Optimal WSN Deployment Models for Air Pollution Monitoring

Ahmed Boubrima, Walid Bechkit, Hervé Rivano

HAL Id: hal-01473393

https://hal.inria.fr/hal-01473393

Submitted on 19 May 2017

Optimal WSN Deployment Models for Air Pollution Monitoring

Ahmed Boubrima, Walid Bechkit and Hervé Rivano

Abstract—Air pollution has become a major issue of modern megalopolis because of industrial emissions and increasing urbanization along with traffic jams and heating/cooling of buildings. Monitoring urban air quality is therefore required by municipalities and by the civil society. Current monitoring systems rely on reference sensing stations that are precise but massive, costly and therefore seldom. In this paper, we focus on an alternative or complementary approach, with a network of low cost and autonomic wireless sensors, aiming at a finer spatiotemporal granularity of sensing. Generic deployment models of the literature are not adapted to the stochastic nature of pollution sensing. Our main contribution is to design integer linear programming models that compute sensor deployments capturing both the coverage of pollution under time-varying weather conditions and the connectivity of the infrastructure. We evaluate our deployment models on a real data set of Greater London. We analyze the performance of the proposed models and show that our joint coverage and connectivity formulation is tight and compact, with a reasonable enough execution time. We also conduct extensive simulations to derive engineering insights for effective deployments of air pollution sensors in an urban environment.

Index Terms—Air pollution monitoring, Detection of threshold crossings, Atmospheric dispersion modeling, Wireless sensor networks (WSN), Deployment, Coverage, Connectivity.

I. INTRODUCTION

AIR pollution affects human health dramatically. According to World Health Organization (WHO), exposure to air pollution is accountable to seven million casualties in 2012 [1]. In 2013, the International Agency for Research on Cancer (IARC) classified particulate matter, the main component of outdoor pollution, as carcinogenic for humans [2]. Air pollution has become a major issue of modern megalopolis, where the majority of world population lives, adding industrial emissions to the consequences of an ever denser urbanization with traffic jams and heating/cooling of buildings. As a consequence, the reduction of pollutant emissions is at the heart of many sustainable development efforts, in particular those of smart cities. Monitoring urban air pollution and detecting pollution peaks is therefore required by both municipalities and the civil society, wanting to design and assess, or ask for, pollution mitigation public policies.

Most of actual air quality monitoring is operated by independent authorities [3]. Conventional measuring stations are equipped with multiple lab quality sensors. These systems are however massive, inflexible and expensive. An alternative – or complementary – solution would be to use wireless sensor networks (WSN) which consist of a set of lower cost nodes that

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.

can measure information from the environment, process and relay them to some base stations, denoted sinks [4]. The main advantages of a WSN infrastructure, namely self-organization and healing as well as energetic autonomy of the nodes, for air pollution monitoring is to obtain a finer spatiotemporal granularity of measurements, thanks to the resulting lighter installation and operational costs [5]. Pollution monitoring may target two objectives: i) the periodic air quality sampling and mapping; and ii) the detection of threshold crossings in order to trigger adequate alerts [6]. In this paper, we focus on the second application where sensors are deployed to control concentrations of pollutants released by pollution sources like factories, sewage treatment plants and urban traffic [7]. We investigate on the computation of minimum-cost optimal deployments that ensure both pollution coverage and network connectivity while considering the phenomenon dispersion.

Minimizing the deployment cost 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 minimizing the deployment cost [8]. The deployment is constrained by the cost of the nodes and sinks, but also by operational costs such as the energy spent by the nodes [9][10][11]. The network is said connected if each sensor can communicate information to at least one sink [12]. Many papers on the deployment issue have assumed that two nodes are able to communicate with each other if the distance between them is less than a radius called the communication range [13]. The coverage issue has often been modeled as a k-coverage problem in which at least k sensors should monitor each point of interest [14]. The majority of deployment approaches 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 [14], [15]. This can be true for some applications like presence sensors but is not suitable for pollution monitoring. Indeed, a pollution sensor detects pollutants that are brought in contact by the wind. The notion of detection range is therefore irrelevant in this context. In order to define a realistic formulation of pollution coverage, we consider pollution propagation models that take into account the inherently stochastic weather conditions.

In this paper, we propose two optimization models for the deployment of WSN for air pollution monitoring based on a preliminary work and few discussions that were presented in [16]. The expected deployment should ensure pollution coverage and network connectivity while minimizing the deployment cost. Based on the pollution dispersion modeling and the related work on ILP formulations of WSN coverage and connectivity, we first propose an optimization model where

pollution coverage and network connectivity are modeled independently. Then, we propose an enhanced and more effective model in which we give a joint formulation of coverage and connectivity using only the flow concept. The enhanced model is compact and tighter than the first one. The computation time is therefore reduced so that scenarios at the scale of large cities are tractable. The two proposed models take into account the probabilistic sensing of pollution sensors and are designed to handle multiple scenarios of weather conditions.

The remainder of this paper is organized as follows. We first review some related works on ILP formulations of WSN coverage and connectivity in section II. Then, we present and analyze the atmospheric dispersion modeling of pollutants in section III. Section IV details our two proposed optimization models while section V presents the simulation dataset and the obtained results. Finally, we conclude and propose some perspectives in section VI.

II. RELATED WORK

Progress in Operational Research has made solvers of Integer Linear Programs so efficient that mathematical modeling is a widely used approach for network optimization despite the theoretical complexity of the underlying combinatorial problems [17]. Several integer linear programming formulations have been proposed in the literature to model wireless sensor networks deployment [9][10][11]. The problem consists in determining the optimal node positions while ensuring the coverage of the deployment field and the connectivity of the network [18]. The objective may be to minimize the deployment financial cost or to maximize the lifetime of the network. Coverage and connectivity constraints have been sometimes modeled separately. In some related works [8], authors assume that coverage involves connectivity by considering that the communication range is a multiple of the detection range. In other works, authors assume that coverage is already ensured with some existing sensors, and thus formulate only the connectivity constraint [13].

The majority of the existing optimization strategies formulate the coverage of points of interest based on the distance between sensor locations and the coordinates of points [8]. This cannot be applied to the air quality monitoring where electrochemical sensors are usually used. In this kind of sensors, the pollutant must touch the sensor in order to be detected. Therefore, works that consider a detection range around sensors cannot be used in our application.

Chakrabarty et al. [15] were the first to give an ILP formulation to the deployment problem of WSN. They represent the deployment field as a two or three dimensional grid of points. They first propose a nonlinear formulation for minimizing the cost of sensor deployment while ensuring complete coverage of the sensor field. 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 quality monitoring since a sensor

positioned at a point A cannot cover a neighboring point B if there is a difference between pollution concentrations at the two points. Altinel et al. [19] 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.

Works that are more recent have targeted the connectivity and multi-objective deployment issues. The authors of [9] 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 [11] 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 [10], 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.

Most of the previously mentioned models trying to be generic, they do not take into account the specific characteristics of the environment in which the sensors are deployed and the nature of the phenomenon they cover. It is often assumed that each node can cover interest points within a given detection range. This hypothesis does not hold in the setting of air pollution monitoring because pollutants cannot be detected if they do not touch sensors. Moreover, the coverage and connectivity constraints are modeled independently in the sense that coverage is formulated by analogy to the Set Covering Problem and connectivity formulation is based on the network flow problem.

In section IV, we address these issues while proposing two ILP formulations based on real pollutants dispersion modeling. While coverage and connectivity are independent in the first formulation, they are jointly formulated as a network flow problem in the second one.

III. ATMOSPHERIC DISPERSION MODELING

The theoretical study of pollution atmospheric dispersion is mainly based on fluid mechanics theory [20]. For the sake of clarity, we focus in this work on steady state dispersion, in particular on Gaussian dispersion. However, our approach can be used with any other dispersion model that is able to take into account pollution sources like crossroads and highways or the effect of buildings on the dispersion of pollutants.

