



**HAL**  
open science

# In situ calibration of low-cost instrumentation for the measurement of ambient quantities: evaluation methodology of the algorithms and diagnosis of drifts

Florentin Delaine

## ► To cite this version:

Florentin Delaine. In situ calibration of low-cost instrumentation for the measurement of ambient quantities: evaluation methodology of the algorithms and diagnosis of drifts. Data Structures and Algorithms [cs.DS]. Institut Polytechnique de Paris, 2020. English. NNT: 2020IPPAX075. tel-03086234

**HAL Id: tel-03086234**

**<https://theses.hal.science/tel-03086234>**

Submitted on 22 Dec 2020

**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.



INSTITUT  
POLYTECHNIQUE  
DE PARIS

NNT : 2020IPPAX075

Thèse de doctorat

 Université  
Gustave Eiffel

 efficacity



# Étalonnage in situ de l'instrumentation bas coût pour la mesure de grandeurs ambiantes : méthode d'évaluation des algorithmes et diagnostic des dérives

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à École Polytechnique

École doctorale n°626 École Doctorale de l'Institut Polytechnique de Paris (ED IP  
Paris)  
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Champs-sur-Marne, le 04/12/2020, par

**FLORENTIN DELAINE**

Composition du Jury :

Gilles Roussel Professeur des Universités, Université du Littoral Côte d'Opale (LISIC)	Président
Jean-Luc Bertrand-Krajewski Professeur des Universités, Université de Lyon, INSA Lyon (DEEP)	Rapporteur
Romain Rouvoy Professeur des Universités, Université de Lille (Spirals)	Rapporteur
Nathalie Redon Maître de conférences, IMT Lille Douai (SAGE)	Examinatrice
Bérengère Lebental Directrice de Recherche, Institut Polytechnique de Paris, École Polytechnique (LPICM)	Directrice de thèse
Hervé Rivano Professeur des Universités, Université de Lyon, INSA Lyon (CITILab)	Co-directeur de thèse
Éric Peirano Directeur général adjoint en charge de la R&D, Efficacity	Invité
Matthieu Puigt Maître de conférences, Université du Littoral Côte d'Opale (LISIC)	Invité



# Acknowledgments

---

3 *To facilitate the understanding of this part by the people it concerns, French is used here.*

4 Pour commencer ces remerciements, je souhaite tout d'abord m'adresser au jury qui a évalué  
5 ces travaux. Merci à Gilles Roussel d'avoir accepté d'examiner cette thèse et d'en présider le  
6 jury. Je remercie également Jean-Luc Bertrand-Krajewski et Romain Rouvoy pour avoir accepté  
7 de rapporter mes travaux de thèse dans le contexte sanitaire actuel qui bouleverse les activités  
8 des enseignants-chercheurs. Merci pour vos rapports et les questions qui y ont figuré, m'aidant à  
9 préparer la soutenance. Merci également à Nathalie Redon d'avoir examiné mes travaux et enfin à  
10 Éric Peirano et Matthieu Puigt d'avoir pris part au jury en tant qu'invités. J'ai particulièrement  
11 apprécié nos échanges, notamment à travers les questions que vous avez chacun pu avoir sur mes  
12 travaux.

13 Je souhaite m'adresser ensuite à Bérengère Lebental et Hervé Rivano, respectivement  
14 directrice et co-directeur de thèse. Nous venons chacun de domaines différents et il y a parfois eu  
15 des incompréhensions entre nous liées à cela, ajouté aux difficultés rencontrées lorsque j'avais une  
16 idée en tête et qu'il était complexe de m'en faire dévier. Ça n'a pas toujours été simple d'avancer,  
17 cette thèse aurait pu ne pas aller à son terme mais nous y sommes parvenus. Aussi, vous m'avez  
18 accordé une grande liberté durant ces trois années et rien que pour cela, un grand merci à vous  
19 deux. J'en tire de nombreux enseignements pour la suite de mon parcours dans la vie.

20 Mes mots vont ensuite à Sio-Song Ieng, qui m'a mis en relation avec Bérengère alors que  
21 je cherchais une thèse suite à l'échec du financement d'un premier projet. Grâce à toi, j'ai pu  
22 poursuivre mon parcours académique et devenir docteur.

23 Je souhaite également m'adresser aux collègues des différentes entités qui ont gravité autour  
24 de cette thèse. Je pense tout d'abord à ceux d'Efficity où j'ai passé la plus grande partie de  
25 mon temps, si l'on excepte cette dernière année perturbée par le COVID-19, et à ceux du LISIS,  
26 dont je n'étais pas très loin en étant dans le même bâtiment. Merci à vous. Il y a toujours eu  
27 quelqu'un avec qui discuter, en particulier pour me remonter le moral lorsque ça n'allait pas  
28 ou pour me laisser évacuer. J'ai une pensée particulière pour Maud et Rémi, mes co-bureaux:  
29 bon courage pour la fin de vos thèses respectives. Merci aux membres de l'équipe Agora du  
30 laboratoire CITI qui m'ont toujours bien accueilli lors de mes déplacements du côté de Lyon et  
31 avec qui j'aurais souhaité passer plus de temps. Trois ans ça file vite !

32 Merci à tous les copains et copines qui ont permis de décompresser régulièrement durant ces  
33 trois années et de sortir la tête de l'eau. Je pense à vous et ne cite personne ici en particulier car  
34 je m'en voudrais de commettre des oublis !

35 Merci à ma famille et à ma belle-famille qui sont également des éléments importants dans  
36 ma vie, au-delà de cette thèse, et que je ne vois que trop peu par manque de temps et nos  
37 éloignements géographiques. Je remercie ma Maman avec un grand M majuscule. Il faut bien  
38 cela pour le pilier qu'elle est pour moi dans la vie et pour tout ce qu'elle a pu m'apporter (et  
39 m'apporte encore parfois) comme bases dans la vie. Je pense aussi à Philippe qui est mon autre  
40 pilier majeur et qui a contribué à ma construction aux côtés de ma Maman.

41 Merci à Clémentine qui partage ma vie depuis quelques temps maintenant. Elle m'a ac-

42 accompagné tout au long de cette thèse jour après jour, partageant mes joies et mes peines, me  
43 supportant tant bien que mal quand le professionnel rejaillissait sur le privé, me remettant en  
44 place quand il le fallait et m'apportant tout son amour pour aller au bout de cette aventure.

45 Enfin, je souhaite remercier les enseignants que j'ai pu avoir tout au long de mon parcours.  
46 Certains et certaines ont été plus déterminants que d'autres bien évidemment mais vous faites  
47 toutes et tous un métier que j'admire, qui est un dont je rêve et que je ne chosis pas pour  
48 le moment. Au quotidien par vos enseignements, vous avez permis que j'en arrive là. Merci.  
49 J'admire tous ceux dans mon entourage qui ont choisi cette voie et qui permettront sans doute à  
50 d'autres de réaliser des parcours auxquels ils n'étaient pas prédestinés à la naissance.

52 In various fields going from agriculture to public health, ambient quantities have to be  
53 monitored in indoors or outdoors areas. For example, temperature, air pollutants, water  
54 pollutants, noise and so on have to be tracked. To better understand these various phenomena,  
55 an increase of the density of measuring instruments is currently necessary. For instance, this  
56 would help to analyse the effective exposure of people to nuisances such as air pollutants.

57 The massive deployment of sensors in the environment is made possible by the decreasing  
58 costs of measuring systems, mainly using sensitive elements based on micro or nano technologies.  
59 The drawback of this type of instrumentation is a low quality of measurement, consequently  
60 lowering the confidence in produced data and/or a drastic increase of the instrumentation costs  
61 due to necessary recalibration procedures or periodical replacement of sensors.

62 There are multiple algorithms in the literature offering the possibility to perform the calibration  
63 of measuring instruments while leaving them deployed in the field, called *in situ* calibration  
64 techniques.

