In situ calibration of cross-sensitive sensors in mobile sensor arrays using fast informed non-negative matrix factorization

Olivier Vu Thanh, Matthieu Puigt, Farouk Yahaya, Gilles Delmaire, Gilles Roussel

IN SITU CALIBRATION OF CROSS-SENSITIVE SENSORS IN MOBILE SENSOR ARRAYS USING FAST INFORMED NON-NEGATIVE MATRIX FACTORIZATION

Olivier Vu thanh[⋆†]*, Matthieu Puigt*[⋆]*, Farouk Yahaya*[⋆]*, Gilles Delmaire*[⋆]*, and Gilles Roussel*[⋆]

[⋆] Univ. Littoral Côte d'Opale, LISIC – EA 4491, F-62228 Calais, France
[†] University of Mons, Belgium

ABSTRACT

In this paper, we assume a set of mobile geolocalized sensor arrays observing an area over time. Each of these arrays consists of heterogeneous and cross-sensitive sensors, *i.e.,* the sensor readings provided by one of such sensors also depends on the readings of the other sensors in the array. We further assume that such arrays are possibly-uncalibrated and we aim to propose an *in situ* calibration method—*i.e.*, a data-driven technique—for such arrays. The novelty of this paper is twofold: we first revisit *in situ* calibration of mobile cross-sensitive sensors as an informed factorization of a partially observed non-negative matrix. A fast informed (semi-)NMF approach is then proposed and found to be well-suited for the considered problem.

***Index Terms*—** Sensor calibration, Mobile sensor array network, Informed non-negative matrix factorization, Missing values, Expectation maximization, Nesterov accelerated gradient

1. INTRODUCTION

Air quality is recognized as a public health issue and is usually monitored using a few highly reliable sensors deployed over a desired area. However, such sensors are costly, bulky, and often static. Hence, their drawbacks are a low spatial sampling rate and their inability to sense many local phenomena. The recent emergence of low-cost mobile sensors is a promising change of paradigm concerning sensor networks [1]. Adding these mobile sensors in a static network is currently a highly investigated scheme to increase the quantity of measurements and the spatial coverage. However, their low-cost aspect is an issue when it comes to the quality of accuracy. In [2], comparisons with reliable monitoring sensors indicate that low-cost sensors (LCS) are not ready for high accuracy usage, in particular because of both calibration and lack of repeatability issues. Precisely mapping the output delivered by a sensor to the measured physical quantity consists of performing a calibration. Predeployment calibration is an option to reduce the error of LCS but—as LCS cannot maintain a stable response over time [3]—a regular post-deployment calibration seems mandatory to maintain the LCS accuracy. Since regularly calibrating numerous LCS in a laboratory is not a reasonable solution, several *in situ* calibration methods have been proposed [4, 5]. They can be grouped into three main families, *i.e.*, macro-calibration, micro-calibration, and transfer calibration. Macro-calibration consists of uniformizing the measurements over the whole sensor network. Micro-calibration consists of exploiting the spatio-temporal vicinity between a reference—or a recently calibrated sensor—and an uncalibrated sensor in order to calibrate the latter. Transfer calibration consists of using a master sensor and non-master sensors: the raw data of the latter are standardized according to those of the former through (robust) linear regression. When the master sensor is calibrated, *e.g.*, by training a nonlinear calibration model to fit with the target phenomenon, it transfers its calibration to the standardized non-master sensors. For more details about the different calibration methods, we refer the reader to [4, 5].

Studies with LCS indicate that they are sensitive to their environment and that they can be cross-sensitive. The latter means that the target measurement can also depend on one or several physical variables that are not intended to be measured. For instance in [6] the NO_2 sensors are also sensitive to the concentration of O_3. A common solution consists of deploying sensor arrays—composed for instance of both a NO_2 and a O_3 sensors—and considering the concentration of O_3 in the calibration model of the NO_2 sensor (and vice versa).

Independently from the above studies, in [7–9], in situ calibration of non-cross-sensitive sensors with affine or polynomial sensor responses were revisited as informed (semi-) Non-negative Matrix Factorization (NMF). These methods consider a micro-calibration hypothesis in a macro-calibration scheme to free itself from any structural assumptions in the sensor network. Such a strategy was found to be particularly interesting according to some recent surveys [4] but it is not able to calibrate heterogeneous sensor arrays. In this paper, we propose to extend the formalism of [7] to sensor arrays. We also design a new fast informed (Semi-)NMF method to solve in situ calibration of cross-sensitive sensors.