The basic Gaussian model estimates the concentrations of a pollutant gas released by a pointwise pollution source in a free space environment [21]. The estimated value C (g/m^3) at a measurement location (x, y, z) is given by Formula (1). Table I details the parameters of the model. The pollution source

is located at the point $(0, 0, h_s)$ and the measurement point location is given according to a 3D coordinate system where the x-axis is oriented in the wind direction D_w. Parameters σ_y and σ_z describe the stability of the atmosphere and can be approximated using Briggs formulas: $\sigma_y = a_y.|x|^{b_y}$ and $\sigma_z = a_z.|x|^{b_z}$. The parameter H, which represents the pollutant effective release height, is equal to the sum of the pollutant source height h_s and the plume rise Δh. The pollution plume is located above the pollution source and Δh is the vertical distance between the source and the center of the pollution plume. Briggs formulas are commonly used for the calculation of the Δh parameter. To simplify the analysis, we only consider the case where the temperature of the pollutant T_s is greater than the ambient air temperature T, which is usually the case. The value of Δh is given by Formula (2) where F, the pollutant gas buoyancy, is computed using Formula (3).

$$C(x, y, z) = \frac{Q}{2\pi V_w \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y}} (e^{-\frac{(z-H)^2}{2\sigma_y}} + e^{-\frac{(z+H)^2}{2\sigma_y}}) \quad (1)$$

$$\Delta h = \frac{1,6 \cdot F^{1/3} \cdot x^{2/3}}{V_w} \quad (2)$$

$$F = \frac{g}{\pi} \cdot V \cdot (\frac{T_s - T}{T_s}) \quad (3)$$

Measurement location	
x	Downwind distance from the pollution source (m)
y	Crosswind distance from the pollution source (m)
z	Hight (m)
Pollution emission parameters	
h_s	Pollutant source height (m)
Q	Mass flow rate at the emission point (g/s)
V	Volumetric flow rate at the emission point (m³/s)
T_s	Pollutant temperature at the emission point (K)
Weather	
T	Ambient air temperature (K)
V_w	Wind velocity (m/s)
D_w	Wind direction (degree)
Constants	
a_y, b_y	Horizontal dispersion coefficients
a_z, b_z	Vertical dispersion coefficients
g	Gravity constant (9.8m/s²)

TABLE I: Parameters of the Gaussian dispersion model

The Gaussian model considers only one scenario of weather conditions at a time to compute pollution concentrations. Simulations of the Gaussian model on three scenarios of weather conditions are depicted in Fig. 1. Pollution concentrations are given in $\mu g/m^3$ and grouped in 5 intervals. Common values used for simulation are depicted in table II. By comparing Fig. 1(a) to Fig. 1(b), we notice that wind direction affects the direction of the pollution plume since pollutants are transported by wind. Moreover, variations of the ambient temperature and the wind velocity between Fig. 1(b) and Fig. 1(c) have affected the concentrations of the pollution plume.

IV. DEPLOYMENT MODELS

A. Basic Model

In smart cities applications, some restrictions on node positions may apply because of authorization or practical

Parameter	value
h_s	$25m$
Q	$5g/s$
V	$1.9mm^3/s$
T_s	30^oC
a_y	1.36
b_y	0.82
a_z	0.275
b_z	0.69

TABLE II: simulation params of the Gaussian model

issues. For instance, in order to alleviate the energy constraints, we may place sensors on lampposts and traffic lights as experimented in CitySense [22]. In the following, we consider a set of a pre-defined potential positions, denoted $\mathcal{P}$, which is obtained using a discretization of the deployment field restricted to allowed positions. In free space environments without deployment restrictions, that would be a regular grid. We denote $\mathcal{N} = |\mathcal{P}|$ the number of potential positions. The locations of pollution sources, e.g. factories, sewage treatment plants, crossroads, highways..., is denoted $\mathcal{I}$. $\mathcal{M}$ denotes the number of pollution sources. The binary decision variables x_p, resp. y_p, define if a sensor, resp. a sink, is placed at position p.

We consider that sinks are equipped with pollution sensors. They are also connected to a backbone network. Deploying a sink is therefore more expensive than a regular sensor node. The cost of deploying a sensor, resp. a sink, at position p is denoted c_p^{sensor}, resp. c_p^{sink}. Since a sink embeds sensing capabilities, a sink and a sensor cannot be deployed at the same potential position p as formulated in constraint 4. Our optimization models minimize the sensors and sinks overall deployment cost. Thus, we get the objective function in formula 5.

$$x_p + y_p \leq 1, \quad p \in \mathcal{P} \quad (4)$$

$$\mathcal{F} = \sum_{p \in \mathcal{P}} c_p^{sensor} \cdot x_p + \sum_{p \in \mathcal{P}} c_p^{sink} \cdot y_p \quad (5)$$

Before tackling the formulation of the coverage and connectivity constraints, we summarize in table III the notations used in the ILP models.

1) Probabilistic pollution coverage: The coverage constraints rely on the modeling of the atmospheric dispersion. We assume that pollution sources release pollutants independently and may have simultaneous release. Our formulation ensures the coverage of threshold crossings in all cases. As shown in section III, pollution concentrations vary depending on weather conditions. Hence, we consider a set of possible weather scenarios $\mathcal{S}$ that can be obtained based on statistical data or weather forecast. A scenario corresponds to a tuple of ambient temperature T, wind velocity V_w and wind direction D_w: $s = (T^s, V_w^s, D_w^s)$. Each scenario s has probability α_s to happen. We assume that $\mathcal{S}$ is a partition of the space of weather conditions, i.e. $\sum_{s \in \mathcal{S}} \alpha_s = 1$ and $s1 \cap s2 = \emptyset, \; \forall s1, s2 \in S$.

Using an atmospheric dispersion model, we determine the set of generated pollution zones. Each zone $\mathcal{Z}_i^s$ corresponds to the geographical area, i.e. set of positions, where the
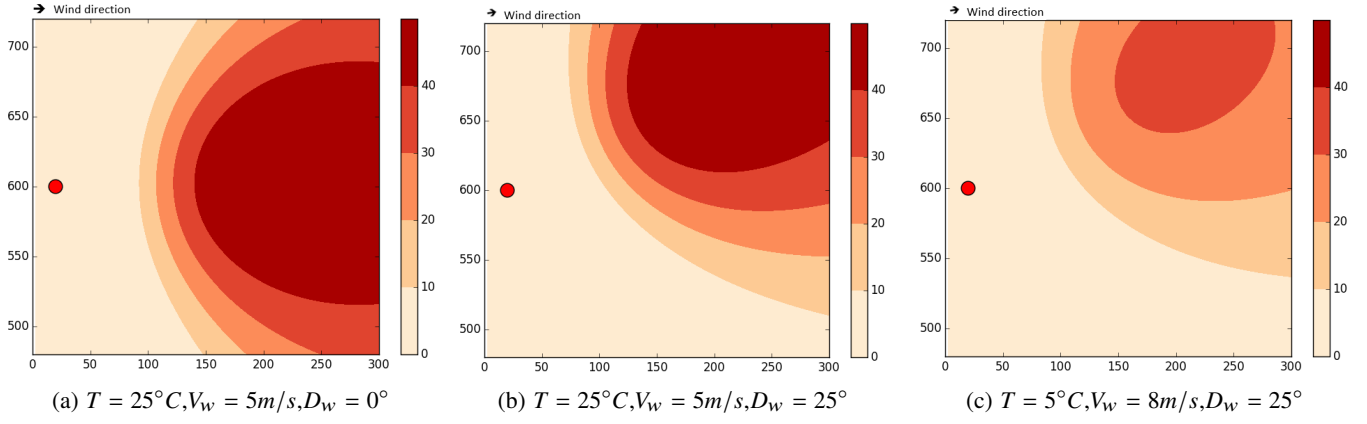
(a) $T = 25°C, V_w = 5m/s, D_w = 0°$ (b) $T = 25°C, V_w = 5m/s, D_w = 25°$ (c) $T = 5°C, V_w = 8m/s, D_w = 25°$

Fig. 1: Simulation of the Gaussian dispersion model with a single pollution source (red point) and different weather conditions.

