CASPA 2019

Session 2 : Capteurs - Données

CARTOGRAPHIE ET ÉTALONNAGE DE CAPTEURS CONJOINTS PAR TRAITEMENT DES DONNÉES ISSUES DE CAPTEURS MOBILES

MATTHIEU PUIGT

Cartographie et étalonnage de capteurs conjoints par traitement des données issues de capteurs mobiles *

M. Puigt[1] C. Dorffer[1,2,‡] F. Yahaya[1,†] G. Delmaire[1] G. Roussel[1]

[1] Univ. Littoral Côte d'Opale, LISIC – EA 4491, F-62228 Calais Cedex, France
[2] ENSTA-Bretagne, Lab-STICC – UMR 6285, CNRS, F-29806 Brest Cedex 9, France

1–4 avril 2019

* Travaux financés par la Région Hauts-de-France (projet "chercheurs-citoyens") et par le CNRS (projet MASTODONS). Les expériences ont été réalisées sur la plateforme de calcul scientifique CALCULCO gérée par le SCoSI de l'ULCO.

† F. Yahaya remercie la Région Hauts-de-France qui finance en partie sa thèse de doctorat.

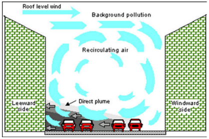
‡ Dr C. Dorffer remercie le Pôle Métropolitain de la Côte d'Opale qui a financé une partie de sa thèse de doctorat.

Context

The limits of current air quality monitoring

- Increasing human activities ➪ pollution peaks


- Impact on health ➪ ≈ 400.000 premature deaths per year in EU
- Emission reduction ➪ monitoring (observation & modeling)

➪ Local effects not sensed and hard to model with a sparsely distributed sensor network

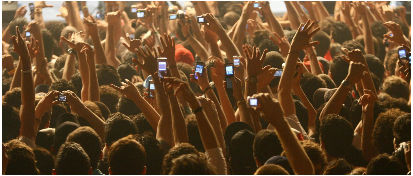
Context

Mobile Crowdsensing

Is it possible to ?

1. create a complementary low-cost air quality sensor network
2. which offers a finer spatial coverage
3. and which involves the public in the sensing procedure

➩ **Mobile crowdsensing** : a crowd of volunteers to sense geolocated and time-stamped measurements



OSCAR : *Observation et Sensibilisation Citoyenne à la surveillance de la qualité de l'Air en Région*

- ATMO Hauts-de-France
- BES
- INRIA Spirals
- LISIC

Objectives

Designing an air quality sensing campaign involving the population to build the sensors and sense air quality using mobile crowdsensing

Technical issues

- Creating a low-cost air quality sensor (LISIC)
- Learning the population to make the sensors (BES)
- Manage data collection campaign (INRIA) and take into account the ASQAA measurements (ATMO HdF)
- Signal processing to make sense of the collected data (LISIC)

OSCAR : *Observation et Sensibilisation Citoyenne à la surveillance de la qualité de l'Air en Région*

- ATMO Hauts-de-France
- BES
- INRIA Spirals
- LISIC

Objectives

Designing an air quality sensing campaign involving the population to build the sensors and sense air quality using mobile crowdsensing

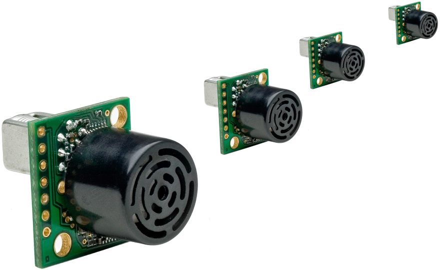
Technical issues

- Creating a low-cost air quality sensor (LISIC)
- Learning the population to make the sensors (BES)
- Manage data collection campaign (INRIA) and take into account the ASQAA measurements (ATMO HdF)
- Signal processing to make sense of the collected data (LISIC)

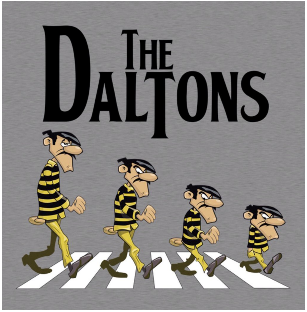
The why of sensor calibration

- Observed phenomenom ➪ voltage
- Voltage ➪ Physical value ?