65 The objective of this thesis is to contribute to the research effort on the improvement of data  
66 quality for low-cost measuring instruments through their *in situ* calibration.

67 In particular, we aim at 1) facilitating the identification of existing *in situ* calibration  
68 strategies applicable to a sensor network depending on its properties and the characteristics of its  
69 instruments; 2) helping to choose the most suitable algorithm depending on the sensor network  
70 and its context of deployment; 3) improving the efficiency of *in situ* calibration strategies through  
71 the diagnosis of instruments that have drifted in a sensor network.

72 Three main contributions are made in this work. First, a unified terminology is proposed  
73 to classify the existing works on *in situ* calibration. The review carried out based on this  
74 taxonomy showed there are numerous contributions on the subject, covering a wide variety of  
75 cases. Nevertheless, the classification of the existing works in terms of performances was difficult  
76 as there is no reference case study for the evaluation of these algorithms.

77 Therefore in a second step, a framework for the simulation of sensors networks is introduced.  
78 It is aimed at evaluating *in situ* calibration algorithms. A detailed case study is provided across  
79 the evaluation of *in situ* calibration algorithms for blind static sensor networks. An analysis  
80 of the influence of the parameters and of the metrics used to derive the results is also carried  
81 out. As the results are case specific, and as most of the algorithms recalibrate instruments  
82 without evaluating first if they actually need it, an identification tool enabling to determine the  
83 instruments that are actually faulty in terms of drift would be valuable.

84 Consequently, the third contribution of this thesis is a diagnosis algorithm targeting drift  
85 faults in sensor networks without making any assumption on the kind of sensor network at stake.  
86 Based on the concept of rendez-vous, the algorithm allows to identify faulty instruments as  
87 long as one instrument at least can be assumed as non-faulty in the sensor network. Across  
88 the investigation of the results of a case study, we propose several means to reduce false results  
89 and guidelines to adjust the parameters of the algorithm. Finally, we show that the proposed  
90 diagnosis approach, combined with a simple calibration technique, enables to improve the quality  
91 of the measurement results. Thus, the diagnosis algorithm opens new perspectives on *in situ*  
92 calibration.



94 Dans de nombreux domaines allant de l'agriculture à la santé publique, des grandeurs  
95 ambiantes doivent être suivies dans des espaces intérieurs ou extérieurs. On peut s'intéresser par  
96 exemple à la température, aux polluants dans l'air ou dans l'eau, au bruit, etc. Afin de mieux  
97 comprendre ces divers phénomènes, il est notamment nécessaire d'augmenter la densité spatiale  
98 d'instruments de mesure. Cela pourrait aider par exemple à l'analyse de l'exposition réelle des  
99 populations aux nuisances comme les polluants atmosphériques.

100 Le déploiement massif de capteurs dans l'environnement est rendu possible par la baisse des  
101 coûts des systèmes de mesure, qui utilisent notamment des éléments sensibles à base de micro ou  
102 nano technologies. L'inconvénient de ce type de dispositifs est une qualité de mesure insuffisante.  
103 Il en résulte un manque de confiance dans les données produites et/ou une hausse drastique des  
104 coûts de l'instrumentation causée par les opérations nécessaires d'étalonnage des instruments ou  
105 de remplacement périodique des capteurs.

106 Il existe dans la littérature de nombreux algorithmes qui offrent la possibilité de réaliser  
107 l'étalonnage des instruments en les laissant déployés sur le terrain, que l'on nomme techniques  
108 d'étalonnage *in situ*.

109 L'objectif de cette thèse est de contribuer à l'effort de recherche visant à améliorer la qualité  
110 des données des instruments de mesure bas coût à travers leur étalonnage *in situ*.

111 En particulier, on vise à 1) faciliter l'identification des techniques existantes d'étalonnage *in*  
112 *situ* applicables à un réseau de capteurs selon ses propriétés et les caractéristiques des instruments  
113 qui le composent ; 2) aider au choix de l'algorithme le plus adapté selon le réseau de capteurs et  
114 son contexte de déploiement ; 3) améliorer l'efficacité des stratégies d'étalonnage *in situ* grâce au  
115 diagnostic des instruments qui ont dérivé dans un réseau de capteurs.

116 Trois contributions principales sont faites dans ces travaux. Tout d'abord, une terminologie  
117 globale est proposée pour classer les travaux existants sur l'étalonnage *in situ*. L'état de l'art  
118 effectué selon cette taxonomie a montré qu'il y a de nombreuses contributions sur le sujet,  
119 couvrant un large spectre de cas. Néanmoins, le classement des travaux existants selon leurs  
120 performances a été difficile puisqu'il n'y a pas d'étude de cas de référence pour l'évaluation de  
121 ces algorithmes.

122 C'est pourquoi dans un second temps, un cadre pour la simulation de réseaux de capteurs  
123 est introduit. Il vise à guider l'évaluation d'algorithmes d'étalonnage *in situ*. Une étude de cas  
124 détaillée est fournie à travers l'évaluation d'algorithmes pour l'étalonnage *in situ* de réseaux  
125 de capteurs statiques et aveugles. Une analyse de l'influence des paramètres et des métriques  
126 utilisées pour extraire les résultats est également menée. Les résultats dépendant de l'étude  
127 de cas, et la plupart des algorithmes réétalonnant les instruments sans évaluer au préalable si  
128 cela est nécessaire, un outil d'identification permettant de déterminer les instruments qui sont  
129 effectivement fautifs en termes de dérive serait précieux.

130 Dès lors, la troisième contribution de cette thèse est un algorithme de diagnostic ciblant les  
131 fautes de dérive dans les réseaux de capteurs sans faire d'hypothèse sur la nature du réseau  
132 de capteurs considéré. Basé sur le concept de rendez-vous, l'algorithme permet d'identifier les  
133 instruments fautifs tant qu'il est possible de supposer qu'un instrument n'est pas fautif dans le  
134 réseau de capteurs. À travers l'analyse des résultats d'une étude de cas, nous proposons différents

135 moyens pour diminuer les faux résultats et des recommandations pour régler les paramètres  
136 de l'algorithme. Enfin, nous montrons que l'algorithme de diagnostic proposé, combiné à une  
137 technique simple d'étalonnage, permet d'améliorer la qualité des résultats de mesure. Ainsi, cet  
138 algorithme de diagnostic ouvre de nouvelles perspectives quant à l'étalonnage *in situ*.

# List of publications

---

## 140 In Journals

141 F. Delaine, B. Lebental and H. Rivano, "In Situ Calibration Algorithms for Environ-  
142 mental Sensor Networks: A Review," in *IEEE Sensors Journal*, vol. 19, no. 15, pp.  
143 5968-5978, 1 Aug.1, 2019, DOI: 10.1109/JSEN.2019.2910317.

144 F. Delaine, B. Lebental and H. Rivano, "Framework for the Simulation of Sensor  
145 Networks Aimed at Evaluating In Situ Calibration Algorithms," in *Sensors*, vol. 20,  
146 no. 16, 2020, DOI: 10.3390/s20164577.

## 147 In conferences without proceedings

148 F. Delaine, B. Lebental and H. Rivano, "Methodical Evaluation of In Situ Calibration  
149 Strategies for Environmental Sensor Networks," Colloque CASPA, Paris, April 2019.

## 150 Online repositories

151 F. Delaine, B. Lebental and H. Rivano, "Example case study applying a "Framework  
152 for the Simulation of Sensor Networks Aimed at Evaluating In Situ Calibration  
153 Algorithms",", Université Gustave Eiffel, 2020, DOI: 10.25578/CJCYMZ.