Sets	
$\mathcal{P}$	Set of potential positions of sensors and sinks
$\mathcal{N}$	Number of sensors and sinks potential positions
$\mathcal{I}$	Set of pollution sources
$\mathcal{M}$	Number of pollution sources
$\mathcal{S}$	Set of weather scenarios
Parameters	
$\mathcal{Z}_i^s$	The pollution zone formed by source i under scenario s
$\mathcal{B}_{ip}^s$	Define whether the position p belongs to the zone $\mathcal{Z}_i^s$ or not
$\Gamma(p)$	The neighborhood of the potential position p
c_p^{sensor}	The cost of deploying a sensor at position p
c_p^{sink}	The cost of deploying a sink at position p
β	Minimum coverage probability to ensure for each zone
$\mathcal{W}_{ip}^s$	The probability of detecting the zone $\mathcal{Z}_i^s$ at position p
δ	Percentage of scenarios that have to be taken into account
α_s	Probability that scenario s is realized
C_0	Pollutant concentration threshold
Variables	
x_p	Define whether a sensor is deployed at position p or not $x_p \in \{0,1\}, \; p \in \mathcal{P}$
y_p	Define whether a sink is deployed at position p or not $y_p \in \{0,1\}, \; p \in \mathcal{P}$
t_i^s	Define whether the zone $\mathcal{Z}_i^s$ is covered or not $t_i^s \in \{0,1\}, \; i \in \mathcal{I}, \; s \in \mathcal{S}$
g_{pq}	Flow quantity transmitted from node p to node q $g_{pq} \in \{0,1,...\}, \; p \in \mathcal{P}, \; q \in \Gamma(p)$
f_{ip}^s	Flow quantity transmitted from zone $\mathcal{Z}_i^s$ to node p $f_{ip}^s \in \{0,1\}, \; i \in \mathcal{I}, \; s \in \mathcal{S}, \; p \in \mathcal{Z}_i^s$

TABLE III: Notations used in deployment models.

pollution threshold is crossed when the pollution source i is releasing pollutants under the weather scenario s. Let the binary parameter $\mathcal{B}_{ip}^s$ denote whether a position p belongs to $\mathcal{Z}_i^s$ or not. A pollution zone $\mathcal{Z}_i^s$ is therefore the set $\{p \in \mathcal{P} \; where \; \mathcal{B}_{ip}^s = 1\}$. When using the pointwise Gaussian dispersion model, the value of $\mathcal{B}_{ip}^s$ is calculated using Formula (6) where σ_y, σ_z, Q and H are the parameters presented in Section III, $p = (x, y, z)$ and C_0 is the threshold of pollutant concentration above which a point is considered as polluted.

$$\mathcal{B}_{ip}^s = \begin{cases} 1 \text{ if } \dfrac{Q}{2\pi V_w^s \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y}} \left(e^{-\frac{(z-H)^2}{2\sigma_y}} + e^{-\frac{(z+H)^2}{2\sigma_y}} \right) \geq C_0 \\ 0 \text{ otherwise} \end{cases} \quad (6)$$

A sensor exposed to a given pollutant will detect its concentration with a probability depending on the sensing accuracy. We denote $\mathcal{W}_{ip}^s \in]0,1[$ the probability of detecting the pollution source i under the weather scenario s at position p, $p \in \mathcal{Z}_i^s$. The $\mathcal{W}_{ip}^s$ parameters are mainly due to the technical characteristics of pollution sensors and are not related to the dispersion model.

Once the pollution zones $\mathcal{Z}_i^s$ are identified and the probability parameters $\mathcal{W}_{ip}^s$ are computed, we formulate the coverage of each pollution source i under each weather scenario s with a probability β in constraint 7.

$$\prod_{p \in \mathcal{Z}_i^s} (1 - \mathcal{W}_{ip}^s \cdot (x_p + y_p)) \leq (1 - \beta), \quad i \in \mathcal{I}, \; s \in \mathcal{S} \quad (7)$$

When a sensor or a sink is placed at position p, i.e. $x_p + y_p = 1$, $1 - \mathcal{W}_{ip}^s \cdot (x_p + y_p)$ is then equal to $1 - \mathcal{W}_{ip}^s$, the probability that the node deployed at p do not cover the pollution zone Z_i^s at position p. Assuming that the detection events are independent among all potential positions, constraint 7 ensures therefore that each zone Z_i^s is covered with a probability $\beta \in]0,1[$.

Partial coverage: Constraint 7 ensures that each pollution source is covered with a probability β under each scenario s. One could want to relax this constraint and ask only for the coverage of each pollution source under δ percent of weather scenarios, with a β probability for each scenario. For that, we introduce the binary variable t_i^s that define whether source i is covered during weather scenario s. Therefore $t_i^s = 1$ if a sufficient number of sensors is placed in the pollution zone Z_i^s. The percentage of weather conditions where i can be detected is the sum of the probabilities that a scenario in which i is detected happens. As a result, the coverage formulation of the partial coverage case is given by the constraints 8 and 9. Constraint 8 should be linearized in order to get an ILP formulation. We first apply the log function to get constraint 11. Since $x_p + y_p$ and t_i^s are binary, the log can be rewritten to get a linear form as shown in constraint 12.

$$\prod_{p \in \mathcal{Z}_i^s} (1 - \mathcal{W}_{ip}^s \cdot (x_p + y_p)) \leq (1 - \beta \cdot t_i^s), \quad i \in \mathcal{I}, s \in \mathcal{S} \quad (8)$$

$$\sum_{s \in \mathcal{S}} t_i^s \cdot \alpha_s \geq \delta, \quad i \in \mathcal{I} \tag{9}$$

$$\log \prod_{p \in \mathcal{Z}_i^s} (1 - \mathcal{W}_{ip}^s \cdot (x_p + y_p)) \leq \log(1 - \beta \cdot t_i^s), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{10}$$

$$\sum_{p \in \mathcal{Z}_i^s} \log(1 - \mathcal{W}_{ip}^s \cdot (x_p + y_p)) \leq \log(1 - \beta \cdot t_i^s), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{11}$$

$$\sum_{p \in \mathcal{Z}_i^s} (x_p + y_p) \cdot \log(1 - \mathcal{W}_{ip}^s) \leq t_i^s \cdot \log(1 - \beta), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{12}$$

Uniform detection probability: If the probabilistic sensing values $\mathcal{W}_{ip}^s$ are identical among all the points that are within the same pollution zone $\mathcal{Z}_i^s$, i.e. $\mathcal{W}_{ip}^s = \mathcal{W}_i^s \ \forall p \in \mathcal{Z}_i^s$, constraint 7 can be simplified as in 13. Using the same binarity argument as before to get the variables out of the log, we get the linear formulation in constraint 17.

$$\prod_{p \in \mathcal{Z}_i^s} (1 - \mathcal{W}_i^s \cdot (x_p + y_p)) \leq (1 - \beta), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{13}$$

$$\prod_{p \in \mathcal{Z}_i^s \text{ where } (x_p + y_p = 1)} (1 - \mathcal{W}_i^s) \leq (1 - \beta), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{14}$$

$$(1 - \mathcal{W}_i^s)^{\left(\sum_{p \in \mathcal{Z}_i^s} (x_p + y_p)\right)} \leq (1 - \beta), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{15}$$

$$\left(\sum_{p \in \mathcal{Z}_i^s} (x_p + y_p)\right) \cdot \log(1 - \mathcal{W}_i^s) \leq \log(1 - \beta), \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{16}$$

$$\sum_{p \in \mathcal{Z}_i^s} (x_p + y_p) \geq \frac{\log(1 - \beta)}{\log(1 - \mathcal{W}_i^s)}, \quad i \in \mathcal{I}, s \in \mathcal{S} \tag{17}$$

2) Connectivity: As in [13], [9] and [23], we formulate in this first model the connectivity constraint as a network flow problem. In contrast to these works, we consider the same potential positions set $\mathcal{P}$ for sensors and sinks and we do not assume that potential positions of sinks are known or different from those of sensors. We first denote by $\Gamma(p)$, $p \in \mathcal{P}$ the set of neighbors of a node deployed at the potential position p. This set can be computed using any adequate propagation models. Then, we define the decision variables g_{pq} as the flow quantity transmitted from a node located at potential position p to another node located at potential position q. We suppose that each sensor of the resulting WSN generates a flow unit in the network, and verify if these units can be recovered by sinks. The following constraints ensure that deployed sensors and sinks form a connected wireless sensor network; i.e. each sensor can communicate with at least one sink.

