



**HAL**  
open science

# Sensor Calibration using Informed Matrix Factorisation for Air Quality Crowdsensing

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, Gilles Roussel

► **To cite this version:**

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, Gilles Roussel. Sensor Calibration using Informed Matrix Factorisation for Air Quality Crowdsensing. 4ème Journée Régionale des Doctorants en Automatique et Traitement du signal, Jun 2017, Béthune, France. 2017. hal-01831329

**HAL Id: hal-01831329**

**<https://hal.science/hal-01831329>**

Submitted on 18 Aug 2018

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

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



# Sensor Calibration using Informed Matrix Factorisation for Air Quality Crowdsensing

Clément DORFFER, Matthieu PUIGT, Gilles DELMAIRE, and Gilles ROUSSEL

Univ. Littoral Côte d'Opale, LISIC – EA 4491, F-62228 Calais, France,

Firstname.LASTNAME@univ-littoral.fr

## Context and Objectives

- **Air quality monitoring** performed by few fixed or mobile devices of the official air quality monitoring organisms (french AASQA) ⇨ poor spatial coverage/important measurement cost
- **Citizen weakly involved**
- **Mobile crowd-sensing**: data acquisition from a crowd of sensors connected to smartphones
- Can we use both mobile crowd-sensing and official data to build dense pollution maps with high precision?
- Can we involve citizens into the monitoring process?

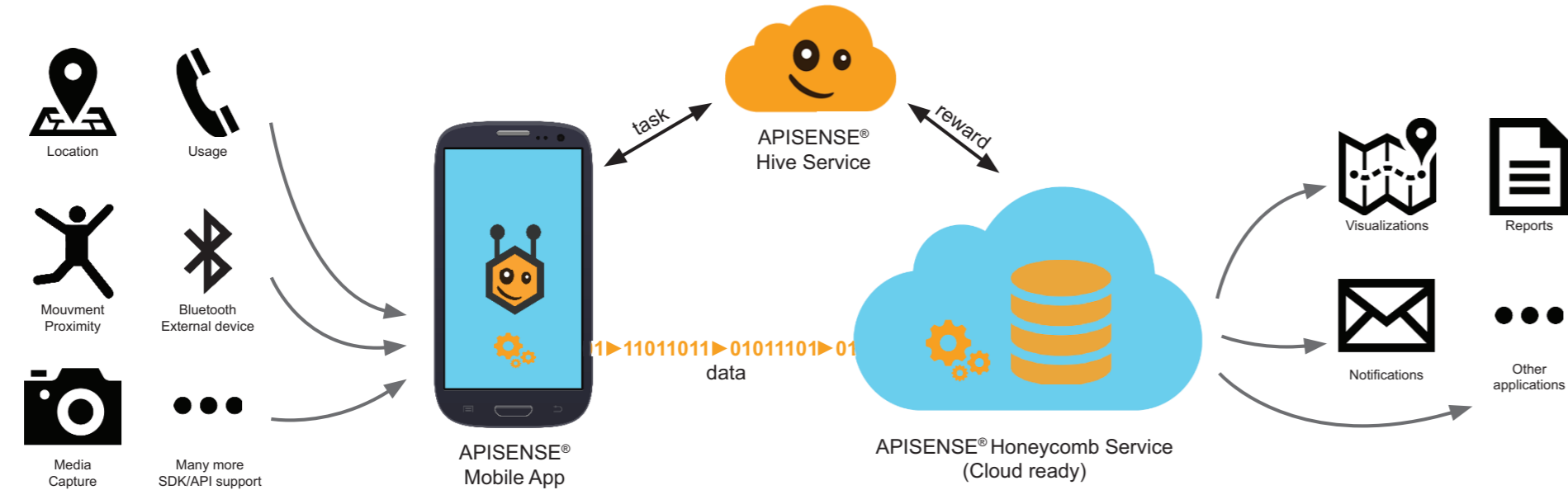
## From mobile crowd-sensing data to accurate estimation

- **Low cost air quality device development**: arduino-based data acquisition device connected to smartphones



1. Devices build by high school students,
2. carried by volunteers who then know the quality of the air they breathe

- **Energy- and privacy-preserving mobile crowd-sensing** acquisition with APISENSE® (<https://apisense.io>) [1]