The remainder of the paper reads as follows. Section 2 introduces the sensor array calibration problem. We then propose an *in situ* calibration method in Section 3 and we investigate its performance in Section 4. We lastly conclude and discuss about future work in Section 5.

2. PROBLEM STATEMENT, DEFINITIONS AND ASSUMPTIONS

In this paper, we assume that a set of m mobile geolocalized and time-stamped sensor arrays are observing an area over time. Each of these arrays consists of p sensors which are cross-sensitive [6], *i.e.*, the output—denoted x^k hereafter—of a given sensor, say Sensor k, depends on various physical phenomena—denoted $g_1, \ldots, g_p$—according to

$$x^k \approx f_0^k + \sum_{i=1}^{p} f_i^k \cdot g_i, \tag{1}$$

where f_0^k represents the offset parameter and f_i^k represents the i-th gain parameter of the k-th sensor of the sensor array, g_i represents the i-th sensed physical variable. Please note that if we assume that

This work was partially funded by the ULCO research pole "Mutations Technologiques et Environnementales" and by the Région Hauts-de-France.

$\forall i \neq k, f_i^k = 0$, this model reduces to a simpler affine calibration model as proposed in [7]. We also assume to get the outputs of fixed reference instruments sensing the same phenomena as the mobile sensors. Following the same strategy as in [7], we model these fixed sensor arrays as the $(m+1)$-th one of the network. We lastly assume that all the sensor arrays are able to send their geolocalized and time-stamped sensor readings to a unique trusted server, which is a common strategy in environmental mobile crowdsensing [9, 10].

We now introduce some definitions needed to explain the proposed approach.

Definition 1 ([11]). *A **rendez-vous** is a spatio-temporal vicinity between two sensors.*

A *rendez-vous* is thus characterized by a distance Δd and a duration ΔT. When two sensors make a *rendez-vous*, the fluctuations of the phenomenon between two locations closer than Δd during a time interval of duration ΔT are negligible. However, both highly depend on the sensed physical phenomenon [11]. As an example, if one observes the variations of temperature and of carbon monoxyde concentrations, the values of Δd and ΔT for the latter will be much lower than for the former [11].

As a consequence, in our considered network, each sensor of the array is associated with its own *rendez-vous* parameters, denoted Δd_k and ΔT_k for Sensor k. We then introduce the definition of a sensor array rendezvous.

Definition 2. *Two sensor arrays make a **rendez-vous** if $\forall k \in [\![1, p]\!]$, their respective k-th sensors make a rendez-vous.*

In practice, two sensor arrays thus make a *rendez-vous* if their distance is below[1] $\Delta d = \min_{1 \leq k \leq p} \Delta d_k$ and the duration between their measurements is below $\Delta T = \min_{1 \leq k \leq p} \Delta T_k$.

A *rendez-vous* can occur between a mobile sensor (respectively, a mobile sensor array) and a fixed or a mobile sensor (respectively, a fixed or a mobile sensor array). Multi-hop calibration for instance considers recently calibrated mobile sensors as virtual references. When an uncalibrated sensor makes a *rendez-vous* with a virtual reference, it is calibrated according to this virtual reference and then becomes a virtual reference for another uncalibrated sensor, hence the risk of error propagation along the multi-hop path. On the contrary, the model in [7] takes advantage of every kind of *rendez-vous*, whether the sensors are mobile or not, whether the sensors are uncalibrated or not. Since our proposed model is an extension of [7], we need to also define a *scene*.

Definition 3 ([7]). *A **scene** $\mathcal{S}$ is a discretized area observed during a time interval of duration ΔT. The size of the spatial pixels is set so that any couple of points inside the same pixel has a distance below Δd.*

According to Definition 3, sensors in *rendez-vous* are located in the same cell of a *scene*. Fig. 1 shows a simple example of a scene. As it is spatially sampled, the spatial samples can be stacked to form an observed matrix $\mathbf{X}^k$, related to the k-th sensor of the arrays. Each row of $\mathbf{X}^k$ represents a spatial pixel of the scene $\mathcal{S}$ as seen by each Sensor k of the different arrays. Let us recall that all the different measurements in a scene are performed during a time interval ΔT. During this duration, the mobile sensors are free to move. This means that the same sensor can provide some readings in several areas in a scene, as seen in Fig. 1. Each column of $\mathbf{X}^k$ then contains

[1] Please note that it may be possible to relax such a constraint, as proposed in [9] for non-cross-sensitive sensor calibration.

