



**HAL**  
open science

# Outlier-Robust Calibration Method for Sensor Networks

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, Gilles Roussel

► **To cite this version:**

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, Gilles Roussel. Outlier-Robust Calibration Method for Sensor Networks. 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics (ECMSM), May 2017, San Sebastian, Spain. pp.1-7, 10.1109/ECMSM.2017.7945907 . hal-01498172

**HAL Id: hal-01498172**

**<https://hal.science/hal-01498172v1>**

Submitted on 24 Apr 2017

**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.

# Outlier-Robust Calibration Method for Sensor Networks

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel

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

clement.dorffer@univ-littoral.fr, matthieu.puigt@univ-littoral.fr  
gilles.delmaire@univ-littoral.fr, gilles.rousseau@univ-littoral.fr

**Abstract**—In this paper, we aim to blindly calibrate the responses of a sensor network whose outputs are possibly corrupted by outliers. In particular, we extend some well-known nullspace-based blind calibration approaches, proposed for fixed sensors with affine responses—i.e., with unknown gain and offset for each sensor—to that difficult case. These state-of-the-art approaches assume that the true data lie in a known lower dimensional subspace, so that in practice sensors can be calibrated by projection of the uncalibrated observations to this subspace. A robust extension was recently proposed in order to provide less sensitivity to noise. In this paper, we show that such methods (including the robust extensions) are very sensitive to outliers and we propose new extensions able to deal with such issues. For that purpose, we assume the outliers to be rare events, which can be modeled as a sparse contribution to the low-rank observed data. Using such an assumption, we separate sparse outliers from the low-rank data, so that we can perform calibration. We show that the proposed approach is able to handle up to 10% of outliers in the data without major impact on the calibration accuracy while state-of-the-art methods are already sensitive to the presence of one unique outlier.

## I. INTRODUCTION

While being increasingly cheap, sensors are now part of our daily lives. Indeed, deployed in large-scale networks, they provide massive quantities of data and allow many applications for, e.g., smart cities [1], agriculture [2], or environment monitoring [3]. In such scenarios, the high number of measurements provided by the numerous sensors compensate the poor sensitivity and quality of an individual sensor (which is due to the fact that such sensors are low-cost) [4].

However, an inherent problem to any sensor network is the sensor calibration step which is necessary to get interpretable and consistent data. Calibration usually consists of estimating the unknown parameters that link the sensors outputs to the corresponding physical inputs. Traditionally, calibration can be performed in a laboratory by determining the calibration parameters to fit a known and controlled input to which sensors are exposed to. However, when the sensors are too numerous or inaccessible, calibrating them in laboratory is no longer possible and specific signal processing methods have been proposed to remotely calibrate the sensors of the network from the collected measurements.

In the current literature, these calibration approaches can be classified into two main families, namely micro- and macro-calibration [5]. Micro-calibration consists of calibrating one single sensor of a network at a time while macro-calibration

operates on the whole set of sensors. Depending whether or not the sensors are mobile in the network, different assumptions have been proposed to solve such problem. When the sensors are mobile, they can be in *rendezvous*, i.e., they are in the same spatio-temporal neighborhood, thus sensing the same phenomenon [6]. Such an assumption was recently used<sup>1</sup> in both micro- [8], [9], [10] and macro-calibration<sup>2</sup> [12], [13], [14]. However, in the case of fixed sensors, other assumptions are needed. For fixed sensors, state-of-the-art calibration approaches assume to know the low-rank subspace where the sensed phenomenon lies [15], [16] or consider a compressed sensing framework [17], [18]. Other approaches use statistical properties<sup>3</sup> of the sensed phenomenon to derive estimates of the calibration parameters [19].

Another issue with sensor networks may happen if the sensors are faulty. In such a case, the readings result in outliers which must be taken into account [20]. In this paper, we investigate the influence of outliers on blind calibration. In particular, we focus on the subspace-based approach [15] and its robust extension [16]. We show that these methods are very sensitive to the presence of even a unique outlier and we propose an outlier-robust extension, which assumes that the number of outliers is low with respect to the mass of generated data, i.e., the outliers can be modeled as a sparse contribution in the observed low-rank data.

The remainder of the paper reads as follows. We introduce the sensor calibration problem in Section II. In Section III, we briefly recall the principles of state-of-the-art methods [15] and we show how sensitive to outliers these methods are. In Section IV, we propose a novel outlier-robust calibration method which combines a preprocessing stage—which finds and removes outliers from the sensor readings—to the nullspace-based strategies. In Section V, we experimentally show the enhancement provided by our proposed method. In particular, the preprocessing stage is shown to be able to filter up to 10% of outliers in the data. Lastly, we conclude and discuss the possible extension of the work in Section VI.