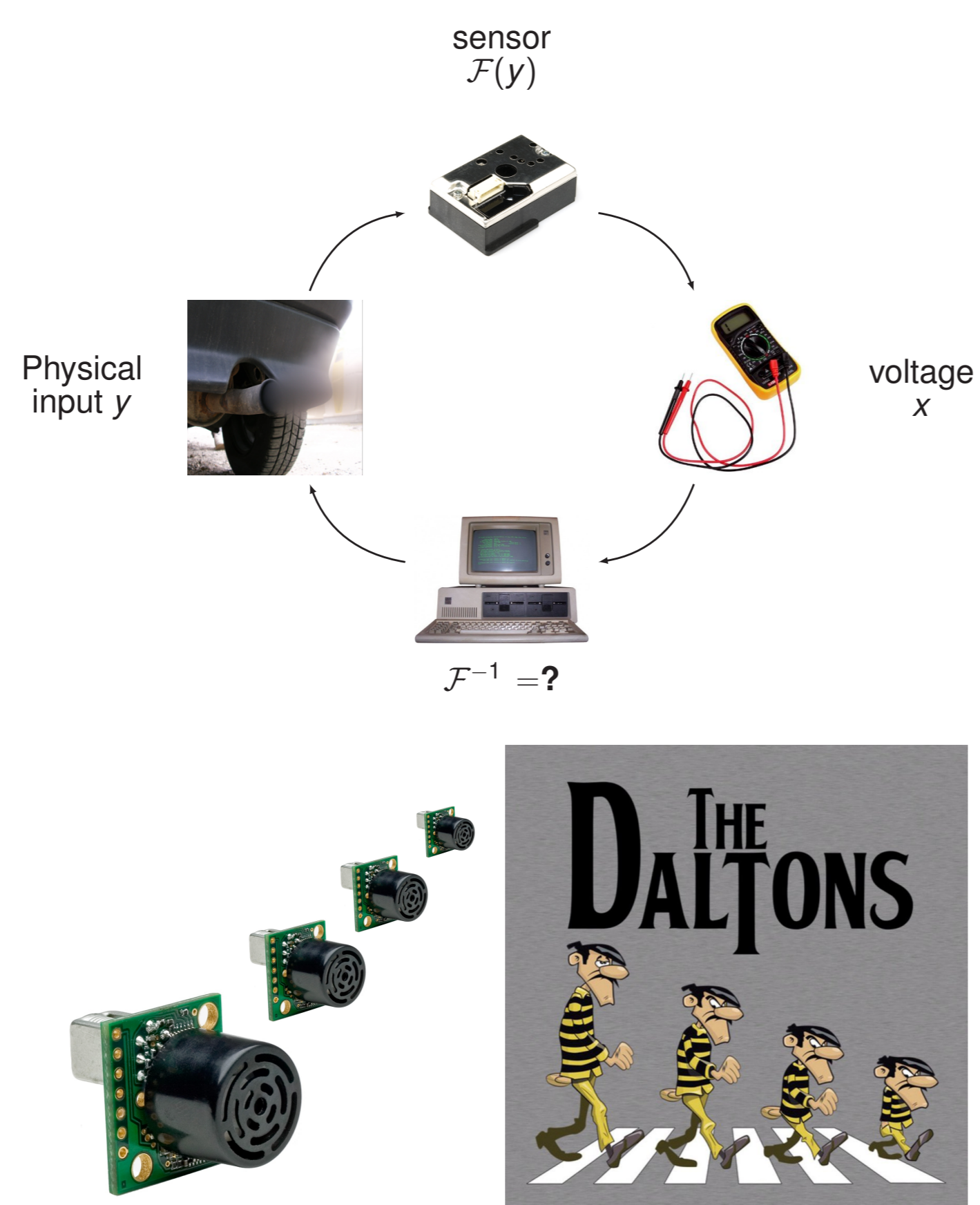
- **Enhance data quality** by fusing data from both the mobile devices and the authoritative sensors (from ATMO in France)

