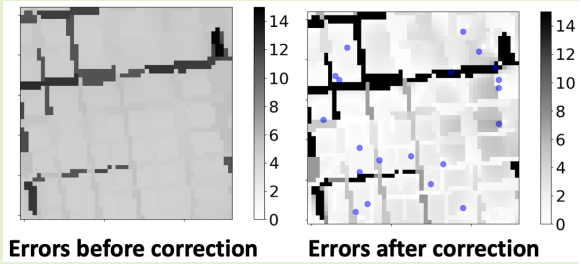
UrMapSens: Joint Mapping-Sensing Design of Wireless Sensor Networks for Urban Environmental Monitoring

Ahmed Boubrima, Walid Bechkit, Hervé Rivano, Lionel Soulhac

UrMapSens: Joint Mapping-Sensing Design of Wireless Sensor Networks for Urban Environmental Monitoring

Ahmed Boubrima, Walid Bechkit, Hervé Rivano, and Lionel Soulhac

***Abstract*— In this paper, we focus on the application of air quality monitoring where wireless sensor networks (WSNs) collect pollution measurements to estimate ground-truth pollution maps. We present UrMapSens, a novel WSN design strategy that aims to optimize the deployment locations of sensor nodes and the estimation of urban pollution maps. Unlike prior work, UrMapSens performs the optimization of both sensor deployment locations and pollution maps' estimation in a joint way. We design our UrMapSens approach based on data assimilation by combining physical models' simulations and expected sensor measurements in the estimation process of air pollution maps. We ensure a joint mapping-sensing design by determining the optimal number and locations of sensor nodes at the same time, in addition to finding the best way of assimilating the measurements that will be collected. Moreover, we use enhanced linearization techniques to provide a convex mathematical model based on mixed integer programming, which can be solved efficiently using numerical solvers. To validate UrMapSens in real-life scenarios, we designed an air pollution monitoring platform and used it to collect a dataset of nitrogen dioxide concentrations in Lyon, France. Based on the experimentally collected dataset, we perform extensive evaluations of the proposed approach and show that our joint mapping-sensing design outperforms the prior work in which the optimization of sensor deployment locations and pollution maps' estimation are performed independently.**



***Index Terms*— Wireless sensor networks, sensor deployment, air pollution monitoring, data assimilation, joint design, experimental validation.**

I. Introduction

AIR pollution is one of the major challenging issues in developed cities which, by trying to provide enhanced public services, highly increase urbanization and industrialization. The harmful effects of air pollution on human health and the environment have been widely established in several studies. This includes for instance major diseases like cancer and asthma, global climate change, acid rains, reduced outdoor visibility, etc. [2]. The first step in tackling air pollution is to ensure effective fine-grained monitoring of harmful pollutants. The main aim of air pollution monitoring, and mainly air pollution mapping, is usually to generate accurate pollution maps in a given area [3].

Traditionally, air pollution monitoring has been operated by environmental regional and governmental authorities, which use conventional measurement stations that are equipped with high-quality sensing technologies allowing them to get accurate measurements of air pollution concentrations. The traditional reference stations are however massive and very expensive to be deployed at the scale of the neighborhood. These limitations of traditional stations pushed industrials to build a more flexible solution through the use of tiny pollution sensors which are less accurate but smaller than the stations and most importantly highly cheaper [4]. Those sensors are designed to be deployed in a large number in a given city and usually communicate wirelessly forming, therefore, a wireless sensor network (WSN) [5].

Although sensor nodes are low-cost and can be deployed in a large scale, the deployment budget is usually limited and thus only the necessary number of nodes should be deployed. As a result, a deployment approach should be used to optimize the number and the locations of sensor nodes while ensuring reliable pollution monitoring [6]. The deployment optimization is a major challenge in WSN design [7], and consists in determining the optimal positions of sensors and sinks so as to cover the environment and ensure the network connectivity while optimizing an objective function such as the deployment cost or the network lifetime [8]. The network is said connected if each sensor can communicate information to at least one

This work is an extended version of an IEEE ICCCN'2018 paper [1] (url: https://ieeexplore.ieee.org/document/8487343), and was supported by the "LABEX IMU" (ANR-10-LABX-0088) of Université de Lyon (Corresponding author: Ahmed Boubrima).

Ahmed Boubrima, Walid Bechkit, and Hervé Rivano are with INRIA, INSA-Lyon, Université de Lyon, CITI-Lab, F-69621, Villeurbanne, France (emails: ahmed.boubrima@inria.fr, walid.bechkit@insa-lyon.fr, herve.rivano@insa-lyon.fr).

Lionel Soulhac is with INSA-Lyon, France (email: lionel.soulhac@ec-lyon.fr).

sink node.

In addition to measurement-based traditional monitoring, air pollution maps can be also obtained using physical models (also called atmospheric dispersion simulators) which simulate the phenomenon of air pollution dispersion. Physical models take as input the locations of pollution sources, the pollutant emission rate of each pollution source, and meteorological data (temperature, wind, etc.) in order to estimate the pollutant concentration at a given location [9].

In this paper, we propose UrMapSens, a novel WSN-based air pollution monitoring strategy that aims to optimize the deployment locations of sensor nodes and the estimation of pollution maps. We use in this paper the terms pollution sensing and pollution mapping to refer to the optimization of sensor deployment locations and pollution maps' estimation respectively. UrMapSens performs the optimization of both sensor deployment locations (i.e. pollution sensing) and pollution maps' estimation (i.e. pollution mapping) in a joint way in contrary to prior work that aims to optimize either sensing [10] [11] [12] or mapping [13] [14] [15] [16] [17] [18].

We design UrMapSens based on data assimilation by combining physical models' simulations and expected sensor measurements in the estimation process of air pollution maps. We ensure a joint mapping-sensing of air pollution monitoring by determining at the same time the optimal number and locations of sensor nodes, in addition to finding the best way of assimilating the measurements that will be collected. To that end, we define an appropriate mathematical formulation of pollution coverage quality in the context of air pollution mapping. Indeed, we formulate the quality of air pollution mapping of a given sensor network, depending on the assimilation error of pollution concentration at locations where no sensor is deployed. The assimilation error is defined as the difference between the unknown ground truth (or real) value of pollution concentration and the concentration obtained by applying an adequate data assimilation method on the measurements of sensor nodes and the outputs of physical models. We then use our formulation of air quality mapping combined with mathematical linearization techniques to define a deployment model using mixed integer linear programming (MILP).

In order to validate our proposed approach in real-life scenarios, we designed an air pollution monitoring platform and used it to collect an accurate dataset of nitrogen dioxide concentrations. The collected dataset allowed us to characterize in a realistic way the inputs of the data assimilation process, which are used in our deployment optimization approach.

This paper is an extension of our preliminary work previously published in [1]. We highlight below the main extensions and contributions compared to [1]

- In the early version of our work, the key idea was to use data assimilation in order to define an appropriate mathematical formulation of coverage quality in the context of air pollution mapping by formulating the quality of air pollution mapping of a given sensor network depending on the assimilation error of pollution concentration at locations where no sensor is deployed. Unlike our preliminary work, we propose in the journal version to optimize jointly and at the same time both the positioning of the sensors and the data assimilation parameters. That is, we determine at the same time the optimal number and locations of sensor nodes in addition to finding the best way of assimilating the measurements that will be collected at the selected locations.
- In order to validate our approach in real-life scenarios, and contrary to our preliminary work, we evaluate the proposed deployment model while considering a data-based characterization of the inputs of the optimization process. To that end, we designed an air pollution monitoring platform and used it to collect an accurate dataset of nitrogen dioxide pollution concentrations.
- We perform in this journal extension multiple evaluations in order to assess the performance of our joint optimization of WSN deployment and pollution data assimilation. Among our results, we show that the joint optimization process outperforms the standard optimization way that we previously proposed in our preliminary work in [1].

The remainder of this paper is organized as follows. We first review the related work on the deployment issue of WSN in section II. Then, we present in detail our mathematical formulation of coverage quality in section III. Next, we present our optimization model in section IV. We then present the evaluation data set and analyze the obtained results in sections V and VI before concluding our paper.

II. Related work

We review in what follows the main prior work on wireless sensor networks' deployment. Prior work, which started in the early 2000s, focuses on the optimal placement of sensor nodes so that the deployment region is covered and the connectivity between nodes is maintained while minimizing the deployment cost of the network or maximizing the network lifetime [8] [19].