155	<b>Acknowledgments</b>	<b>i</b>
156	<b>Abstract</b>	<b>iii</b>
157	<b>Résumé</b>	<b>v</b>
158	<b>List of publications</b>	<b>vii</b>
159	<b>Contents</b>	<b>ix</b>
160	<b>List of Figures</b>	<b>xiii</b>
161	<b>List of Tables</b>	<b>xvii</b>
162	<b>List of Acronyms and Abbreviations</b>	<b>xix</b>
163	<b>List of Notations</b>	<b>xxi</b>
164	<b>General Introduction</b>	<b>1</b>
165	1    Context of the thesis . . . . .	1
166	2    Contributions . . . . .	2
167	3    Organisation of the manuscript . . . . .	3
168	<b>1 Low-cost Measuring Instruments for Air Quality Monitoring: Description,</b>	
169	<b>Performances and Challenges</b>	<b>5</b>
170	Introduction . . . . .	6
171	1    Measurement of ambient quantities with low-cost instruments . . . . .	6
172	1.1    Definition of a measuring instrument . . . . .	6
173	1.2    Measuring chain of an instrument . . . . .	6
174	1.3    Low-cost instruments . . . . .	8
175	2    Threats to data quality for measuring instruments . . . . .	11
176	2.1    Introduction . . . . .	11
177	2.2    Faults . . . . .	12
178	2.3    Discussion . . . . .	15
179	3    Calibration of measuring instruments . . . . .	15
180	3.1    Definition . . . . .	16
181	3.2    Analysis . . . . .	17
182	3.3 <i>In situ</i> calibration . . . . .	17
183	3.4    Discussion . . . . .	18
184	4    Sensor networks . . . . .	19
185	4.1    Definition . . . . .	19
186	4.2    Characteristics of interest in this work . . . . .	19

187	4.3	Conclusion . . . . .	21
188	5	Problem statement . . . . .	21
189	5.1	Motivations . . . . .	21
190	5.2	Objectives of the thesis . . . . .	21
191	<b>2</b>	<b>In Situ Calibration Algorithms for Environmental Sensor Networks</b>	<b>23</b>
192		Introduction . . . . .	24
193	1	Scope of the taxonomy . . . . .	24
194	2	Taxonomy for the classification of the algorithms . . . . .	25
195	2.1	Use of reference instruments . . . . .	25
196	2.2	Mobility of the instruments . . . . .	26
197	2.3	Calibration relationships . . . . .	26
198	2.4	Instrument grouping strategies . . . . .	27
199	3	Comparison to other taxonomies . . . . .	28
200	4	Review of the literature based on this classification . . . . .	30
201	4.1	Overview . . . . .	30
202	4.2	Mobile and static nodes . . . . .	32
203	4.3	Calibration relationships . . . . .	33
204	4.4	Pairwise strategies . . . . .	33
205	4.5	Blind macro calibration . . . . .	35
206	4.6	Group strategies . . . . .	36
207	4.7	Comment regarding other surveys . . . . .	37
208	5	Conclusion . . . . .	37
209	<b>3</b>	<b>Framework for the Simulation of Sensor Networks Aimed at Evaluating In Situ Calibration Algorithms</b>	<b>39</b>
210			
211	1	Challenges for the comparison of <i>in situ</i> calibration algorithms . . . . .	41
212	2	Description of the framework . . . . .	43
213	2.1	Simulation-based strategy . . . . .	43
214	2.2	Functional decomposition . . . . .	44
215	3	Comparison of <i>in situ</i> calibration strategies for blind static sensor networks . . .	46
216	3.1	Frame of the study . . . . .	46
217	3.2	Application of the framework . . . . .	47
218	3.3	Results . . . . .	51
219	3.4	Conclusions . . . . .	53
220	4	Evaluation of measurements after correction . . . . .	55
221	4.1	Problem statement . . . . .	55
222	4.2	Evaluation with an error model . . . . .	56
223	4.3	Means of visualisation . . . . .	60
224	4.4	Conclusion . . . . .	61
225	5	Sensitivity of the calibration algorithms to the specificities of the case study . . .	62
226	5.1	Using a more realistic model of the true values . . . . .	62
227	5.2	Density of the sensor network . . . . .	67
228	5.3	Instrument modelling . . . . .	71
229	5.4	Parameters of calibration strategies . . . . .	77
230	5.5	Summary of the results . . . . .	79
231	6	Discussion and conclusion . . . . .	80

232	<b>4</b>	<b>Diagnosis of Drift Faults in Sensor Networks</b>	<b>83</b>
233	1	Motivations . . . . .	85
234	2	State-of-the-art of fault diagnosis applied to sensor networks . . . . .	86
235	2.1	Overview of fault diagnosis for sensor networks through existing reviews . . . . .	86
236	2.2	Existing methods of diagnosis compatible with the concept of rendez-vous addressing any fault . . . . .	87
237			
238	2.3	Positioning of the contribution . . . . .	89
239	3	Definition of concepts for a drift diagnosis algorithm based on rendez-vous . . . . .	89
240	3.1	Validity of measurement results . . . . .	90
241	3.2	Compatibility of measurement results . . . . .	90
242	3.3	Rendez-vous . . . . .	91
243	4	Algorithm for the diagnosis of calibration issues in a sensor network . . . . .	92
244	4.1	General idea . . . . .	92
245	4.2	Procedure for the diagnosis of all the instruments in a sensor network . . . . .	94
246	4.3	Improvements and extensions of the presented algorithm . . . . .	95
247	4.4	Conclusion . . . . .	95
248	5	Application of the algorithm to a first case study . . . . .	97
249	5.1	Definition of the case study . . . . .	97
250	5.2	Configuration of the diagnosis algorithm . . . . .	99
251	5.3	Definition of the true status of an instrument . . . . .	99
252	5.4	Metrics for the evaluation of performances of the diagnosis algorithm . . . . .	99
253	5.5	Results . . . . .	101
254	5.6	Explanations of false results . . . . .	107
255	5.7	On the parameters of the diagnosis algorithm . . . . .	108
256	5.8	Conclusion . . . . .	110
257	6	On the assumption regarding the top-class instruments being always predicted as non-faulty . . . . .	111
258			
259	6.1	Theoretical discussion . . . . .	111
260	6.2	Case study with the instrument of class $c_{max}$ drifting . . . . .	111
261	6.3	Conclusion . . . . .	113
262	7	Means to reduce false results . . . . .	113
263	7.1	Keep the predicted status of instruments unchanged once they are predicted as faulty . . . . .	113
264			
265	7.2	Alternate definition for rates used to decide of the status of the instruments . . . . .	113
266	7.3	Adjustment of the maximal tolerated values for the different rates used to determine the statuses of the instruments . . . . .	115
267			
268	7.4	Conclusion . . . . .	116
269	8	Adjustment of the minimal size required for a set of valid rendez-vous to allow a prediction between the statuses faulty and non-faulty . . . . .	117
270			
271	8.1	Algorithm determining an upper boundary for $ \Phi_v _{min}$ . . . . .	117
272	8.2	Application to the case study . . . . .	120
273	8.3	Conclusion . . . . .	122
274	9	Sensitivity of the algorithm to changes in the case study . . . . .	122
275	9.1	Influence of the values of drift . . . . .	122
276	9.2	Influence of the true values . . . . .	123
277	9.3	Influence of the model used for the true values . . . . .	124
278	9.4	Influence of the density of instruments . . . . .	125
279	9.5	Influence of other faults . . . . .	131
280	9.6	Conclusion . . . . .	131
281	10	Combination with a simple calibration approach . . . . .	133