## Calibrating the sensor network responses

1. An electronic sensor breathes the air concentration  $y$  and provides an output voltage  $x$  linked to  $y$  through a calibration function  $\mathcal{F}$
2.  $\mathcal{F}$  is an unknown calibrating function to be estimated for each sensor (in laboratory)
- ⇨ Polynomial approximation hypothesis of  $\mathcal{F}$

$$x \approx \mathcal{F}(y) \approx \sum_{k=0}^N f_k \cdot y^k$$

3.  $m$  mobile sensors →  $m$  calibrating functions  $\mathcal{F}_j$  ( $j = 1 \dots m$ ) to estimate
- ⇨ Aim: to get homogeneous physical measurements from different sensors for the same stimulus (Not a Daltons' distribution)
4. If  $m$  is large, impossible to perform calibration in a laboratory (too expensive). How to do it remotely?



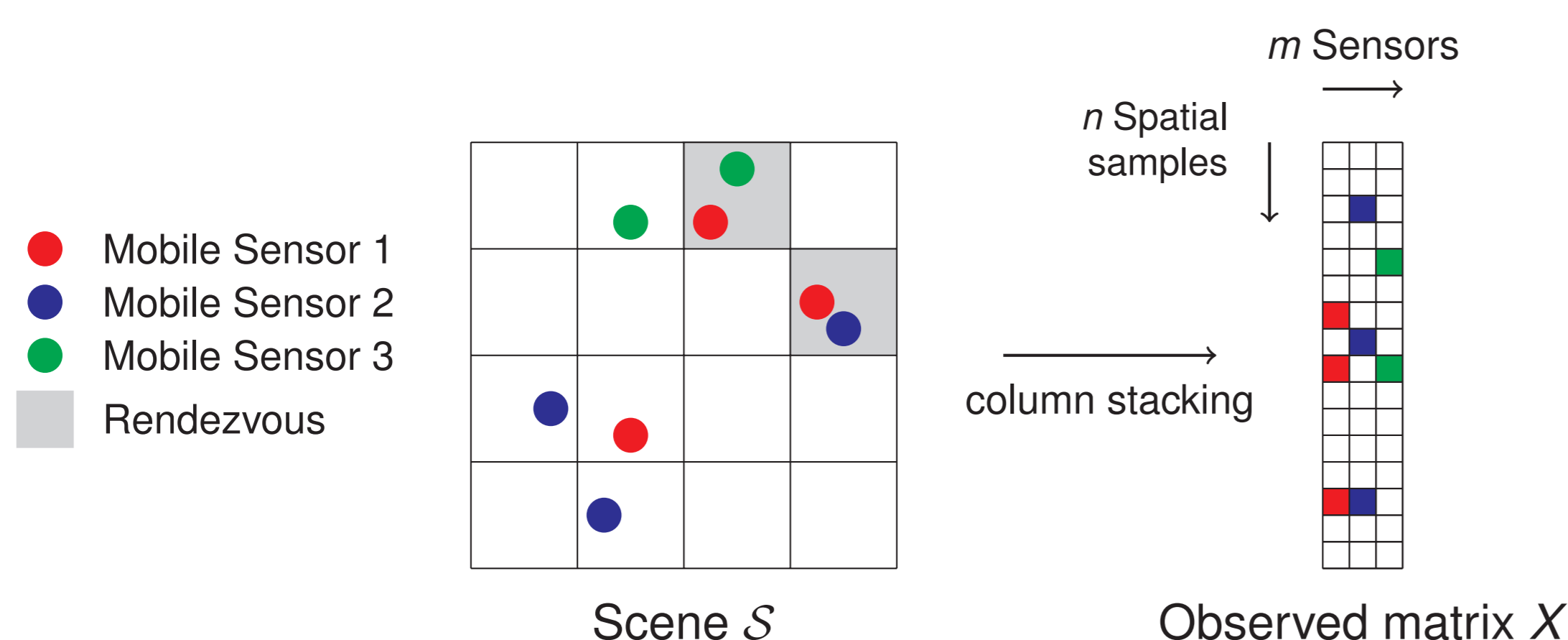
## Rendezvous principle

### Rendezvous [2]

A rendezvous is a spatial and temporal vicinity between two sensors.

### Scene [3]

A Scene is a discretized area observed during a time interval  $[t, t + \Delta_t]$ . A spatial pixel has a size lower than  $\Delta_d$ , where  $\Delta_t$  and  $\Delta_d$  defines the vicinity of the rendezvous.