For the sake of readability, the notations used in this paper are defined in Table I.

<sup>1</sup>Some approaches, e.g., [7], do not use the rendezvous assumption.

<sup>2</sup>Please note that in that case, macro-calibration has similarities with low-rank matrix completion [11].

<sup>3</sup>Moment-based calibration methods were also proposed for mobile sensor networks, e.g., in [7].

TABLE I  
NOTATIONS USED IN THIS PAPER.

Symbol	Definition
$a$	a scalar
$\underline{a}$	a column vector
$A$	a matrix
$\mathbf{A}$	a diagonal matrix
variables $x$ , $\underline{x}$ , or $X$	the sensed physical phenomenon
variables $y$ , $\underline{y}$ , or $Y$	the outlier-free contributions of the sensor readings
variables $z$ , $\underline{z}$ , or $Z$	the sensor readings
$\mathcal{P}(\cdot)$	the projection operator on the nullspace of $\underline{x}(t)$

## II. PROBLEM STATEMENT

In this paper, we assume to observe the sensed signals  $\underline{z}(t) \in \mathbb{R}^n$  provided by a synchronized network composed of  $n$  fixed sensors. Moreover, we also assume that the sensors are possibly defecting. Sensor faults might be due to, e.g., communication errors between sensors, signal subsampling due to desynchronization (assuming that missing samples are set to zero), sparse or continuous spurious sensor readings (see Fig. 1). In that case, the corrupted sensor readings  $\underline{z}(t)$ , observed at time  $t$  read

$$\underline{z}(t) = \underline{y}(t) + \underline{o}(t), \quad (1)$$

where  $\underline{y}(t)$  and  $\underline{o}(t)$  are the vectors of outlier-free sensor readings and outliers, respectively. In this paper, we assume the outliers to be sparse, i.e., most values in  $\underline{o}(t)$  are null.

If one consider that sensor readings are outlier-free, i.e.,  $\underline{o}(t) = 0$  for any time  $t$ , then one can directly observe  $\underline{y}(t)$ . If we assume that the calibration functions for each sensor of the network to be affine, then the true input phenomenon—denoted  $x_i(t)$ —reads

$$x_i(t) = y_i(t) \cdot \alpha_i + \beta_i, \quad (2)$$

where  $y_i(t)$  is the signal observed by the  $i$ -th sensor at time  $t$ , and  $\alpha_i$  and  $\beta_i$  are the unknown sensor gain and offset associated with the sensor  $i$ , respectively. Then, by defining  $\underline{x}(t) \in \mathbb{R}^n$  as the column vector containing the input phenomenon observed by the  $n$  sensors and  $\mathbf{Y}(t)$  the diagonal matrix of  $\underline{y}(t)$ , one can write

$$\underline{x}(t) = \mathbf{Y}(t) \cdot \underline{\alpha} + \underline{\beta}, \quad (3)$$

where  $\underline{\alpha}$  and  $\underline{\beta}$  are the vectors containing the sensor gains and offsets, respectively. Calibrating the sensor network then consists of estimating  $\underline{\alpha}$  and  $\underline{\beta}$ .

For that purpose, we briefly recall the assumptions made in [15], upon which we build our proposed method. We first assume that the sensor network is oversampling the observed signal, i.e.,  $\underline{x}(t)$  lies in a low dimensional subspace  $\mathcal{S}$  which is assumed to be spanned by a set of  $r$  basis vectors, where  $r \ll n$  is the rank of  $\mathcal{S}$ . These basis vectors are assumed to be known. Such an assumption might look strong. However, it arises in several situations, e.g., for bandlimited signals with sensor distances lower than required by the Shannon-Nyquist sampling rate [15]. It is also valid in a machine learning framework—where the sensors are calibrated in their early

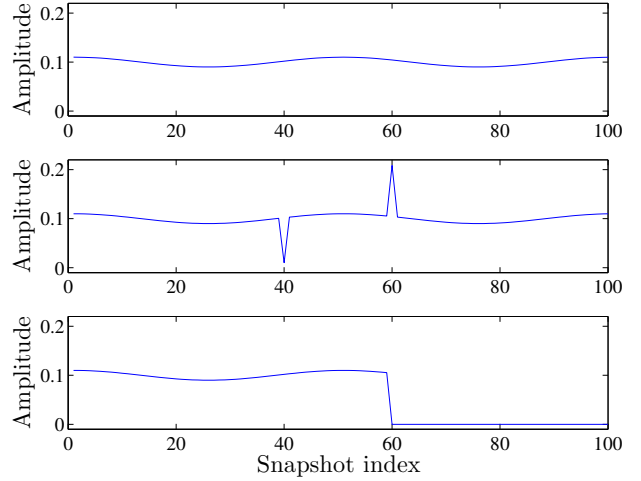


Fig. 1. Examples of sensor faults: (top) outlier-free sensor reading, (middle) presence of 2 local sensor faults (e.g., constant or bottom-noise fault), and (bottom) fault due to defective sensor or communication lost.