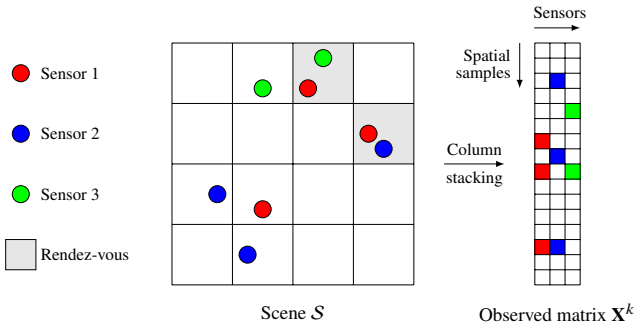

Fig. 1. A scene $\mathcal{S}$ measured by 3 sensors and its associated observed matrix $\mathbf{X}^k$. The sensors in a grey square are on a *rendez-vous*. By construction, the sensors on the same line of $\mathbf{X}^k$ make a *rendez-vous*.

all the measurements performed by a sensor. As an arbitrary choice, the last column of $\mathbf{X}^k$ always contains the measurements done by the reference sensors. In fact, since the reference sensors are static, they are modeled as only one mobile reference sensor which can only make some measurements where the fixed references are located. This explains why several reference sensors are characterized by only one column—*i.e.*, the last one—of $\mathbf{X}^k$.

Let us now assume that all the arrays are providing readings in all the spatial pixels of a scene. In particular, for the k-th sensor, combining Eq. (1) with the above scene formulation leads to

$$\mathbf{X}^k_{\text{theo}} \approx \mathbf{G} \cdot \mathbf{F}^k \tag{2}$$

with

$$\mathbf{X}^k_{\text{theo}} = \begin{pmatrix} x^k_{1,1} & \dots & x^k_{1,m} & g_{1,k} \\ \vdots & & \vdots & \vdots \\ x^k_{n,1} & \dots & x^k_{n,m} & g_{n,k} \end{pmatrix}, \tag{3}$$

where $x^k_{i,j}$ denotes the sensor reading provided by Sensor k of Array j in Pixel i of the *scene*,

$$\mathbf{G} = \begin{pmatrix} \mathbb{1}_n & \mathbf{g}_1 & \dots & \mathbf{g}_p \end{pmatrix}, \tag{4}$$

where $\mathbf{g}_k = \begin{pmatrix} g_{1,k} & \dots & g_{n,k} \end{pmatrix}^T$ is the vector of the k-th physical phenomenon sensed in the n different spatial pixels of the scene, $\mathbb{1}_n$ is a $n \times 1$ vector of ones, and

$$\mathbf{F}^k = \begin{pmatrix} f^k_{0,1} & \dots & f^k_{0,m} & 0 \\ f^k_{1,1} & \dots & f^k_{1,m} & \delta^k_1 \\ \vdots & & \vdots & \vdots \\ f^k_{p,1} & \dots & f^k_{p,m} & \delta^k_p \end{pmatrix}, \tag{5}$$

where $f^k_{i,j}$ denotes the i-th calibration parameter of Sensor k of Array j, and where $\delta^k_l = 1$ if $l = k$, 0 otherwise. $\mathbf{F}^k$ contains the calibration parameters of every sensor measuring the k-th physical variable among the p physical variables that are considered in Eq. (1). $\mathbf{G}$ contains the physical variables considered in Eq. (1) in every spatial pixel, and a column of ones for the offset. Each column of $\mathbf{F}^k$ provides the calibration parameters of the k-th sensor of one of the $m+1$ sensor arrays. By construction, as the last column of $\mathbf{X}^k_{\text{theo}}$ represents the measurements done by the reference sensors, the last column of $\mathbf{F}^k$ represents the calibration parameters of these sensors. Since we can trust the calibration process of the reference sensors, we can assume that their output values are directly the physical variables they are supposed to measure, hence the last column of $\mathbf{X}^k_{\text{theo}}$

being equal to $\mathbf{g}_k$. This implies the absence of an offset and the presence of Kronecker functions δ_l^k in the last column of $\mathbf{F}^k$.

