Wireless Sensor Networks Deployment for Air Pollution Monitoring

Ahmed Boubrima, Walid Bechkit, Hervé Rivano, Anne Ruas

Wireless Sensor Networks Deployment for Air Pollution Monitoring

Ahmed Boubrima[1], Walid Bechkit[1], Hervé Rivano[1], and Anne Ruas[2]

[1] Univ Lyon, Inria, INSA Lyon, CITI, F-69621 Villeurbanne, France
[2] IFSTTAR, Paris, France

Abstract. Recently, air pollution monitoring emerges as a major issue of the development of smart cities and the well-being of citizens. Air pollution is traditionally monitored using some measuring stations that are accurate but expensive, big and inflexible. This leads to bad global estimations of pollution concentrations. The emergence of air quality sensors, which are less expensive, allows to consider a new pollution monitoring paradigm based on the wireless interconnection between these sensors. This allows to ensure a district-wide air pollution monitoring. In this paper, we tackle the minimum-cost node positioning issue for the detection of air pollution thresholds. We propose two models for wireless sensors deployment while taking into account the air pollution modelling and the probabilistic sensing of nodes. We evaluate our deployment models on a real data set of Greater London and conduct extensive simulations to study the impact of some parameters, among which sensors' height. Results show that the deployment cost depends on the dispersion of pollutants in the area of interest and can be minimized by placing sensors at a height close to the one of pollution sources.
Keywords— Air pollution monitoring, Detection of threshold crossings of pollutants, Atmospheric dispersion modeling, Wireless sensor networks (WSN), Deployment, Coverage, Connectivity.

1 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 [17]. In 2013, the International Agency for Research on Cancer (IARC) classified particulate matter, the main component of outdoor pollution, as carcinogenic for humans [14]. 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. Conventional measuring stations are equipped with multiple lab quality sensors

[1]. 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 can measure information from the environment, process and relay them to some base stations, denoted sinks [18]. 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 a to obtain a finer spatiotemporal granularity of measurements, thanks to the resulting lighter installation and operational costs [2]. 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 [5]. In this work, 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 [20]. 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 [8][11][10]. The network is said connected if each sensor can communicate information to at least one sink [19]. 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 [3]. The coverage issue has often been modeled as a k-coverage problem in which at least k sensors should monitor each point of interest [6]. 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 [6]. 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. 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 a more effective model in which we give a joint formulation of coverage and connectivity using only the flow concept. The second model is compact and tighter than the first one. In both of the two models, we take into account the probabilistic sensing of pollution sensors.

The remainder of this paper is organized as follows. We first present and analyze the atmospheric dispersion modeling of pollutants in section 2. Then, section 3 details our two proposed optimization models while section 4 presents the simulation dataset and the obtained results. Finally, we conclude and propose some perspectives in section 5.

2 Atmospheric dispersion modeling

As claimed in the previous section, generic detection models using the notion of detection range are not suitable to the sensing of air pollutants. In particular, a realistic formulation of pollution coverage has to capture the dispersion in the atmosphere. The atmospheric dispersion of pollutants was extensively studied in the literature and several models have been proposed and validated. These models are of major interest for many applications such as weather forecasting, assessment of contamination, poisoning, etc. The theoretical study of pollutant atmospheric dispersion is mainly based on fluid mechanics theory [4]. For the sake of clarity, we focus in this work on steady state dispersion models, in particular on Gaussian dispersion. Our approach is however more generic and can be adapted to more sophisticated dispersion models.

The basic Gaussian model estimates the concentrations of a pollutant gas released by a pointwise pollution source in a free space environment [12]. The estimated value C (g/m^3) at a measurement location (x, y, z) is given by Formula (1). Table 1 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. 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. In this case, the value of Δh is given by Formula (2) where F, which denotes the pollutant gas buoyancy, is computed using Formula (3).

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

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

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

Formula (1) takes into account only pointwise pollution sources, and thus cannot be applied to area sources like crossroads and line sources like highways. Multiple extensions were proposed in the literature to deal with these kinds of

Pollution source	
h_s	Pollutant source height (m)
Δh	Plume rise (m)
H	Pollutant effective release height (m)
Measurement location	
x	Downwind distance from the pollution source (m)
y	Crosswind distance from the pollution source (m)
z	Hight (m)
Pollutant gas	
Q	Mass flow rate at the emission point (g/s)
V	Volumetric flow rate at the emission point (m^3/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 ($9.8m/s^2$)

Table 1: Parameters of the Gaussian dispersion model

pollution sources. In addition, many enhanced systems were developed based on the Gaussian model to take into account complex meteorological data, effect of obstacles on pollution dispersion, etc.