Chakrabarty et al. were among the first authors to tackle the WSN deployment problem as they proposed in [20] [21] to represent the deployment region as a grid of points. They formulated coverage while assuming that each sensor has a circular detection area, which defines the points that a sensor can cover. In [22], Meguerdichian and Potkonjak proposed to leverage the analogy between coverage and the Set Cover Problem [23], an NP-complete optimization problem, in order to reduce the complexity of the method proposed by Chakrabarty et al. [20]. Both prior works assumed uniform coverage between all the target grid points as opposed to Altinel et al., who proposed in [24] to use integer linear programming in order to consider different coverage requirements among the sensor field points. This was followed by the works in [25] [26] that integrated data fusion in the coverage definition in order to take into account the sensors' collaborative detection of targets while relying on probabilistic sensing models.

More recent methods improved upon the aforementioned ones by integrating the communication constraints in the design of the proposed deployment methods as in [7] [27] [28] where authors used mixed integer programming to define the communication paths between the network nodes and

ensure the connectivity of the proposed solutions. Moreover, recently published works as in [29] [30] [31] [32] [33] [34] proposed dedicated heuristics, which provide near-optimal solutions in minimal time. Furthermore, authors in [35] [36] [37] address also the energy efficiency of the deployed nodes by maximizing, for instance, the number of cover sets while ensuring energy-aware connectivity.

The works mentioned so far target mainly the application of event detection where sensor networks are mostly used to detect harmful events (also called targets) that might occur in the environment. Nevertheless, they do not fit applications where sensors are characterized by inherent correlations and hence measure correlated values depending on their spatial and temporal location as in our application of regular air quality monitoring. Indeed, air pollution sensors need to come into contact with the pollutants in order to measure their concentrations.

Authors in [10] [11] [12] [38] [39] proposed to address the specific category of correlation-based sensor applications by first assuming the existence of a set of data snapshots characterizing the spatio-temporal variability of the phenomenon to monitor, and then formulating the problem of finding the best locations of sensors to reconstruct the data of the whole phenomenon. However, these existing works are generic and are not adapted to the application of regular air quality monitoring as they assume that the physical phenomenon to be monitored is either Gaussian or can be described with a stationary process. Indeed, the dispersion of pollutants in the air is more complicated than the Gaussian distribution and is usually handled using physical models that take into account the impact of weather conditions and pollution emissions [40].

In addition to being generic and not well adapted to the application of air pollution monitoring, existing WSN deployment methods are designed to optimize only sensor locations without optimizing the mapping of the phenomenon of interest. On the contrary, we propose and validate in the next sections a novel WSN-based air pollution monitoring strategy that aims to optimize the deployment locations of sensor nodes and the estimation of pollution maps in a joint way by incorporating the variables of the pollution estimation process in the formulation of the optimal sensor locations.

III. Coverage Quality Metric Formulation

In this section, we use data assimilation fundamentals in order to define the coverage metric of our optimization approach. Given a set of space locations with associated simulated pollution concentrations in addition to a limited number of measurements, we estimate air pollution concentrations by combining both the simulations and the sensor measurements. We focus on the Best Linear Unbiased Estimation ($BLUE$) method. The efficiency of $BLUE$ in providing accurate estimations compared to other simulations' correction methods has been demonstrated in many literature works [16] [41] [42]. The main notations used in this section are summarized in table I.

	Description	Dim.
$\boldsymbol{p}$	Space points	$l \times 1$
$\boldsymbol{g}$	*Unknown* ground truth vector	$l \times 1$
$\boldsymbol{s}$	Simulations' vector	$l \times 1$
$\boldsymbol{s^*}$	Unbiased simulations' vector	$l \times 1$
$\boldsymbol{z}$	Measurements' vector	$n \times 1$
$\boldsymbol{c}$	Corrected simulations' vector	$l \times 1$
H	Measurements-to-Space mapping matrix	$n \times l$
$\boldsymbol{\delta}$	Vector of random simulation errors	$l \times 1$
$\boldsymbol{\theta}$	Vector of random sensing errors	$n \times 1$
$\boldsymbol{\gamma}$	Mean vector of simulation errors	$l \times 1$
B	Covariance matrix of simulation errors	$l \times l$
F	Covariance matrix of corrected simulations' errors	$l \times l$
R	Covariance matrix of sensing errors	$n \times n$
W	Correction weights matrix	$l \times n$

TABLE I: Main notations

A. Inputs

Deployment region.: Let $\boldsymbol{p}$ be a vector of l discrete points approximating the space in 2D or 3D. **i.e.** $\boldsymbol{p} = [p_1, p_2, ..., p_l]^T$ where $p_i = (x_i, y_i, z_i)$. Without loss of generality, we assume that sensors can be deployed at any location within $\boldsymbol{p}$. We use $\mathbf{g} \in \mathbb{R}^l$ to denote the unknown ground truth pollution concentrations at the l points of space. **i.e.** $\mathbf{g} = [g_1, g_2, ..., g_l]^T$ where g_i is the pollution concentration at point i.

Sensor measurement locations.: Let $\boldsymbol{z} \in \mathbb{R}^n$ be the set of measurements performed using a set of sensors that are located at n different locations (n being much less than l) in space $\boldsymbol{p}$. **i.e.** $\boldsymbol{z} = [z_1, z_2, ..., z_n]^T$ where z_i is measurement number i. To map measurements to space locations, we define a matrix $H \in \mathbb{R}^{n \times l}$ where each matrix element h_{ij} is a Boolean set to 1 if measurement number i is performed at point j. Note that H is a matrix where the sum of each row is equal to 1.

Sensing errors.: We assume that each measurement i has a random error θ_i. With these definitions, $\boldsymbol{z}$ can be written as a function of $\boldsymbol{g}$ as in formula 1. In matrix form, this is equivalent to formula 2 where $\boldsymbol{\theta}$ is the vector of measurement errors.

$$z_i = \Big(\sum_{j \in [1,l]} h_{ij} \cdot g_j\Big) + \theta_i \tag{1}$$

$$\boldsymbol{z} = H\boldsymbol{g} + \boldsymbol{\theta} \tag{2}$$

We assume that θ_i has a zero mean and a variance equal to r_i. This is usually the case when pollution sensors are properly calibrated, in which case the variance of measurement errors is given by the manufacturer [43]. We also assume that measurement errors are uncorrelated because they mainly depend on the electronics of the sensing mechanism. Hence, the covariance matrix $R \in \mathbb{R}^{n \times n}$ of sensing errors is diagonal.

Physical models' simulations.: We denote by $\boldsymbol{s} \in \mathbb{R}^l$ the vector of simulated concentrations where s_i, the simulated value at point i, has an error δ_i with respect to the ground truth value g_i as in formula 3.

$$\boldsymbol{s} = \boldsymbol{g} + \boldsymbol{\delta} \tag{3}$$

Simulation errors can be of any distribution but we assume that their covariance matrix B and their mean vector $\boldsymbol{\gamma}$ are

known. We denote by $\boldsymbol{s}^*$ the vector of unbiased simulations as follows:

$$\boldsymbol{s}^* = \boldsymbol{s} - \boldsymbol{\gamma} \tag{4}$$

Note that Matrix B and vector $\boldsymbol{\gamma}$ are specific to a given atmospheric dispersion simulator. They can therefore be approximated using existing simulations together with measurements data collected from other deployments that exist in other regions of interest. Indeed, actual sensor data in addition to pollution simulation results can be used to provide a statistical sample of simulation errors and therefore approximate the mean error values in $\boldsymbol{\gamma}$ and covariances in B.

In case there are no existing deployments that can be leveraged, our approach can still be used in practice following a short pre-deployment phase where a few sensors are used to collect real-world data that is used to sample simulation errors and then approximate mean errors and covariances.

In order to approximate matrix B using the obtained sample of simulation errors, note that the covariance can be defined as a function of variances and spatial correlations when dealing with spatial phenomena such as air pollution [44]. Based on that, we use the obtained sample of pollution simulation errors in order to first calculate separately the variances and the correlations of simulation errors. We provide an example of this process in section V as we use sensor data obtained from a 4-month pre-deployment phase in Lyon, France to characterize the simulation errors of atmospheric pollution simulators.