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \geq x_p - \mathcal{N} \cdot y_p, \quad p \in \mathcal{P} \tag{18}$$

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \leq x_p, \quad p \in \mathcal{P} \tag{19}$$

$$\sum_{q \in \Gamma(p)} g_{pq} \leq \mathcal{N} \cdot x_p, \quad p \in \mathcal{P} \tag{20}$$

$$\sum_{p \in \mathcal{P}} \sum_{q \in \Gamma(p)} g_{pq} = \sum_{p \in \mathcal{P}} \sum_{q \in \Gamma(p)} g_{qp} \tag{21}$$

Constraints 18 and 19 are designed to ensure that each deployed sensor, i.e. such that $x_p = 1$, generates a flow unit in the network. These constraints are equivalent to the following.

$$\sum_{q \in \Gamma(p)} g_{pq} - \sum_{q \in \Gamma(p)} g_{qp} \begin{cases} = 1 & if \ x_p = 1, y_p = 0 \\ = 0 & if \ x_p = y_p = 0 \\ \leq 0, \geq -\mathcal{N} & if \ x_p = 0, y_p = 1 \end{cases}$$

The first case corresponds to deployed sensors that should generate, each one of them, a flow unit. The second case, combined with constraint 20, ensures that absent nodes, i.e. $x_p = y_p = 0$, do not participate in the communication. The third case concerns deployed sinks, and ensures that each sink cannot receive more than $\mathcal{N}$ units. The case $x_p = y_p = 1$ is not possible because of constraint 4. Constraint 20 ensures also that deployed sinks cannot transmit flow units, and only act as receivers. Constraint 21 means that the overall flow is conservative. The flow sent by deployed sensors has to be received by deployed sinks.

3) ILP model: At the end, our general basic optimization model can be written as follows.

[*Basic model*]

$$\begin{aligned} Minimize \quad & (5) \\ Subject \ to. \quad & (4), (9), (12), (18), (19), (20) \\ & and \ (21) \end{aligned}$$

B. Enhanced Model

The basic formulation cannot deal with large-scale instances. One of the main reasons is that the two sub-problems, namely connectivity and coverage, are formulated as set of constraints of different natures. To cope with this, we propose in this section a more efficient modeling. By considering pollution sources as a part of the network, we obtain a homogeneous coverage/connectivity formulation as a network flow problem. In the second model, each pollution source i should transmit some flow units to potential nodes p which are located within the pollution zone corresponding to each weather scenario s i.e. $p \in \mathcal{Z}_i^s$. In addition, sensors are flow conservative and the sinks receive the flow units generated by pollution sources. Therefore, the definition of the joint coverage/connectivity is to ensure that sinks will be informed each time that a threshold crossing occurs. In this regard, a sensor has to receive at most one unit from a given pollution zone. We hence define the binary decision variable f_{ip}^s as the flow quantity from the pollution source i to the potential node p in the case of weather scenario s. The following constraints ensure coverage and connectivity for pollution monitoring.

$$\sum_{s \in \mathcal{S}} t_i^s \cdot \alpha_s \geq \delta, \quad i \in \mathcal{I} \tag{22}$$

$$\sum_{p \in \mathcal{Z}_i^s} f_{ip}^s \cdot \log(1 - \mathcal{W}_{ip}^s) \leq t_i^s \cdot \log(1 - \beta), i \in \mathcal{I}, s \in \mathcal{S} \tag{23}$$

$$\sum_{i \in \mathcal{I}, s \in \mathcal{S}: p \in \mathcal{Z}_i^s} f_{ip}^s + \sum_{q \in \Gamma(p)} g_{qp} - g_{pq} \leq \mathcal{N}\mathcal{M}|\mathcal{S}|y_p, \quad p \in \mathcal{P} \tag{24}$$

$$\sum_{i \in \mathcal{I}, s \in \mathcal{S}: p \in \mathcal{Z}_i^s} f_{ip}^s + \sum_{q \in \Gamma(p)} g_{qp} - g_{pq} \geq 0, \quad p \in \mathcal{P} \qquad (25)$$

$$\sum_{q \in \Gamma(p)} g_{pq} \leq \mathcal{N}\mathcal{M}|\mathcal{S}|x_p, \quad p \in \mathcal{P} \qquad (26)$$

$$f_{ip}^s \leq x_p + y_p, \quad p \in \mathcal{P}, i \in \mathcal{I}, s \in \mathcal{S} \qquad (27)$$

Coverage is formulated in constraints 23 and 23. Constraint 23 is similar to the basic model and ensures coverage of each pollution source under a δ percentage of weather scenarios. Constraint 23 is derived from constraint 12 and ensures that each pollution source i generates a sufficient number of flow units in the network. Constraint 27 enforces that all the flow units are received by deployed nodes. Thanks to constraints 24 and 25, when a sensor is deployed on point p (case $y_p = 0$ and $x_p=1$), we ensure that the inflow of sensor p is equal to its outflow; i.e. The flow is conservative on deployed sensors. In addition, constraints 24 and 25 also ensure that the sinks are allowed to gather all the flow units that are generated in the network (case $y_p = 1$ and $x_p=0$). Constraints 26 and 27 combined with constraints 24 and 25 ensure that absent nodes do not participate in the communication. As a results, the deployed sensors have to send the flow units gathered from pollution sources to the sinks in order to get the connectivity constraints verified. The enhanced optimization model can then be written as follows.

[*Enhanced model*]

$$Minimize \quad (5)$$
$$Subject\ to. \quad (4), (23), (23), (24), (25),$$
$$(26)\ and\ (27)$$

Particular cases: As in the basic model, some simplifications can be made in some particular cases. First, binary variables t_i^s can be avoided when covering all the pollution zones is required, in which case all t_i^s variables will be replaced by 1. Then, when considering homogeneous probabilistic sensing among potential positions of sensors, i.e. $\mathcal{W}_{ip}^s = \mathcal{W}_i^s$ for all $p \in \mathcal{P}$, constraint 23 can be transformed to constraint 28 using the same transformations done with the basic model.

$$\sum_{p \in \mathcal{P}} f_{ip}^s \geq \frac{\log(1 - \beta)}{\log(1 - \mathcal{W}_i^s)}, \quad i \in \mathcal{I}, s \in \mathcal{S} \qquad (28)$$

Finally, considering the case of mono sink WSN, constraints 24 and 25 can be grouped and replaced by the following constraint 29.

$$\sum_{i \in \mathcal{I}, s \in \mathcal{S}: p \in \mathcal{Z}_i^s} f_{ip}^s + \sum_{q \in \Gamma(p)} g_{qp} - g_{pq} = \sum_{i \in \mathcal{I}, s \in \mathcal{S}} K_i^s \cdot y_p, \quad p \in \mathcal{P} \quad (29)$$

V. SIMULATION RESULTS

A. Greater London dataset

We evaluate our deployment models on a data set provided by the Greater London community [24]. London is one of the most polluted cities in Europe [25]. The data set corresponds to the locations of urban pollution sources. In this data set, mostly urban facilities have the potential to affect the air quality such as petrol stations, waste oil burners, cement works, etc. The set of pollution sources is spread over the 32 boroughs of Greater London. Overall, 1090 pollution sources are considered. Pollution sources are depicted in Fig. 2. The distribution of pollution sources per borough depends on the surface of the borough and ranges from 6 sources to 161 sources.
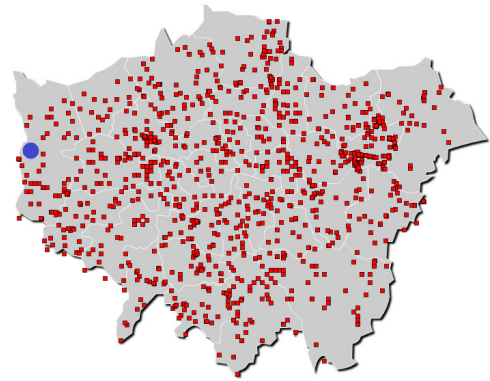

Fig. 2: Pollution sources (squares) and weather station (disk).