- General modelling of the sensor network observations [4]

$$x_{ij} \approx \mathcal{F}_j(y_i) \approx \sum_{k=0}^N f_{kj} \cdot y_i^k \quad \Leftrightarrow \quad X \approx G \cdot F$$

where  $X$ ,  $G$ , and  $F$  respectively contain the sensor outputs, the calibration model (Vandermonde matrix), and the sensor parameters, i.e.,

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}, \quad G = \begin{bmatrix} 1 & y_1 & \dots & y_1^N \\ \vdots & \vdots & \dots & \vdots \\ 1 & y_n & \dots & y_n^N \end{bmatrix}, \quad F = \begin{bmatrix} f_{01} & f_{02} & \dots & f_{0m} \\ \vdots & \vdots & \dots & \vdots \\ f_{N1} & f_{N2} & \dots & f_{Nm} \end{bmatrix}$$

- In practice, missing data in  $X$  (irregular sampling) ⇨ **binary mask  $W$**

### Hypotheses:

- Dense network
- $X$  well conditioned
- Enough diversity in  $W \circ X$
- $X \geq 0$  (voltages)
- $G \geq 0$  (concentrations)
- $F$  of unconstrained sign

### Extra-information:

- Known values in  $G$  (**model + reference data**)
- Known values in  $F$  (**reference sensors**)

## Resolution process

- **Weighted semi-nonnegative matrix factorization (Semi-NMF) problem with structured matrices**

$$\{\hat{G}, \hat{F}\} = \arg \min_{G, F} \frac{1}{2} \|W \circ (X - G \cdot F)\|_F^2, \quad \text{s.t.} \quad \begin{cases} G \text{ is Vandermonde} \\ G = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G \geq 0 \\ F = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F \end{cases} \quad (1)$$

where  $\Omega_G \circ \Phi_G$  et  $\bar{\Omega}_G \circ \Delta_G$  stand for the **known** and **free** parts of  $\hat{G}$ , respectively, and  $\|\cdot\|_F$  is the Frobenius norm.

### Example

$$\begin{bmatrix} x_{1,1} & \dots & x_{1,3} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ x_{3,1} & \dots & x_{3,3} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ x_{4,2} & \dots & y_4 & \dots \end{bmatrix} = W \circ \begin{bmatrix} 1 & y_1 & y_1^2 \\ 1 & y_2 & y_2^2 \\ 1 & y_3 & y_3^2 \\ 1 & y_4 & y_4^2 \end{bmatrix} \cdot \begin{bmatrix} f_{0,1} & f_{0,2} & f_{0,3} & 0 \\ f_{1,1} & f_{1,2} & f_{1,3} & 1 \\ f_{2,1} & f_{2,2} & f_{2,3} & 0 \end{bmatrix}$$

with

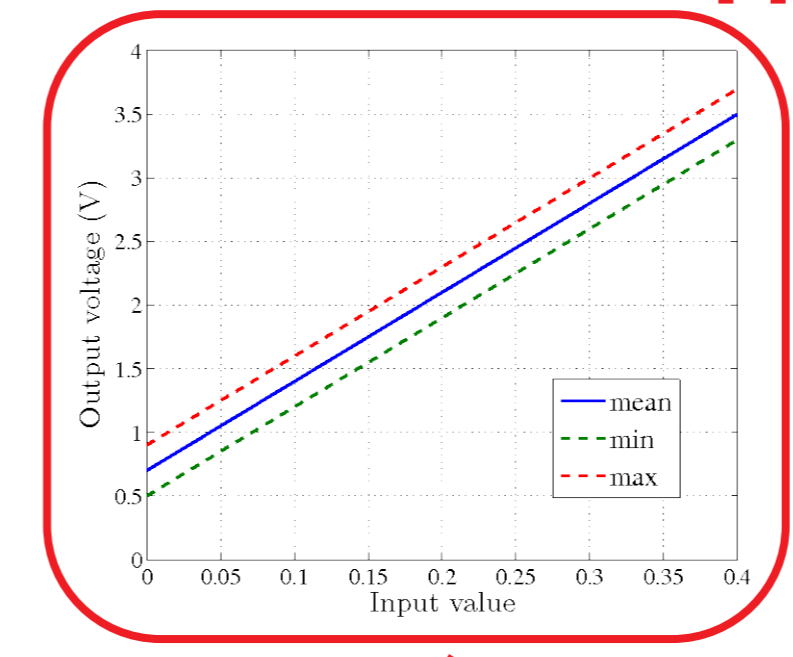
- **3 mobile sensors (Sensor 1 to 3)**
- **1 reference (Sensor 4)** ⇨  $x_{i,4} = y_i \quad \forall i$
- polynomial sensor response of degree  $N = 2$
- 2 measurements per sensor over a  $2 \times 2$  scene