Equation (2) alone means that only the desired physical variable is measured by the mobile sensors. Considering sensor arrays instead—that are constituted of sensors measuring every physical variable in Eq. (1)—a general relationship between the cross-sensitive sensors of the different sensor arrays can be obtained by concatenating the observed matrices $\mathbf{X}_{\text{theo}}^k$, *i.e.*,

$$\mathbf{X}_{\text{theo}} = \begin{pmatrix} \mathbf{X}_{\text{theo}}^1 & \dots & \mathbf{X}_{\text{theo}}^p \end{pmatrix}. \tag{6}$$

Combining Eqs. (2) and (6) provides

$$\mathbf{X}_{\text{theo}} \approx \mathbf{G} \cdot \mathbf{F}, \tag{7}$$

where $\mathbf{G}$ is defined in (4) and

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^1 & \dots & \mathbf{F}^p \end{pmatrix}. \tag{8}$$

At this stage, Eq. (7) is a structured matrix factorization problem where $\mathbf{X}_{\text{theo}}$ is non-negative—as it contains voltage outputs and/or concentrations—$\mathbf{G}$ is non-negative, as it contains contration values and a column of ones. Finally, if the calibration parameters in $\mathbf{F}$ are positive—which is not a far fetched hypothesis[2]—then Eq. (7) is in fact a specific NMF problem. Solving it is equivalent to performing sensor array calibration since $\mathbf{F}$ contains all the calibration parameters of the sensor arrays. The quality of the calibration is then implied by the error of estimation upon the matrix $\mathbf{F}$ itself. In a real case, $\mathbf{X}_{\text{theo}}$ is not accessible. Only its noisy projection $\mathbf{X}$ on the observation space Ω_X is observed, through the operator $\mathcal{P}_{\Omega_\mathbf{X}}$, *i.e.*,

$$\mathbf{X} \approx \mathcal{P}_{\Omega_\mathbf{X}}(\mathbf{X}_{\text{theo}}). \tag{9}$$

In practice, each sensor is associated with its own confidence measure, denoted ρ_j^k for Sensor k of Array j. Combining Eqs. (9) and (7) yields an NMF problem with missing entries, *i.e.*,

$$\min \frac{1}{2} \|\mathbf{W} \circ (\mathbf{X} - \mathbf{G} \cdot \mathbf{F})\|_\mathcal{F}^2, \tag{10}$$

where $\circ$ denotes the Hadamard product and $\mathbf{W}$ is defined such that its (i, j)-th entry—with $k \cdot (m+1) < i < (k+1) \cdot (m+1)$—reads

$$w_{i,j} = \begin{cases} \rho_j^k & \text{if the } (i,j)\text{-th value of } \mathbf{X} \text{ is measured,} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The above definition of $\mathbf{W}$ allows to take into account some difference of confidence in the accuracy of the sensed data. If this confidence is the same for all observed sensor readings, then $\mathbf{W}$ can be replaced by a binary matrix. Without additional constraints, solving NMF—whether there are some missing values or not—can lead to multiple solutions [12], because of the well-known scale and permutation ambiguities. In the considered context of sensor calibration, such ambiguities are not allowed since they might degrade the calibration accuracy. In our model, these ambiguities are cleared up by the imposed structure on $\mathbf{G}$ and $\mathbf{F}$. To take them into account in the factorization, we use the same parameterization as in [7, 13]: we rewrite both $\mathbf{G}$ and $\mathbf{F}$ as the sum of their free and fixed parts, *i.e.*,

$$\mathbf{G} = \Omega_\mathbf{G} \circ \Phi_\mathbf{G} + \overline{\Omega}_\mathbf{G} \circ \Delta_\mathbf{G}, \quad \mathbf{F} = \Omega_\mathbf{F} \circ \Phi_\mathbf{F} + \overline{\Omega}_\mathbf{F} \circ \Delta_\mathbf{F}, \tag{12}$$