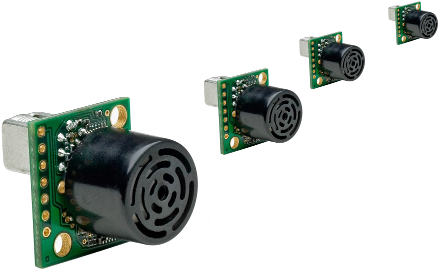
The why of sensor calibration



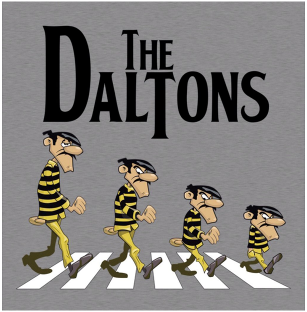
- Observed phenomenom ➪ voltage
- Voltage ➪ Physical value ?



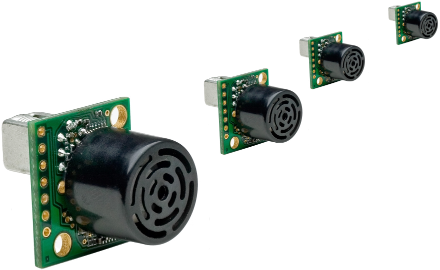
The why of sensor calibration

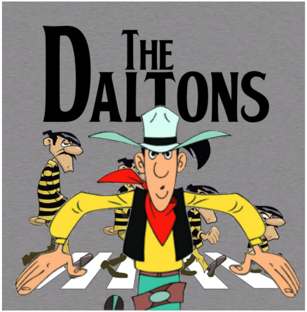


- Observed phenomenom ➪ voltage
- Voltage ➪ Physical value ?
  - Sensor calibration cannot be performed in lab
  - ➪ Data-driven approaches (a.k.a. "blind" or "self" calibration techniques)



The why of sensor calibration





- Observed phenomenom ➪ voltage
- Voltage ➪ Physical value ?
  - Sensor calibration cannot be performed in lab
  - ➪ Data-driven approaches (a.k.a. "blind" or "self" calibration techniques)
  - Presence of reference data (ATMO HdF)

The how of sensor calibration

A data processing taxonomy :

Numerous techniques

- **Supervised**
  - In-lab regression
- **Weakly supervised**
  - Learning & Correcting
  - (Balzano *et al.*, 2007-2014)
- **Unsupervised**
  - **Macro-calibration**
    - Moments (Wang *et al.*, 2008)
    - Graph (Lee *et al. 2014*)
  - **Micro-calibration**
    - Rendez-vous
    - (Saukh *et al.*, 2013-2017)

For *mobile-crowdsensing*

- Graph-based methods (macro-calibration)
- Multi-hop techniques (micro-calibration)

➮ Problem-specific techniques
➮ Not necessarily applicable to OSCAR network

➮ Novel macro-calibration techniques using micro-calibration assumptions

The how of sensor calibration

A data processing taxonomy :