3 Deployment models

In this paper, we propose two integer programming formulations for optimal deployment of wireless sensor networks to efficiently monitor air pollution. Our formulations rely on a pollution dispersion model. For the sake of clarity, we use the aforesaid Gaussian dispersion model in this work. Our approach is however more generic and can be adapted to more sophisticated dispersion models and environments. We consider pollutants that can be released by industrial sources like factories, sewage treatment plants, etc, as well as traffic sources such as highways and crossroads. In the first model, we formulate pollution coverage and network connectivity independently. In the second one, we propose a more sophisticated formulation in which coverage and connectivity are joint in a common network flow problem. Table 2 depicts the notations used in the integer programming formulations.

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
Parameters	
$\mathcal{Z}_i$	The pollution zone formed by source i
$\mathcal{B}_{ip}$	Define whether the position p belongs to the zone Z_i 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	Coverage requirements of each zone
$\mathcal{W}_i$	The probability of detecting the zone Z_i
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 $y_p \in \{0, 1\},\ p \in \mathcal{P}$
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}	Flow quantity transmitted from zone $\mathcal{Z}_i$ to node p $f_{ip} \in \{0, 1\},\ i \in \mathcal{I},\ \ p \in \mathcal{Z}_i$

Table 2: Notations used in deployment models.

3.1 Model 1

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 lampposts and traffic lights as experimented in CitySense [9]. 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}. Our optimization models minimize the sensors and sinks overall deployment cost. Thus, the objective function is the following.

$$\mathcal{F} = \sum_{p \in \mathcal{P}} c_p^{sensor} * x_p + \sum_{p \in \mathcal{P}} c_p^{sink} * y_p \quad (4)$$

Since a sink embeds sensing capabilities, a sink and a sensor cannot be deployed at the same potential position p as formulated in constraint 5.

$$x_p + y_p \leq 1, \quad p \in \mathcal{P} \quad (5)$$

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.

Using an atmospheric dispersion model, we determine the set of generated pollution zones. Each zone $\mathcal{Z}_i$ corresponds to the geographical area, i.e. set of positions, where the pollution threshold is crossed when the pollution source i is releasing pollutants. Let the binary parameter $\mathcal{B}_{ip}$ denote whether a position p belongs to $\mathcal{Z}_i$ or not. A pollution zone $\mathcal{Z}_i$ is therefore the set $\{p \in \mathcal{P} \ where \ \mathcal{B}_{ip} = 1\}$. When using the pointwise Gaussian dispersion model, the value of $\mathcal{B}_{ip}$ is calculated using Formula (6) where σ_y, σ_z, Q and H are the parameters presented in Section 2, $p = (x, y, z)$ and C_0 is the threshold of pollutant concentration above which a point is considered as polluted.

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

We assume that a sensor exposed to a given pollutant will detect its concentration with a probability that is different from a pollution zone to another. We denote $\mathcal{W}_i \in]0, 1[$ the probability of detecting the pollution source i by a sensor located within its zone.

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

$$\prod_{p \in \mathcal{Z}_i} (1 - \mathcal{W}_i * (x_p + y_p)) \leq (1 - \beta), \quad i \in \mathcal{I} \quad (7)$$

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

Constraint 7 should be linearized in order to get an ILP formulation. The process of linearization is done through formulas 8, 9 and 10. Hence, we get the linear form in constraint 11.

$$\prod_{p\in\mathcal{Z}_i \text{ where } (x_p+y_p=1)} (1-\mathcal{W}_i) \leq (1-\beta), \quad i \in \mathcal{I} \tag{8}$$

$$(1-\mathcal{W}_i)^{(\sum_{p\in\mathcal{Z}_i}(x_p+y_p))} \leq (1-\beta), \quad i \in \mathcal{I} \tag{9}$$

$$(\sum_{p\in\mathcal{Z}_i}(x_p+y_p)) * log(1-\mathcal{W}_i) \leq log(1-\beta), \quad i \in \mathcal{I} \tag{10}$$

$$\sum_{p\in\mathcal{Z}_i}(x_p+y_p) \geq \frac{log(1-\beta)}{log(1-\mathcal{W}_i)}, \quad i \in \mathcal{I} \tag{11}$$

Connectivity We formulate in this first model the connectivity constraint as a network flow problem. 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.

