K}} x_p^k, p \in \widehat{\mathcal{P}} \quad (39)
\end{aligned}$$

$$\begin{aligned}
\sum_{p \in \widehat{\mathcal{P}}} \sum_{q \in \widehat{\mathcal{P}}} g_{pq} &= \sum_{p \in \widehat{\mathcal{P}}} \sum_{q \in \widehat{\mathcal{P}}} g_{qp} \quad (40) \\
\sum_{p \in \widehat{\mathcal{P}}} y_p &\leq I \quad (41)
\end{aligned}$$

Constraint 37 (respectively constraint 38) forces to 0 the flow from p to q if neither a sensor nor a sink is positioned at point p (respectively at point q) or a sensor of type k is positioned at point p (respectively at point q) but cannot communicate with its neighboring point q (respectively point p). The assumption of sensors that should generate, each of which, a flow unit is guaranteed thanks to constraint 39, which is relaxed if a sink is deployed at point p ; that is a sink can receive up to N flow units, which is the maximum number of sensors that we can have in the network. The overall flow is conservative thanks to constraint 40, this means that sinks recover all the flow units that are generated by sensor nodes. Finally, constraint 41 allows us to fix the maximum number of sinks of the resulting network that we denote I . This constraint can be relaxed by setting I to N , in which case the mixed integer programming model optimizes the number of sink nodes in addition to the optimization of their positions.

We provide in Fig. 1 an example of a pollution sensor network in order to show how the connectivity constraints are executed. Here we have 5 sensor nodes that have been already located using the coverage constraints and we already have a potential location of the sink node. Our objective is to verify if the network is connected or not using the flow concept.

- In Fig. 1, the dotted lines are not valid links because sensors which are involved in those lines are not close enough. Therefore, flow units are not sent over those links thanks to constraints 37 and 38.
- We force all the sensors to send each a flow unit as stated in the connectivity constraint 39. Note that some nodes send their flow unit directly to the sink node whereas the others pass through their neighbors. Also note that the flow is conservative on each node, for instance a node which receives a unit from a neighbor has to send to the sink two units: its unit plus the one of its neighbor.
- Finally, note that the overall flow is conservative thanks to constraint 40: the sink node receives all the 5 units generated by the 5 sensor nodes.

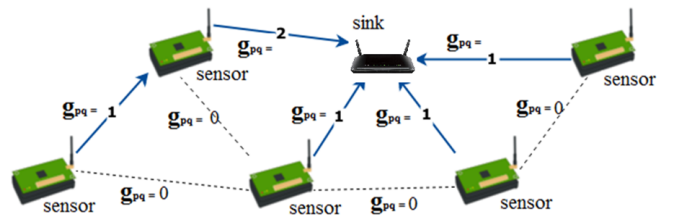


Fig. 1: Illustrative example of the flow concept.

4) *MILP model*: Finally, the mixed integer programming model that allows us to minimize the network deployment cost while ensuring a required coverage quality is denoted MIN_COST and can be written as follows:

[MIN_COST]

Objective: minimize \mathcal{F}

Pollution mapping constraints: (19), (21), (23), (cf)*, (29), (30), (33), (34), (35), (36)

Connectivity constraints: (37), (38), (39), (40), (41)

Decision variables: $x_p^k, y_p \in \{0, 1\}$

Auxiliary variables: $e_p, g_{pq} \in \mathbb{R}^+$; $h_{pq}^{kt}, l_{pq}^{kt} \in \mathbb{R}$

* : as explained in the beginning of section IV-A2, (cf) should be replaced by either (24) or (25)

B. MIN_ERROR: Estimation error minimization under deployment budget constraint

In the second MILP model, we propose to minimize the estimation error in the deployment region f defined in formulas 24 and 25. The deployment cost of the network should be constrained by the budget J as in constraint 42.

$$\mathcal{F} \leq J \quad (42)$$

Therefore, the second MILP model is denoted MIN_ERROR and can be written as follows:

[MIN_ERROR]

Objective: minimize f

Budget constraint: (42)

Pollution mapping constraints: (19), (21), (cf)*, (29), (30), (33), (34), (35), (36)

Connectivity constraints: (37), (38), (39), (40), (41)

Decision variables: $x_p^k, y_p \in \{0, 1\}$

Auxiliary variables: $e_p, g_{pq} \in \mathbb{R}^+$; $h_{pq}^{kt}, l_{pq}^{kt} \in \mathbb{R}$

* : as explained in the beginning of section IV-A2, (cf) should be replaced by either (24) or (25)