B. Formulation of simulations' correction

Given an unbiased simulation vector $\boldsymbol{s}^*$, a measurement vector $\boldsymbol{z}$ and a matrix H defining measurement locations, the objective is to get an estimation vector $\boldsymbol{c} \in \mathbb{R}^l$ by combining both measurements and simulation outputs; i.e. $\boldsymbol{c} = f(\boldsymbol{s}^*, \boldsymbol{z})$. In the case of the Best Linear Unbiased Estimation (BLUE), $\boldsymbol{c}$ is defined as in formulas 5 and 6.

$$c_i = s_i^* + \sum_{j \in [1,n]} w_{ij} \cdot (z_j - \sum_{k \in [1,l]} h_{jk} \cdot s_k^*) \tag{5}$$

$$\boldsymbol{c} = \boldsymbol{s}^* + W(\boldsymbol{z} - H\boldsymbol{s}^*) \tag{6}$$

Note that c_i is therefore a correction of the simulated value s_i^* and this correction is obtained by subtracting a weighted combination of the drifts between measurements and simulations. The correction weights are defined by the matrix W. In order to minimize the average variance of corrected simulations' errors, the weights in matrix W are calculated using the formula 7 [45]. Note that in this case, W depends neither on the measurements $\boldsymbol{z}$ nor on the simulations $\boldsymbol{s}^*$, but is rather a function of simulation and measurement error covariances that are defined by matrices B and R respectively.

$$W = BH^T(R + HBH^T)^{-1} \tag{7}$$

Let η_i be the error of the corrected simulation with respect to the unknown ground truth value at point i defined as in formula 8. By replacing $\boldsymbol{c}$ using formula 6 and after simplification, we can write $\boldsymbol{\eta}$ as a function of random measurement errors $\boldsymbol{\theta}$ and random simulation errors $\boldsymbol{\delta}$ as in formula 9.

$$\boldsymbol{\eta} = \boldsymbol{c} - \boldsymbol{g} \tag{8}$$

$$\boldsymbol{\eta} = (\boldsymbol{\delta} - \boldsymbol{\gamma}) + W(\boldsymbol{\theta} - H(\boldsymbol{\delta} - \boldsymbol{\gamma})) \tag{9}$$

Note that the mean of corrected simulation errors $\boldsymbol{\eta}$ is 0 because measurement errors $\boldsymbol{\theta}$ are unbiased and $\boldsymbol{\gamma}$ is the mean of simulation errors $\boldsymbol{\delta}$. In this case, the covariance matrix F of corrected simulations' errors is a function of matrices W, B and H as demonstrated in [45] as:

$$F = (I_l - WH)B, \tag{10}$$

where I_l is the identity matrix. Formula (10) is equivalent to formula (11) after replacing W based on formula (7).

$$F = (I_l - BH^T(R + HBH^T)^{-1}H)B \tag{11}$$

C. Pollution coverage metric

We recall that H defines the locations of pollution sensors and hence a given sensor network topology. To take into account the quality of physical simulations' correction in the sensors' deployment process, we define our pollution coverage metric as the overall variance of corrected simulations' errors, which corresponds to the trace of matrix F. We use in the following section our pollution coverage metric to constrain the deployment optimization of sensor nodes to the requirements of the assimilation of pollution measurements in order to improve the quality of physical simulations.

IV. UrMapSens: Joint optimization of WSN deployment and simulations' correction

In this section, we propose a WSN deployment mixed integer programming model where the objective is to reduce air pollution simulation errors given a certain deployment budget. Sensor nodes should be therefore deployed in such a way to minimize the trace of the matrix F defined in the previous section. In addition to optimizing the location of sensor nodes, we also optimize the pollution simulations' correction parameters (i.e. the correction weights matrix W). Indeed, and in contrast to the data assimilation literature works which focus on only optimizing the correction weights W [45], we aim to determine jointly and at the same time the optimal positioning of sensor nodes in addition to finding the best way of assimilating the measurements that will be collected at the selected locations.

A. Optimization model: minimization of simulations' errors with budget and connectivity constraints

Using a fine characterization of air pollution simulation errors (matrix B and vector $\boldsymbol{\gamma}$) and sensing errors (matrix R), we aim to find the optimal sensor positions and assimilation weights that correct the simulations $\boldsymbol{s}$ in the best way while respecting a given deployment budget. In addition, we ensure that the deployed sensors form a connected network. That is,

each sensor is capable of communicating its collected data to the network sink node either directly or through neighboring sensor nodes.

1) Objective, coverage constraints and main variables: The main decision variables of the optimization model are the matrix H which defines the optimal sensor locations in addition to the matrix W which defines the parameters of the physical simulations' correction process. We define our objective function as in formula 12 in order to ensure the minimization of the overall variance of corrected simulations' errors $f_{ii},\ i \in [1,l]$.

Moreover, we derive from formulas 7 and 11 presented in the previous section the scalar form of simulation errors' variances f_{ii} as in constraints 13 and 14 where b_{ij}, h_{ij}, w_{ij}, and r_{ij} are the elements of matrices B, H, W, and R respectively.

Let parameters $cost_{ij}$ define the cost of deploying a node i at point j. Based on that, we define constraint 15 to ensure that the overall cost of nodes does not exceed the maximum allowed deployment budget, denoted using the input parameter $BUDGET$.

In addition to the main coverage decision variables (h_{ij}, w_{ij}, and f_{ii}), and in order to take into the network connectivity constraints that are discussed in the next subsection, we use extra decision variables y_i (denoting the optimal location of the network sink node) and g_{ij} (denoting the communication traffic between nodes i and j).

Objective: minimize $$\sum_{i\in[1,l]} f_{ii} \tag{12}$$

Subject to:

Pollution estimation constraints:

$$\forall i_1 \in [1,l], j_1 \in [1,n], \sum_{i_2\in[1,l]} b_{i_1 i_2}\cdot h_{j_1 i_2} = \sum_{j_2\in[1,n]} w_{i_1 j_2}\cdot\Big(\sum_{i_2\in[1,l]} r_{j_2 i_2} + \sum_{i_3\in[1,l]} h_{j_2 i_2}\cdot h_{j_1 i_3}\cdot b_{i_2 i_3}\Big) \tag{13}$$

$$\forall i \in [1,l],\ f_{ii} = b_{ii}\cdot\Big(1 - \sum_{j\in[1,n]} w_{ij}\cdot h_{ji}\Big) - \sum_{i'\in[1,l]\ |\ i'\neq i,\ j\in[1,n]} w_{ij}\cdot h_{ji'}\cdot b_{i'i} \tag{14}$$

Deployment budget constraint:

$$\sum_{i\in[1,n],\ j\in[1,l]} h_{ij}\cdot cost_{ij} \leq BUDGET \tag{15}$$

Auxiliary constraints:

$$\forall j \in [1,l], \sum_{i\in[1,n]} h_{ij} \leq 1 \tag{16}$$

$$\forall i \in [1,n], \sum_{j\in[1,l]} h_{ij} \leq 1 \tag{17}$$

Decision variables:

$$h_{ij}, y_i \in \{0,1\};\ w_{ij} \in [0,1];\ f_{ii}, g_{ij} \in \mathbb{R}^+$$

2) Connectivity constraints: We formulate the pollution sensor network connectivity as a flow problem. We consider the same potential positions set for both sensors and the network sink node while using binary variables y_i to define the optimal location of the sink node in the deployment region. We denote by $\Gamma(i),\ i \in [1,l]$, the set of neighbors of a node deployed at the potential position i. This set can be determined using sophisticated path loss models. It can also be determined using the binary disc model, in which case $\Gamma(i) = \{j \in [1,l]\ where\ j \in Disc(i,R)\}$ where R is the communication range of sensors. Then, we define the decision variables g_{ij} as the flow quantity transmitted from a node located at potential position i to another node located at potential position j.

In order to formulate the network connectivity constraints, we suppose that each sensor of the resulting WSN generates a flow unit in the network, and then we verify if these generated units can be recovered by sink nodes. Constraints 18 and 19 are designed to ensure that each deployed sensor generates a flow unit in the network. In addition, constraint 20 is designed to ensure that absent nodes (i.e. corresponding to locations with no deployed sensors) do not participate in the communication process. Finally, Constraint 21 ensures that the overall flow is conservative, i.e. the flow sent by the deployed sensors has to be received by the deployed sink node.