$$\sum_{q\in\Gamma(p)} g_{pq} - \sum_{q\in\Gamma(p)} g_{qp} \geq x_p - \mathcal{N} * y_p, \quad p \in \mathcal{P} \tag{12}$$

$$\sum_{q\in\Gamma(p)} g_{pq} - \sum_{q\in\Gamma(p)} g_{qp} \leq x_p, \quad p \in \mathcal{P} \tag{13}$$

$$\sum_{q\in\Gamma(p)} g_{pq} \leq N * x_p, \quad p \in \mathcal{P} \tag{14}$$

$$\sum_{p\in\mathcal{P}}\sum_{q\in\Gamma(p)} g_{pq} = \sum_{p\in\mathcal{P}}\sum_{q\in\Gamma(p)} g_{qp} \tag{15}$$

Constraints 12 and 13 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 14, 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 5. Constraint 14 ensures also that deployed sinks cannot transmit flow units, and only act as receivers. Constraint 15 means that the overall flow is conservative. The flow sent by deployed sensors has to be received by deployed sinks.

ILP Model At the end, our first optimization model can be written as follows.

$$
\begin{aligned}
&[Model 1]\\
&\qquad\qquad Minimize\ (4)\\
&\qquad\qquad Subject\ to.\ (5), (11), (12), (13), (14) and\ (15)
\end{aligned}
$$

3.2 Model 2

Despite the huge progress made in solving ILPs, the first 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. The underlying writing of coverage is an instance of set cover, a very complex combinatorial problem that does not combine efficiently with the network flow problem of connectivity. In particular, solvers might not be able to use their computation optimization tailored for network flow. 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 this second model, each pollution source i should transmit some flow units to potential nodes p which are located within its pollution zone. 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} as the flow quantity from the pollution source i to the potential node p. The following constraints ensure coverage and connectivity for pollution monitoring.

$$\sum_{p \in \mathcal{P}} f_{ip} \geq \frac{log(1-\beta)}{log(1-\mathcal{W}_i)}, \quad i \in \mathcal{I} \tag{16}$$

$$\sum_{i \in \mathcal{I}: p \in \mathcal{Z}_i} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - g_{pq} \leq \mathcal{NM} y_p, \quad p \in \mathcal{P} \tag{17}$$

$$\sum_{i \in \mathcal{I}: p \in \mathcal{Z}_i} f_{ip} + \sum_{q \in \Gamma(p)} g_{qp} - g_{pq} \geq 0, \quad p \in \mathcal{P} \tag{18}$$

$$\sum_{q \in \Gamma(p)} g_{pq} <= \mathcal{NM} x_p, \quad p \in \mathcal{P} \tag{19}$$

$$f_{ip} <= x_p + y_p, \quad p \in \mathcal{P}, i \in \mathcal{I} \tag{20}$$

Coverage is formulated in constraint 16, which ensures that each pollution source i generates a sufficient number of flow units in the network. Constraint 20 enforces that all the flow units are received by deployed nodes. The flow is conservative on deployed sensors thanks to constraints 17 and 18. These two constraints combined with constraints 19 and 20 also ensure that absent nodes do not participate in the communication. Flow conservation on sensors ensures that the deployed sinks will receive all the flow units generated by the pollution sources. The second optimization model can then be written as follows.

$$\begin{aligned}
&[Model2] \\
&\qquad Minimize\ (4) \\
&\qquad Subject\ to.\ (5), (16), (17), (18), (19)\ and\ (20)
\end{aligned}$$

4 Simulation results

In this section, we present the simulations that we have performed to evaluate our deployment models. We first present the data set of Greater London that we used in simulation. Next, as a proof of concept we apply our models to the London Borough of Camden. Then, we investigate the performance of the two optimization models in terms of computational burden. Finally, we investigate some engineering insights while we study the impact of the height, the number and the coverage requirements of pollution sources on the deployment cost.

4.1 Greater London dataset

We evaluate our deployment models on a data set provided by the Greater London community [15]. London is one of the most polluted cities in Europe [16]. 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 distribution per borough depends on the surface of the borough and ranges from 6 sources to 161 sources.

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 3. We simulate the Gaussian dispersion model with the parameters depicted in table 4. Moreover, we define the nodes neighboring Γ based on a given transmission range. We assume that the cost of nodes is independent from the position of the node, i.e. $c_p^{sensor} = c^{sensor}$ and $c_p^{sink} = c^{sink}$.

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)	0.90
Detection sensitivity of sensors ($\mathcal{W}_i$)	0.80
Pollution threshold (C_0)	$20\mu g/m^3$

Table 3: Summary of default simulation parameters of the ILP models.

Parameter	**value**
h_s	$25m$
Q	$5g/s$
V	$1.9mm^3/s$
T_s	30^oC
D_w	$225°$
V_w	$5m/s$
T	7^oC
a_y	1.36
b_y	0.82
a_z	0.275
b_z	0.69

Table 4: Default simulation parameters of the Gaussian model

4.2 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 [13]. Camden is spread over an area of around $8km \times 6km$ and contains 19 pollution sources. Fig. 1 depicts the pollution zones obtained by running of the Gaussian dispersion model. Fig. 1 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.