V. RESOLUTION OF THE OPTIMIZATION MODELS

A. Exact MILP solvers and theoretical complexity

The proposed optimization models are based on integer linear programming that can be solved using the exact MILP solvers. In this paper we use the cplex exact solver from IBM but only to solve small instances of our deployment problem. Indeed, the MILP models are proved in the literature to be NP-hard, i.e. the execution time of the MILP solvers increases exponentially with the size of the problem. In our models, the size of the problem depends on the number of points approximating the deployment region $|\mathcal{P}|$, the number of time snapshots $|\mathcal{T}|$ and the number of sensor types $|\mathcal{K}|$. In fact, what makes the MILP models NP-hard is the number of binary variables which causes an exponential increase in the number of iterations when using the exact MILP solvers. Note that the proposed models have the same size since MIN_ERROR is obtained by changing the objective function of MIN_COST

and constraining the network deployment cost \mathcal{F} instead of the estimation error f . In both of our models, the number of binary variables is equal to $|\mathcal{P}| \cdot |\mathcal{K}|$. That is, the complexity of the models is mainly due to the number of points and the number of sensor types.

B. Linear-relaxation based heuristic

In order to solve our optimization models on large instances in a reasonable time while getting near-optimal solutions, we propose to use the concept of linear relaxation to design a resolution algorithm for the model MIN_COST. As for solving the model MIN_ERROR, we propose to use the concept of binary search in addition to linear relaxation. Thanks to linear relaxation, the binary variables are eliminated and our heuristics run in polynomial time as we show later in this section.

1) *Solving MIN_COST:* We first define the linear programming model LP1 while considering the same objective function and constraints as MIN_COST and relaxing all the binary variables x_p^k and y_p ; i.e. binary variables are considered in the range of $[0, 1]$, this means that the solutions of the LP1 model are not necessarily binary. Note that in a given solution of LP1 where placement variables x_p^k and y_p are fractional, the variable having the maximum value (i.e. the closest binary variable to 1) corresponds to the most important node in the satisfaction of coverage and connectivity constraints. Based on this fact, we propose in each iteration of our heuristic algorithm presented in Algorithm 1 to set a sensor of type k_0 at point p_0 where $x_{p_0}^{k_0}$ is the closest variable to 1 or to set a sink at point p_0 if y_{p_0} is the closest variable to 1. The loop, which performs iterative rounding, stops once the placement variables are equal to either 0 or 1 and all the coverage and connectivity constraints are ensured.

Algorithm 1 Heuristic algorithm to solve MIN_COST

Inputs: \mathcal{P}, \mathcal{T}

Outputs: $\{x_p^k\}, \{y_p\}$

repeat

Solve the LP model

Let v be the maximum fractional variable among x_p^k and y variables

Add constraint $v = 1$ to the LP model

until all the variables are binary

The theoretical complexity of Algorithm 1 mainly depends on the number of iterations in the relaxation loop since solving the LP1 model by the exact solvers runs in polynomial time. Note that the number of iterations is at most equal to the number of points \mathcal{P} , which happens when a node has to be deployed at each point. As a result, Algorithm 1 runs in polynomial time.

2) *Solving MIN_ERROR:* We recall that MIN_ERROR is in fact the dual problem of MIN_COST where the objective function and the main coverage constraint have been exchanged. Let J be the deployment cost that we obtain by

solving MIN_COST when the tolerated estimation error is set to E . In this case, E can be considered as a solution to the model MIN_ERROR when the deployment budget is set to J . We base on this duality and use the principle of binary search in order to define Algorithm 2 that allows us to solve the problem MIN_ERROR. The idea is to explore the interval of the possible values of the tolerated estimation error E which are in the range $[U, V]$. U and V define the bounds on the estimation error that we can get with the budget J . Note that the estimation error never exceeds $\max_{p \in \mathcal{P}, t \in \mathcal{T}} \mathcal{M}_p^t + 2 * m_p^t$ according to our definition given in section III. The interval $[U, V]$ is tightened in every iteration and the main loop stops once the length of the interval is sufficiently small compared to a given threshold TH .