282	11	Conclusion . . . . .	138
283		<b>Conclusion and perspectives</b>	<b>141</b>
284	1	Conclusion . . . . .	141
285	2	Perspectives . . . . .	142
286		<b>A Diagnosis Algorithm for Drift Faults in Sensor Networks: Extensions</b>	<b>145</b>
287	1	Formulation reducing the number of iterations of the algorithm . . . . .	145
288	2	Diagnosis with the prediction as non-faulty based on the highest sufficient class .	146
289	3	From a centralised to a decentralised computation . . . . .	149
290	4	Multiple measurands . . . . .	152
291	4.1	General idea . . . . .	152
292	4.2	Diagnosis of drift faults in a sub-network with instruments having influence	
293		quantities . . . . .	152
294	4.3	Conclusion . . . . .	153
295	5	Diagnosis algorithm for drift faults in sensor networks with an event-based formulation	153
296	6	Real-time diagnosis algorithm of drift faults in sensor networks . . . . .	154
297	6.1	Choice of an event-based approach . . . . .	155
298	6.2	General idea . . . . .	155
299	6.3	Initialisation of the algorithm . . . . .	156
300	6.4	Allowed changes of status . . . . .	156
301	6.5	Conclusion . . . . .	156
302		<b>B On the Reason Why Assuming the Instruments of Class <math>c_{max}</math> are not Drift-</b>	
303		<b>ing is Acceptable</b>	<b>157</b>
304		<b>C Sensitivity of the Diagnosis Algorithm: Case Study with Values Related</b>	
305		<b>Over Time for the Parameters of the True Values' Model</b>	<b>159</b>
306	1	Introduction . . . . .	159
307	2	Model used . . . . .	159
308	3	Results . . . . .	159
309	4	Conclusion . . . . .	160

# List of Figures

---

311	1.1.1	Description of a mercury-in-glass thermometer . . . . .	7
312	1.1.2	Schematic description of the measuring chain of a measuring instrument . . . . .	8
313	1.2.1	Example of situation where a fault from a system-centric point of view generates a	
314		fault and then a failure from a data-centric perspective. . . . .	12
315	1.2.2	Considered faults for measuring instruments in the manuscript (continued) . . . . .	13
316	1.3.1	Schematic description of the steps of a calibration operation . . . . .	16
317	1.4.1	Examples of sensor networks with a different number of reference instruments and	
318		different relationships between the instruments. . . . .	19
319	1.4.2	Illustration of the principle of static, static and mobile, and mobile sensor networks.	20
320	2.2.1	Proposed taxonomy for the classification of <i>in situ</i> calibration algorithms . . . . .	25
321	2.2.2	Examples of cases for grouping strategies from the point of view of an instrument	
322		to calibrate . . . . .	28
323	2.2.3	Taxonomy for the classification of self-calibration algorithms according Barcelo-	
324		Ordinas <i>et al.</i> [8] . . . . .	29
325	3.1.1	Usual evaluation process performed across papers . . . . .	42
326	3.2.1	Schematic diagram of the methodology proposed for <i>in situ</i> calibration strategies	
327		evaluation . . . . .	44
328	3.3.1	Examples of maps of concentration $C$ used . . . . .	48
329	3.3.2	Evolution of $A$ and $FWHM$ for the modelling of the concentration of pollutant .	49
330	3.3.3	Positions of the 16 measuring instruments considered in the case study, deployed	
331		uniformly in the field . . . . .	49
332	3.3.4	True values, measured values and corrected values with the strategies considered	
333		for a particular instrument $s_1 = 2$ between $t = 5424h$ and $t = 5471h$ . . . . .	52
334	3.3.5	Evolution of the MAE computed for each week of the drift period between the	
335		drifted values and the true values, and between the corrected values for each	
336		strategy and the true values for a particular instrument, after the start of drift . .	54
337	3.4.1	Evolution of the slope, intercept and score of the error model of $s_1$ , computed on	
338		each week of the drift period . . . . .	59
339	3.4.2	Box plots of the MAE, computed on the entire time interval of drift, of the 16	
340		nodes of the network without calibration and with <b>SM-SVD</b> . . . . .	60
341	3.4.3	Matrix of the MAE, computed on the entire time interval of drift, of the 16 nodes	
342		of the network without calibration and with <b>SM-SVD</b> . . . . .	60
343	3.4.4	Target plot of the 16 nodes as a function of their slope and intercept in the error	
344		model for each calibration strategy . . . . .	61
345	3.5.1	Example of concentration map used and based on the Gaussian plume model . . .	63
346	3.5.2	True values, measured values and corrected values with the strategies considered	
347		for a particular instrument between $t = 5424h$ and $t = 5471h$ with the Gaussian	
348		plume model . . . . .	63

349	3.5.3	Target plot of the 16 nodes as a function of their slope and intercept in the error model, computed on the entire time interval of drift, for each calibration strategy and with the Gaussian plume model . . . . .	65
350			
351			
352	3.5.4	Evolution of the mean of the MAE, computed on the entire time interval of study, of all the nodes of the network, as a function of the number of nodes $ S $ , for the 2D Gauss model . . . . .	67
353			
354			
355	3.5.5	Target plot of the 16 nodes as a function of their slope and intercept in the error model, computed on the entire time interval of study, for each calibration strategy and with the Gaussian plume model . . . . .	70
356			
357			
358	3.5.6	Evolution of the mean of MAE, computed on the entire time interval of study, over the network for the strategies SM-SVD, SM-LS and SM-TLS as a function of $w$ .	78
359			
360	3.5.7	Evolution of the mean of MAE, computed on the entire time interval of study, over the network for the strategies SM-SVD, SM-LS and SM-TLS as a function of $w$ with a different signal subspace than in Figure 3.5.6 . . . . .	78
361			
362			
363	3.5.8	Summary of the results observed by changing the parameters of the baseline case study which was built with a 2D Gaussian function to model the measurand, with the WGLI drift model applied to the measured values of the instruments, with 16 nodes in the network and the values indicated in Section 3 for the parameters of the <i>in situ</i> calibration algorithms . . . . .	79
364			
365			
366			
367			
368	4.4.1	Example where the measurement results $m(s_i, t)$ and $m(s_j, t')$ are compatible with each other but where $m(s_i, t)$ is not compatible with its true value . . . . .	93
369			
370	4.5.1	Map of the 100 positions available for the instruments in the case study. . . . .	97
371	4.5.2	Evolution of the true status and of the predicted status for several instruments . .	102
372	4.5.3	Evolution of the metrics computed for each diagnosis procedure as a function of the current diagnosis procedure id . . . . .	104
373			
374	4.5.4	Representation of an example set of measurement results $M(s_i, (t_d, \Delta t))$ . . . . .	106
375	4.5.5	Evolution of the rates $r_{true}$ and $r_{\tilde{\Phi}_v}$ for $s_1$ in the case study of Section 5 . . . . .	108
376	4.6.1	Evolution of the true status and of the predicted status of $s_9$ , the instrument of class 1, when it is drifting . . . . .	112
377			
378	4.7.1	Evolution of the true status and of the predicted status of $s_1$ while keeping $\hat{\Omega}(s_1, t) = F$ for $t \geq t_d$ such as $\hat{\Omega}(s_1, t_d) = F$ for the first time. . . . .	114
379			
380	4.7.2	Evolution for $s_1$ in the case study of Section 5 (a) of the rates $r_{true}$ and $r_{\tilde{\Phi}_v}$ (both definitions) ; (b) of its true status and of its predicted status with the alternate definition for $r_{\tilde{\Phi}_v}$ ; while keeping $\hat{\Omega}(s_1, t) = F$ for $t \geq t_d$ such as $\hat{\Omega}(s_1, t_d) = F$ for the first time. . . . .	115
381			
382			
383			
384	4.8.1	Evolution of $\max  \Phi_v _{min}$ as a function of the number of faulty instruments $\lambda$ in the network, based on the matrix $\mathbf{Mat}_\Phi(D, \Delta t)$ derived from the database used for the case study of Section 5 . . . . .	120
385			
386			
387	4.9.1	Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with $ \Phi_v _{min} = 10$ . . . . .	127
388			
389	4.9.2	Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with $ \Phi_v _{min} = 6$ . . . . .	129
390			
391	4.10.1	Evolution of true and predicted status for instrument $s_1$ for different choices of calibration . . . . .	134
392			
393	4.10.2	Evolution of the true values, of the values without recalibration and with recalibration with a linear regression for $s_1$ . . . . .	135
394			
395	4.10.3	Target plot of the 9 nodes of class 0 of the network as a function of their slope and intercept following the error model of Chapter 3, computed on the entire time interval of study, for each calibration strategy. . . . .	136
396			
397			