where $\Omega_\mathbf{G}$ and $\Omega_\mathbf{F}$ (respectively $\overline{\Omega}_\mathbf{G}$ and $\overline{\Omega}_\mathbf{F}$) are the binary matrices informing of the presence (respectively the absence) of constraints upon $\mathbf{G}$ and $\mathbf{F}$, and $\Phi_\mathbf{G}$ and $\Phi_\mathbf{F}$ (respectively $\Delta_\mathbf{G}$ and $\Delta_\mathbf{F}$) are the matrices containing the constrained values (respectively the free values) of $\mathbf{G}$ and $\mathbf{F}$. In what follows, to simplify the notations, let $\Phi_\mathbf{G} = \Omega_\mathbf{G} \circ \Phi_\mathbf{G}$, $\Phi_\mathbf{F} = \Omega_\mathbf{F} \circ \Phi_\mathbf{F}$, $\Delta_\mathbf{G} = \overline{\Omega}_\mathbf{G} \circ \Delta_\mathbf{G}$ and $\Delta_\mathbf{F} = \overline{\Omega}_\mathbf{F} \circ \Delta_\mathbf{F}$. Finally, sensor array calibration leads to solve

$$\begin{aligned} (\tilde{\mathbf{G}}, \tilde{\mathbf{F}}) = \underset{\mathbf{G} \geq 0, \mathbf{F} \geq 0}{\arg\min} \quad & \frac{1}{2} \|\mathbf{W} \circ (\mathbf{X} - \mathbf{G} \cdot \mathbf{F})\|_\mathcal{F}^2 \\ \text{subject to} \quad & \mathbf{G} = \Phi_\mathbf{F} + \Delta_\mathbf{G}, \\ & \mathbf{F} = \Phi_\mathbf{F} + \Delta_\mathbf{F}. \end{aligned} \tag{13}$$

It should be noticed that Eq. (13) is similar to the cost function proposed in [7], except that the dimensions of the matrices are here much larger than those in [7]. This implies that even if the calibration method proposed in [7]—named IN-Cal in [9]—is found to be accurate for the considered application, it may not be applicable in the configuration we here consider, because of the increased matrix sizes and of the slow convergence rate of its update rules. As a consequence, we propose a fast extension of IN-Cal—named F-IN-Cal in the remainder of the paper—to solve Eq. (13).

3. PROPOSED CALIBRATION METHOD

To solve NMF with missing entries, different strategies have been proposed in the literature, *i.e.*, minimizing the cost function by using only the observed data—either by considering $\mathbf{W}$ in the cost function [14] or by using a stochastic gradient descent in the case of binary weights [15]—or through an Expectation-Maximization (EM) framework [16]. The latter consists of a two-stage approach. The expectation step (E-step) consists of estimating the missing entries in $\mathbf{X}_{\text{theo}}$ as $\mathbf{X}^{\text{comp}} = \mathbf{W} \circ \mathbf{X} + \overline{\mathbf{W}} \circ (\mathbf{G} \cdot \mathbf{F})$, where $\overline{\mathbf{W}} = \mathbb{1}_{n,p(m+1)} - \mathbf{W}$ and $\mathbb{1}_{n,p(m+1)}$ is the $n \times p(m+1)$ of ones. The maximization step (M-step) consists of estimating $\mathbf{G}$ and $\mathbf{F}$ from $\mathbf{X}^{\text{comp}}$ with any NMF solver. The authors in [17] proposed to combine such an EM framework with the Nesterov accelerated gradient [18] to solve NMF with missing entries. They found this strategy to be much more efficient than using Nesterov gradient descent with the original weighted NMF optimization problem. Our proposed method extends [17] by also tackling the equality constraints of $\Phi_\mathbf{G}$ and $\Phi_\mathbf{F}$. Then, as for most NMF algorithms, the M-step consists of alternatively solving

$$\begin{aligned} \tilde{\mathbf{G}} = \underset{\mathbf{G} \geq 0}{\arg\min} \quad & \frac{1}{2} \left\|\mathbf{X}^{\text{comp}} - \mathbf{G} \cdot \mathbf{F}\right\|_\mathcal{F}^2 \\ \text{subject to} \quad & \mathbf{G} = \Phi_\mathbf{F} + \Delta_\mathbf{G}, \end{aligned} \tag{14}$$