Connectivity constraints:

$$\sum_{j\in\Gamma(i)} g_{ij} - \sum_{j\in\Gamma(i)} g_{ji} \geq \sum_{k\in[1,n]} h_{ki} - (l+1)\cdot y_i, i \in [1,l] \tag{18}$$

$$\sum_{j\in\Gamma(p)} g_{ij} - \sum_{j\in\Gamma(p)} g_{ji} \leq \sum_{k\in[1,n]} h_{ki}, \quad i \in [1,l] \tag{19}$$

$$\sum_{j\in\Gamma(i)} g_{ij} \leq l\cdot \sum_{k\in[1,n]} h_{ki}, \quad i \in [1,l] \tag{20}$$

$$\sum_{i\in[1,l]}\sum_{j\in\Gamma(i)} g_{ij} = \sum_{i\in[1,l]}\sum_{j\in\Gamma(i)} g_{ji} \tag{21}$$

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

- In Fig. 1, the dotted lines are not valid links because the sensors that are involved in those lines are not close enough. Therefore, flow units are not sent over those links thanks to constraints 18, 19 and 20.
- We force all the sensors to send each a flow unit. Note that some nodes send their flow unit directly to the sink node whereas the others go through their neighbors. Also note that the flow is conservative on each node, for instance, a node that receives a unit from a neighbor has to send to the sink two units: its unit plus the one of its neighbor.

- Finally, note that the overall flow is conservative thanks to constraint 21: the sink node receives all the 5 units generated by the 5 sensor nodes.

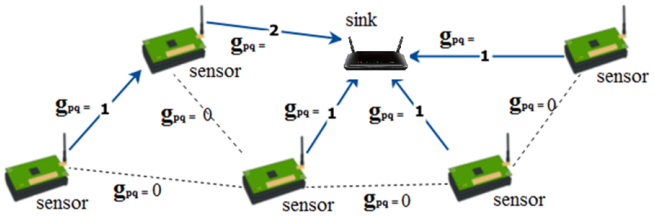


Fig. 1: Illustrative example of the flow concept.

3) Linearization of the optimization model: Due to the product of decision variables h and w in constraints 13 and 14, the optimization model is not convex and hence cannot be solved efficiently with state-of-the-art optimization solvers. To address that, we use linearization techniques [46] that allow us to transform the variables' products into linear constraints and therefore get a convex and mixed integer linear programming model (MILP), which can be solved efficiently in practice [47].

We linearize each product of two binary variables $h_{ab} \cdot h_{cd}$ by replacing it with a new variable v_{abcd} that is defined using the following constraints:

$$\begin{aligned} & v_{abcd} \leq h_{ab} \\ & v_{abcd} \leq h_{cd} \\ & v_{abcd} \geq h_{ab} + h_{cd} - 1 \end{aligned} \tag{22}$$

Furthermore, we linearize each product of a real variable and a binary variable $w_{ab} \cdot h_{cd}$ by replacing it with a new variable o_{abcd} that is defined as in the following constraints:

$$\begin{aligned} & o_{abcd} \leq w_{ab} \\ & o_{abcd} \leq h_{cd} \\ & o_{abcd} \geq w_{ab} - (1 - h_{cd}) \cdot \\ & o_{abcd} \geq 0 \end{aligned} \tag{23}$$

B. Dual model: minimization of the deployment cost with coverage constraints

In addition to the optimization model previously presented in this section, we also consider in this paper the case of the dual problem where the objective is to minimize the deployment cost while ensuring that the corrected simulation errors' overall variance does not exceed a given threshold. In this case, the objective function to minimize is defined using the budget formulation in formula 15. In addition, we define a new pollution coverage constraint in order to restrict the trace of matrix F to the threshold input value of corrected simulation errors' variance.

C. Resolution of the optimization model

The proposed optimization model is based on mixed integer linear programming that can be solved using exact MILP solvers. In terms of complexity, the execution time of the MILP solvers increases exponentially with the size of the problem. Indeed, the complexity of MILP models is mainly due to the number of binary variables which causes an exponential increase in the number of iterations when using the exact MILP solvers. Based on this fact, we use the MILP form of the proposed formulation only to solve small and medium-sized instances. For large problem instances, we design a heuristic algorithm that allows us to reduce the execution time of the deployment approach while providing near-optimal deployment solutions.

Tuning the exact MILP solvers while using an input integrality gap value is a common technique to get feasible solutions of the MILP models within a reasonable execution time. This integrality gap value defines the quality gap between the theoretical optimal solution and the current solution of the MILP solver during its execution time.

In order to solve our optimization model on large instances in a reasonable time while getting near-optimal solutions, we propose in this paper to leverage the well-known concept of linear relaxation (Algorithm 1). Linear relaxation is an approximation technique that allows us to reduce the complexity of our designed integer programming model by relaxing the integrality constraints of binary decision variables h_{ij} and y_j, thus resulting in a linear programming model that can be solved efficiently using commercial solvers. The fractional solutions of the resulting linear model are then rounded to either 0 or 1 in order to obtain the desired solutions of the original integer programming model [48].

Algorithm 1 Heuristic algorithm

Inputs: $\boldsymbol{p}$
Outputs: $\{h_{ij}\}, \{y_j\}$

repeat

Solve the assimilation-based LP deployment model
Let q be the maximum fractional variable among h_{ij} and y_j variables
Add constraint $q = 1$ to the LP model

until all the variables are binary

Algorithm: We first define the linear programming model LP while considering the same objective function and constraints as our initial assimilation-based deployment model and relaxing all the binary variables h and y. Note that since binary variables are considered in the range of $[0, 1]$, the solutions of the LP model are not necessarily binary. Moreover, note that in a given LP solution where deployment variables h and y are fractional, the variable having the maximum value (i.e. the closest binary variable to 1) corresponds to the most important node in the satisfaction of coverage and connectivity constraints. Based on this fact, we propose in each iteration of our heuristic algorithm to deploy a sensor at point $\bar{j}$ where $h_{i\bar{j}}$ is the closet variable to 1 or to deploy the sink node at point $\bar{j}$ if $y_{\bar{j}}$ is the closest variable to 1. The loop, which performs

iterative rounding after running the optimization model in each iteration, stops once the deployment variables are equal to either 0 or 1, and all the coverage and connectivity constraints are ensured.

V. Experimental characterization of simulation errors

The main input of the WSN deployment model that we propose in this paper is the covariance matrix of air pollution simulation errors, denoted B, which is the main parameter in our pollution coverage metric formulation presented in formula 11. To validate our approach in real-life scenarios, we aim in this section to experimentally characterize the covariance matrix B. To that end, we collect a neighborhood-scale air pollution dataset using a custom-built platform while focusing on the nitrogen dioxide pollutant.

Moreover, and without loss of generality, we focus in the evaluation part of this paper on the case of pollution simulations that are generated by Sirane [40] [49] [50], which is an enhanced atmospheric dispersion simulator that is designed for urban areas and takes into account the impact of street canyons on the pollution dispersion phenomenon.

In the following, we first present our air pollution monitoring platform. Then, we discuss the context of the collected dataset and finally, we present the characterization process of the covariance matrix of air pollution simulation errors.

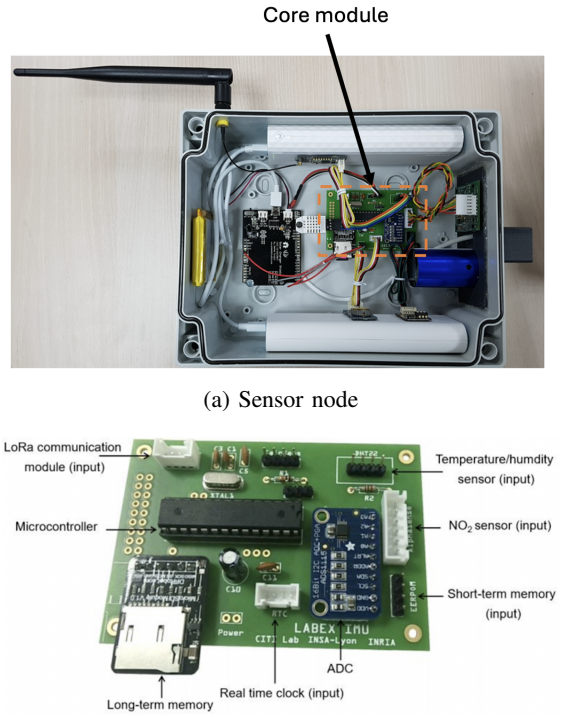
A. UrPolSens: a WSN-based pollution sensing platform

In order to collect accurate pollution measurements at the neighborhood scale, we designed a low-cost WSN-based air quality monitoring platform while focusing on the energy consumption of nodes compared to existing platforms. In our platform, which is named UrPolSens and illustrated in Fig. 2, sensor nodes measure the nitrogen dioxide pollutant (NO2) at a $1Hz$ frequency (using $B4$ electrochemical sensing probes that are developed by Alphasense Inc.; each NO2 probe is connected to a 16-bit $ads1115$ ADC running at gain 4). In addition, we also measure temperature and humidity using $DHT22$ probes. NO2, humidity, and temperature sensing probes are connected to an ATMEGA328P-PU, which is an 8-bit low-power microcontroller operating on 5V.