Numerous techniques

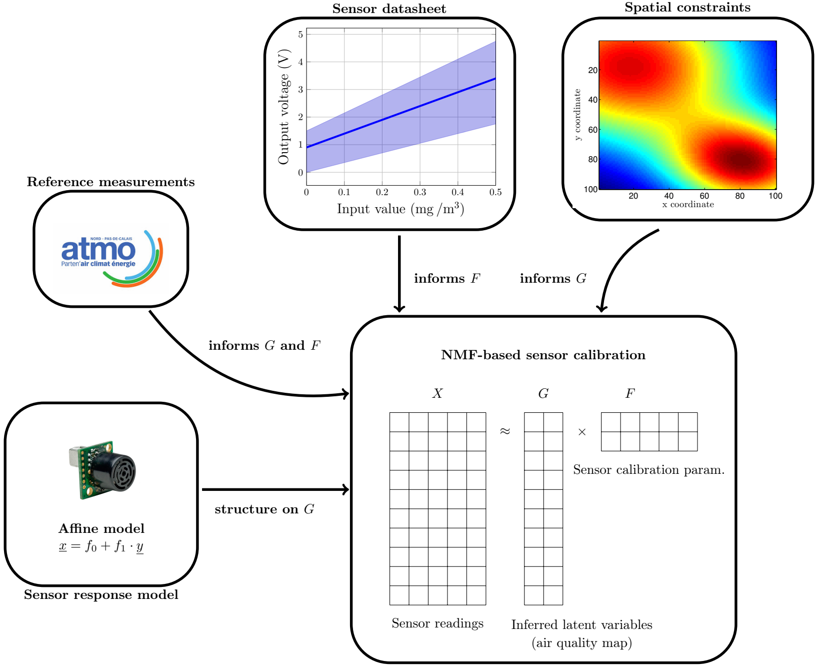
- **Supervised**
  In-lab regression
- **Weakly supervised**
  Learning & Correcting
  (Balzano *et al.*, 2007-2014)
- **Unsupervised**
  - **Macro-calibration**
    Moments (Wang *et al.*, 2008)
    Graph (Lee *et al. 2014*)
  - **Micro-calibration**
    Rendez-vous
    (Saukh *et al.*, 2013-2017)

A must-read paper

B. Maag, Z. Zhou, and L. Thiele : *A Survey on Sensor Calibration in Air Pollution Monitoring Deployments*, In IEEE Internet of Things Journal, to appear, `http://doi.org/10.1109/JIOT.2018.2853660`

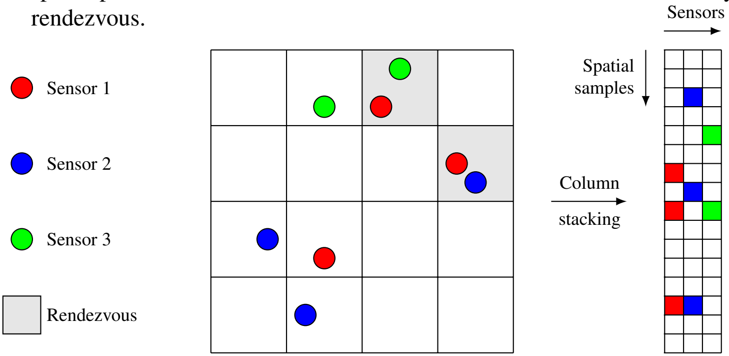
*"Currently, **there is no one-for-all network calibration solution available**. Recent research efforts investigate the **possibility of a general applicable network calibration method**, e.g., by combining different aspects from the three methods. Some theoretical investigations already provide mixtures of different models. For instance, Dorffer et al. [72]–[74] combine the two ideas of blind and collaborative network calibration to increase the possibilities for sensor re-calibration."*

The Big Picture



Definitions

- A rendezvous is a temporal and spatial vicinity between two sensors (Saukh *et al.*, 2013).
- A scene $\mathcal{S}$ is a discretized area observed during a time interval $[t, t+\Delta t)$. A spatial pixel has a size lower than Δd, where Δt and Δd define the vicinity of the rendezvous.