life—so that  $\mathcal{S}$  can be learned from early sensor readings. Then, while sensors are aging, their responses drift with time, i.e., their calibration parameters vary. This implies that at a given time  $t$ , the sensor readings follow the calibration model (3). Besides, as  $\mathcal{S}$  is known, one can compute  $\mathcal{P}$ , the linear projection on its orthogonal subspace. The key idea used in [15] (and in its robust extension [16]) consists of noticing that  $\underline{x}(t)$  is in the nullspace of  $\mathcal{P}$ , i.e.,

$$\mathcal{P}(\underline{x}(t)) = \mathcal{P}(\mathbf{Y}(t) \cdot \underline{\alpha} + \underline{\beta}) = 0. \quad (4)$$

In practice, we assume to collect synchronized snapshots  $\underline{y}(t_1), \underline{y}(t_2), \dots, \underline{y}(t_k)$  from the sensor network, where  $t_j$  denotes the time of the  $j$ -th snapshot. In order to perform calibration, the number  $k$  of snapshots satisfies [15]

$$k \geq \left\lceil \frac{n-1}{n-r} \right\rceil + 1. \quad (5)$$

Defining  $\mathbf{Y}_j \triangleq \mathbf{Y}(t_j)$ , the nullspace blind calibration problem thus aims to solve

$$\{\hat{\underline{\alpha}}, \hat{\underline{\beta}}\} = \arg \min_{\underline{\alpha}, \underline{\beta}} |\mathcal{P}(\mathbf{Y}_j \cdot \underline{\alpha} + \underline{\beta})|, \quad \forall j \in \llbracket 1, k \rrbracket. \quad (6)$$

In the next sections, we briefly introduce the principles of the nullspace-based calibration methods [15], [16]. We then show that they are very sensitive to the presence of outliers in the observed signals—i.e., when  $\underline{o}(t) \neq 0$ —and we propose a novel nullspace calibration method which introduces a preprocessing stage to remove outliers.

## III. NULLSPACE-BASED SENSOR CALIBRATION

We now briefly recall the principles of the nullspace-based calibration methods. As  $\mathcal{P}$  is linear and defining

$$\bar{\mathbf{Y}} \triangleq \frac{1}{k} \sum_{j=1}^k \mathbf{Y}_j, \quad (7)$$

the centered version of Eq. (4) reads

$$\mathcal{P}((\mathbf{Y}_j - \bar{\mathbf{Y}}) \cdot \underline{\alpha}) = 0, \quad \forall j \in \llbracket 1, k \rrbracket, \quad (8)$$

which yields

$$\mathcal{P}(\mathbf{Y}_j - \bar{\mathbf{Y}}) \cdot \underline{\alpha} = 0, \quad \forall j \in \llbracket 1, k \rrbracket. \quad (9)$$

The authors in [15] then defined a matrix  $C$  by stacking the projection in each snapshot, i.e.,

$$C \triangleq \begin{bmatrix} \mathcal{P}(\mathbf{Y}_1 - \bar{\mathbf{Y}}) \\ \vdots \\ \mathcal{P}(\mathbf{Y}_k - \bar{\mathbf{Y}}) \end{bmatrix}. \quad (10)$$

Combining Eqs. (9) and (10) implies that the unknown gain vector  $\underline{\alpha}$  is solution of

$$\hat{\underline{\alpha}} = \arg \min_{\underline{\alpha}} \|C \cdot \underline{\alpha}\|_2, \quad (11)$$

which can be solved using ordinary Least Square (LS) or SVD—noticing that the last eigenvector is proportional to  $\underline{\alpha}$ —as proposed in [15]. However, the solutions of Eq. (11) are not unique: indeed, if  $\underline{\alpha}$  is a solution to Eq. (11), then  $\lambda \cdot \underline{\alpha}$  is also a solution. The authors in [15], [16] removed the scale ambiguity by setting the first element<sup>4</sup> of  $\underline{\alpha}$  to 1.