In Fig. 3, we provide a block diagram of our lab-designed sensor nodes. Collected data is timestamped every second using a real-time clock module before being stored in an EEPROM short-term memory. Then, for every period of 10 minutes, the content of the EEPROM is transferred into an SD card. Using the EEPROM memory allows us to minimize the power consumption of the SD module, which is turned on only for a few milliseconds every period of 10 minutes.

Sensors also transmit averaged collected data to a gateway every period of 10 minutes using low-power long-range (LoRa) communication modules (Semtech SX1276). The gateway node (which is connected to a power outlet) forwards the sensors' data to the decision center using a cellular 4G connection.

In terms of power consumption, and using the aforementioned sensing and communication periods, our sensor nodes consume on average $75mW$ (which corresponds to $15mA$ on $5V$). Each node is powered using 2 USB batteries (totaling a $30000mAh$ capacity at $3.7V$) allowing it to run for over 2 consecutive months thanks to the low-power hardware components used in our design.



(a) Sensor node

(b) Core module

Fig. 2: An illustration of our lab-designed sensor nodes (a) and the designed core module part (b).

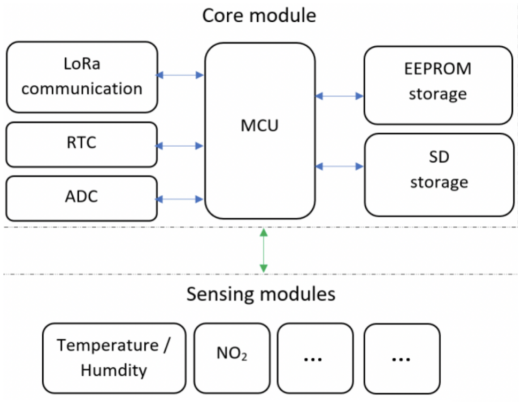


Fig. 3: A block diagram of our sensor nodes.

B. Measurement campaign and collected dataset

We deployed a set of 12 UrPolSens sensors in the center of Lyon, France from July to October 2018. The deployment locations are depicted in Fig. 4, which highlights the 350m x 150m deployment area. In order to ensure the accuracy of the collected measurements, our sensor nodes were calibrated based on reference sensors (blue cylinders in Fig. 2, which are

$Cairsens$ sensors provided by Cairpol Inc.) during 3 consecutive days before the main measurement campaign. Thanks to this proper calibration, we concluded that our sensors have less than 10% error compared to reference sensors.

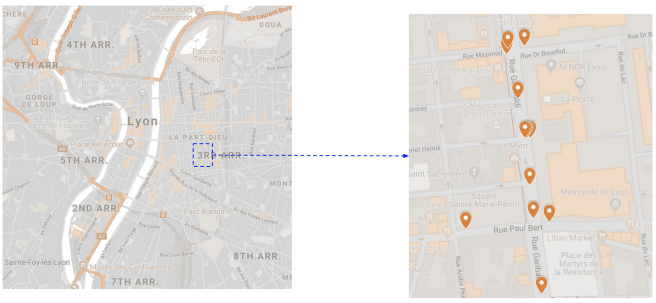


Fig. 4: Deployment area of the UrPolSens platform.

C. Characterization of simulation errors

In order to evaluate our deployment approach in a real-life scenario, and without loss of generality, we approximate the B matrix of nitrogen dioxide pollution simulations that are generated by Sirane, and this by building an empirical covariance function using our collected measurements in addition to typical simulation outputs of Sirane. Mainly, we use monthly pollution simulations that correspond to the same 4 months during which our data was collected (July - October 2018). We therefore obtain a set of 48 monthly simulation values (4 monthly values at each of our 12 sensing locations) in addition to the initial set of 48 monthly measurements. Then, by subtracting simulations and measurements, we obtain a dataset of real-world sensing errors of pollution simulations.

The covariance can be defined as a function of variances and spatial correlations when dealing with spatial phenomena such as air pollution [44]. We therefore use the aforementioned real-world dataset of pollution simulation errors in order to first characterize separately the variance and the correlation functions of these errors before deriving their covariance function (matrix B).

Characterization of the variance. We plot the variance of simulation errors as a function of Sirane's simulation values. The corresponding results are depicted in Fig. 5 and show that the higher the simulated value, the higher the variance of Sirane's errors. This results in a good linear fit with 95% accuracy.

Characterization of the correlations. In addition to the variance function characterization, we also characterize the correlations between simulation errors at each 2 spatial locations using our pollution dataset. Mainly, we characterize spatial correlations using the Euclidean distance between sensing locations, and this is based on the fact that pollution is a spatially distributed phenomenon [51]. We depict the corresponding results in Fig. 6, which shows that our spatial correlations decrease as the distance between locations increases. In addition, we show that using a Gaussian fit, we achieve 90% accuracy.

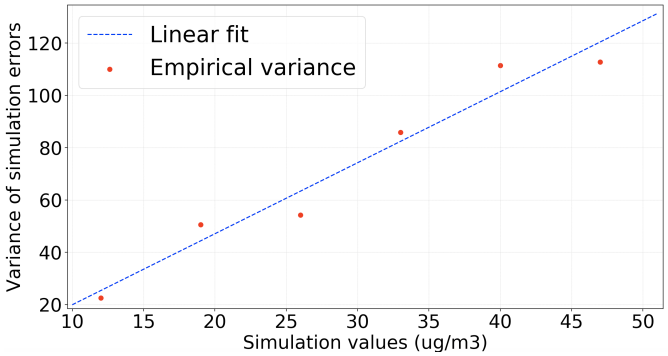


Fig. 5: Characterization of the variance function of Sirane's simulation errors.

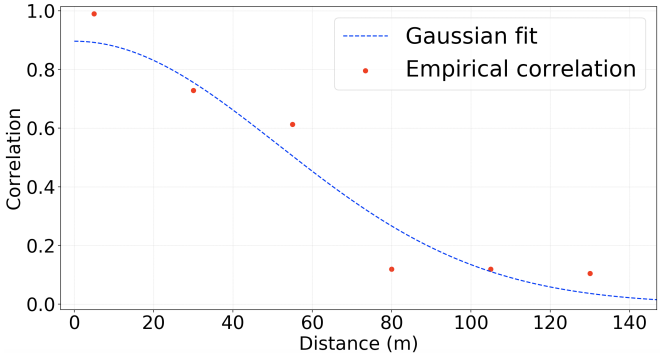


Fig. 6: Characterization of the spatial correlation function of Sirane's simulation errors.

VI. Evaluation Results

Our proposed optimization deployment model is a sensor deployment strategy that ensures -by mathematical construction- that the selected deployment locations of sensor nodes are the optimal ones for improving air pollution simulations. That is, given a set of air pollution simulations, our deployment model is formulated to determine the best locations where sensors should be placed in order to reduce the errors of air pollution simulations.

In this section, we present the evaluations that we have performed in order to assess the performance of our proposed WSN deployment model. We consider as a potential deployment region (space area denoted $\boldsymbol{p}$) an area of 500m x 500m located in Lyon, France. We discretize our deployment area while considering a $30m$ spatial granularity, which allows us to consider 289 potential deployment locations of sensor nodes.

We run the empirical variance function (linear fit in Fig. 5) to build a variance map of simulation errors on top of our potential deployment region and depict the output in Fig. 7. We then use this variance map together with our characterization of spatial correlations (Gaussian fit in Fig. 6) in order to build the B matrix corresponding to the potential deployment region.

In terms of nodes' characteristics, we define matrix R by setting the default variance of pollution sensing errors to $1[\mu g]^2/m^6$ (provided by the nitrogen dioxide sensing cell manufacturer). We also set the default communication range of nodes to $100m$.

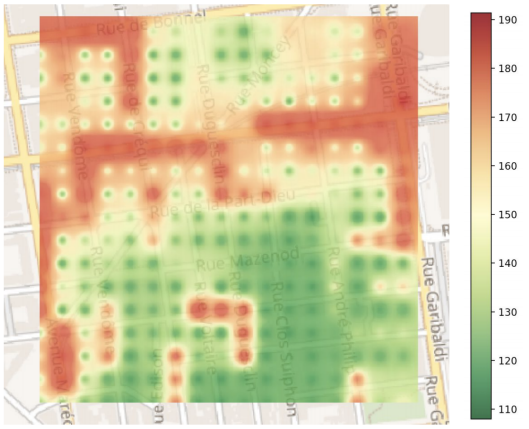


Fig. 7: Variance map of the deployment region in $[\mu g]^2/m^6$.