Scene $\mathcal{S}$ Observed matrix X

Assumptions (1)

- Sensor response (calibration function $\mathcal{F}(.)$ of Sensor j)

$$\underbrace{x(i,j)}_{\text{sensor-output voltage}} \simeq \mathcal{F}_j(y(i))$$

$$\simeq \underbrace{(y(i)}_{\text{physical phenomenom}} \cdot \overbrace{f_{1,j}) + f_{0,j}}^{\text{unknown gain and offset}}$$

➪ Matrix form (if **each** of the m sensor senses **all** the scene)

$$\underbrace{\begin{bmatrix} x(1,1) & \cdots & x(1,m) \\ \vdots & & \vdots \\ x(n,1) & \cdots & x(n,m) \end{bmatrix}}_{X} \simeq \underbrace{\begin{bmatrix} 1 & y(1) \\ \vdots & \vdots \\ 1 & y(n) \end{bmatrix}}_{G} \cdot \underbrace{\begin{bmatrix} f_{0,1} & f_{0,2} & \cdots & f_{0,m} \\ f_{1,1} & f_{1,2} & \cdots & f_{1,m} \end{bmatrix}}_{F}$$

- In practice, irregular sampling : $W \circ X$ with

$$W(i,j) \triangleq \begin{cases} 0 & \text{if } x(i,j) \text{ is not available,} \\ \rho_j & \text{otherwise,} \end{cases}$$

where ρ_j is a weight coefficient associated with Sensor j

Assumptions (2)

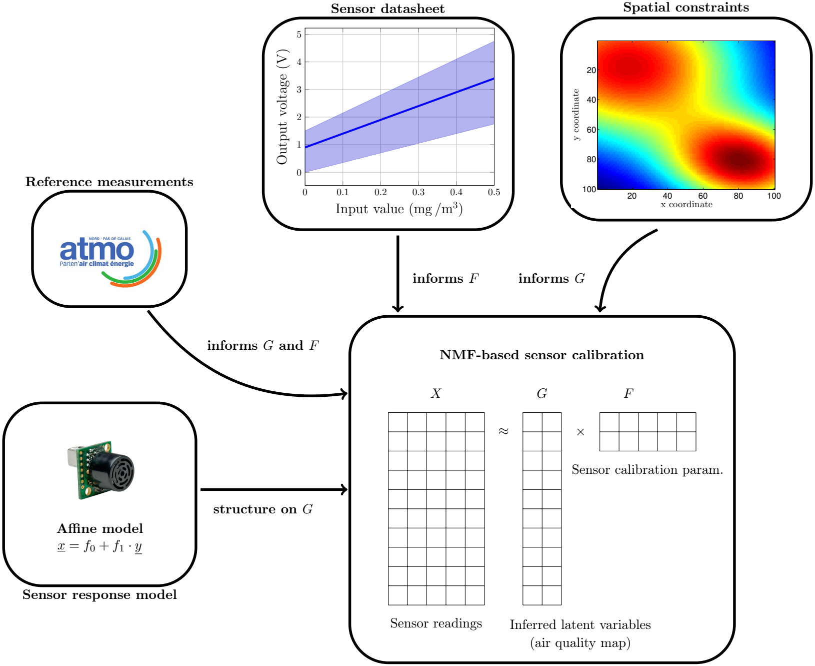
- X, G, and F are nonnegative (air quality application)
- A known reference
- ➩ $\forall i = 1, \ldots, n, \quad x(i,m) = y(i)$ (i.e., $f_{1,m} = 1, f_{0,m} = 0$)
- ➩ Blind calibration revisited as a weighted nonnegative matrix factorization problem

$$W \circ \underbrace{\begin{bmatrix} x(1,1) & \cdots & x(1,m-1) & y(1) \\ x(2,1) & \cdots & x(2,m-1) & y(2) \\ \vdots & & \vdots & \vdots \\ x(n,1) & \cdots & x(n,m-1) & y(n) \end{bmatrix}}_{X} \simeq W \circ \left(\underbrace{\begin{bmatrix} 1 & y(1) \\ 1 & y(2) \\ \vdots & \vdots \\ 1 & y(n) \end{bmatrix}}_{G} \cdot \underbrace{\begin{bmatrix} f_{0,1} & f_{0,2} & \cdots & f_{0,m-1} & 0 \\ f_{1,1} & f_{1,2} & \cdots & f_{1,m-1} & 1 \end{bmatrix}}_{F} \right)$$