$$\begin{aligned} \tilde{\mathbf{F}} = \underset{\mathbf{F} \geq 0}{\arg\min} \quad & \frac{1}{2} \left\|\mathbf{X}^{\text{comp}} - \mathbf{G} \cdot \mathbf{F}\right\|_\mathcal{F}^2 \\ \text{subject to} \quad & \mathbf{F} = \Phi_\mathbf{F} + \Delta_\mathbf{F}, \end{aligned} \tag{15}$$

with Nesterov iterations to update $\mathbf{G}$ and $\mathbf{F}$. Denoting $\mathcal{J}(\mathbf{G}, \mathbf{F}) = \frac{1}{2} \|\mathbf{X}^{\text{comp}} - \mathbf{G} \cdot \mathbf{F}\|_\mathcal{F}^2$, the algorithm to update $\mathbf{F}$ is described in Alg. 1, where $\frac{\partial \mathcal{J}}{\partial \Delta_\mathbf{F}}(\mathbf{G}, \Delta_\mathbf{F} + \Phi_\mathbf{F}) = \mathbf{G}^T\mathbf{G}\Delta_\mathbf{F} + \mathbf{G}^T\mathbf{G}\Phi_\mathbf{F} - \mathbf{G}^T\mathbf{X}^{\text{comp}}$ is the derivative of $\mathcal{J}$ according to $\Delta_\mathbf{F}$, $\mathcal{P}$ is the endomorphism projecting negative values on zero (to respect the non-negativity constraint[3]), and the Hadamard product with $\overline{\Omega}_\mathbf{F}$ ensures that $\Delta_\mathbf{F} \circ \Omega_\mathbf{F} = 0$ (to respect the equality constraint).

[2] Considering calibration parameters that can be negative would result in a semi-NMF which is still solvable by the method proposed in Section 3.

[3] Let us recall that by removing $\mathcal{P}$ from the update rules of $\mathbf{F}$, we can perform Semi-NMF and tackle the case of negative calibration parameters.

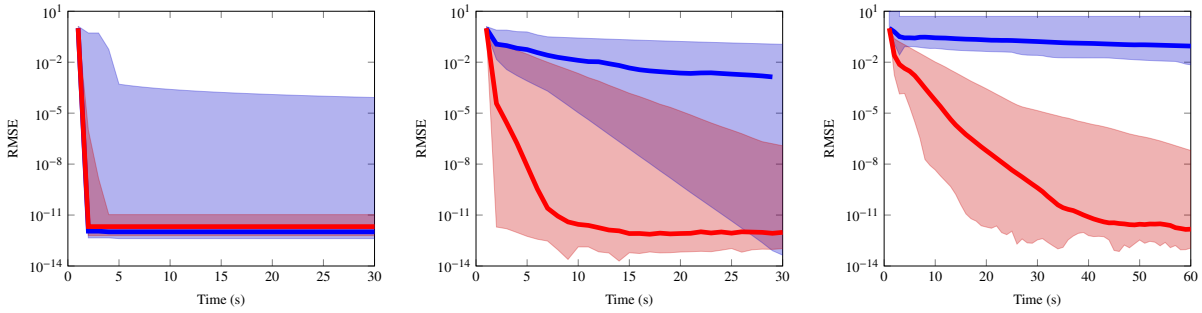

Fig. 2. Envelope and median RMSE with respect to time for 20 simulations. In blue, IN-Cal [7]. In red, the proposed F-IN-Cal method. (a) $n = 100,\ m = 25,\ p = 1$, 4 references, (b) $n = 100,\ m = 25,\ p = 2$, 4 references (c) $n = 400,\ m = 100,\ p = 2$, 10 references.

Algorithm 1: Update of $\mathbf{F}$ with Nesterov gradient

Data: $\mathbf{G}^t, \mathbf{F}^t$

$\mathbf{Y}_0 = \Delta_{\mathbf{F}}{}^t,\ \alpha_0 = 1,\ L = \|\mathbf{G}^{tT}\mathbf{G}^t\|_2,\ k = 0;$

while *Stopping criteria not reached* **do**

  $\Delta_{\mathbf{F}k} = \mathcal{P}(\overline{\Omega}_{\mathbf{F}} \circ (\mathbf{Y}_k - \frac{1}{L}\frac{\partial \mathcal{J}}{\partial \Delta_{\mathbf{F}}}(\mathbf{G}^t, \mathbf{Y}_k + \Phi_{\mathbf{F}})));$