Algorithm 2 Heuristic algorithm to solve MIN_ERROR

Inputs: \mathcal{P}, \mathcal{T}

Outputs: $\{x_p^k\}, \{y_p\}$

$U \leftarrow 0$

$V \leftarrow \max_{p \in \mathcal{P}, t \in \mathcal{T}} \mathcal{M}_p^t + 2 * m_p^t$

repeat

$H \leftarrow \frac{U+V}{2}$

if $OBJ_MIN_COST(H) \leq J$ **then**

$V \leftarrow H$

else

$U \leftarrow H$

until $V - U \leq TH$

return $SOLUTION_MIN_COST(V)$

The number of iterations is a key factor in the complexity of Algorithm 2 and is equal to $\gamma = \lceil \log_2(V_0/TH) \rceil$ where $V_0 = \max_{p \in \mathcal{P}, t \in \mathcal{T}} \mathcal{M}_p^t + 2 * m_p^t$. The value of the constant γ mainly depends on the nature of pollutants that defines the value of V_0 . The theoretical complexity of Algorithm 2 is equal to γ multiplied by the complexity of solving MIN_COST that occurs in each iteration. As a result, Algorithm 2 runs in polynomial time when using our relaxation based algorithm to solve MIN_COST in the main resolution loop.

VI. SIMULATION RESULTS

A. Dataset

In order to consider the real dispersion of air pollutants in the simulated pollution concentrations \mathcal{M}_p^t , we perform our simulations on a set of 12 monthly and one annual pollution snapshots. The dataset corresponds to the 2008 Nitrogen Dioxide (NO_2) concentrations in the Lyon district of La-Part-Dieu, which is the heart of the Lyon City. Pollution snapshots are generated by an enhanced atmospheric dispersion simulator called SIRANE [33], which is designed for urban areas and takes into account the impact of street canyons on pollution dispersion. The dataset has been provided by LMFA, which is a research lab specialized in fluid mechanics in the Lyon city, France.

We depict in Fig. 2b the pollution map that corresponds to the annual mean of 2008. The deployment region has a spatial resolution of 50 meters and is depicted in Fig. 2a. We consider

as potential positions of nodes all the grid points. In order to show how we take into account simulation errors, we consider a map of randomly generated errors, which is depicted in Fig. 2c.

We used as MILP exact solver the IBM ILOG CPLEX solver. Default simulation parameters are summarized in TABLE II. When it is not precised in simulation scenarios, we suppose that sensing is perfect. In addition, we fix the maximum number of sinks to 1 in order to get mono-sink networks.

Parameter	Notation	Value
Number of discrete points	\mathcal{N}	225
Maximum correlation distance	d	100m
Attenuation coefficient of correlation	α	1
Communication range of sensor nodes	\mathcal{R}	150m
The tolerated estimation error	\mathcal{E}	$5\mu g/m^3$
The maximum number of sinks	I	1
The cost of deploying a sensor at point p	δ_p	1
The cost of deploying a sink at point p	ψ_p	10
Sensing error at point p	s_p	0
Threshold of wind direction	$\mathcal{T}\mathcal{H}_\beta$	5 degrees

TABLE II: Default values of main simulation parameters.

B. Proof-of-concept

In order to validate our formulation of pollution-aware coverage quality, we run the model MIN_ERROR to minimize the maximum estimation error while considering 3 values for the tolerated estimation error: 5, 8 and 10 $\mu g/m^3$. We depict in Fig. 3 the obtained positions of sensors and sinks for the three simulation cases. We also evaluate at each point of the map the estimated concentration and then we calculate the resulting estimation error. The obtained errors are also depicted in Fig. 3. We notice that less sensors are used when the tolerated estimation error increases. This is expected since better deployment precision needs more sensor nodes. In addition, Fig. 3 shows that the maximum value in the error map in each simulation case is bounded by the tolerated estimation error, which fits our coverage formulation. Moreover, the obtained nodes form a connected network as formulated in our connectivity constraint.

C. Evaluation of the proposed heuristics

We evaluate the experimental complexity of our optimization models and the performance of our heuristic algorithms. We assume in this simulation case that simulation models of pollution concentrations are perfect and we focus on solving the model MIN_COST while considering the minimization of the deployment cost of a connected network without exceeding a tolerated estimation error equal to $5\mu g/m^3$. TABLE III presents the results that we obtained while considering the Lyon district of La-part-dieu with different monthly snapshots of pollution concentrations. Results in TABLE III are sorted according to the execution time of the MILP solver. Note that this sort fits well with pollution variability (the range length of pollution concentrations corresponding to each snapshot).