IN-Cal (Informed NMF-based Calibration)

1. Start with initial matrices F^0 and G^0
2. Look for the “best” F
3. Look for the “best” G
4. Go back to step 2.

The Big Picture



Ajout de contraintes spatiales

$$W \circ X = \begin{bmatrix} x_{1,1} & - & - & - \\ - & x_{2,2} & x_{2,3} & - \\ x_{3,1} & - & x_{3,3} & - \\ x_{4,1} & x_{4,2} & - & - \end{bmatrix}$$

- ambiguïté d'echelle
- ➪ étalonnage relatif
- ➪ normalisation (ACIN-Cal)

Ajout de contraintes spatiales

$$\text{Si } W \circ X = \begin{bmatrix} x_{1,1} & - & - \\ x_{2,1} & - & - \\ - & - & y_3 \\ - & - & y_4 \\ - & x_{5,2} & - \\ - & x_{6,2} & - \end{bmatrix}$$

➮ pas d'étalonnage relatif...

Ajout de contraintes spatiales

$$\text{Si } W \circ X = \begin{bmatrix} x_{1,1} & - & \tilde{\boldsymbol{y}}_1 \\ x_{2,1} & - & \tilde{\boldsymbol{y}}_2 \\ - & - & y_3 \\ - & - & y_4 \\ - & x_{5,2} & \tilde{\boldsymbol{y}}_5 \\ - & x_{6,2} & \tilde{\boldsymbol{y}}_6 \end{bmatrix}$$

Vers des rendez-vous virtuels

- Rajouter un a priori spatial sur le phénomène physique
- Lier les mesures par rendez-vous virtuel

Interpolation de $\boldsymbol{y}$

- *a priori* de parcimonie : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \mathcal{D} \cdot a$
- régression : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \mathcal{Y}(\boldsymbol{\theta}, \boldsymbol{p})$
- calage de fonctions : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \sum_j c_j \cdot \Phi_j(\boldsymbol{p})$
- modèle physique : $\boldsymbol{y} \approx \tilde{y} = \mathcal{Y}(\boldsymbol{p}, \text{vent}, \text{traffic}, \ldots)$
- modèle géostatistique : $y_i \approx \tilde{y}_i = \sum_{j \neq i} \omega_{ij} \cdot y_j$

Ajout de contraintes spatiales

$$\text{Si } W \circ X = \begin{bmatrix} x_{1,1} & - & \tilde{\boldsymbol{y}}_1 \\ x_{2,1} & - & \tilde{\boldsymbol{y}}_2 \\ - & - & y_3 \\ - & - & y_4 \\ - & x_{5,2} & \tilde{\boldsymbol{y}}_5 \\ - & x_{6,2} & \tilde{\boldsymbol{y}}_6 \end{bmatrix}$$

Vers des rendez-vous virtuels

- Rajouter un a priori spatial sur le phénomène physique
- Lier les mesures par rendez-vous virtuel

Interpolation de $\boldsymbol{y}$

- *a priori* de parcimonie : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \mathcal{D} \cdot a$
- régression : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \mathcal{Y}(\boldsymbol{\theta}, \boldsymbol{p})$
- calage de fonctions : $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \sum_j c_j \cdot \Phi_j(\boldsymbol{p})$
- modèle physique : $\boldsymbol{y} \approx \tilde{y} = \mathcal{Y}(\boldsymbol{p}, \text{vent,traffic,...})$
- modèle géostatistique : $y_i \approx \tilde{y}_i = \sum_{j \neq i} \omega_{ij} \cdot y_j$

Ajout de contraintes spatiales

Hypothèses (H1)