(a) Pollution sources and zones

(b) WSN obtained by application of deployment models. Sensors are depicted in green and sinks in red.

Fig. 1: Application of our deployment models to the London Borough of Camden

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.

4.3 Computational comparison between the proposed models

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.

After executing the two models, we got the same objective values; this was expected since the second model is derived from the first one. We depict in table 5 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 notice that the instances that are more complex take more time to be resolved when using both of the two models. Moreover, the joint formulation allows to enhance the total mean execution time with a factor of around 8. This is due to the fact that in the second model, coverage and connectivity are modeled in joint way.

Area of interest (km^2)	**CPU time (seconds)**	
	Model 1	**Model 2**
$[0.00 - 0.20[$	7.460	0.890
$[0.20 - 0.45[$	20.400	2.810
$[0.45 - 0.70[$	29.830	3.360
$[0.70 - 0.95[$	68.200	8.820
Mean	**31.470**	**3.970**

Table 5: Model 1 VS Model 2.

4.4 Engineering insights

Impact of the height of pollution sources We now study the impact of the height of pollution sources on the deployment cost. We assume that nodes are placed on a height equal to $10m$, and all the pollution sources have the same height, which is considered in the range from 0 meters to 25 meters. We plot in Fig. 2 the sensors and sinks overall deployment cost depending on the height of pollution sources while considering two different values of the wind direction. The results are averaged over all the London blocks. On the 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 the hight of the sources 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 when pollution sources are $5m$ above sensors is less than the deployment cost when pollution sources are $5m$ below sensors. Fig. 2 also shows the impact of wind on the deployment cost which is different in the two considered cases. Indeed, weather conditions impact 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.

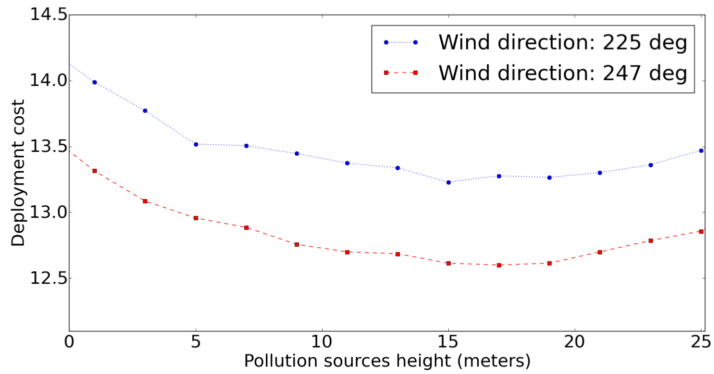

Fig. 2: Deployment cost average depending on pollution sources height while considering different wind directions.

Impact of the number of pollution sources 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 where the wind direction is equal to 225°. The results are 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. 3 the deployment cost variations depending on the pollution sources height while considering three different densities: 4, 5 and 6 pollution sources per instance. Fig. 3 shows that the more there are pollution sources in the environment, the 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.

Impact of the coverage requirements of pollution sources The coverage requirements of pollution sensors is one of the most important factors that affect the topology of sensor networks used for pollution monitoring. Fig. 4 depicts the average cost of the resulting deployments of the block instances while considering two values of the minimum requirement of coverage probability: $\beta = 0.90$ and $\beta = 0.98$. As expected, the deployment cost increases with the coverage requirements. We notice that the ratio between the two curves is around 1.1. This

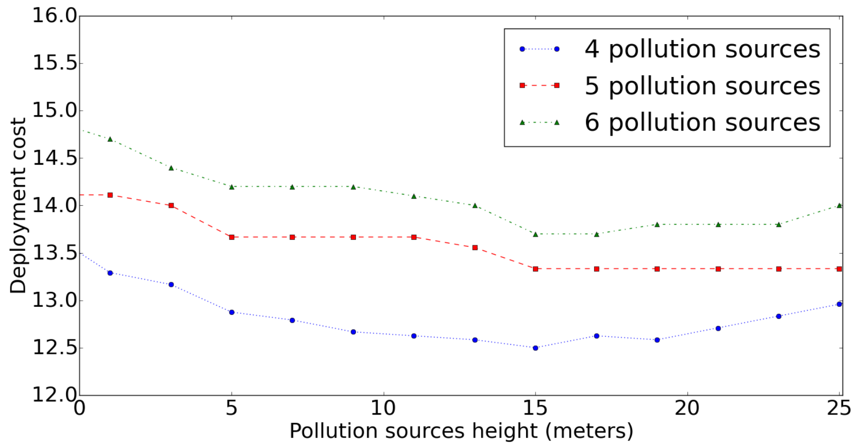

Fig. 3: Deployment cost average depending on the height and the density of pollution sources.