  $\alpha_{k+1} = 0.5 \cdot (1 + \sqrt{4\alpha_k^2 + 1});$

  $\mathbf{Y}_{k+1} = \Delta_{\mathbf{F}k} + \frac{\alpha_k - 1}{\alpha_{k+1}}(\Delta_{\mathbf{F}k} - \Delta_{\mathbf{F}k-1});$

  $k \leftarrow k + 1;$

$\mathbf{F}^{t+1} = \Phi_{\mathbf{F}} + \Delta_{\mathbf{F}k};$

4. EXPERIMENTAL VALIDATION

In this section, we investigate the performance reached by our proposed F-IN-Cal method in different sensor calibration simulations. Let us first recall that we showed in Section 2 that calibrating a network of sensor arrays could be revisited as an informed (semi-)NMF problem, which is also the case of single-sensor network calibration [7]. We thus propose to compare the enhancement provided by both our proposed F-IN-Cal method and IN-Cal[4] [7]—an informed NMF method using multiplicative updates—on several configurations. To simulate the physical phenomena in a scene, several normal distributions with random parameters are summed and projected into intervals that could represent air pollutant concentrations. The calibration parameters are also randomly chosen in an interval that can be provided by a manufacturer. Some parameters are fixed and their influence will not be investigated in this paper, namely the missing values proportion—*i.e.*, the proportion of zeros in $\mathbf{W}$—is fixed to 0.5, the proportion of *rendez-vous* between reference and uncalibrated sensors is fixed to 0.3, the sensor cross-sensitivity—considered as an input signal-to-interference ratio [19]—is set to 15 dB, and the signal-to-noise ratio is fixed to infinity. Moreover, each sensor makes at most one *rendez-vous* with a reference sensor, which is a difficult scenario as state-of-the-art multi-hop techniques [20, 21] require many of such *rendez-vous* to work. Both methods are stopped after the same amount of CPU time and are tested on the same set of simulations with the same initialization. Each simulation is repeated 20 times, so that we can derive some statistics for each method. The calibration accuracy is evaluated with the root mean-square error (RMSE) computed over one line[5] of $\mathbf{F}^k$ if the physical variable of interest is supposed to be measured by the k-th sensor of a sensor array. The tests were run on a laptop with an Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz, 2592 MHz, 4 cores, 8 threads and 8.00 GB of RAM.

We first confront the methods in a test where IN-Cal is known for performing well according to [7], namely scenes of size 10×10 with 25 mobile sensors that are not cross-sensitive and 4 references. Our tests presented in Fig. 2(a) show that both methods converge in less than 2 seconds to a median RMSE around 10^{-12}, even if the high part of the IN-Cal envelope is slowly decreasing along time. We then consider 10×10 scenes observed by 25 mobile sensor arrays—where each array is composed of 2 cross-sensitive sensors—and 4 reference sensor arrays. Figure 2(b) shows the convergence speed of IN-Cal to be drastically reduced, because the quantity of information has doubled as we aim to calibrate 2 sensors per array. After 30 s, the median RMSE reached by IN-Cal is equal to $1.3 \cdot 10^{-3}$. On the contrary, F-IN-Cal still converges in 15 s to reach a median RMSE around 10^{-12}. Lastly, we consider a larger problem with simulations of scenes of size 20×20, observed by 100 mobile sensor arrays and 10 reference sensor arrays. In Fig. 2(c), our method still converges in a reasonable amount of time although there are more than sixteen times more data to process in this test than in the tests plot in Fig. 2(b). On the contrary, the median RMSE provided by IN-Cal is slowly decreasing to reach 0.08 after 1 min. These simulations show the relevance of the proposed approach.

5. CONCLUSION AND DISCUSSION

In this paper, we proposed an extension of the calibration model in [7] to tackle mobile sensor arrays. We first revisited sensor array calibration as a matrix factorization problem with missing entries and equality constraints. We then derived a fast informed NMF technique using an EM framework and a Nesterov gradient. Our simulations show that our proposed approach is much faster than IN-Cal [7], especially when the size of the scene and/or of the sensor network increase. Future work will focus on the enhancement provided by random projections [22, 23] in order to increase even more the convergence speed.