TABLE III shows that the less the pollution variability in the deployment region, the higher is the execution time of the

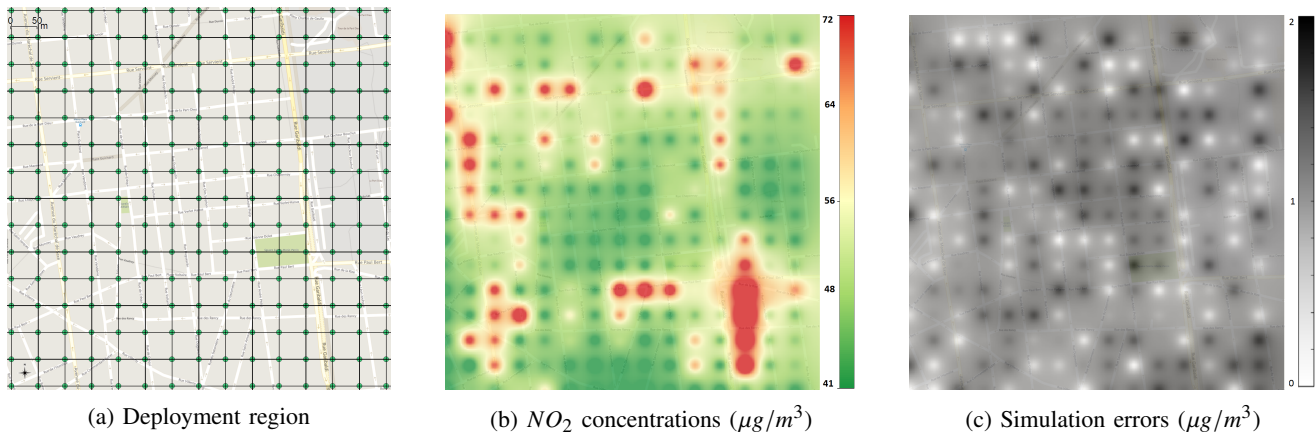


Fig. 2: Deployment region, simulation of 2008 annual concentrations of NO_2 and simulation errors corresponding to the district of La-part-dieu, Lyon, France.

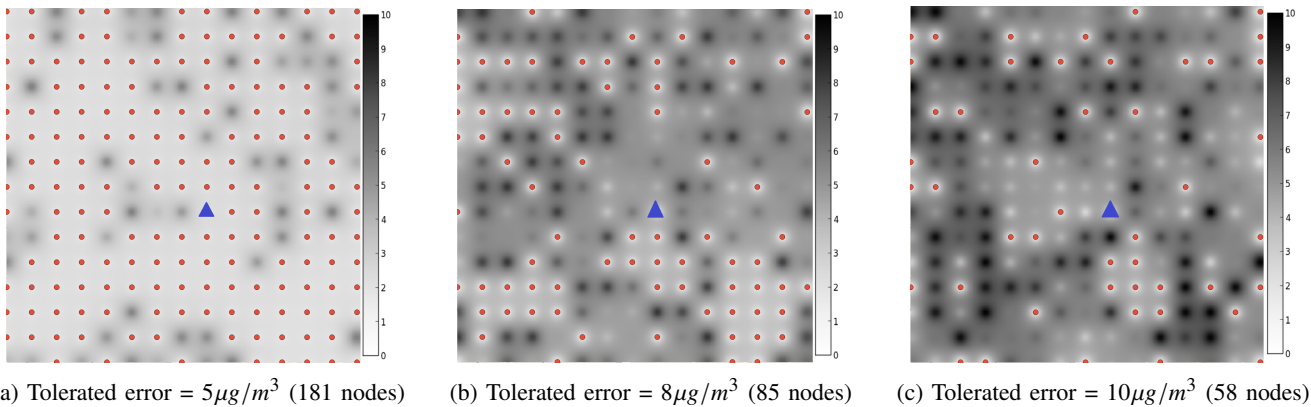


Fig. 3: Proof-of-concept: Optimal WSN topology and the corresponding estimation errors ($\mu g/m^3$) while considering different values of the maximum tolerated error. Sensors (respectively sinks) are depicted in red circles (blue triangles).

exact MILP solver. For example, the MILP solver takes more than one hour to solve the problem instance corresponding to January. However, the execution time of our heuristic algorithm is stable regarding the pollution variability and is less than one minute in most of the simulations. As a result, the time gain factor of our heuristic increases with the complexity of the problem and is higher than 10 for most of the instances.

As for the value of the objective function, we notice that the optimal deployment cost increases with pollution variability; i.e. more sensors are needed when pollution variability is higher. In order to compare the optimal solutions to the approximated ones given by our heuristic, we define the percentage drift which is equal to $\frac{o_2 - o_1}{o_1}$ where o_1 the optimal objective value and o_2 is the approximated value. We notice in TABLE III that the percentage drift ranges between 23% and 34% with a mean equal to 30%.