398	B.0.1	Evolution of the true status of $s_0$ , computed with the relative uncertainty $\Delta_r v$ on	
399		the results of the instruments of class 0 instead of its own $\Delta_r v$ , and its predicted	
400		status when it is drifting . . . . .	158



# List of Tables

---

402	1.1.1	Minimum and maximal values of the coefficient of determination $R^2$ reported in the cited publications for $PM_{10}$ , $PM_{2.5}$ , $PM_1$ , $O_3$ , $NO_2$ , $NO$ , $CO$ and $SO_2$ . . .	10
403			
404	2.4.1	<i>In situ</i> calibration strategies for sensor networks . . . . .	31
405	3.3.1	Mean and standard deviation of each metric, computed on the entire time interval of drift, over the 16 nodes of the network. . . . .	52
406			
407	3.3.2	Values of the metrics, computed on the entire time interval of drift, for two particular instruments of the network $s_1 = 2$ and $s_2 = 10$ . . . . .	53
408			
409	3.3.3	Mean and standard deviation of the weekly mean of each metric for $s_1$ during the drift period . . . . .	54
410			
411	3.4.1	Mean and standard deviation of the parameters of the error model and the regression score, computed on the entire time interval of drift, over the 16 nodes of the network	57
412			
413	3.4.2	Values of the parameters of the error model and the regression score, computed on the entire time interval of drift, for two particular instruments of the network $s_1 = 2$ and $s_2 = 10$ . . . . .	57
414			
415			
416	3.4.3	Mean and standard deviation of the parameters of the error model and the regression score for $s_1$ , computed on each week of the drift period . . . . .	58
417			
418	3.5.1	Mean and standard deviation of each metric, computed on the entire time interval of drift, over the 16 nodes of the network with the Gaussian plume model . . . . .	64
419			
420	3.5.2	Values of the metrics, computed on the entire time interval of drift, for two particular instruments of the network $s_1 = 2$ and $s_2 = 10$ , with the Gaussian plume model . . . . .	66
421			
422			
423	3.5.3	Mean and standard deviation of each metric, computed on the entire time interval of study, over the 100 nodes of the network with the Gaussian plume model . . .	68
424			
425	3.5.4	Values of the metrics, computed on the entire time interval of study, for two particular instruments of the network $s_1 = 2$ and $s_2 = 10$ , with the Gaussian plume model . . . . .	69
426			
427			
428	3.5.5	Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the RGOI drift model . . . . .	73
429			
430	3.5.6	Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the CGOI drift model . . . . .	74
431			
432	3.5.7	Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the WGLI drift model, spikes and noise . . . . .	76
433			
434			
435	4.5.1	Values of the parameters of the case study . . . . .	98
436	4.5.2	Contingency table of the different primary metrics . . . . .	100
437	4.5.3	Metrics derived from the metrics $P$ , $N$ , $TP$ , $TN$ , $FP$ , $FN$ , $NDP$ and $NDN$ . .	100
438	4.5.4	Confusion matrix of the case study . . . . .	103
439	4.5.5	Statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study . . . . .	103

LIST OF TABLES

---

440	4.6.1	Confusion matrix of the case study when the instrument of class 1 is drifting . . .	112
441	4.6.2	Statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study when the	
442		instrument of class 1 is drifting . . . . .	113
443	4.7.1	Confusion matrix of the case study with the predicted statuses kept as faulty from	
444		one diagnosis procedure to another once the status of an instrument is predicted	
445		as faulty for the first time . . . . .	114
446	4.7.2	Statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study with the predicted	
447		statuses kept as faulty from one diagnosis procedure to another once the status of	
448		an instrument is predicted as faulty for the first time . . . . .	115
449	4.7.3	Confusion matrix of the case study with the predicted statuses kept as faulty from	
450		one diagnosis procedure to another once the status of an instrument is predicted	
451		as faulty for the first time and with the alternate definition for $r_{\tilde{\Phi}_v}$ . . . . .	116
452	4.8.1	Confusion matrix of the case study with the predicted statuses kept as faulty from	
453		one diagnosis procedure to another once the status of an instrument is predicted	
454		as faulty for the first time and $ \Phi_v _{min} = 10$ . . . . .	121
455	4.8.2	Statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study with the predicted	
456		statuses kept as faulty from one diagnosis procedure to another once the status of	
457		an instrument is predicted as faulty for the first time, $ \Phi_v _{min} = 10$ . . . . .	121
458	4.9.1	Confusion matrix for 100 simulations of the case study with drift values drawn	
459		again for each one . . . . .	123
460	4.9.2	Confusion matrix for 100 simulations of the case study with true values following a	
461		2D Gaussian model which parameters are randomly drawn for each simulation . .	124
462	4.9.3	Confusion matrix for 100 simulations of the case study with true values randomly	
463		drawn following a uniform law . . . . .	125
464	4.9.4	Confusion matrix of the case study with the predicted statuses kept as faulty from	
465		one diagnosis procedure to another once the status of an instrument is predicted as	
466		faulty for the first time, $ \Phi_v _{min} = 10$ and spikes and noise added to the measured	
467		values . . . . .	132
468	4.9.5	Statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study with the predicted	
469		statuses kept as faulty from one diagnosis procedure to another once the status of	
470		an instrument is predicted as faulty for the first time, $ \Phi_v _{min} = 10$ and spikes and	
471		noise added to the measured values . . . . .	132
472	4.10.1	Confusion matrix of the case study on calibration without recalibration . . . . .	135
473	4.10.2	Confusion matrix of the case study on calibration with an oracle-based calibration	136
474	4.10.3	Confusion matrix of the case study on calibration with calibration based on linear	
475		regression . . . . .	137
476	4.10.4	Statistics of the slope, intercept and score of the error model of Chapter 3 computed	
477		for the nodes of class 0 of the network on the entire time interval of study and for	
478		each calibration strategy. . . . .	137
479	4.11.1	Recommendations for the different parameters of the diagnosis algorithm . . . . .	139
480	A.6.1	Possible cases when a new rendez-vous occurs between $s_i$ and $s_j$ for a real-time	
481		diagnosis . . . . .	155
482	C.3.1	Confusion matrix of the case study with true values based on the 2D Gauss model	
483		with values of its parameters derived from a real dataset . . . . .	160

# List of Acronyms and Abbreviations

---

485	<b>VIM</b>	International Vocabulary of Metrology
486	<b>ARMA</b>	Autoregressive moving average
487	<b>SM-SVD</b>	Subspace model + Singular value decomposition (calibration strategy used in Chapter 3)
488		3)
489	<b>SM-LS</b>	Subspace model + Least squares (calibration strategy used in Chapter 3)
490	<b>SM-TLS</b>	Subspace model + Total least squares (calibration strategy used in Chapter 3)
491	<b>AB-DT</b>	Average Based + Difference trigger (calibration strategy used in Chapter 3)
492	<b>AB-KF</b>	Average Based + Kalman filter (calibration strategy used in Chapter 3)
493	<b>SVR-KF</b>	Support Vector Regression + Kalman filter (calibration strategy used in Chapter 3)
494	<b>K-KF</b>	Kriging + Kalman filter (calibration strategy used in Chapter 3)
495	<b>PCA</b>	Principal components analysis
496	<b>WGLI</b>	Weekly gain linear increase (drift model defined in Chapter 3 Section 3)
497	<b>RGOI</b>	Random gain and offset increase (drift model defined in Chapter 3 Section 5.3.1)
498	<b>CGOI</b>	Continuous gain and offset increase (drift model defined in Chapter 3 Section 5.3.1)



500 The notations are organised into several categories to quickly identify a set of notations used  
501 in a particular context.