In addition to pollution sources locations, we compute the weather scenarios leveraging statistical data gathered by a weather station of London [26]. The location of this station is depicted in Fig. 2. We consider weather conditions of each month of the year averaged over the last past 10 years. The set of weather conditions is depicted in table IV, each scenario corresponds to values of ambient temperature, wind direction and wind velocity. As proof of concept and without loss of generality, we assume that weather scenarios provided by the considered weather station are homogeneous in all the area of Greater London.

ILP formulations are implemented using the IBM ILOG CPLEX Optimization Studio and executed on a PC with Intel Xeon E5649 processor under Linux. The ILP solver is executed with a time limit of 30 minutes. The default values of simulation parameters are summarized in Table V. We generate the pollution inputs of our deployment models using the Gaussian model presented in section III while considering the same pollutants characteristics as discussed in section III. We define the nodes neighboring Γ based on a given transmission range. Moreover, we assume that the cost of nodes is independent of the position of the node, i.e. $c_p^{sensor} = c^{sensor}$ and $c_p^{sink} = c^{sink}$. Furthermore, we investigate the coverage of pollution sources with respect to all the considered scenarios, i.e. $\delta = 1.0$. In addition, we consider that the probabilistic sensing value $\mathcal{W}_{ip}^s$ is constant and equal to 90%.

Month of year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Ambient temperature (^oC)	7	7	9	13	15	19	21	20	18	14	9	7
Wind velocity (m/s)	5	5	5	5	5	5	5	5	5	5	5	5
Wind direction (degree)	225	247	270	270	270	225	225	225	225	202	225	247

TABLE IV: Weather statistics of London.

Parameter	Value
Nodes transmission range	$100m$
Nodes height	$10m$
Sensors cost (c_p^{sensor})	1 monetary unit
Sinks cost (c_p^{sink})	10 monetary units
Coverage requirements of pollution zones (β)	0.98
Coverage requirements of weather conditions (δ)	1.0
Detection sensitivity of sensors ($\mathcal{W}_{ip}^s$)	0.9
Ambient Temperature (T)	7^oC
Wind velocity (V_w)	$5m/s$
Wind direction (D_w)	225°
Pollution threshold (C_0)	$20\mu g/m^3$

TABLE V: Summary of default simulation parameters.

B. Application to the London Borough of Camden

As a proof of concept, we first execute our models on the London Borough of Camden. We use streetlights as potential positions of sensors in order to alleviate the energy constraints. The streetlights data set was provided by the Camden DataStore [27]. Camden is spread over an area of around $8km \times 6km$ and contains 19 pollution sources. Fig. 3 depicts the pollution zones obtained by running of the Gaussian dispersion model while taking into account weather conditions of the month of January. Fig. 3 also shows the obtained positions of wireless sensor network nodes computed by the deployment models. We notice that sensors are placed at the intersections of the different pollution zones in order to minimize the coverage deployment cost. Moreover, the resulting network consists of 7 sub-networks, a sink is deployed in each one and some sensors are added to ensure connectivity.

The following results have been obtained by running our deployment models on a hundred of $1200m \times 1200m$ blocks extracted from the Greater London map. The density of pollution sources varies between 3 and 18 sources per block. We discretize each block with a resolution of $100m$ to get a $2D$ grid of points that we consider as potential positions of WSN nodes.

C. Tractability evaluation

1) Compactness and tightness: Despite the fact that the models are to run offline, a better model formulation allows the execution on large-scale instances. Integer programming formulations are usually compared in terms of compactness and tightness [28], [29]. The compactness of a model is given by the size of the instances mainly the number of variables and the number of constraints. A compact formulation, i.e. a formulation that allows for small size instances, is not sufficient to get better performances. A good formulation should be compact and tight at the same time. The tightness of a model allows for a reduced search space and is usually measured using the integrality gap [28]. This is the gap between the optimal objective value of the ILP Z_{ILP} and the optimal objective value of the relaxed formulation Z_{LP}. In the case of a minimization problem, the integrality gap is equal to $(Z_{ILP} - Z_{LP})/Z_{ILP}$. When considering two models with more or less the same compactness, the fastest model is the one that has the lowest integrality gap.

2) Metric of computational complexity: In order to show the impact of the complexity of the block instances on the tractability of our models, we consider the area of interest as a complexity metric. For a given block b, let C_b be the set of potential positions of sensors that are at least within a pollution zone generated by the block pollution sources under the weather scenarios that are considered. The metric value is defined as the area of the convex envelope of the set C_b. This means that the area of interest includes all the potential positions needed for pollution coverage, i.e. C_b, and also the area where relay nodes may be placed. Indeed, neither coverage sensors nor relay nodes will be placed in the block area that is not included in the area of interest.

3) Comparison results: As mentioned before, we execute the deployment models using formulations with respect to the simulation parameters. The two models gave the same objective values; this was expected since the enhanced model is derived from the basic one. We depict in table VI the compactness, the tightness and the execution time of the models depending on the area of interest of block instances. Results have been averaged with respect to the complexity metric class of each instance. We also plot in Fig. 4 the distribution of execution time within the classes of the area of interest using a boxplot.

We notice that the instances that are more complex take more time to be resolved when using both of the two models. Moreover, Basic Model and Enhanced Model have nearly the same compactness, the difference is due to f_{ip}^s variables added to the enhanced formulation to define a link between pollution sources and WSN nodes. However, the enhanced model is much tighter than the basic one with lower integrality gaps. This difference in tightness impacts well CPU time and allows to enhance the total mean execution time with a factor of around 8.

We now plot in Fig. 5 the histogram of the ratio of execution time between the enhanced model and the basic model. We notice that for around 40% of the block instances, the enhanced model is at least 10 times faster than the basic model. Fig. 5 also shows the density function of the ratio of integrality gap between the enhanced model and the basic model. The density function of integrality gap ratio fits well with the histogram of execution time ratio; this is explained by the fact that the integrality gap and the execution time are highly correlated.

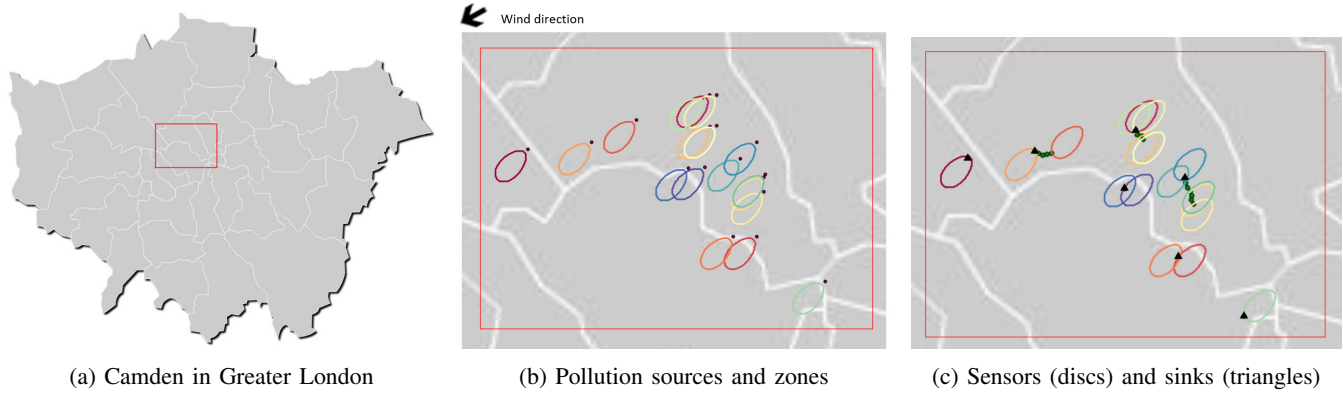Based on the results obtained in this section, we execute

(a) Camden in Greater London (b) Pollution sources and zones (c) Sensors (discs) and sinks (triangles)