The authors in [16] then proposed an extension of [15] which uses Total Least Square (TLS) to solve Eq. (11). TLS mainly consists of a weighted SVD where the weights allow to take into account the accuracy of each sensor.

Once the gain vector is estimated, the authors in [15] estimate the offsets<sup>5</sup>, assuming the observed phenomenon is zero-mean (or at least its average value is known). In that case, one easily derive

$$\underline{\beta} = -\bar{\mathbf{Y}} \cdot \underline{\alpha}. \quad (12)$$

Finally, the calibration process is summarized in Algorithm 1 and provides an interesting performance while remaining easy to implement. However, the projection step makes

---

**Algorithm 1** Nullspace-based calibration methods [15], [16].

---

- Center the measurement matrices  $\mathbf{Y}_j$  (see Eq. (9)) ;
  - Construct  $C$  (see Eq. (10)) ;
  - Estimate  $\underline{\alpha}$  from Eq. (11) using LS, SVD or TLS ;
  - Estimate  $\underline{\beta}$  (see Eq. (12)).
- 

the method very sensitive to outliers. Indeed, in a realistic framework, the measurements provided by the sensor network can be corrupted by outliers, as explained in Section II. The outlier-free vectors  $\underline{y}(t_j)$  are thus inaccessible and one must deal with corrupted sensor readings  $\underline{z}(t_j)$  instead. If

<sup>4</sup>Please note that other strategies—e.g., assuming to know the sum of the values of the unknown calibration parameters [18]—have been proposed in the literature and can be also applied with these methods instead. In the remainder of the paper, we follow the same strategy as in [15], [16].

<sup>5</sup>An approach for estimating  $\underline{\beta}$  when  $n \ll m$  offsets are known was also proposed in [15].

one directly applies the above nullspace-based calibration techniques on  $\underline{z}(t_j)$ , then

$$\mathcal{P}(\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta}) = \mathcal{P}((\mathbf{Y}_j + \mathbf{O}_j) \cdot \underline{\alpha} + \underline{\beta}), \quad (13)$$

$$= \mathcal{P}(\mathbf{Y}_j \cdot \underline{\alpha} + \underline{\beta}) + \mathcal{P}(\mathbf{O}_j \cdot \underline{\alpha}), \quad (14)$$

where  $\mathbf{Z}_j \triangleq \text{diag}(\underline{z}(t_j))$  and  $\mathbf{O}_j \triangleq \text{diag}(\underline{o}(t_j))$ . It turns out that the estimation of  $\underline{\alpha}$  is perturbed by  $\mathbf{O}_j$ , which also implies a poor estimation accuracy of  $\underline{\beta}$ . Please note that the TLS method should be robust to outliers if the latter are present in very few sensors. Indeed, by lowering the weights associated to the relevant sensors, it should be still possible to perform accurate calibration on the outlier-free sensors. However, TLS should provide a very limited enhancement (or even no enhancement at all) if the outliers are distributed over all the sensors readings, which is more likely to happen in practice.

It should be noticed that the above methods were also extended to a partially blind calibration problem in [15], [16]. In that case, some sensor calibration parameters are assumed to be known, thus improving the calibration performance. However, even these methods are sensitive to outliers as they are also based on nullspace projection.

As a consequence, a novel approach able to handle outliers must be proposed, which we introduce in the next section.

#### IV. PROPOSED OUTLIER-ROBUST CALIBRATION METHOD

The outliers that may corrupt the data can be seen as sparse noisy data. In sensor networks, it usually corresponds to spurious readings or transmission faults.

At this stage, it should be noticed that  $\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta}$  in Eq. (13) can also be expressed as

$$\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta} = \mathbf{A} \cdot \underline{z}(t_j) + \mathbf{B}, \quad (15)$$

where  $\mathbf{A} \triangleq \text{diag}(\underline{\alpha})$  and  $\mathbf{B} \triangleq \text{diag}(\underline{\beta})$ , respectively. We now define the matrix  $Z \triangleq [\underline{z}(t_1), \dots, \underline{z}(t_k)]$ , i.e., whose  $(i, j)$ -th element reads

$$(Z)_{ij} \triangleq z_i(t_j). \quad (16)$$

From Eq. (1) and Eq. (16), we can derive

$$Z = Y + O, \quad (17)$$

where  $Y$  and  $O$  contain the entries of  $\underline{y}(t_j)$  and  $\underline{o}(t_j)$ , respectively.