502 *Note from the author: Conflicts exist due to identical notations that are used for different*  
503 *objects depending on the context. It concerns notably the measurand models (2D Gaussian Model*  
504 *and Gaussian Plume model) and the parameters of the calibration algorithms to remain consistent*  
505 *with the usual notations in the literature.*

## 506 Common objects

507	$t$	Instant of time
508	$\Delta t$	Time duration
509	$\omega$	Angular frequency
510	$E$	Expectation
511	$\mathcal{N}(\alpha, \beta)$	Normal law of mean $\alpha$ and standard deviation $\beta$ .
512	$\mathcal{U}(\alpha, \beta)$	Uniform law on the range $[\alpha; \beta]$
513	$\mu$	Average
514	$\sigma$	Standard deviation

## 515 Measuring instruments

516	$s_i, s_j$	Measuring instruments (or systems if it is mentioned)
517	$c(s_i)$	Accuracy class of $s_i$
518	$S$	Set of measuring instruments. $S^k$ , $S^{k+}$ and $S^{k-}$ are respectively the sets of
519		measuring instruments where $c(s_i) = k$ , $c(s_i) \geq k$ and $c(s_i) \leq k$
520	$m(s_i, t)$	Measurement result of $s_i$ at $t$
521	$M(s_i, (t, \Delta t))$	Set of measurement results for $s_i$ , over the time range $[t - \Delta t; t]$
522	$v(s_i, t)$	Measured value of $s_i$ at $t$
523	$V(s_i, (t, \Delta t))$	Set of measured values for $s_i$ , over the time range $[t - \Delta t; t]$
524	$v_{true}(s_i, t)$	True value that should be measured by $s_i$ at $t$ if it were ideal
525	$\Delta v(s_i, t)$	Measurement uncertainty of the value measured by $s_i$ at $t$ . It can be a
526		constant or not

527	$\Delta v_r(s_i, t)$	Relative measurement uncertainty of the value measured by $s_i$ at $t$ .
528	$v_{min}(k)$	is the detection limit of instruments of class $k$
529	$\zeta(s_i, t)$	Indication of $s_i$ at $t$
530	$\mathcal{F}$	Function representing the measuring chain of an instrument
531	$\hat{\mathcal{F}}$	Estimate of the measuring chain of an instrument
532	$\hat{\mathcal{F}}^{-1}$	Inverse of the estimate of the measuring chain of an instrument
533	$\mathbf{q}(t)$	Vector of values of influence quantities
534	$\hat{\mathbf{q}}(t)$	Estimate of the vector of values of influence quantities
535	$\mathcal{H}$	Relationship built during an <i>in situ</i> calibration
536	$\mathcal{H}^{-1}$	Calibration relationship derived from an <i>in situ</i> calibration

537 **Fault models**

538	$G$	Gain drift
539	$O$	Offset drift
540	$\varepsilon$	Noise
541	$\psi$	Spike value
542	$p_\psi$	Spike probability

543 **2D Gaussian model**

544	$C$	Concentration
545	$x, y, z$	Coordinates
546	$A$	Amplitude
547	$\sigma$	Standard deviation
548	$FWHM$	Full width at half maximum

549 **Gaussian Plume model**

550	$Q$	Emission rate at the source
551	$V_w$	Wind speed
552	$\sigma_y, \sigma_z$	Horizontal and vertical dispersions
	$H$	Pollutant effective release

$$H = h_s + \Delta h(t)$$

with  $h_s$ , the pollutant source height, and  $\Delta h$ :

$$\Delta h(t) = \frac{1.6F^{\frac{1}{3}}x^{\frac{2}{3}}}{V_w}$$

with  $F$  such as

$$F = \frac{g}{\pi} D \left( \frac{\theta_s - \theta(t)}{\theta} \right)$$

553	$g$	gravity constant
554	$D$	volumetric flow
555	$\theta$	ambient temperature
556	$\theta_s$	source temperature

### 557 **Parameters of the calibration algorithms**

558	$\nu$	Number of components of the PCA for the algorithms SM-X
559	$w$	Parameter for the algorithms SM-X
560	$R, Q$	Parameter for the algorithms X-KF
561	$C$	Penalty parameter for the algorithm SVR-KF
562	$a, c_0, c_1$	Kriging parameters for the algorithm K-KF

### 563 **Usual metrics**

564	$MAE$	Mean absolute error
565	$MAPE$	Mean absolute percentage error
566	$RMSE$	Root-mean-square error
567	$\rho$	Pearson correlation coefficient
568	$R^2$	Coefficient of determination

### 569 **Error model**

570	$\mathcal{F}$	Function
571	$a$	Slope
572	$b$	Intercept
573	$\varepsilon$	Additive error

### 574 **Metrics for the evaluation of the performances of the diagnosis algorithm**

575	$TN$	True negative
576	$FN$	False negative
577	$NDN$	Non-determined negative
578	$FP$	False positive
579	$TP$	True positive
580	$NDP$	Non-determined positive
581	$P$	Number of positives

582	$N$	Number of negatives
583	$Prev$	Prevalence
584	$TPR$	True positive rate
585	$TNR$	True negative rate
586	$FPR$	False positive rate
587	$FNR$	False negative rate
588	$NDPR$	Non-determined positive rate
589	$NDNR$	Non-determined negative rate
590	$NDR$	Non-determined rate
591	$PPV$	Positive predictive value
592	$FDR$	False discovery rate
593	$NPV$	Negative predictive value
594	$FOR$	False omission rate
595	$ACC$	Accuracy
596	$\Delta\mathcal{D}(s_i)$	Delay of first positive detection for an instrument $s_i$
597	<b>Diagnosis algorithm</b>	
598	$D$	Set of diagnosis procedures
599	$d$	Diagnosis procedure
600	$t_d$	Instant of the diagnosis procedure $d$
601	$\Delta c_{min}^{\mathcal{D}}(k)$	Minimal relative difference of class required for instruments of class $k$ with their diagnoser instruments
602		
603	$c_{min}^{\mathcal{D}}(s_i)$	Minimal class of sensors allowed to diagnose $s_i$ .
604	$\Omega(s_i, t)$	True status of $s_i$ at $t$ , equal either to $NF$ (non-faulty) or $F$ (faulty)
605	$\hat{\Omega}(s_i, t)$	Predicted status of $s_i$ at $t$ , equal either to $NF$ (non-faulty), $F$ (faulty) or $A$ (ambiguous)
606		
607	$\tilde{\Omega}(s_i, t)$	Actualised status of $s_i$ at $t$ , equal either to $NF$ (non-faulty), $F$ (faulty) or $A$ (ambiguous)
608		
609	$S^{\mathcal{D}}$	Set of sensors to diagnose
610	$S^{NF}(t)$	Set of sensors where $\hat{\Omega}(s_i, t) = NF$
611	$S^F(t)$	Set of sensors where $\hat{\Omega}(s_i, t) = F$
612	$\mathcal{C}_F(S, \lambda)$	Set of combination of $\lambda$ sensors in $S$ that can be diagnosed as faulty