- $\boldsymbol{y}$ admet une décomposition parcimonieuse suivant $\mathcal{D} \in \mathbb{R}^{m,l}$

➭ $\exists\, \boldsymbol{a} \in \mathbb{R}^{l}$ parcimonieux, tel que $\boldsymbol{y} \approx \tilde{y} = \mathcal{D} \cdot \boldsymbol{a}$

- $\boldsymbol{y}$ admet une approximation polynômiale

➭ $\exists\, \boldsymbol{\theta} \in \mathbb{R}^{N}$ tel que $\boldsymbol{y} \approx \tilde{\boldsymbol{y}} = \mathcal{Y}(\boldsymbol{\theta}, \boldsymbol{p})$

Nouveau problème d'optimisation

$$\begin{aligned}
\arg \min_{\hat{G}, \hat{F} \geq 0, \tilde{\boldsymbol{y}}} \quad & ||W \circ (X - G \cdot F)||_f^2 + \lambda \cdot ||g_2 - \tilde{\boldsymbol{y}}||_f^2 \\
s.c. \quad & G = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G \\
& F = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F \\
& \tilde{\boldsymbol{y}} \text{ vérifie } (H1)
\end{aligned}$$

- Non-convexe par rapport à G, F et $\boldsymbol{y}$

➭ stratégie de mises à jour alternées

1. MàJ de F (comme IN-Cal)
2. Estimation de $\tilde{\boldsymbol{y}}$
3. MàJ de G

Ajout de contraintes spatiales

Estimation de $\boldsymbol{y}$

$$\begin{aligned} \tilde{\boldsymbol{y}} \quad &= \quad \arg\min_{\tilde{\boldsymbol{y}}} ||g_2 - \tilde{\boldsymbol{y}}||_f^2 \\ s.c. \quad & \tilde{\boldsymbol{y}} = \mathcal{D} \cdot \boldsymbol{a} \\ & \boldsymbol{a} \quad \text{parcimonieux} \end{aligned}$$

➯ $\boldsymbol{a}$ = meilleure décomposition parcimonieuse de g_2 (SpIN-Cal)

$$\begin{aligned} \tilde{\boldsymbol{y}} \quad &= \quad \arg\min_{\tilde{\boldsymbol{y}}} ||g_2 - \tilde{\boldsymbol{y}}||_f^2 \\ s.c. \quad & \tilde{\boldsymbol{y}} \approx \mathcal{Y}(\boldsymbol{\theta}, \boldsymbol{p}) \end{aligned}$$

➯ $\tilde{\boldsymbol{y}}$ = meilleure approximation polynômiale de g_2 (PolIN-Cal)

KIN-Cal (Kriging-based extension of IN-Cal)

We also proposed in C. Dorffer's PhD thesis a spatial constraint based on Kriging :

- Not exactly following the above procedure
  - ➯ Kriging is independently applied on each sensor to create virtual rendez-vous
  - ➯ Update of the factor matrices are then derived as in IN-Cal
- More flexibility than PolIN-Cal and SpIN-Cal (no parameters to tune or dictionnary to estimate)
- Privacy-preserving strategy

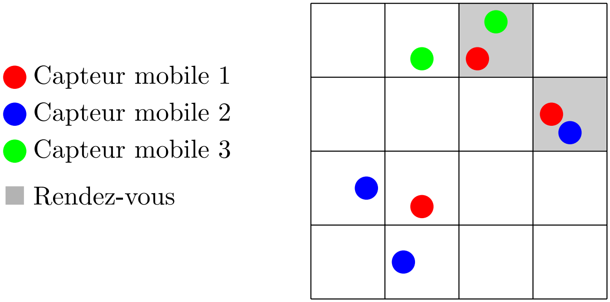
Discussion (1)

Apport des contraintes spatiales

- Connection des capteurs avec le dictionnaire ou le modèle
  - ➪ Plus besoin de "rendez-vous exacts" (au sens de la littérature)