Using the same formalism as in Eq. (15), Eq. (3) provides

$$X = \mathbf{A} \cdot Y + \mathbf{B}, \quad (18)$$

where  $X \triangleq [\underline{x}(t_1), \dots, \underline{x}(t_k)]$ . Since  $\underline{x}(t)$  lies in a low dimensional subspace  $\mathcal{S}$  defined by  $r \ll \min(n, k)$  basis functions, it turns out that  $X$  is a low-rank matrix, and as a consequence,  $Y$  is also low-rank. As  $O$  is assumed to be sparse, it turns out that the sensor readings  $Z$  are written as the sum of a low-rank and a sparse matrix, which has been intensively studied under the name of Robust Principal

Component Analysis (RPCA) since the pioneering work in [21], [22]. RPCA aims to solve

$$\min_{Y,O} \|Y\|_*^2 + \lambda \|O\|_1, \quad \text{s.t. } Z = Y + O, \quad (19)$$

where  $\|\cdot\|_*$  is the nuclear norm of a matrix—i.e., the sum of its eigenvalues— $\|\cdot\|_1$  is the  $\ell_1$  norm, and  $\lambda$  a positive weighting parameter. RPCA became a popular tool as it finds numerous applications, e.g., in video surveillance, face recognition, or ranking and collaborative filtering [22].

Numerous RPCA algorithms have been proposed in the literature and can thus be applied to  $Z$ , in order to estimate  $Y$  and  $O$ . In some preliminary tests, we found the Inexact Augmented Lagrange Multiplier (IALM) method [23] to be fast and accurate for the considered application. As most RPCA algorithms, the IALM-RPCA technique is iterative and computes partial SVDs of the size of  $Z$  to estimate  $O$  and to derive  $\hat{Y}$ , an estimation of  $Y$ .

Then, we can easily extract the vectors  $\hat{y}(t_j)$  and build the diagonal matrices  $\hat{Y}_j, \forall j \in \llbracket 1, k \rrbracket$ . At this stage, it is straightforward to perform blind calibration by applying any of the above projection-based calibration techniques to  $\hat{C}$ , the matrix defined in Eq. (10) using the estimated  $\hat{Y}_j$ . The structure of the proposed methods is provided in Algorithm 2. Let us stress that using RPCA as a preprocessing stage of

---

**Algorithm 2** Proposed outlier-robust calibration methods.

---

- Construct  $Z$  (see Eq. (16))
  - Estimate  $\hat{Y}$  using IALM-RPCA [23]
  - Center the matrices  $\hat{Y}_j$  (see Eq. (9)) ;
  - Construct  $\hat{C}$  (see Eq. (10)) ;
  - Estimate  $\underline{\alpha}$  from Eq. (11) using LS, SVD or TLS ;
  - Estimate  $\underline{\beta}$  (see Eq. (12)).
- 

the nullspace-based calibration does not involve any extra-assumption. Moreover, in a machine learning framework, it can also be applied to early readings—where the sensors are assumed to be calibrated—as sensor faults or communication errors might happen. In that case, the estimation of  $\mathcal{S}$  might be improved as well. Lastly, such a preprocessing can also be applied to the partially blind extensions of the above nullspace-based blind calibration methods, as proposed in [15], [16].

In the next section, we investigate the enhancement provided by our proposed method with respect to the state-of-the-art ones, in both blind and partially blind scenarios.

## V. EXPERIMENTAL VALIDATION AND DISCUSSION

In this section, we investigate the performance of the proposed outlier-robust calibration methods. To that end, we simulate a low-dimensional subspace  $\mathcal{S}$  of rank  $r = 20$  and we derive a set of  $n = 100$  sensors by randomly picking 100 gains and offsets. We then simulate  $k = \lceil 3 \cdot r \cdot \log(n) \rceil$  signals from the subspace  $\mathcal{S}$ —as done in [16]—and we derive the outlier-free sensor responses associated to each signal—using Eq. (2)—thus providing the vectors  $y_j, \forall j \in \llbracket 1, k \rrbracket$ .

In the tests below, we explore the effects of both the subspace error and the number of outliers on the calibration performance. Indeed, as the subspace is assumed to be learned prior to the calibration, it may be not perfectly known. Its estimation error is simulated by adding random zero-mean Gaussian perturbations to the theoretical subspace with increasing variance [15]. Such an error is then estimated as

$$\epsilon_{\mathcal{S}} \triangleq \frac{1}{k} \cdot \sum_{i=1}^k \frac{\|P(x_i)\|_2}{\|x_i\|_2}. \quad (20)$$

In the tests reported below, the values of  $\epsilon_{\mathcal{S}}$  range from 0 to 0.45. A few outliers are then randomly added to the observed data matrix  $Y$ . Their proportion  $\rho_o$  ranges between 0 and 20% of the number of data points in  $Z$ . For each pair of parameters  $(\epsilon_{\mathcal{S}}, \rho_o)$ , we apply the LS [15] and TLS [16] nullspace-based calibration methods to the sensor readings corrupted by outliers. The weights used in TLS are all set to 1, so that TLS behaves as SVD. We also apply our proposed extension, where IALM-RPCA is applied to  $Z$  before applying LS or TLS calibration.