---

613	$M^{\approx}(s_i, (t, \Delta t))$	Set of measurement results coherent with true values for $s_i$ , over the time range $[t - \Delta t; t]$
614		
615	$M^+(s_i, (t, \Delta t))$	Set of measurement results upper non-coherent with true values for $s_i$ , over the time range $[t - \Delta t; t]$
616		
617	$M^-(s_i, (t, \Delta t))$	Set of measurement results lower non-coherent with true values for $s_i$ , over the time range $[t - \Delta t; t]$
618		
619	$M^*(s_i, (t, \Delta t))$	Set of measurement results for $s_i$ , over the time range $[t - \Delta t; t]$ that are metrologically valid
620		
621	$\varphi(s_i \rightarrow s_j, t)$	Rendez-vous between $s_i$ and $s_j$ at $t$
622	$\Phi(s_i \rightarrow s_j, (t, \Delta t))$	Set of rendez-vous between $s_i$ and $s_j$ over the time range $[t - \Delta t; t]$ . $\Phi(s_i \rightarrow S, (t, \Delta t))$ is the set of rendez-vous between $s_i$ and any other instrument of $S$ over the time range $[t - \Delta t; t]$ . It is also denoted $\Phi(s_i, (t, \Delta t))$
623		
624		
625	$\Phi^{\approx}(s_i \rightarrow s_j, (t, \Delta t))$	Set of coherent rendez-vous between $s_i$ and $s_j$ over the time range $[t - \Delta t; t]$
626	$\Phi^+(s_i \rightarrow s_j, (t, \Delta t))$	Set of upper non-coherent rendez-vous between $s_i$ and $s_j$ over the time range $[t - \Delta t; t]$
627		
628	$\Phi^-(s_i \rightarrow s_j, (t, \Delta t))$	Set of lower non-coherent rendez-vous between $s_i$ and $s_j$ over the time range $[t - \Delta t; t]$
629		
630	$\Phi(s_i \rightarrow S, (t, \Delta t)) _k$	Set of rendez-vous between $s_i$ and all $s_j \in S$ such as $c(s_j) \geq k$ over the time range $[t - \Delta t; t]$
631		
632	$\mathbf{Mat}_{\Phi}(D, \Delta t)$	Matrix of the minimal number of rendez-vous between two instruments on a duration $\Delta t$ and over a set of diagnosis procedures $D$
633		
634	$\Phi_v(s_i \rightarrow S, (t, \Delta t))$	Set of rendez-vous between $s_i$ and $s_j \in S$ over the time range $[t - \Delta t; t]$ that are valid
635		
636	$ \Phi_v _{min}$	Minimal number of valid rendez-vous required to conduct successfully a diagnosis
637		
638	$r_{\varphi_v}^{\approx}(s_i, (t, \Delta t))$	Rate of coherent rendez-vous in the set of valid rendez-vous of $s_i$ over the time range $[t - \Delta t; t]$
639		
640	$r_{\varphi_v}^+(s_i, (t, \Delta t))$	Rate of upper non-coherent rendez-vous in the set of valid rendez-vous of $s_i$ over the time range $[t - \Delta t; t]$
641		
642	$r_{\varphi_v}^-(s_i, (t, \Delta t))$	Rate of lower non-coherent rendez-vous in the set of valid rendez-vous of $s_i$ over the time range $[t - \Delta t; t]$
643		
644	$(r_{\varphi_v}^+)_{max}$	Maximal tolerated value for $r_{\varphi_v}^+(s_i, (t, \Delta t))$ for any $t$
645	$(r_{\varphi_v}^-)_{max}$	Maximal tolerated value for $r_{\varphi_v}^-(s_i, (t, \Delta t))$ for any $t$
646	$(r_{\varphi_v}^+ + r_{\varphi_v}^-)_{max}$	Maximal tolerated value for $(r_{\varphi_v}^+ + r_{\varphi_v}^-)(s_i, (t, \Delta t)) = 1 - r_{\varphi_v}^{\approx}(s_i, (t, \Delta t))$ for any $t$
647		
648	$r_{true}(s_i, (t, \Delta t))$	Rate of measurement results coherent with true values of $s_i$ over the time range $[t - \Delta t; t]$
649		

---

650  $r_{true}(s_i, (t, \Delta t))|_{\Phi}$  Rate of measurement results coherent with true values in the set of rendez-  
 651 vous of  $s_i$  over the time range  $[t - \Delta t; t]$

652  $r_{true}(s_i, (t, \Delta t))|_{\Phi_v}$  Rate of measurement results coherent with true values in the set of valid  
 653 rendez-vous of  $s_i$  over the time range  $[t - \Delta t; t]$

$w$  Weight operator for rendez-vous such as:

$$w(\varphi(s_i \rightarrow s_j, t)) = \frac{1}{|\Phi(s_i, t)|}$$

654 with  $\Phi(s_i, t) = \Phi((s_i, S), t|0)$

655 **Practical definition of rendez-vous**

656  $l(s_i, t)$  Position of  $s_i$  at  $t$

657  $\Delta l_{\varphi}$  Maximal distance between two instruments to consider them in rendez-vous  
 658 (spatial condition)

659  $\Delta t_{\varphi}$  Maximal difference between the timestamps of the measurement results of  
 660 two instruments to consider them in rendez-vous (temporal condition)

661  $a(s_i, t)$  Representativity area of the instrument  $s_i$  at  $t$

## 1 Context of the thesis

663

664 The monitoring of ambient quantities is a fundamental operation for the digitisation of our  
665 environment [59]. From the level of precipitation [32] to the concentration of pollutants in the air  
666 [111], more and more quantities need to be monitored. The information brought by measurements  
667 is used in various fields from agriculture [117] to public health [46] for instance. Such a monitoring  
668 may have a major impact because these quantities can have significant effects on human lives.  
669 For instance, heatwaves, corresponding to high temperatures at both day and night, induce an  
670 abnormal mortality rate that can be extremely high [81]. Combined with drought, they can also  
671 be disastrous for agriculture and generate water stress [159]. Thus, the comprehension of these  
672 phenomena is particularly important to explain and predict their evolution over time but also to  
673 react to them.

674

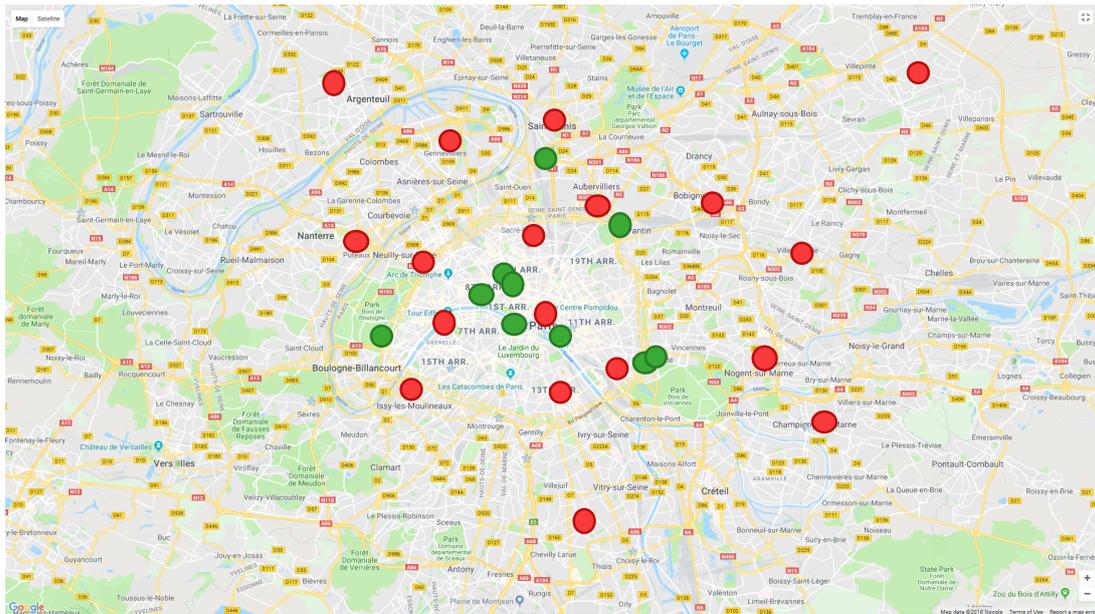
675 This activity of monitoring has evolved over time, following the progress in science and  
676 technologies. Nowadays, the observation of ambient quantities is carried out with high quality  
677 measuring systems, at least for strategic or regulatory applications like meteorology [174] or air  
678 quality monitoring [158]. They can be deployed on the ground, in the sea, in the air or even in  
679 space via satellites. Regulatory-grade instruments are usually expensive. Regarding those on the  
680 ground, they are often deployed at fixed positions. Depending on the target area to monitor, on  
681 the spatial variability of the measurand and on the spatio-temporal resolution desired, hundreds  
682 of devices may be required, as shown for instance by [21] and [101] in the field of air quality  
683 monitoring. Thus, it may not be possible to have a high spatial resolution with such instruments  
684 as few of them are deployed together most of the time due to their cost [78]. For instance, about  
685 30 measuring stations are deployed across Paris and its inner suburbs to monitor air pollution [2]  
686 as shown in the figure page 2. Moreover, all of them are not monitoring the same quantities. For  
687 instance, only a third of these stations are monitoring  $O_3$ . Consequently, the spatial resolution is  
688 even smaller for some measurands.