D. Evaluation of the coverage results

In this section, we evaluate the optimal coverage results of our deployment models. First, we depict in Fig. 4a the deployment cost depending on the estimation tolerated error while considering two error functions: the maximum error and the mean error. Results show that the less the tolerated error,

the higher is the deployment cost, which is reasonable since reducing the estimation error requires more information and hence more sensors. Also, we notice that when considering the same tolerated value on the x-axis, we deploy less sensors for the mean error than the maximum error. Indeed, the maximum error constraint is stronger than the constraint of the mean error since outliers do not appear in the mean function.

We also evaluate the impact of the sensing (s_p) and simulation (m_p) errors on the final estimation error while considering 4 scenarios, in each of which we vary the values of sensing and simulation errors. We depict the obtained results in Fig. 4b. Results show that the estimation error is shifted by $2 * m + s$, which is explained by our definition of the estimation error in formula 17 where m occurs two times and s occurs only one time.

Finally, we evaluate the impact of sensing heterogeneity while considering two different types of sensors: low accuracy and high accuracy ones. We vary the cost ratio of the two sensor types and consider two cases in the quality difference between the two types. Results are depicted in Fig. 4c and show that the more the accurate sensors are expensive, the higher is the deployment cost. Indeed, accurate sensors are usually used more than low accuracy sensors. However, when the cost of the accurate sensors reaches the double of the low

Time snapshot (month)	Pollution variability ($\mu\text{g}/\text{m}^3$)	Execution time (seconds)			Objective function (deployment cost)		
		MILP solver	Heuristic	Gain factor	MILP solver	Heuristic	Drift (%)
September	34.877	183.914	40.850	4.502	95.000	117.000	23.2
April	34.630	188.911	42.459	4.449	92.000	121.000	31.5
March	31.827	471.260	42.534	11.080	84.000	113.000	34.5
November	29.483	822.781	47.396	17.360	76.000	102.000	34.2
February	28.024	1495.090	60.790	24.594	72.000	95.000	31.9
January	27.717	4596.050	47.295	97.178	75.000	100.000	33.3
Mean	32.162	1112.852	46.351	22.843	86.143	111.429	30.0

TABLE III: MILP solver VS Linear relaxation-based heuristic.

accuracy ones, only these latter are used because their number can compensate their quality.

E. Evaluation of the connectivity results

In the final simulation scenario, we evaluate the connectivity results while considering two different types of nodes depending on their communication capabilities: nodes with a communication range equal to 150m that we consider as short range communication nodes (like 802.15.4 for instance); and nodes with a communication range equal to 500m that we consider as long-range communication nodes (like Lora Sigfox for instance). We vary the cost ratio between the two node types and depict the optimal network deployment cost in Fig. 5. Results show that only short-range communications are used when long range communications are very expensive; i.e. starting from a ratio equal to the double on the x-axis. Indeed, using several short range communication nodes can replace long range nodes when these latter are much expensive than short range nodes. In addition, when the tolerated estimation error is lower, the curve converges faster as shown in Fig. 5. The main reason behind this is that the network is denser when the tolerated error is lower.

VII. DISCUSSIONS

We discuss in this section the main characteristics of our deployment solutions regarding the application case of air pollution. We also provide some details to show how our models can be extended in order to take into account the network lifetime and nodes' mobility.

A. Pollution-aware WSN deployment

Our deployment approach targets specifically the application of air pollution monitoring which is characterized by two main features compared to many other WSN applications: the availability of sophisticated air pollution simulation models and the issue of pollution measurements' reliability. In this work, we leverage the use of state-of-the-art simulation models in order to characterize the ground truth concentrations with bounded errors and then define an appropriate coverage formulation that is adapted to air pollution monitoring. In order to deal with the uncertainty in pollution sensors' measurements, our coverage formulation takes into account the impact of weather conditions on sensing errors to filter out bad measurements in the interpolation process.

B. Formulation of network connectivity

The network connectivity is formulated in our deployment models using the flow concept based on prior knowledge regarding communication links' quality. Indeed, only good quality links are used by sensors to communicate their data to sink nodes. Although we consider in this paper mainly binary communication links, the extension of our work to take into account probabilistic communication links is one of our main perspectives.