➪ Réduction de la discrétisation spatiale possible (rendez-vous *virtuels*)

- Mais discrétisation temporelle toujours nécessaire



Scène $\mathcal{S}$

Discussion (1)

Apport des contraintes spatiales

- Connection des capteurs avec le dictionnaire ou le modèle
  - ➪ Plus besoin de "rendez-vous exacts" (au sens de la littérature)

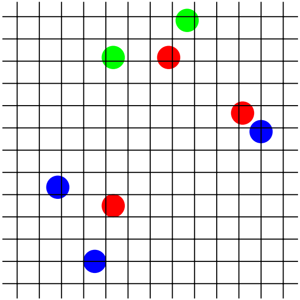
➪ Réduction de la discrétisation spatiale possible (rendez-vous *virtuels*)

- Mais discrétisation temporelle toujours nécessaire

Capteur mobile 1

Capteur mobile 2

Capteur mobile 3



Scène $\mathcal{S}$

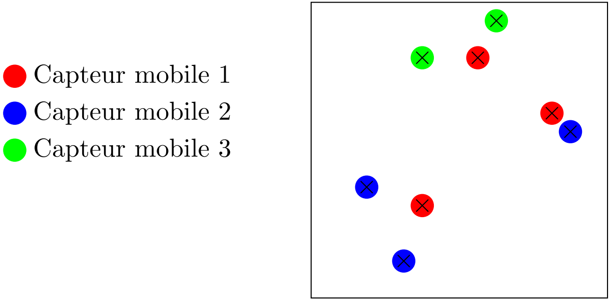
Discussion (1)

Apport des contraintes spatiales

- Connection des capteurs avec le dictionnaire ou le modèle
  - ➯ Plus besoin de "rendez-vous exacts" (au sens de la littérature)

➯ Réduction de la discrétisation spatiale possible (rendez-vous *virtuels*)

- Mais discrétisation temporelle toujours nécessaire



Scène $\mathcal{S}$

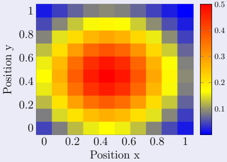
simulation des données

Le réseau

- Simulation de 25 capteurs de PM aux paramètres obtenus par tirages aléatoires $+\begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow F_{theo} \in \mathbb{R}^{2\times 26}$

La scène

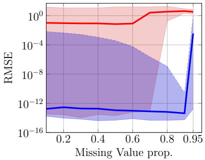
- Scène = 10×10 ($m = 100$) ➪ $G_{theo} \in \mathbb{R}^{100\times 2}$
- Dictionnaire de 62 atomes **décorrélés** t.q. $\boldsymbol{y}$ 2-parcimonieux



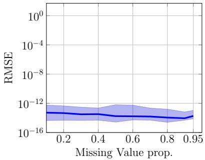
Les paramètres de simulation

- 4 références
- prop. données manquantes (ρ_{DM}) : $10\% \leq \rho_{DM} \leq 95\%$
- prop. rendez-vous (ρ_{RV}) : $0\% \leq \rho_{RV} \leq 100\%$
- bruit d'entrée : $20\,\mathrm{dB} \leq \mathrm{SNR} \leq \infty\mathrm{dB}$