Fig. 3: Application of our deployment models to the London Borough of Camden

Area of interest (km^2)	Compactness (nb. of vars. and consts)		Tightness (int. gap)		CPU time (seconds)	
	Basic Model	Enhanced Model	Basic Model	Enhanced Model	Basic Model	Enhanced Model
[0.00 − 0.20[	2193	2275	0.880	0.070	7.460	0.890
[0.20 − 0.45[	2194	2354	0.870	0.080	20.400	2.810
[0.45 − 0.70[	2195	2429	0.850	0.120	29.830	3.360
[0.70 − 0.95[	2196	2474	0.830	0.180	68.200	8.820
Mean	**2194.5**	**2383**	**0.860**	**0.110**	**31.470**	**3.970**

TABLE VI: Basic Model VS Enhanced Model.
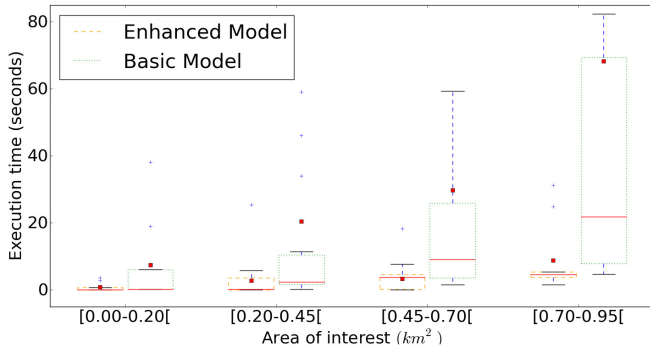
Fig. 4: Impact of the area of interest on the execution time. Red squares represent the means of the classes and "plus" signs represent the outliers.
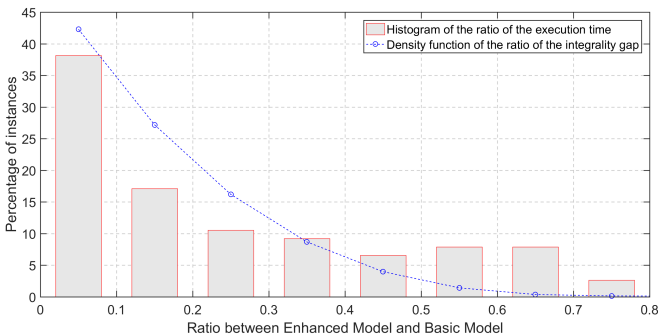
Fig. 5: Relationship between the integrality gap and the execution time.

only the enhanced model in the following simulations where we investigate the impact of some decision parameters on the deployment results.

D. Analysis of the network connectivity

1) Evaluation of the number of nodes: In this simulation case, we analyze the number of sinks and sensors in the resulting networks while varying the ratio between sink cost and sensor cost. We plot in Fig. 6 the impact of the cost ratio on the optimal number of sensors and sinks. The cost ratio ranges from 1 to 12 and the results are averaged over all the London blocks defined in the previous section. On one hand, we notice that sensors are less used when their cost is close to the sinks cost. For instance, only sinks are used when the cost ratio is equal to 1. On the other hand, when the cost ratio increases, more sensors are used and the number of required sinks tends to one. As a result, the network is usually formed by only one sink when the cost ratio is greater than 10. This is explained by the fact that adding some relay sensor nodes to ensure connectivity has a less cost than using a lot of sinks that are equipped with pollution sensors.

In the following simulations, we execute the deployment models with a default value of sink/sensor cost ratio equal to 10 as shown in table V. Thus, we use formulations corresponding to the mono-sink case.

2) Evaluation of the number of hops to sink nodes: In this simulation case, we evaluate the obtained networks in terms of the number of hops to sink nodes, which is a measure of the network lifetime and communication delay [30]. As formulated in our connectivity constraints, the positioning of sink nodes does not take into account the length of sensor-to-sink paths. However, sinks can be relocated on the obtained network when the network is monosink, which is the case as shown in the previous subsection. The sink node can be relocated in such a way that the maximum distance to the sink in terms of hops is minimized. This distance is called the radius of the network and describes how much the network
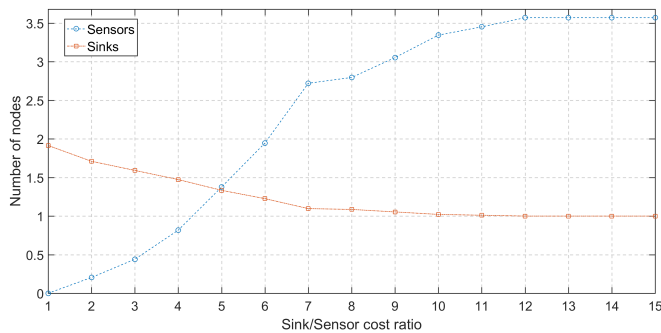
Fig. 6: Number of nodes depending on the cost ratio.

is well connected when considering the best position of the sink node. If the sink position is given randomly, the maximum distance to the sink is bounded by the diameter of the network, which is the distance to the sink node when choosing the worst position of the sink.

We depict in Fig. 7 the cumulative distribution functions of the network radius and diameter. Results show that the network radius is as at most equal to 5 for more than 96% of the instances. This means that the number of hops to the sink node after relocation is at most equal to 5 in almost all instances. Moreover, we notice that the network radius is nearly equal to the diameter of the network for the rest of the instances, which means that the sink relocation does not improve much the network connectivity in this case.
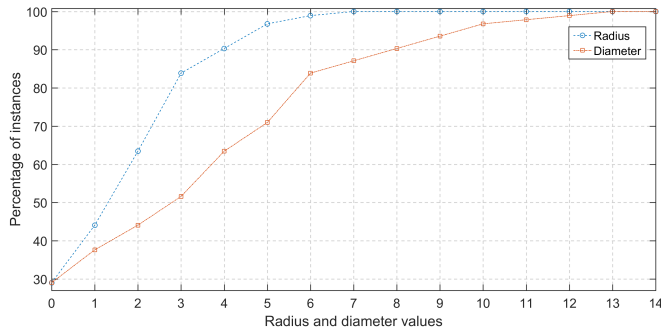

Fig. 7: Cumulative distribution functions of the network radius and diameter.

E. Analysis of the coverage results

1) Impact of nodes height: We now study the impact of the height at which are placed sensors and sinks on the deployment cost. We assume that the height of pollution sources is equal to $25m$, and all the sensors and sinks are deployed at the same height, which is considered in the range from 5 to 40 meters. We plot in Fig. 8 the sensors and sinks overall deployment cost depending on their height when applying two different weather scenarios, those corresponding to January and December. The results are averaged over all the London blocks. On one hand, we notice that the deployment cost is minimal when the nodes height is close to the effective release height of pollution sources H, which is nearly equal to 25.1 in our case. This is explained by the fact that pollution concentration gets

the highest values when being near to the pollutant effective release height H. On the other hand, pollutants are more likely to drop than to increase, which is due to gravitation. Indeed, the deployment cost at $40m$ is much greater than the deployment cost at $5m$. Fig. 8 also shows that when using different weather scenarios, the deployment cost is not the same. Indeed, weather conditions impacts the disposition of pollution zones allowing for more or less intersections. As a result, the obtained WSN topology depends on the weather conditions taken into account.
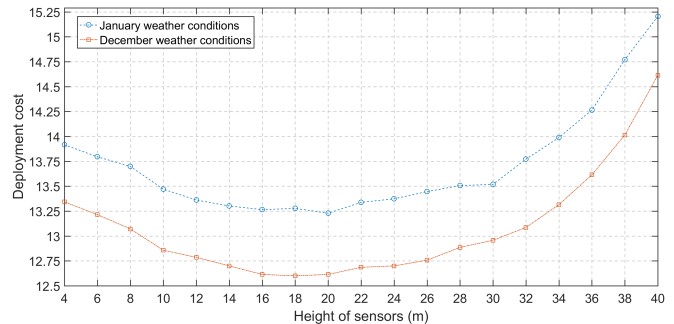

Fig. 8: Deployment cost average depending on nodes height with different weather conditions.