As a performance criterion, we compute the error between the estimated parameters  $\underline{\hat{\alpha}}$  and the true ones  $\underline{\alpha}$  using

$$\text{Error} = \frac{\|\underline{\alpha} - \underline{\hat{\alpha}}\|_2}{\|\underline{\alpha}\|_2}. \quad (21)$$

Figure 2 shows the median performance of the tested methods obtained over 100 trials. Let us first focus on the enhancement reached by the state-of-the-art methods. When the proportion of outliers is set to 0, the TLS calibration approach outperforms the LS one, which is consistent with [16]. While the median calibration error reached using LS is above 0.9 when the subspace error  $\epsilon_{\mathcal{S}}$  is above 0.26, the TLS-based technique provides median calibration errors always below 0.30. However, its spread (shown by considering minimum and maximum errors) is higher than with LS. As an example, the errors reached when  $\epsilon_{\mathcal{S}} = 0.45$  are between 0.17 and 0.92 with TLS calibration while they are between 0.94 and 0.98 with LS calibration. When some outliers are added to the data, the calibration error reached by both methods significantly increases. However, their behaviour is not similar. On the one hand, the median error reached with LS calibration remains close to 1 for almost any tested value  $\epsilon_{\mathcal{S}}$  of the subspace error (it is always above 0.9 when only  $\rho_o = 0.2\%$  of outliers are present in the data, for any initial subspace error). On the contrary, the median TLS calibration error remains always below 0.6 when  $\rho_o \leq 0.2\%$ . However, for higher values of  $\rho_o$ , the median TLS calibration error drastically increases (it is always above 10 when  $\rho_o \geq 2\%$ , for any tested value of  $\epsilon_{\mathcal{S}}$ ), thus showing that such a robust approach is actually very sensitive to the presence of outliers.

On the contrary, both our proposed methods—which combine the IALM-RPCA preprocessing with the above calibration techniques—provide almost constant performance up to  $\rho_o = 10\%$  of outliers. Interestingly, the preprocessing does not affect the calibration performance when no outlier is