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.

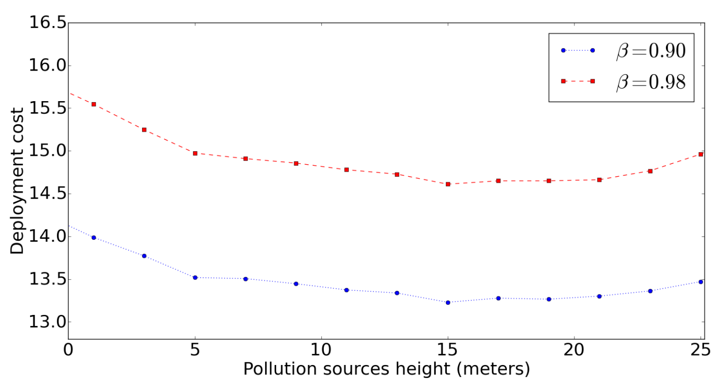

Fig. 4: Deployment cost average depending on the height and the coverage requirements of pollution sources.

5 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 paper, we focused on a new paradigm of pollution monitoring based on a set of interconnected and tiny sensor nodes. 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 do not take into account the pollution propagation, we based on atmospheric dispersion modeling to take into account the nature of the addressed phenomenon.

We investigated some engineering insights on the deployment of sensor nodes while 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 coverage requirements of pollution sources and have shown that the higher is this parameter, the higher is the deployment cost.

As a future work, we plan to consider the energy consumption in our models while maintaining their tractability. Another perspective would be to consider the impact of the nature of pollutants and the urban topography on the coverage results.

Acknowledgment

This work has been supported by the "LABEX IMU" (ANR-10-LABX-0088) and the "Programme Avenir Lyon Saint-Etienne" of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

References

1. AirParif: The air quality monitoring organization of the paris agglomeration, http://www.airparif.fr/en/index/index
2. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer networks 38(4), 393–422 (2002)
3. Cardei, M., Pervaiz, M.O., Cardei, I.: Energy-efficient range assignment in heterogeneous wireless sensor networks. In: ICWMC'06. International Conference on. pp. 11–11. IEEE (2006)
4. Daly, A., Zannetti, P.: Air pollution modeling–an overview. Ambient air pollution (2007)
5. 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: 2016-01-20
6. Ghosh, A., Das, S.K.: Coverage and connectivity issues in wireless sensor networks. Mobile, wireless, and sensor networks: Technology, applications, and future directions pp. 221–256 (2006)
7. Kadri, A., Yaacoub, E., Mushtaha, M., Abu-Dayya, A.: Wireless sensor network for real-time air pollution monitoring. In: ICCSPA, 2013 1st International Conference on. pp. 1–5. IEEE (2013)

8. Keskin, M.E., Altınel, İ.K., Aras, N., Ersoy, C.: Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility. Ad Hoc Networks 17, 18–36 (2014)
9. Murty, R.N., Mainland, G., Rose, I., Chowdhury, A.R., Gosain, A., Bers, J., Welsh, M.: Citysense: An urban-scale wireless sensor network and testbed. In: Technologies for Homeland Security, 2008 IEEE Conference on. pp. 583–588. IEEE (2008)
10. Rebai, M., Snoussi, H., Hnaien, F., Khoukhi, L., et al.: Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks. Computers & Operations Research 59, 11–21 (2015)
11. Sengupta, S., Das, S., Nasir, M., Panigrahi, B.K.: 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 26(1), 405–416 (2013)
12. Stockie, J.M.: The mathematics of atmospheric dispersion modeling. Siam Review 53(2), 349–372 (2011)
13. Camden Data: Camden lighting point, on http://www.camdendata.info/Pages/Home.aspx [2015-07-15]
14. 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]
15. London DataStore: London atmospheric emissions inventory 2010, on http://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2010 [2016-01-15]
16. 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]
17. 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]
18. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Computer networks 52(12), 2292–2330 (2008)
19. Younis, M., Akkaya, K.: Strategies and techniques for node placement in wireless sensor networks: A survey. Ad Hoc Networks 6(4), 621–655 (2008)
20. Zhu, C., Zheng, C., Shu, L., Han, G.: A survey on coverage and connectivity issues in wireless sensor networks. Journal of Network and Computer Applications 35(2), 619–632 (2012)