C. Optimization of the network lifetime

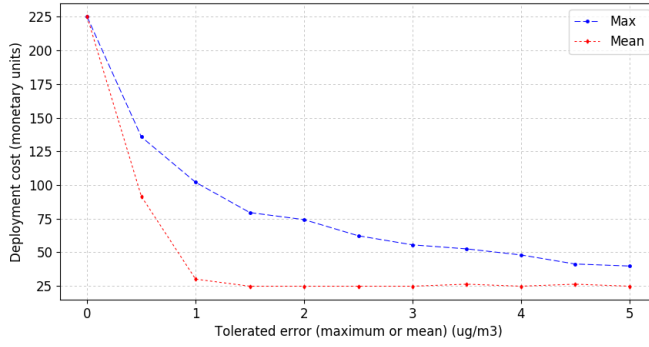
We believe that our optimization models and heuristic algorithms can be extended in order to take into account the network lifetime. One possible extension is to do as in the work of [18] where authors define the energy constraints based on the flow concept. The idea is to split the lifetime of the network into a sequence of timeframes (a sequence of minutes for instance) and enforce each sensor node to send a flow unit in each timeframe. Then, a new constraint should be added to the model in order to ensure that for each sensor, the sum of the energy that is consumed in the set of timeframes is less than the energy of the battery of the sensor. The lifetime objective function corresponds to the maximum number of timeframes where the network is operating. This solution is viable for our proposal since our models are also based on the flow concept.

D. Mobile deployment

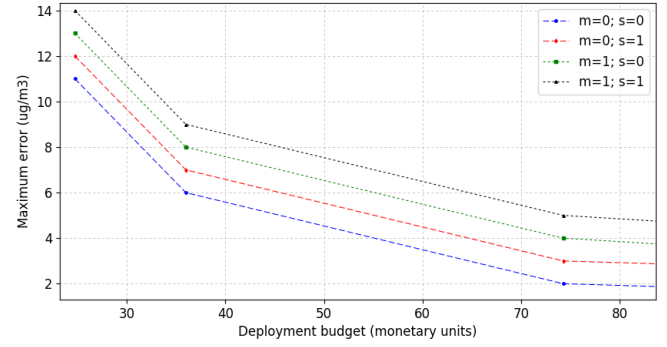
Our models are designed for static networks because pollution sensors operate well when they are static [36]. However, sinks can be considered as mobile nodes and this can be integrated in our model based on existing mobile-sinks formulations such as the work of [18]. The idea is to consider a set of timeframes as in the model extension of network lifetime. The mobile sink changes its positions in each timeframe. Then, the flow constraints should be formulated in order to ensure that the flow is conservative in the network in each timeframe. This means that in each timeframe, each sensor generates a flow unit and the sink node receives all the generated units.

VIII. CONCLUSION

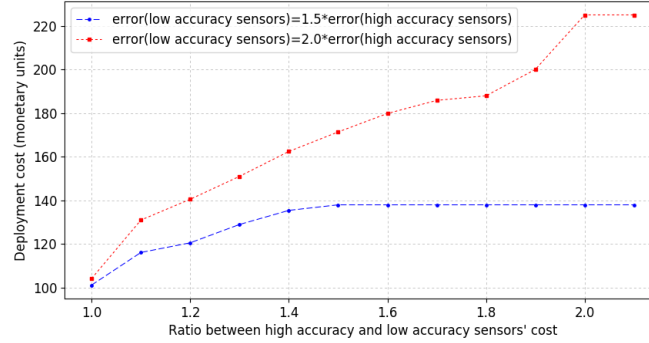
In this paper, we tackle the deployment issue of heterogeneous sensor networks and propose mixed integer programming models and heuristic algorithms taking into account the network deployment cost and the air pollution mapping quality while ensuring the network connectivity. Our main



(a) Deployment cost vs. tolerated error.



(b) Impact of sensing and simulation errors.



(c) Impact of sensing heterogeneity.

Fig. 4: Coverage results.

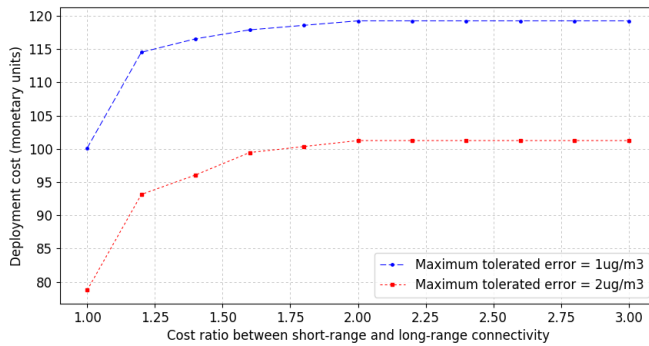


Fig. 5: Impact of communication heterogeneity.

contribution is to define an appropriate coverage formulation for air pollution regular mapping and then derive optimal deployment models and approximate resolution algorithms. We applied our models and algorithms on a dataset of the Lyon City, France and evaluated the computational complexity of our proposal. In addition, we assessed the impact of the deployment requirements on the coverage and connectivity results. We have shown that a heterogeneous topology in both sensing and communication may be interesting or not depending on the cost of the accurate nodes.