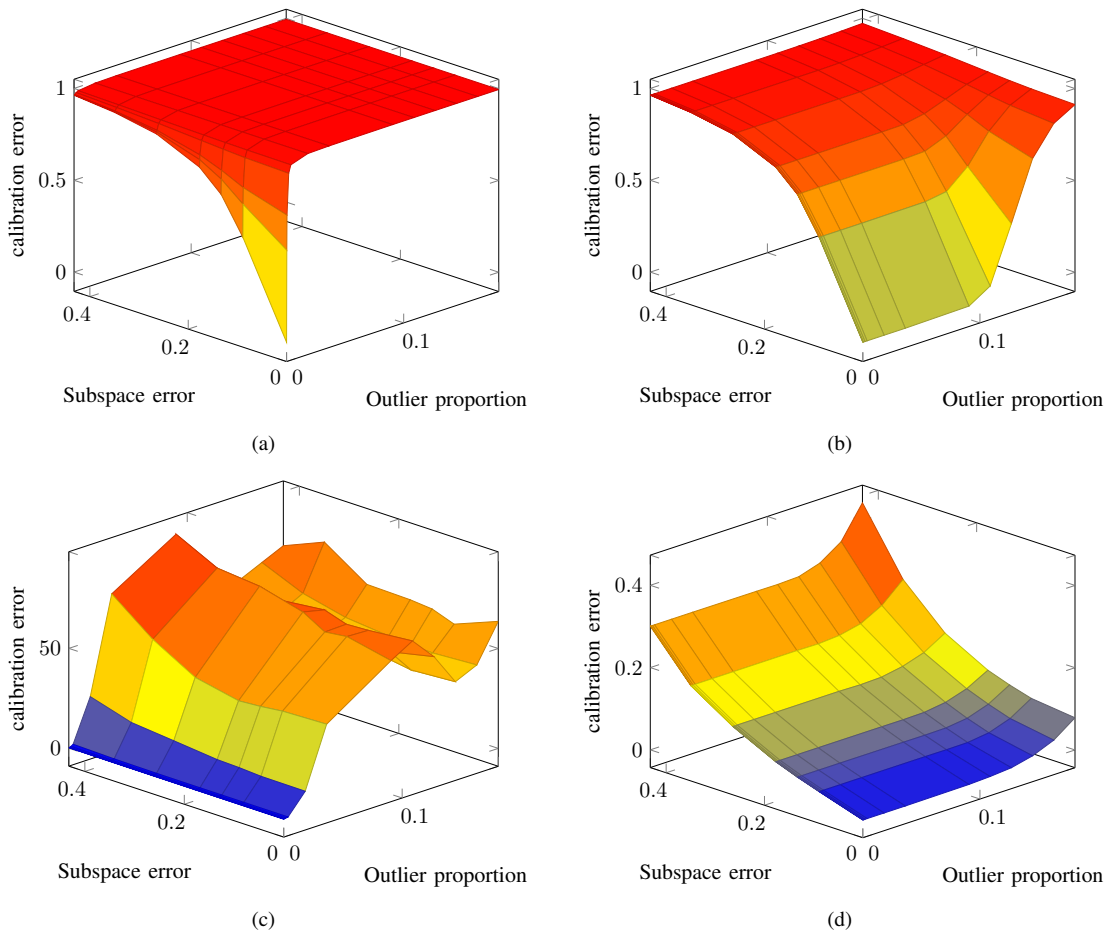


Fig. 2. Performance of the nullspace-based calibration methods with respect to the subspace estimation error  $\epsilon_S$  and the outlier proportion  $\rho_o$ . The nullspace in Eq. (11) is estimated using: (a) LS [15], (b) RPCA preprocessing + LS, (c) TLS [16], and (d) RPCA preprocessing + TLS.

present in the data. Indeed, in that case, the performance reached by the state-of-the-art calibration methods and their respective proposed extensions are almost equal. When the outlier proportion is higher than 20% (the situation is not plotted on Fig. 2), the outlier matrix  $O$  is not very sparse and the IALM-RPCA method fails in accurately estimating it, thus leading to a degradation of the calibration performance.

We also investigated the case of partially blind (PB) calibration. We reproduced the same experiment as above, except that we considered a set of 5 sensors to be perfectly calibrated, i.e., their calibration parameters are assumed to be known. Due to space consideration, we cannot show the plots on the paper but the performance reached by the state-of-the-art and the proposed methods are consistent with the blind case. Indeed, the PB state-of-the-art techniques using LS and TLS show the same sensitivity to outliers than the blind LS and TLS approached tested above. The use of well calibrated sensors is then not sufficient to make these methods robust to outliers. Using RPCA as a preprocessing for PB calibration methods allows some robustness to the presence of up to 10% of outliers while it does not affect the performance of the methods in outlier-free simulations.

It should be noticed that in the nullspace-based calibration

problem, the rank of  $Y$  is assumed to be known, as it was estimated with  $\mathcal{S}$ . This knowledge is not fully taken into account in the IALM-RPCA algorithm, as it aims to find a matrix of minimal rank but it does not take into account the known rank  $r + 1$  of matrix  $Y$ . Keeping the rank equality constraint in RPCA might allow to filter additive noise in the sensor readings, and thus to improve the performance reached by the proposed techniques.

## VI. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a robust extension of nullspace-based blind calibration methods. We reordered the collected data in a low-rank matrix that we filtered using an RPCA algorithm. While not adding any extra-assumption, such a preprocessing stage allows removing up to 10% of outliers without any loss of calibration performance. This work opens many perspectives. We recently proposed some mobile sensor calibration techniques based on informed matrix factorization [12], [13], [14]. In future work, we will investigate some outlier-robust extensions of these approaches, using a similar low-rank modeling. We will compare such a formalism to robust informed matrix factorization using parametric divergences [24], [25] or the Huber norm [26], that we recently

proposed for another application [27]. Moreover, the proposed techniques need to know the low-rank subspace where the sensed phenomenon lie. We will investigate the use of machine learning to learn  $\mathcal{S}$  and to perform sensor calibration.

#### ACKNOWLEDGEMENT

This work was funded partly by the ‘‘OSCAR’’ project within the R egion Hauts-de-France ‘‘Chercheurs Citoyens’’ Program, and partly by the CNRS MASTODONS project under the grant ‘‘DoMaSQ’Air 2017’’.