In the following, we first provide a proof-of-concept in order to explain how the WSN deployment process is performed. Then, we evaluate our pollution coverage efficiency which is based on a joint optimization of sensor positioning and simulations' correction while providing a comparison to prior work. Finally, we assess the impact of pollution estimation requirements on network connectivity.

A. Proof of concept

In order to provide a proof of concept of our joint WSN deployment and air pollution simulations' correction, we run our optimization model to minimize the deployment budget while ensuring the network connectivity and an average assimilation errors' variance that is below a threshold equal to $100[ug]^2/m^6$. We report the selected deployment locations in Fig. 8 where sensors are depicted using blue circles and relay nodes (i.e. nodes deployed to maintain the network connectivity) are depicted using orange circles.

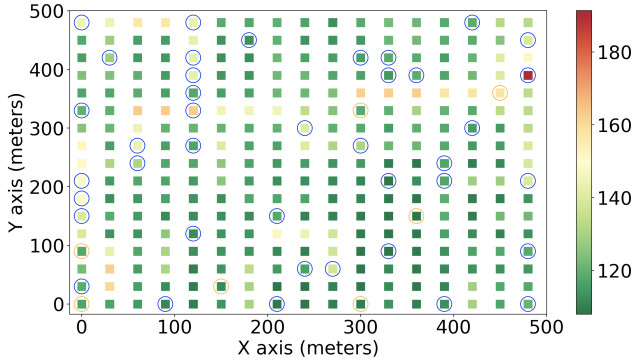


Fig. 8: Optimal deployment locations on top of the variance map of simulation errors. Blue circles denote sensors and orange circles denote relay nodes.

Fig. 8 shows that 40 sensors and 7 relay nodes are deployed in total in order to ensure that the average variance of corrected simulations is less than $100[ug]^2/m^6$. Mainly, most sensors are positioned at locations where the variance of simulation errors is very high (greater than $150[ug]^2/m^6$). This is because the optimized correction error's variance is a function of simulation errors' variance as already defined in formula 11. Moreover, in high variance areas, nodes appear to be deployed right next to each other (i.e. high density). This is indeed due to the fact that the correlation between space locations drops considerably beyond $50m$ according to our experimental characterization in Fig. 6 (we recall that the spatial resolution of the space grid in Fig. 8 is 30m).

B. Analysis of coverage results

We now evaluate the performance of our proposed deployment model in terms of the accuracy of corrected simulations that can be obtained using our optimal deployment. We vary the deployment budget and report in Fig. 9 the improvement factor of the corrected simulations compared to Sirane's simulated values (i.e. ratio of the average variance of initial simulation errors and corrected simulation errors). Moreover, we consider two different cases in terms of sensing accuracy: (ii) low sensing errors where the sensing errors' variance is set to $1[ug]^2/m^6$; (ii) and mid-quality sensing errors where the sensing error's variance is set to $25[ug]^2/m^6$.

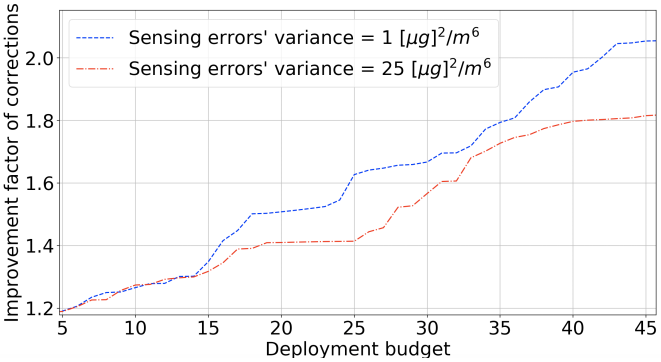


Fig. 9: Performance improvement factor of our optimal deployment model depending on the deployment budget and sensing errors.

We first notice that the higher the number of nodes, the higher the performance improvement. This is indeed expected as deploying more sensors leads to the correction of the simulations of more neighboring locations. Interestingly, the results show that less than 50 sensors (out of 289 potential locations) can reduce the variance of the errors of the Sirane simulator by a factor of 2.

Figure 9 also shows that the performance improvement achieved by our optimal deployment depends on the accuracy of the sensing nodes. For instance, using only 45 nodes with low sensing errors allows us to achieve more than a 2 improvement factor whereas the performance improvement of the same number of mid-quality sensors is around 1.8. Nevertheless, the difference in terms of performance improvement between both cases remains overall low, especially when only a few sensor nodes are available to deploy.

C. Comparison to existing works

To analyze the performance benefit of our joint optimization of WSN deployment and pollution simulations' correction, we

provide now a comparison between UrMapSens and existing deployment methods where pollution mapping estimation parameters are not optimized. We use as a baseline the optimization method presented in [1] and [19] where sensing positions are optimized without optimizing pollution data assimilation. We refer to the latter method by standard deployment optimization and report in Fig. 10 the obtained results in terms of corrected simulation errors' average variance.

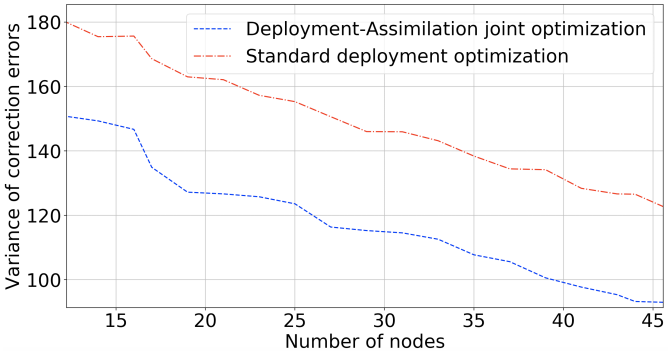


Fig. 10: Our joint optimization of WSN deployment and simulations' correction compared to standard WSN deployment.

Results show that overall, we achieve on average 25% errors' variance reduction compared to the standard WSN deployment model whether the number of deployed nodes is high or low. This is because the parameters of simulation correction (matrix W) are highly connected to the optimal locations of sensor nodes. As a result, a joint optimization of both outputs allows the best simulation correction performance.

D. Analysis of connectivity results

Finally, we evaluate the impact of the connectivity range on the use of the deployment budget and the quality of data assimilation. While assuming that nodes are equipped with long-range-capable radio modules (LoRa), we consider two different scenarios depending on the communication range of nodes: (i) a communication range equal to 100m; (ii) and a longer communication range equal to 500m. We vary the deployment budget and depict the resulting variance of assimilation errors in Fig. 11.

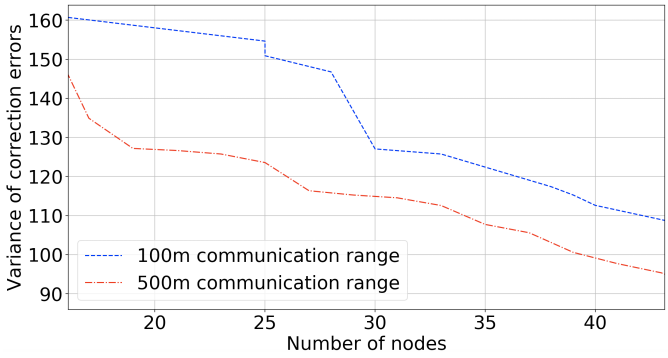


Fig. 11: Impact of the deployment budget on coverage results depending on nodes' communication range.

Results show that using long-range communications leads to less assimilation error when compared to short-range communications for the same value of the deployment budget. This is explained by the relay nodes that are added just to ensure connectivity in the case of short-range communications, which means that the deployment optimization of some nodes is performed to improve connectivity but not necessarily coverage.

Fig. 11 also shows that when the deployment budget increases, the difference between the two communication technologies decreases. This is due to the deployment of coverage nodes becoming so dense that the network is usually almost connected even when using short-range communications.

VII. Discussion: Time Synchronization in Pollution Sensor Networks

Large-scale air pollution monitoring using low-cost wireless pollution sensors requires efficient time synchronization in order to ensure that pollution packets are accurately ordered once they arrive at the sink node. Indeed, pollution assimilation techniques, in general, and our proposed UrMapSens approach, in particular, rely on data fusion to combine sensor data and physical simulations in the pollution mapping process. We assume that the construction of a pollution map at a given time relies on measurements and simulations that correspond also to the same and exact timestamp. As a result, when sensor clocks are not synchronized, the pollution mapping results obtained using our proposed optimization approach are not guaranteed to be optimal.