- Parameterization of the known and free parts:

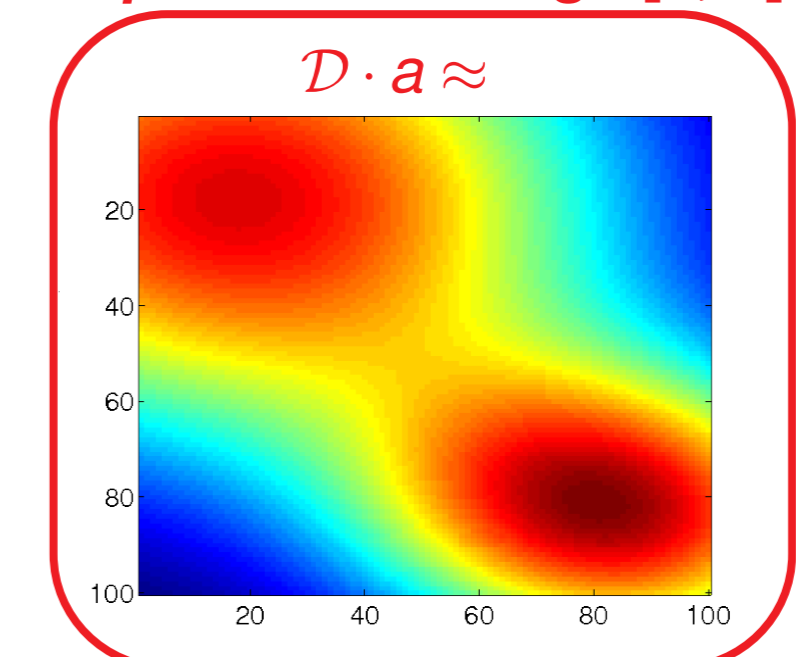
$$G = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 0 & 0 \\ 1 & y_2 & y_2^2 \\ 1 & 0 & 0 \\ 1 & y_4 & y_4^2 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \circ \begin{bmatrix} 0 & y_1 & y_1^2 \\ 0 & 0 & 0 \\ 0 & y_3 & y_3^2 \\ 0 & 0 & 0 \end{bmatrix}$$