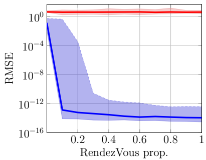
Performances IN-Cal et SpIN-Cal



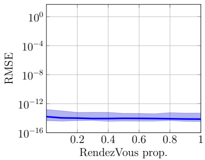
(a) IN-Cal



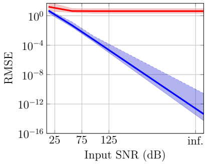
(b) SpIN-Cal



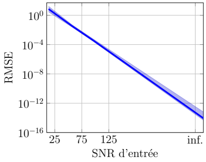
(c) IN-Cal



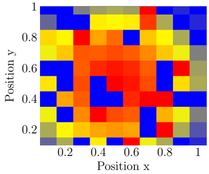
(d) SpIN-Cal



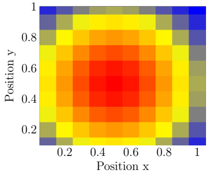
(e) IN-Cal



(f) SpIN-Cal



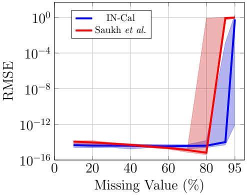
(g) IN-Cal



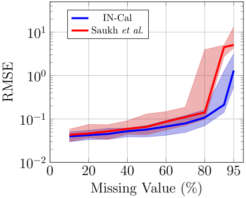
(h) SpIN-Cal

Performances IN-Cal et SpIN-Cal

- Comparison with multi-hop technique (Saukh *et al.* 2014)
  - calibration is sequentially performed
  - at least two rendez-vous between calibrated and uncalibrated sensors
- We compute the product between G^{theo} and F^{theo} to generate X^{theo} and we randomly remove data



(i) SNR=∞dB

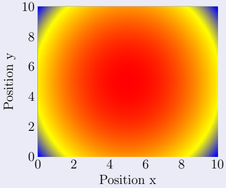


(j) SNR≈ 30 dB

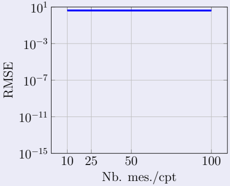
Simulation des données (2)

Nouvelle simulation

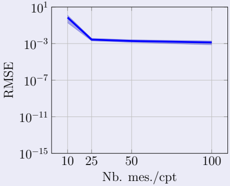
- Scène continue
- ➮ Discrétisation à faire avec IN-Cal (rendez-vous)
- ➮ Dictionnaire avec discrétisation plus fine
- Paramètres de tests : nb. mesures par capteur



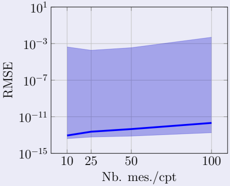
Résultats



(k) IN-Cal



(l) SpIN-Cal



(m) PolIN-Cal

To conclude

Conclusion

- Blind mobile sensor calibration revisited as informed matrix factorization
  - Specific constraints due to the problem (parameterization, sparse priors, known average calibration parameters)
- Proposed approaches robust to the number of missing entries and of rendezvous (no spatial discretization is required with the sparse approximation)
- We also proposed methods for nonlinear calibration models

Current and future work

✔ Faster-than-fast NMF techniques

✔✍ Replacing the dictionary by a model

✍ Multi-scene processing

✍ Case of cross-sensitive sensors

To get more information

1. **Informed NMF for sensor calibration :**
   - C. Dorffer, M. Puigt, G. Delmaire, G. Roussel, *Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration*, IEEE Transactions on Signal and Information Processing over Networks, Volume 4, Issue 4, pp. 667-682, December 2018.
   - C. Dorffer, *Méthodes informées de factorisation matricielle pour l'étalonnage de réseaux de capteurs mobiles et la cartographie de champs de pollution*, Ph.D. thesis, Dec. 2017.
2. **NMF for big data (to be applied to mobile crowdsensing) :**
   - C. Dorffer, M. Puigt, G. Delmaire, G. Roussel, *Fast nonnegative matrix factorization and completion using Nesterov iterations*, in Proceedings of the 13th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2017), Springer International Publishing AG, vol. LNCS 10179, pp. 26-35, Grenoble, France, February 21-24, 2017.
   - F. Yahaya, M. Puigt, G. Delmaire, G. Roussel, *Faster-than-fast NMF using random projections and Nesterov iterations*, in Proceedings of iTWIST : international Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques, Marseille, France, November 21-23, 2018.

Give them a try

Matlab codes available at : `https://gogs.univ-littoral.fr/puigt/Informed_NMF_Mobile_Sensor_Calibration`