Fortunately, the specifics of time-synchronization in resource-constrained WSN networks have been recently investigated in the literature. Among the potential synchronization approaches that fit our UrMapSens design is the BATS scheme (Beaconless Asymmetric energy-efficient Time Synchronization) [52]. Indeed, and adequately with the UrMapSens assumptions, BATS relies on the sink nodes to carry out the computational overhead of the time synchronization process while assuming that sinks are equipped with powerful processors and are not energy-constrained. Running the synchronization procedure on sink nodes allows therefore to (i) reduce the energy consumption of resource-constrained sensor nodes and (ii) also reduce the synchronization errors that can result from using the limited hardware of sensors. Furthermore, and in contrast to traditional flooding-based synchronization schemes where nodes broadcast beacon messages, BATS reduces the communication overhead by relying instead on reverse one-way message dissemination. This is achieved by reversing the direction of synchronization messages from $sink \rightarrow sensors$ to $sensors \rightarrow sink$. As a result, the BATS scheme highly reduces the energy consumption that is due to the time synchronization process.

VIII. Conclusion

Wireless sensor networks (WSN) are widely used in environmental applications where the aim is to sense a physical phenomenon such as temperature, humidity, air pollution, etc. Air pollution is one of the main physical phenomena that still

need to be studied and characterized because it highly depends on other phenomena such as temperature and wind variations.

In this paper, we tackle the deployment problem of wireless pollution sensor networks and propose an application-aware approach where both the positioning of pollution sensors and the data assimilation parameters are optimized in a joint way using mixed integer linear programming. In order to experimentally validate our proposed approach, we applied it to a real-world pollution dataset collected using custom-built nitrogen dioxide sensors. This allowed us to assess the impact of the optimization inputs on the results of our joint optimization of WSN deployment and pollution data assimilation. As a perspective, we plan to evaluate our approach using other datasets that will be collected within different urban areas.

Acknowledgment

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

References

[1] Ahmed Boubrima, Walid Bechkit, Hervé Rivano, and Lionel Soulhac. "Leveraging the Potential of WSN for an Efficient Correction of Air Pollution Fine-Grained Simulations". In: *Computer Communication and Networks (ICCCN), 2018 27th International Conference on*. IEEE. 2018.

[2] Bernardo Beckerman, Michael Jerrett, Jeffrey R Brook, Dave K Verma, Muhammad A Arain, and Murray M Finkelstein. "Correlation of nitrogen dioxide with other traffic pollutants near a major expressway". In: *Atmospheric Environment* 42.2 (2008), pp. 275–290.

[3] Ali Marjovi, Adrian Arfire, and Alcherio Martinoli. "High resolution air pollution maps in urban environments using mobile sensor networks". In: *Distributed Computing in Sensor Systems (DCOSS), 2015 International Conference on*. IEEE. 2015, pp. 11–20.

[4] Ahmed Boubrima, Zhambyl Shaikhanov, and Edward W Knightly. "Toward Accurate Environmental Mapping using Balloon-based UAVs". In: *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*. IEEE. 2024, pp. 135–142.

[5] Kasimir Aula, Eemil Lagerspetz, Petteri Nurmi, and Sasu Tarkoma. "Evaluation of Low-Cost Air Quality Sensor Calibration Models". In: *ACM Transactions on Sensor Networks* 18.4 (2022), pp. 1–32.

[6] Erik Blasch, Tien Pham, Chee-Yee Chong, Wolfgang Koch, Henry Leung, Dave Braines, and Tarek Abdelzaher. "Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges". In: *IEEE Aerospace and Electronic Systems Magazine* 36.7 (2021), pp. 80–93.

[7] M Emre Keskin, İ Kuban Altınel, Necati Aras, and Cem Ersoy. "Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility". In: *Ad Hoc Networks* 17 (2014), pp. 18–36.

[8] Mohammed Farsi, Mostafa A Elhosseini, Mahmoud Badawy, Hesham Arafat Ali, and Hanaa Zain Eldin. "Deployment techniques in wireless sensor networks, coverage and connectivity: A survey". In: *Ieee Access* 7 (2019), pp. 28940–28954.

[9] Michael Jerrett, Altaf Arain, Pavlos Kanaroglou, Bernardo Beckerman, Dimitri Potoglou, Talar Sahsuvaroglu, Jason Morrison, and Chris Giovis. "A review and evaluation of intraurban air pollution exposure models". In: *Journal of Exposure Science and Environmental Epidemiology* 15.2 (2005), pp. 185–204.

[10] Zhiwen Hu, Zixuan Bai, Kaigui Bian, Tao Wang, and Lingyang Song. "Real-time fine-grained air quality sensing networks in smart city: design, implementation, and optimization". In: *IEEE Internet of Things Journal* 6.5 (2019), pp. 7526–7542.

[11] Venkat Roy, Andrea Simonetto, and Geert Leus. "Spatio-temporal sensor management for environmental field estimation". In: *Signal Processing* 128 (2016), pp. 369–381.

[12] Keigo Yamada, Yuji Saito, Koki Nankai, Taku Nonomura, Keisuke Asai, and Daisuke Tsubakino. "Fast greedy optimization of sensor selection in measurement with correlated noise". In: *Mechanical Systems and Signal Processing* 158 (2021), p. 107619.

[13] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, Tabita Arn, Jan Beutel, and Lothar Thiele. "Deriving high-resolution urban air pollution maps using mobile sensor nodes". In: *Pervasive and Mobile Computing* 16 (2015), pp. 268–285.

[14] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, and Lothar Thiele. "Pushing the spatio-temporal resolution limit of urban air pollution maps". In: *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE. 2014, pp. 69–77.

[15] Yiwen Hu, Zengliang Zang, Dan Chen, Xiaoyan Ma, Yanfei Liang, Wei You, Xiaobin Pan, Liqiong Wang, Daichun Wang, and Zhendong Zhang. "Optimization and evaluation of SO2 emissions based on WRF-Chem and 3DVAR data assimilation". In: *Remote Sensing* 14.1 (2022), p. 220.

[16] Anne Tilloy, Vivien Mallet, David Poulet, Céline Pesin, and Fabien Brocheton. "BLUE-based NO2 data assimilation at urban scale". In: *Journal of Geophysical Research: Atmospheres* 118.4 (2013), pp. 2031–2040.

[17] Stefan Tsokov, Milena Lazarova, and Adelina Aleksieva-Petrova. "A hybrid spatiotemporal deep model based on CNN and LSTM for air pollution prediction". In: *Sustainability* 14.9 (2022), p. 5104.

[18] Jia Xing, Siwei Li, Dian Ding, James T Kelly, Shuxiao Wang, Carey Jang, Yun Zhu, and Jiming Hao. "Data assimilation of ambient concentrations of multiple air pollutants using an emission-concentration response modeling framework". In: *Atmosphere* 11.12 (2020), p. 1289.

[19] Ahmed Boubrima. "Deployment and scheduling of wireless sensor networks for air pollution monitoring". PhD thesis. Lyon, 2019.

[20] Krishnendu Chakrabarty, S Sitharama Iyengar, Hairong Qi, and Eungchun Cho. "Coding theory framework for target location in distributed sensor networks". In: *Information Technology: Coding and Computing, 2001. Proceedings. International Conference on*. IEEE. 2001, pp. 130–134.

[21] Krishnendu Chakrabarty, S Sitharama Iyengar, Hairong Qi, and Eungchun Cho. "Grid coverage for surveillance and target location in distributed sensor networks". In: *Computers, IEEE Transactions on* 51.12 (2002), pp. 1448–1453.

[22] Seapahn Meguerdichian and Miodrag Potkonjak. *Low power 0/1 coverage and scheduling techniques in sensor networks*. Tech. rep. UCLA Technical Reports 030001, 2003.

[23] Vasek Chvatal. "A greedy heuristic for the set-covering problem". In: *Mathematics of operations research* 4.3 (1979), pp. 233–235.

[24] İ Kuban Altınel, Necati Aras, Evren Güney, and Cem Ersoy. "Binary integer programming formulation and heuristics for differentiated coverage in heterogeneous sensor networks". In: *Computer Networks* 52.12 (2008), pp. 2419–2431.