2) Impact of pollution sources density: In this scenario, we study the impact of pollution sources density on the deployment cost. For this purpose, we take the results of the previous scenario corresponding to January weather conditions and averaged with respect to the number of pollution sources of each instance, i.e. the number of pollution sources within each block instance. We plot in Fig. 9 the deployment cost variations depending on the nodes height while considering three different densities: 4, 5 and 6 pollution sources per instance. Fig. 9 shows that more there are pollution sources in the environment, more there are sensors required and thus higher is the deployment cost. This can be explained by the number of pollution zones that increases with the number of pollution sources, and thus requires much sensors to ensure the coverage requirements. In addition, the increasing in the deployment cost from 5 sources density to 6 sources density is less than the increasing from 4 sources density to 5 sources density. This is because when the number of pollution sources increases, more intersections between pollution zones appear and affect the increasing of the deployment cost.

3) Impact of probabilistic sensing: The probabilistic sensing of pollution sensors is one of the most important factors that affect the topology of sensor networks used for pollution monitoring. Fig. 10 depicts the average cost of the resulting deployments of the block instances while considering two values of the detection sensitivity of sensors: $W_{ip}^s = 0.9$ and $W_{ip}^s = 0.8$. As expected, using sensors with better detection sensitivity yields less deployment cost. We notice that the ratio between the two curves is around 1.1. This is explained by the intersection existence between the different polluted zones, which means that in some cases a sensor can monitor more than one pollution source.

4) Impact of the number of weather conditions: In this simulation case, we study the impact of using a small number
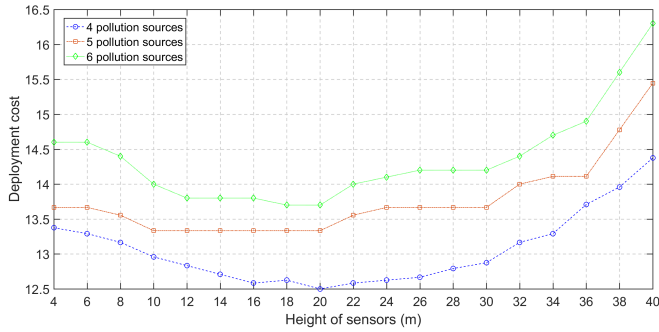
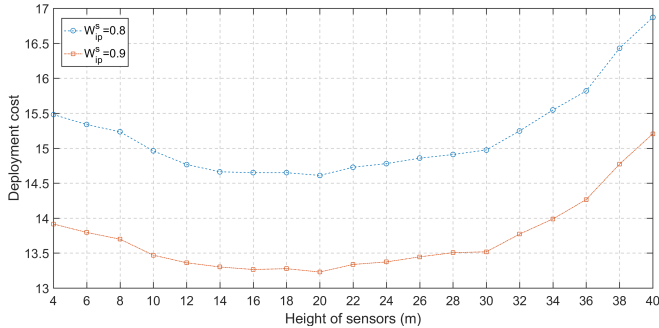Fig. 9: Deployment cost average depending on nodes height and pollution sources density.

Fig. 11: Average percentage of missed pollution zones depending on weather scenarios.

Fig. 10: Deployment cost average depending on nodes height and probabilistic sensing values.

shown in Fig. 11, considering only weather scenarios from December to May meaning only 8 scenarios allows to cover the whole set of pollution zones in contrary to the scenarios set from January to October that requires 10 scenarios.

5) Impact of the spatiotemporal variation of weather conditions: In this last simulation case, we study the impact of spatial and temporal variations of weather conditions on the deployment results. In order to do so, we use the data of a network of 12 weather stations distributed in all the Greater London [31]. Raw data consist of hourly values of temperature, wind direction (12 possible directions in the range from 0° to 360°) and wind velocity of June 2016. We use this data in order to construct the set of weather scenarios of our deployment model. For each weather station, we have got 672 weather scenarios (28 days and 24 scenarios for each day).

In order to take into account the spatial variation of weather conditions, we propose to consider different weather conditions for each pollution source. When applying the pollution dispersion model on a given pollution source, we use the weather conditions corresponding to the weather station that is the nearest to the given pollution source. Indeed, the Gaussian dispersion model assumes that weather conditions are homogeneous in each pollution zone. We mention that the diameter of pollution zones is at most equal to 2 km as shown in Figure 1.

Before the execution of our deployment models, we first eliminate, for each weather station, the weather scenarios that occur more than one time while leaving only one occurrence. We plot in Figure 12(a) the number of weather scenarios of some weather stations while varying the aggregation window of data (in the range from 1 hour to 24 hours). We use the mean function for aggregation. We notice that station 3 has the highest temporal variation and station 4 has the lowest one. We did not depict the curves of all the 12 stations for clarity reasons.

We consider in simulation three cases of spatial variation of weather conditions: i) heterogeneous case: we use the data of all the 12 weather stations; ii) 1st homogeneous case: we use only the data of the station that has lowest temporal variation (station 4); iii) 2nd homogeneous case: we use only the data of the station that has highest temporal variation (station 3). We depict in Figure 12(b) the mean deployment cost of block instances depending on the temporal aggregation window

of weather scenarios on the deployment results. It is clear that when considering all the possible weather scenarios, the resulting WSN ensures better pollution monitoring. However, when there is a huge number of weather scenarios, considering a less number of these scenarios alleviates the deployment models allowing their application on large-scale instances.

We recall that $\mathcal{S}$ is the set of the monthly weather scenarios presented in table IV. Given a subset $\mathcal{S}\prime$ of $\mathcal{S}$, we define the missed pollution zones percentage as the percentage of pollution zones that cannot be covered by the WSN resulting from executing the models under only weather scenarios $\mathcal{S}\prime$. Fig. 11 illustrates the variations of the missed pollution zones percentage depending on the number of weather scenarios taken into account starting from January weather scenario in the first curve and starting from December weather scenario is the second one. Fig. 11 shows that the percentage of missed pollution zones usually decreases when considering more weather scenarios, this is expected since the number of pollution zones depends on the weather scenarios set. However, in some cases, the missed pollution zones percentage remains the same when considering additional scenarios of weather conditions. Indeed, additional weather scenarios involve new pollution zones that may be, in some cases, already included in the set of pollution zones formed without taking into account the additional scenarios. This may happen for instance when additional weather scenarios are slightly different from those already taken into account.

In addition to the impact of the number of weather scenarios, their similarity has also to be taken into account. As
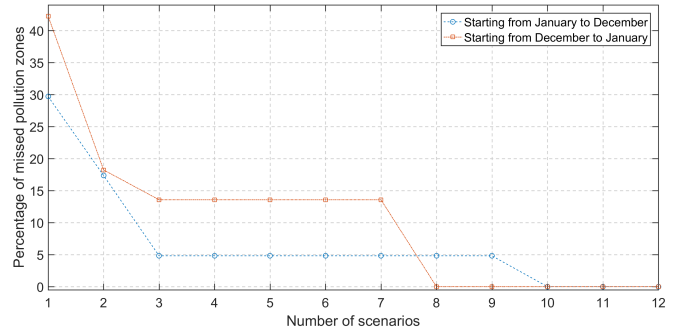
while considering the 3 spatial cases. Results show that, for all the spatial cases, larger aggregation window involves less number of sensors. This is because the number of weather scenarios is reduced when performing aggregation on larger temporal windows as shown in Figure 12(a), i.e. some weather information may be deleted when aggregation is performed on larger temporal windows.

Results also show that on the one hand, when considering the weather data of all the stations (the heterogeneous case), we place more sensors than when considering only the data of the station that has the lowest weather variability (1st homogeneous case, station 4). On the other hand, when considering only the data of the station that has the highest weather variability (2nd homogeneous case, station 3), we place more sensors than when considering the weather data of all the stations (the heterogeneous case). This is due to the fact that station 3 (respectively station 4) involves the maximum (respectively minimum) number of weather scenarios as shown in Figure 12(a).