## Extra information

### Datasheet information [6]



### A priori knowledge [5, 6]

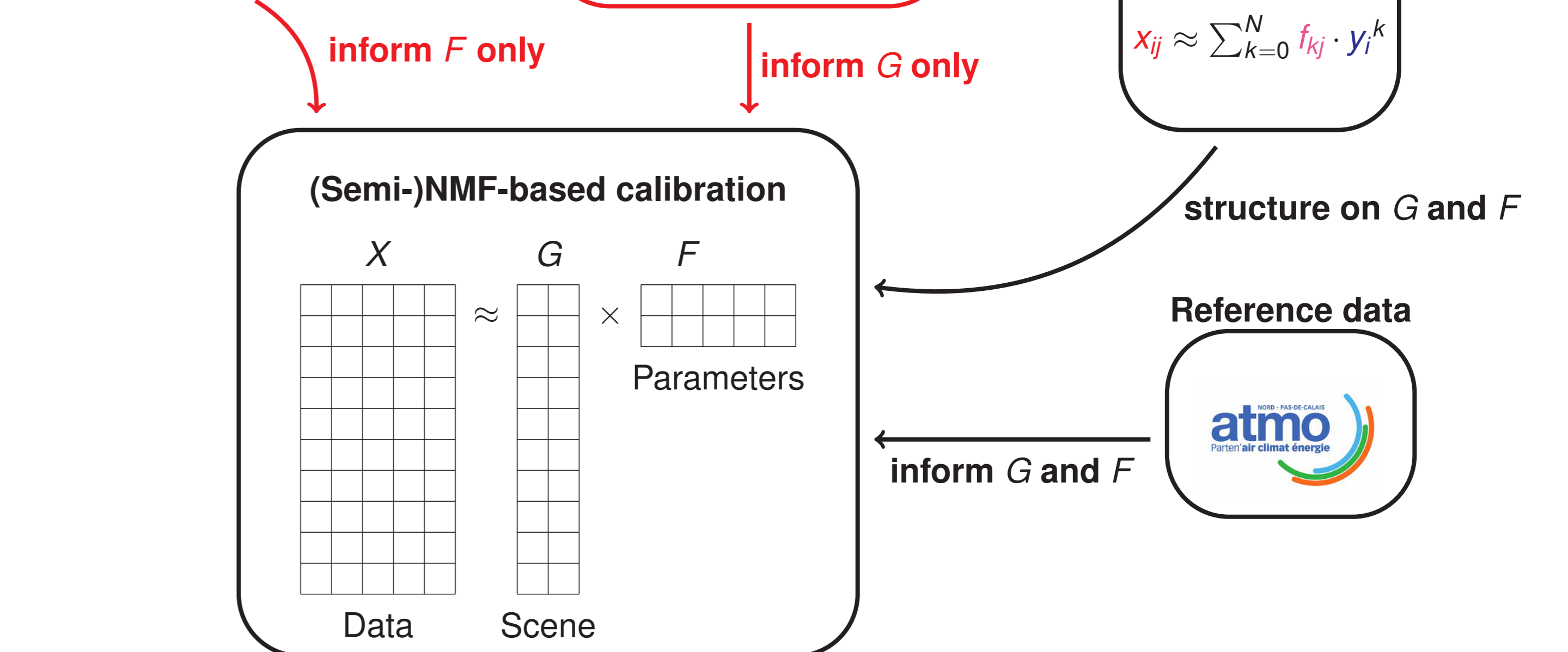


### Sensor response model

$$x_{ij} \approx \sum_{k=0}^N f_{kj} \cdot y_i^k$$

structure on  $G$  and  $F$

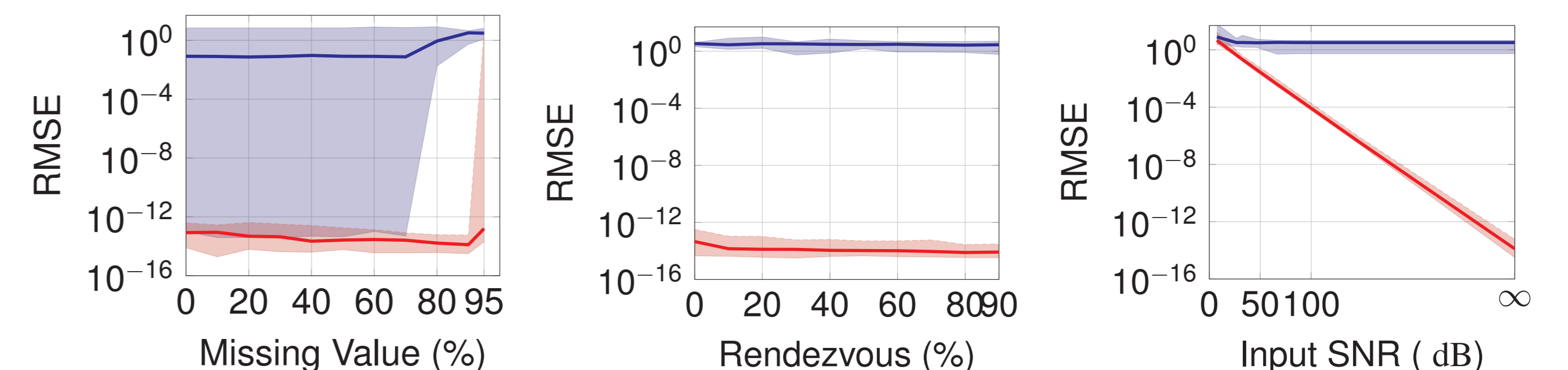
Reference data



## Experimentations

### Calibration performance (estimation of $F$ )

- Simulations of crowdsensing-like air quality monitoring ( $N = 1$ , affine model,  $F$  nonnegative in this case)
- $m = 25$  sensors (+ 1 calibrated one)
- Performance criterion = RMSE between theoretical and estimated parameters
- (blue): calibration using matrix completion and (red): proposed approach [5]
- Default parameters: prop. Rendezvous= 0.3; prop. Missing Value= 0.9; Input SNR=  $\infty$