We provided a discussion regarding the extension of our models and how mobile deployment and network lifetime can be taken into account in our deployment approaches. We also plan to consider the impact of the different urban parameters, such as the structure of streets, on the deployment results.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their comments and feedback. This work has been supported by the "LABEX IMU" (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

REFERENCES

- [1] A. Boubrima, W. Bechkit, and H. Rivano, "Error-bounded air quality mapping using wireless sensor networks," in *Local Computer Networks (LCN), 2016 IEEE 41st Conference on*. IEEE, 2016, pp. 380–388.
- [2] Z. Li, N. Wang, A. Franzen, P. Taher, C. Godsey, H. Zhang, and X. Li, "Practical deployment of an in-field soil property wireless sensor network," *Computer Standards & Interfaces*, vol. 36, no. 2, pp. 278–287, 2014.
- [3] Air Rhône-Alpes, "The air quality monitoring organization of the lyon agglomeration," <http://www.air-rhonealpes.fr> [2017-07-27].
- [4] A. Kumar, H. Kim, and G. P. Hancke, "Environmental monitoring systems: a review," *Sensors Journal, IEEE*, vol. 13, no. 4, pp. 1329–1339, 2013.
- [5] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
- [6] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks *et al.*, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," *Atmospheric Environment*, vol. 70, pp. 186–203, 2013.
- [7] ETSI, "Electromagnetic compatibility and radio spectrum matters (erm); system reference document (srdoc): Spectrum requirements for short range device, metropolitan mesh machine networks (m3n) and smart metering (sm) applications," available on http://www.etsi.org/deliver/etsi_tr/103000_103099/103055/01.01.01_60/tr_103055v010101p.pdf, last access: 2017-07-01.