VI. CONCLUSION AND FUTURE WORK

Air pollution is becoming a major problem of smart cities due to the increasing industrialization and the massive urbanization. In this work, we focused on the use of wireless sensor networks for air pollution monitoring and in particular the detection of threshold crossings. We addressed the deployment issue and proposed two optimization models ensuring pollution coverage and network connectivity with the minimum cost. Unlike the inadequate related works, which rely on a simple and generic detection model, we based on atmospheric dispersion modeling to take into account the nature of the addressed phenomenon. Our deployment models are designed in such a way to handle multiple weather scenarios and take into account the probabilistic sensing of nodes. In addition to the coverage formulation based on atmospheric dispersion, we proposed in the second model a joint formulation of coverage and connectivity based only on flows. This allows reducing the computational burden according to our simulation results.

We evaluated the impact of the model parameters on the deployment results. We concluded that sensors should be placed at a height close to the one of pollution sources. We also studied the impact of the weather scenarios set on the coverage quality and showed that this latter depends on the similarity of scenarios, their number, the aggregation temporal window and the spatial heterogeneity of weather conditions. In addition to the weather heterogeneity, the emission variability would also affect the deployment results: The higher the emission rates, the more intersections and hence the less number of sensors.

Future work

In this paper, we tried to focus on what the literature lacks in WSN deployment for air pollution monitoring, which is an appropriate coverage formulation of thresholds detection. We have also shown how connectivity and coverage can be merged into one constraint in order to improve the execution time. Our models can be extended in order to take into account the network lifetime and mobile deployment.

a) Optimization of the network lifetime: One possible solution is to base on the work of [9] where authors define the energy constraints based on the flow concept. This solution is viable since our model is also 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.

b) Mobile deployment: Indeed, our models are designed for static networks, which is argued by the fact that pollution sensors operate well when they are static [32]. 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 [9]. 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.

Another perspective of our work would be to consider the impact of the nature of pollutants and the urban topography on the coverage results. We are also working on the validation of our deployment models.

ACKNOWLEDGMENT

REFERENCES

[1] World Health Organization, "Burden of disease from household air pollution for 2012," on http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf [2016-02-27].

[2] International Agency for Research on Cancer, "Iarc: Outdoor air pollution a leading environmental cause of cancer deaths," on http://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf [2016-02-20].

[3] AirParif, "The air quality monitoring organization of the paris agglomeration," http://www.airparif.fr/en/index/index.

[4] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[5] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[6] 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," Tech. Rep., 2009.

[7] A. Kadri, E. Yaacoub, M. Mushtaha, and A. Abu-Dayya, "Wireless sensor network for real-time air pollution monitoring," in *ICCSPA, 2013 1st International Conference on*. IEEE, 2013, pp. 1–5.

[8] 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.
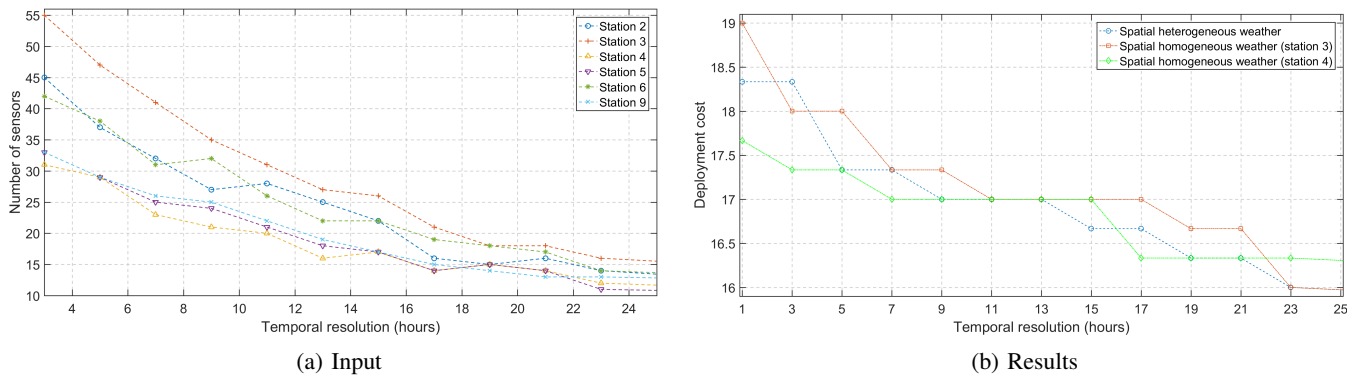
(a) Input

(b) Results

Fig. 12: Impact of the heterogeneity of weather conditions on the deployment results

[9] 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.

[10] S. Sengupta, S. Das, M. Nasir, and B. K. Panigrahi, "Multi-objective node deployment in wsns: 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.

[11] M. Rebai, H. Snoussi, F. Hnaien, 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.

[12] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 6, no. 4, pp. 621–655, 2008.

[13] M. Cardei, M. O. Pervaiz, and I. Cardei, "Energy-efficient range assignment in heterogeneous wireless sensor networks," in *ICWMC'06. International Conference on*. IEEE, 2006, pp. 11–11.

[14] A. Ghosh and S. K. Das, "Coverage and connectivity issues in wireless sensor networks," *Mobile, wireless, and sensor networks: Technology, applications, and future directions*, pp. 221–256, 2006.

[15] 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.

[16] A. Boubrima, F. Matigot, W. Bechkit, H. Rivano, and A. Ruas, "Optimal deployment of wireless sensor networks for air pollution monitoring," in *ICCCN'15. 24th International Conference on Computer Communications and Networks*. IEEE, 2015, pp. 1–7.

[17] T. Koch, T. Achterberg, E. Andersen, O. Bastert, T. Berthold, R. E. Bixby, E. Danna, G. Gamrath, A. M. Gleixner, S. Heinz *et al.*, "Miplib 2010," *Mathematical Programming Computation*, vol. 3, no. 2, pp. 103–163, 2011.

[18] 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.

[19] 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.

[20] A. Daly and P. Zannetti, "Air pollution modeling–an overview," *Ambient air pollution*, 2007.

[21] J. M. Stockie, "The mathematics of atmospheric dispersion modeling," *Siam Review*, vol. 53, no. 2, pp. 349–372, 2011.

[22] R. N. Murty, G. Mainland, I. Rose, A. R. Chowdhury, A. Gosain, J. Bers, and M. Welsh, "Citysense: An urban-scale wireless sensor network and testbed," in *Technologies for Homeland Security, 2008 IEEE Conference on*. IEEE, 2008, pp. 583–588.

[23] M. Patel, R. Chandrasekaran, and S. Venkatesan, "Energy efficient sensor, relay and base station placements for coverage, connectivity and routing," in *IPCCC 2005. 24th IEEE International*. IEEE, 2005, pp. 581–586.

[24] London DataStore, "London atmospheric emissions inventory 2010," on http://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2010 [2016-01-15].

[25] The guardian, "London ranks among worst european cities for air pollution," on http://www.theguardian.com/environment/2011/sep/07/london-worst-european-cities-air-pollution [2016-02-15].

[26] Wind Finder, "Wind and weather statistics london-heathrow," on http://www.windfinder.com/windstatistics/london-heathrow [2016-01-15].

[27] Camden Data, "Camden lighting point," on http://www.camdendata.info/Pages/Home.aspx [2015-07-15].

[28] G. Morales-España, J. M. Latorre, and A. Ramos, "Tight and compact milp formulation of start-up and shut-down ramping in unit commitment," *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 1288–1296, 2013.

[29] ——, "Tight and compact milp formulation for the thermal unit commitment problem," *Power Systems, IEEE Transactions on*, vol. 28, no. 4, pp. 4897–4908, 2013.

[30] S. R. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, "Energy efficient schemes for wireless sensor networks with multiple mobile base stations," in *Global telecommunications conference, 2003. GLOBECOM'03. IEEE*, vol. 1. IEEE, 2003, pp. 377–381.

[31] LGFL, "London grid for learning," on http://weather.lgfl.org.uk/ [2016-06-20].

[32] 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 combinatorial optimization and networks design.

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-time researcher at INRIA, France and the head of the Agora team, which focuses on the use of wireless networks in smart cities. Prior to that, he had been a CNRS researcher from October 2004 to January 2011. 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.