[25] Xiangmao Chang, Rui Tan, Guoliang Xing, Zhaohui Yuan, Chenyang Lu, Yixin Chen, and Yixian Yang. "Sensor placement algorithms for fusion-based surveillance networks". In: *IEEE Transactions on Parallel and Distributed Systems* 22.8 (2011), pp. 1407–1414.

[26] Zhaohui Yuan, Rui Tan, Guoliang Xing, Chenyang Lu, Yixin Chen, and Jianping Wang. "Fast sensor placement algorithms for fusion-based target detection". In: *Real-Time Systems Symposium, 2008*. IEEE. 2008, pp. 103–112.

[27] Maher Rebai, Hichem Snoussi, Faicel Hnaien, Lyes Khoukhi, et al. "Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks". In: *Computers & Operations Research* 59 (2015), pp. 11–21.

[28] Soumyadip Sengupta, Swagatam Das, MD Nasir, and Bijaya K Panigrahi. "Multi-objective node deployment in WSNs: In search of an optimal trade-off among coverage, lifetime, energy consumption, and connectivity". In: *Engineering Applications of Artificial Intelligence* 26.1 (2013), pp. 405–416.

[29] Jun Guo and Hamid Jafarkhani. "Movement-efficient sensor deployment in wireless sensor networks with limited communication range". In: *IEEE Transactions on Wireless Communications* 18.7 (2019), pp. 3469–3484.

[30] Nguyen Thi Hanh, Huynh Thi Thanh Binh, Nguyen Xuan Hoai, and Marimuthu Swami Palaniswami. "An efficient genetic algorithm for maximizing area coverage in wireless sensor networks". In: *Information Sciences* 488 (2019), pp. 58–75.

[31] Chunfeng Liu, Zhao Zhao, Wenyu Qu, Tie Qiu, and Arun Kumar Sangaiah. "A distributed node deployment algorithm for underwater wireless sensor networks based on virtual forces". In: *Journal of Systems Architecture* 97 (2019), pp. 9–19.

[32] Hanaa ZainEldin, Mahmoud Badawy, Mostafa Elhosseini, Hesham Arafat, and Ajith Abraham. "An improved dynamic deployment technique based-on genetic algorithm (IDDT-GA) for maximizing coverage in wireless sensor networks". In: *Journal of Ambient Intelligence and Humanized Computing* 11.10 (2020), pp. 4177–4194.

[33] Huynh Thi Thanh Binh, Nguyen Thi Hanh, Nguyen Phuc Tan, La Van Quan, Dang The Ngoc, Nguyen Huu Nhat Minh, and Huynh Cong Phap. "A heuristic node placement strategy for extending network lifetime and ensuring target coverage in mobile wireless sensor networks". In: *Evolutionary Intelligence* (2024), pp. 1–18.

[34] Meng Xie, Dechang Pi, Chenglong Dai, and Yue Xu. "A metaheuristic-based algorithm for optimizing node deployment in wireless sensor network". In: *Neural Computing and Applications* (2024), pp. 1–23.

[35] Saeed Karimi-Bidhendi, Jun Guo, and Hamid Jafarkhani. "Energy-efficient deployment in static and mobile heterogeneous multi-hop wireless sensor networks". In: *IEEE Transactions on Wireless Communications* 21.7 (2021), pp. 4973–4988.

[36] Sanjai Prasada Rao Banoth, Praveen Kumar Donta, and Tarachand Amgoth. "Target-aware distributed coverage and connectivity algorithm for wireless sensor networks". In: *Wireless Networks* 29.4 (2023), pp. 1815–1830.

[37] Li Cao, Zihui Wang, Zihao Wang, Xiangkun Wang, and Yinggao Yue. "An energy-saving and efficient deployment strategy for heterogeneous wireless sensor networks based on improved seagull optimization algorithm". In: *Biomimetics* 8.2 (2023), p. 231.

[38] Jie Hao, Baoxian Zhang, Zhenzhen Jiao, and Shiwen Mao. "Adaptive compressive sensing based sample scheduling mechanism for wireless sensor networks". In: *Pervasive and Mobile Computing* 22 (2015), pp. 113–125.

[39] Xiaopei Wu and Mingyan Liu. "In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing". In: *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM. 2012, pp. 1–12.

[40] Lionel Soulhac, Chi Vuong Nguyen, P Volta, and P Salizzoni. "The model SIRANE for atmospheric urban pollutant dispersion. PART III: Validation against NO2 yearly concentration measurements in a large urban agglomeration". In: *Atmospheric environment* 167 (2017), pp. 377–388.

[41] Bruce Denby, Martijn Schaap, Arjo Segers, Peter Builtjes, and Jan Horálek. "Comparison of two data assimilation methods for assessing PM10 exceedances on the

European scale". In: *Atmospheric Environment* 42.30 (2008), pp. 7122–7134.

[42] Philipp Schneider, Nuria Castell, Matthias Vogt, Franck R Dauge, William A Lahoz, and Alena Bartonova. "Mapping urban air quality in near real-time using observations from low-cost sensors and model information". In: *Environment international* 106 (2017), pp. 234–247.

[43] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. "Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide". In: *Sensors and Actuators B: Chemical* 215 (2015), pp. 249–257.

[44] William Revelle. *An introduction to psychometric theory with applications in R*. 2009.

[45] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.

[46] Warren P Adams and Hanif D Sherali. "Linearization strategies for a class of zero-one mixed integer programming problems". In: *Operations Research* 38.2 (1990), pp. 217–226.

[47] Ahmed Boubrima, Walid Bechkit, and Hervé Rivano. "Optimal WSN Deployment Models for Air Pollution Monitoring". In: *IEEE Transactions on Wireless Communications* 16.5 (2017), pp. 2723–2735.

[48] Tomas Werner. "Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.8 (2009), pp. 1474–1488.

[49] Lionel Soulhac, Pietro Salizzoni, F-X Cierco, and Richard Perkins. "The model SIRANE for atmospheric urban pollutant dispersion; part I, presentation of the model". In: *Atmospheric environment* 45.39 (2011), pp. 7379–7395.

[50] Lionel Soulhac, Pietro Salizzoni, Patrick Mejean, D Didier, and I Rios. "The model SIRANE for atmospheric urban pollutant dispersion; part II, validation of the model on a real case study". In: *Atmospheric Environment* 49 (2012), pp. 320–337.

[51] Sang-Il Lee. "Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I". In: *Journal of geographical systems* 3.4 (2001), pp. 369–385.

[52] Xintao Huan, Kyeong Soo Kim, Sanghyuk Lee, Eng Gee Lim, and Alan Marshall. "A beaconless asymmetric energy-efficient time synchronization scheme for resource-constrained multi-hop wireless sensor networks". In: *IEEE Transactions on Communications* 68.3 (2019), pp. 1716–1730.

Ahmed Boubrima is a Tenured Faculty Researcher at INRIA in France. He graduated from ESI-Algiers, Algeria in 2015 and got his PhD from INSA-Lyon, France in 2019. During his PhD, he was also a visiting research student at the University of Ottawa, Canada. Before his current position, he was the Lead Postdoc of the ASTRO project at Rice University (Texas, USA), where he worked on autonomous data-driven mobile sensing via networked drones. Ahmed's research interests are focused on sensor and drone networks with an emphasis on wireless and environmental sensing.

Walid Bechkit is an Associate Professor at the Telecommunications Department, INSA-Lyon, France, and is also a member of the INRIA Agora Team, CITI Laboratory. He received an Engineering degree in computer science from the École Nationale Supérieure d'Informatique, Algiers, in 2009, and a Ph.D. degree in system and information technology from the Compiegne University of Technology (UTC), France, in 2012. His main research interests include wireless sensor networks, the IoT for smart city applications, air quality monitoring, data analysis, machine and deep learning, LPWAN technologies, and security.

Hervé Rivano is a Full Professor at INSA-Lyon, France, and the Head of the INRIA Agora Team, which focuses on the use of wireless networks in smart cities. He graduated from the Ecole Normale Supérieure de Lyon, France, and received the Ph.D. degree from the University of Nice-Sophia Antipolis, in November 2003. He was an INRIA and CNRS Researcher, from October 2004 to August 2017. His research interests include combinatorial optimization and approximation algorithms applied to network design and provisioning.

Lionel Soulhac is a Full Professor at INSA-Lyon, France. His research activity mainly focuses on pollutant dispersion modeling in urban areas. He led different projects in this field such as the development of the two models SIRANE and SIRANERISK, which are dedicated to air quality assessment after stationary or non-stationary pollutant releases.