- [8] A. Boubrima, F. Matigot, W. Bechkit, H. Rivano, and A. Ruas, "Optimal deployment of wireless sensor networks for air pollution monitoring," in *Computer Communication and Networks (ICCCN), 2015 24th International Conference on*. IEEE, 2015, pp. 1–7.
- [9] A. Boubrima, W. Bechkit, and H. Rivano, "Optimal wsn deployment models for air pollution monitoring," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2723–2735, 2017.
- [10] C. Zhu, C. Zheng, L. Shu, and G. Han, "A survey on coverage and connectivity issues in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 619–632, 2012.
- [11] K. Chakrabarty, S. S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," *Computers, IEEE Transactions on*, vol. 51, no. 12, pp. 1448–1453, 2002.
- [12] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, and L. Thiele, "Pushing the spatio-temporal resolution limit of urban air pollution maps," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014, pp. 69–77.
- [13] A. Marjovi, A. Arfire, and A. Martinoli, "High resolution air pollution maps in urban environments using mobile sensor networks," in *Distributed Computing in Sensor Systems (DCOSS), 2015 International Conference on*. IEEE, 2015, pp. 11–20.
- [14] S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, "Real-time air quality monitoring through mobile sensing in metropolitan areas," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM, 2013, p. 15.
- [15] B. Liu, O. Dousse, P. Nain, and D. Towsley, "Dynamic coverage of mobile sensor networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 2, pp. 301–311, 2013.
- [16] İ. K. Altınel, N. Aras, E. Güney, and C. Ersoy, "Binary integer programming formulation and heuristics for differentiated coverage in heterogeneous sensor networks," *Computer Networks*, vol. 52, no. 12, pp. 2419–2431, 2008.
- [17] X. Chang, R. Tan, G. Xing, Z. Yuan, C. Lu, Y. Chen, and Y. Yang, "Sensor placement algorithms for fusion-based surveillance networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1407–1414, 2011.
- [18] M. E. Keskin, İ. K. Altınel, N. Aras, and C. Ersoy, "Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility," *Ad Hoc Networks*, vol. 17, pp. 18–36, 2014.
- [19] M. Rebai, H. Snoussi, F. Hnaïen, L. Khoukhi *et al.*, "Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks," *Computers & Operations Research*, vol. 59, pp. 11–21, 2015.
- [20] S. Sengupta, S. Das, M. Nasir, and B. K. Panigrahi, "Multi-objective node deployment in wsn: In search of an optimal trade-off among coverage, lifetime, energy consumption, and connectivity," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 405–416, 2013.
- [21] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Transactions on signal processing*, vol. 62, no. 5, pp. 1135–1146, 2014.
- [22] V. Roy, A. Simonetto, and G. Leus, "Spatio-temporal sensor management for environmental field estimation," *Signal Processing*, vol. 128, pp. 369–381, 2016.
- [23] P. G. Liaskovitis and C. Schurgers, "Leveraging redundancy in sampling-interpolation applications for sensor networks: A spectral approach," *ACM Transactions on Sensor Networks (TOSN)*, vol. 7, no. 2, p. 12, 2010.
- [24] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, no. Feb, pp. 235–284, 2008.
- [25] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Robust sensor placements at informative and communication-efficient locations," *ACM Transactions on Sensor Networks (TOSN)*, vol. 7, no. 4, p. 31, 2011.
- [26] W. Du, Z. Xing, M. Li, B. He, L. H. C. Chua, and H. Miao, "Sensor placement and measurement of wind for water quality studies in urban reservoirs," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 3, p. 41, 2015.
- [27] X. Wang, X. Wang, G. Xing, J. Chen, C.-X. Lin, and Y. Chen, "Intelligent sensor placement for hot server detection in data centers," *IEEE Transactions on parallel and distributed systems*, vol. 24, no. 8, pp. 1577–1588, 2013.
- [28] V. Gallart, S. Felici-Castell, M. Delamo, A. Foster, and J. J. Perez, "Evaluation of a real, low cost, urban wsn deployment for accurate environmental monitoring," in *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*. IEEE, 2011, pp. 634–639.
- [29] D. W. Wong, L. Yuan, and S. A. Perlin, "Comparison of spatial interpolation methods for the estimation of air quality data," *Journal of Exposure Science and Environmental Epidemiology*, vol. 14, no. 5, pp. 404–415, 2004.
- [30] L. Contreras and C. Ferri, "Wind-sensitive interpolation of urban air pollution forecasts," *Procedia Computer Science*, vol. 80, pp. 313–323, 2016.
- [31] A. Tilloy, V. Mallet, D. Poulet, C. Pesin, and F. Brocheton, "Blue-based no2 data assimilation at urban scale," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 4, pp. 2031–2040, 2013.
- [32] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons, 2009, vol. 501.
- [33] L. Souhac, P. Salizzoni, P. Mejean, D. Didier, and I. Rios, "The model sirane for atmospheric urban pollutant dispersion; part II, validation of the model on a real case study," *Atmospheric Environment*, vol. 49, pp. 320–337, 2012.
- [34] O. A. Popoola, G. B. Stewart, M. I. Mead, and R. L. Jones, "Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability," *Atmospheric Environment*, vol. 147, pp. 330–343, 2016.
- [35] L. Spinelle, M. Gerboles, M. G. Villani, M. Alexandre, and F. Bonavitaola, "Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide," *Sensors and Actuators B: Chemical*, vol. 215, pp. 249–257, 2015.
- [36] A. Velasco, R. Ferrero, F. Gandino, B. Montrucchio, and M. Rebaudengo, "A mobile and low-cost system for environmental monitoring: A case study," *Sensors*, vol. 16, no. 5, p. 710, 2016.



Ahmed Boubrima is a PhD student at INRIA, France and a member of the Agora team within the CITI Lab of INSA-Lyon. He received his masters and engineering degrees in Computer Science with highest distinction from the "Ecole nationale Supérieure d'Informatique", Algiers in 2015. His PhD thesis is part of the French interdisciplinary project "UrPolSens", which focuses on the use of wireless sensor networks for air pollution monitoring. His research interests include mathematical programming and approximation algorithms applied

to network design issues.



Walid Bechkit is an associate professor within the Telecommunications department of INSA-Lyon, France. He is also a member of the INRIA Agora team of the CITI laboratory. He obtained his Ph.D. in System and Information Technology from the Compiegne University of Technology (U.T.C.), France in 2012 and his engineering degree in Computer Science from the "Ecole nationale Supérieure d'Informatique", Algiers in 2009. His main research interests include deployment, reliability, security and energy saving issues in wireless sensor networks.



Hervé Rivano is a full professor at INSA-Lyon, France and the head of the INRIA Agora team, which focuses on the use of wireless networks in smart cities. Prior to that, he had been an INRIA and CNRS researcher from October 2004 to August 2017. He obtained his PhD in November 2003 from the University of Nice-Sophia Antipolis after graduated from the "Ecole Normale Supérieure de Lyon", France. His research interests include combinatorial optimization and approximation algorithms applied to network design and provisioning.