[4] See its implementation at `https://gogs.univ-littoral.fr/puigt/Informed_NMF_Mobile_Sensor_Calibration/`.

[5] In our tests, we never encountered a situation where the RMSE computed over a line of $\mathbf{F}^k$ was satisfying while those computed over the other lines of $\mathbf{F}^k$ or of another submatrix of $\mathbf{F}$ were not. As a consequence, for the sake of concision, we only show the performance obtained for one of the $p \cdot (p+1)$ computed RMSEs.

6. REFERENCES

[1] Emily G Snyder, Timothy H Watkins, Paul A Solomon, Eben D Thoma, Ronald W Williams, Gayle SW Hagler, David Shelow, David A Hindin, Vasu J Kilaru, and Peter W Preuss, "The changing paradigm of air pollution monitoring," 2013.

[2] Nuria Castell, Franck R Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak Fishbain, David Broday, and Alena Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?," *Environment international*, vol. 99, pp. 293–302, 2017.

[3] Michael Mueller, Jonas Meyer, and Christoph Hueglin, "Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of zurich," *Atmospheric Measurement Techniques*, vol. 10, no. 10, pp. 3783, 2017.

[4] Balz Maag, Zimu Zhou, and Lothar Thiele, "A survey on sensor calibration in air pollution monitoring deployments," *IEEE Internet of Things Journal*, vol. 5, pp. 4857–4870, 2018.

[5] Florentin Delaine, Bérengère Lebental, and Hervé Rivano, "In situ calibration algorithms for environmental sensor networks: A review," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 5968–5978, 2019.

[6] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele, "Pre-deployment testing, augmentation and calibration of cross-sensitive sensors.," in *EWSN*, 2016, pp. 169–180.

[7] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "Blind calibration of mobile sensors using informed nonnegative matrix factorization," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 497–505.

[8] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a vandermonde factor," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2016, pp. 1–5.

[9] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "Informed nonnegative matrix factorization methods for mobile sensor network calibration," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 4, pp. 667–682, 2018.

[10] Nicolas Haderer, Romain Rouvoy, and Lionel Seinturier, "Dynamic deployment of sensing experiments in the wild using smartphones," in *IFIP International Conference on Distributed Applications and Interoperable Systems*. Springer, 2013, pp. 43–56.

[11] Olga Saukh, David Hasenfratz, Christoph Walser, and Lothar Thiele, *On Rendezvous in Mobile Sensing Networks*, vol. 281, pp. 29–42, 01 2014.

[12] Yu-Xiong Wang and Yu-Jin Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.

[13] Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, and Dominique Courcot, "Non-negative matrix factorization under equality constraints—a study of industrial source identification," *Applied Numerical Mathematics*, vol. 85, pp. 1–15, 2014.

[14] Ngoc-Diep Ho, *Nonnegative matrix factorization algorithms and applications*, Ph.D. thesis, PhD thesis, Université catholique de Louvain, 2008.

[15] Hsiang-Fu Yu, Cho-Jui Hsieh, Si Si, and Inderjit Dhillon, "Scalable coordinate descent approaches to parallel matrix factorization for recommender systems," in *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012, pp. 765–774.

[16] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon, "Learning from incomplete ratings using non-negative matrix factorization," in *Proceedings of the 2006 SIAM international conference on data mining*. SIAM, 2006, pp. 549–553.

[17] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "Fast nonnegative matrix factorization and completion using Nesterov iterations," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2017, pp. 26–35.

[18] Yurii E Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," in *Dokl. akad. nauk Sssr*, 1983, vol. 269, pp. 543–547.

[19] Yannick Deville and Matthieu Puigt, "Temporal and time-frequency correlation-based blind source separation methods. Part I: Determined and underdetermined linear instantaneous mixtures," *Signal Processing*, vol. 87, no. 3, pp. 374–407, 2007.

[20] Olga Saukh, David Hasenfratz, and Lothar Thiele, "Reducing multi-hop calibration errors in large-scale mobile sensor networks," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 2015, pp. 274–285.

[21] Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele, "Scan: Multi-hop calibration for mobile sensor arrays," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–21, 2017.

[22] Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "How to apply random projections to nonnegative matrix factorization with missing entries?," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[23] Farouk Yahaya, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel, "Random projection streams for (weighted) nonnegative matrix factorization," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, accepted.