#### REFERENCES

- [1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, ‘‘Internet of things for smart cities,’’ *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
- [2] A. U. Rehman, A. Z. Abbasi, N. Islam, and Z. A. Shaikh, ‘‘A review of wireless sensors and networks’ applications in agriculture,’’ *Computer Standards & Interfaces*, vol. 36, no. 2, pp. 263–270, 2014.
- [3] M. Puigt, G. Delmaire, and G. Roussel, ‘‘Environmental signal processing: new trends and applications,’’ in *Proc. of ESANN*, 2017.
- [4] E. Cochran, J. Lawrence, A. Kaiser, B. Fry, A. Chung, and C. Christensen, ‘‘Comparison between low-cost and traditional MEMS accelerometers: a case study from the M7.1 Darfield, New Zealand, aftershock deployment,’’ *Annals of Geophysics*, vol. 54, no. 6, 2012.
- [5] K. Whitehouse and D. Culler, ‘‘Macro-calibration in sensor/actuator networks,’’ *Mobile Networks and Applications*, vol. 8, no. 4, pp. 463–472, Aug. 2003.
- [6] O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele, ‘‘On rendezvous in mobile sensing networks,’’ in *Proc. of REALWSN*, 2014, vol. 281 of LNCS, pp. 29–42.
- [7] C. Wang, P. Ramanathan, and K.K. Saluja, ‘‘Moments based blind calibration in mobile sensor networks,’’ in *Proc. of ICC ’08*, May 2008, pp. 896–900.
- [8] O. Saukh, D. Hasenfratz, and L. Thiele, ‘‘Reducing multi-hop calibration errors in large-scale mobile sensor networks,’’ in *Proc. of IPSN*, 2015.
- [9] E. Miluzzo, N.D. Lane, A.T. Campbell, and R. Olfati-Saber, ‘‘CaliBree: A self-calibration system for mobile sensor networks,’’ in *Proc. of DCOSS*, 2008, vol. 5067 of LNCS, pp. 314–331.
- [10] B.-T. Lee, S.-C. Son, and K. Kang, ‘‘A blind calibration scheme exploiting mutual calibration relationships for a dense mobile sensor network,’’ *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1518–1526, May 2014.
- [11] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, ‘‘Fast nonnegative matrix factorization and completion using nesterov iterations,’’ in *Proc. of LVA-ICA*, vol. 10179 of LNCS, pp. 26–35. 2017.
- [12] 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 of LNCS, pp. 497–505. 2015.
- [13] 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, pp. 2941–2945.
- [14] 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*, 2016.
- [15] L. Balzano and R. Nowak, ‘‘Blind calibration of sensor networks,’’ in *Proc. of IPSN*, 2007, pp. 79–88.
- [16] J. Lipor and L. Balzano, ‘‘Robust blind calibration via total least squares,’’ in *Proc. of ICASSP*, May 2014, pp. 4244–4248.
- [17] C. Schulke, F. Caltagirone, F. Krzakala, and L. Zdeborova, ‘‘Blind calibration in compressed sensing using message passing algorithms,’’ in *NIPS* 26, 2013, pp. 566–574.
- [18] C. Bilen, G. Puy, R. Gribonval, and L. Daudet, ‘‘Convex optimization approaches for blind sensor calibration using sparsity,’’ *IEEE Trans. on Signal Processing*, vol. 62, no. 18, pp. 4847–4856, Sept 2014.
- [19] H. Carfantan and J. Idier, ‘‘Statistical linear destriping of satellite-based pushbroom-type images,’’ *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1860–1871, April 2010.
- [20] Y. Zhang, N. Meratnia, and P. Havinga, ‘‘Outlier detection techniques for wireless sensor networks: A survey,’’ *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.
- [21] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.
- [22] E. J. Cand es, X. Li, Y. Ma, and J. Wright, ‘‘Robust principal component analysis?,’’ *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.
- [23] Z. Lin, M. Chen, and Y. Ma, ‘‘The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,’’ UIUC Technical Report ULLU-ENG-09-2215, 2009.
- [24] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, ‘‘Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment,’’ in *Proc. of MLSP*, 2013.
- [25] R. Chreiky, G. Delmaire, C. Dorffer, M. Puigt, G. Roussel, and A. Abche, ‘‘Robust informed split gradient nmf using alpha beta-divergence for source apportionment,’’ in *Proc. of MLSP*, 2016.
- [26] R. Chreiky, G. Delmaire, M. Puigt, G. Roussel, and A. Abche, ‘‘Informed split gradient non-negative matrix factorization using huber cost function for source apportionment,’’ in *Proc. of ISSPIT*, 2016.
- [27] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, ‘‘Non-negative matrix factorization under equality constraints—a study of industrial source identification,’’ *Applied Numerical Mathematics*, vol. 85, pp. 1–15, November 2014.