### Field reconstruction (estimation of $G$ )

- Simulation parameters
- Theoretical field with  $100 \times 100$  cells observed by 50 mobile sensors
- 90% of missing value in the data matrix  $X$ ; 10 reference measurements
- Simulated dictionary composed of 60 atoms

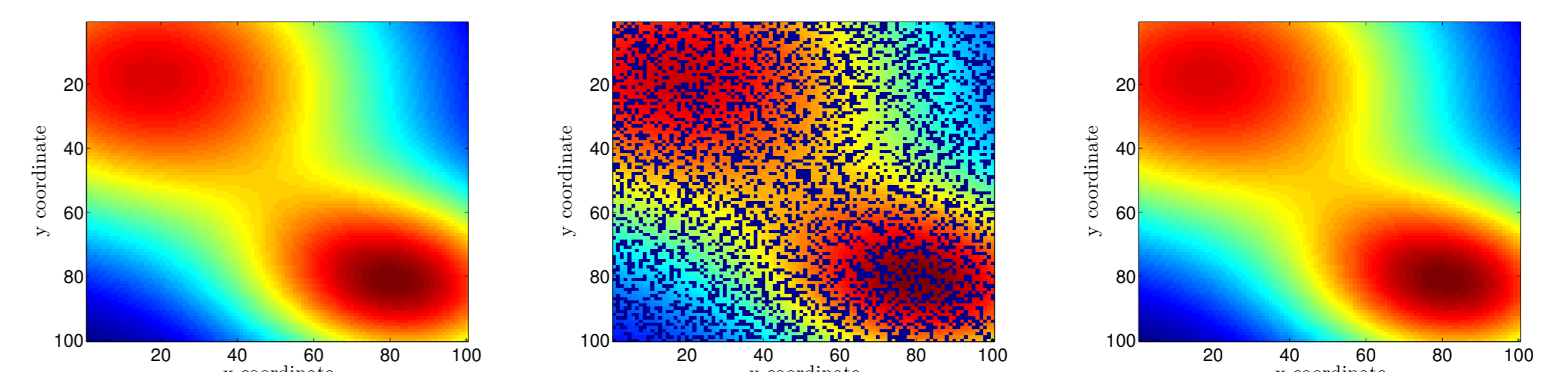


Figure: (left) theoretical field; (middle) Matrix Factorization (MF) based calibration; (right) MF and dictionary calibration

## References and acknowledgments

- [1] N. Haderer, R. Rouvoy, C. Ribeiro, and L. Seinturier. "APISENSE: Crowd-Sensing Made Easy". In: *ERCIM News*. Special theme: Mobile Computing 93 (2013), pp. 28–29.
- [2] O. Saukh, D. Hasenfratz, C. Walsler, and L. Thiele. "On rendezvous in mobile sensing networks". In: *Proc. of REALWSN*. Vol. 281. LNCS, 2014, pp. 29–42.
- [3] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. "Blind calibration of mobile sensors using informed nonnegative matrix factorization". In: *Proc. of LVA-ICA*. Vol. 9237. LNCS, 2015, pp. 497–505.
- [4] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. "Nonlinear Mobile Sensor Calibration Using Informed Semi-Nonnegative Matrix Factorization With a Vandermonde Factor". In: *Proc. of SAM*. July 2016.
- [5] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. "Blind mobile sensor calibration using a nonnegative matrix factorization with a relaxed rendezvous model". In: *Proc. of ICASSP*. Mar. 2016.
- [6] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel. *Informed nonnegative matrix factorization methods for mobile sensor calibration*. Submitted.

This work was funded partly by the "OSCAR" project within the Région Hauts-de-France "Chercheurs Citoyens" Program, and partly by the CNRS MASTODONS project under the grant "DoMaSQ Air 2017".