689

690 Following up the scientific advances in micro and nano-electronics [90, 135], it is now possible  
691 to imagine a dense monitoring of ambient quantities with a fine granularity, both temporally  
692 and spatially, notably for air quality monitoring [110, 135] which is the practical context chosen  
693 for this thesis. Indeed, small low-cost measuring instruments have been designed in recent  
694 years. Fostered by the emergence of the Internet of Things, the interest for such devices has  
695 been growing significantly because they open up new possibilities for environmental sensing  
696 [62]. Without actually replacing regulatory grade instruments, it is hoped they can complement  
697 existing networks of measuring instruments. As they are affordable, they can also be owned by  
698 everybody. In this way, low-cost sensors are tools that could be used for individual awareness  
699 and education, particularly concerning both indoor and outdoor air quality [27].

700

701 Nevertheless, these low-cost sensing technologies are still young, and several challenges are yet  
702 to be tackled [62, 127]. One of them concerns the quality of the measurement results they produce.  
703 The particular issue degrading the results addressed in this thesis is the instrumental drift which  
704 these instruments are prone to. The classical approach to solve this problem is to recalibrate the  
705 instruments in a calibration facility. This requires taking the instrument out of service, bringing



Symbol	Definition
	Permanent urban station
	Permanent traffic station

Map of permanent air pollution monitoring stations of Airparif for Paris and its inner suburbs (December 2018)

703 it to a facility, calibrating it, bringing it back to the field and putting it back in service. This  
 704 operation has both economically and technically a definite cost. Therefore, when hundreds of  
 705 measuring instruments are deployed, performing this task is particularly challenging because it  
 706 has to be carried out more frequently than with high-quality instruments. However, because  
 707 numerous instruments are deployed and forming a sensor network, there may be relationships  
 708 between their measured values that could be exploited to manage this issue.

709 In the literature, multiple solutions have been proposed to mitigate the impact of the drift for  
 710 environmental low-cost sensors by performing what we call an *in situ* calibration. It means  
 711 the calibration of measuring instruments while leaving them in the field, preferably without any  
 712 physical intervention. To avoid these physical interventions, the idea is to use the values provided  
 713 by other instruments deployed.

## 714 2 Contributions

715 The objective of this thesis is to contribute to the research effort on the improvement of data  
 716 quality for low-cost measuring instruments through their *in situ* calibration. In particular, we  
 717 aim at:

- 718 • facilitating the identification of existing *in situ* calibration strategies applicable to a sensor  
 719 network depending on its properties and the characteristics of its instruments.
- 720 • helping to choose the most suitable algorithm depending on the sensor network and its  
 721 context of deployment.
- 722 • improving the efficiency of *in situ* calibration strategies through the diagnosis of instruments

723 that have drifted in a sensor network.

724 Toward this goal, three main contributions are made in this work. First, a unified terminology  
725 is proposed to classify the existing works on *in situ* calibration. Indeed there is no shared  
726 vocabulary in the scientific community enabling a precise description of the main characteristics  
727 of the algorithms. The review carried out based on this taxonomy showed there are numerous  
728 contributions on the subject, covering a wide variety of cases. Due to the type of sensor network  
729 deployed, in terms of properties of the measuring instruments or how these devices can interact  
730 between them, different approaches may be considered. Nevertheless, the classification of the  
731 existing works in terms of performances was difficult as there is no reference case study for the  
732 evaluation of these algorithms.

733 Therefore in a second step, a framework for the simulation of sensor networks is introduced.  
734 It is aimed at evaluating *in situ* calibration algorithms. It details all the aspects to take into  
735 account when designing a case study. A detailed case study is provided across the evaluation of  
736 *in situ* calibration algorithms for blind static sensor networks. An analysis of the influence of  
737 the parameters and of the metrics used to derive the results is also carried out. As the results  
738 are case specific, and as most of the algorithms recalibrate instruments without evaluating first  
739 if they actually need it, an identification tool enabling to determine the instruments that are  
740 actually faulty in terms of drift would be valuable.

741 Thus, the third contribution of this thesis is the design of a diagnosis algorithm targeting  
742 drift faults in sensor networks without making any assumption on the kind of sensor network at  
743 stake. Based on the concept of rendez-vous, the algorithm allows identifying faulty instruments  
744 as long as one instrument at least can be assumed as non-faulty in the sensor network. Across  
745 the investigation of the results of a case study, we propose several means to reduce false results  
746 and guidelines to adjust the parameters of the algorithm. Finally, we show that the proposed  
747 diagnosis approach, combined with a simple calibration technique, enables to improve the quality  
748 of the measurement results.

### 749 **3 Organisation of the manuscript**

750 The manuscript is organised as follows.

751 In Chapter 1, concepts related to measuring instruments are defined and the performances  
752 of low-cost measuring instruments in the context of air quality monitoring are reviewed. Then,  
753 the issue of data quality for these devices and how calibration can mitigate drift problems are  
754 discussed. The concept of *in situ* calibration is introduced, followed by the interesting properties  
755 of sensor networks for this application. At the end of the chapter, the problem statement is  
756 recalled and detailed.

757 In Chapter 2, the taxonomy for the classification of *in situ* calibration strategies is introduced,  
758 followed by the review of existing works.

759 Chapter 3 introduces the framework for the simulation of sensor networks designed for the  
760 evaluation of *in situ* calibration algorithms. A case study concerning blind static sensor networks  
761 is developed, and multiple derived cases are investigated to determine the influence of parameters  
762 and of the metrics used on the interpretations of the results.

763 Afterwards, the diagnosis algorithm for drift faults in sensor networks is introduced in Chapter  
764 4. A case study is developed and means to improve its results are discussed. The sensitivity of  
765 the diagnosis algorithm to the choices made during the design of the case study is investigated.  
766 It is followed by the combination of the diagnosis algorithm with a simple calibration approach  
767 that exploits information build during the diagnosis, applied to the preceding case study.

768 Finally, the contributions are summarised in the last chapter and a general conclusion is  
769 provided, along with perspectives regarding this work.

770 Several appendices are also provided in this manuscript. Appendix A provides variations and  
771 extensions of the diagnosis algorithm presented in Chapter 4. They require further studies but  
772 the basis are introduced towards future work. Appendix B extends the discussion on the main  
773 assumption made for the diagnosis algorithm. Lastly, Appendix C presents an additional case  
774 study related to the sensitivity of the diagnosis algorithm.

Chapter  
**1**

**Low-cost Measuring Instruments for  
Air Quality Monitoring: Description,  
Performances and Challenges**

**Contents**

Introduction . . . . . 6

1 Measurement of ambient quantities with low-cost instruments . . . . . 6

    1.1 Definition of a measuring instrument . . . . . 6

    1.2 Measuring chain of an instrument . . . . . 6

    1.3 Low-cost instruments . . . . . 8

        1.3.1 Definition . . . . . 8

        1.3.2 Challenges of environmental monitoring . . . . . 9

        1.3.3 Performances of low-cost instruments in the literature for air  
            quality monitoring . . . . . 9

2 Threats to data quality for measuring instruments . . . . . 11

    2.1 Introduction . . . . . 11

    2.2 Faults . . . . . 12

    2.3 Discussion . . . . . 15

3 Calibration of measuring instruments . . . . . 15

    3.1 Definition . . . . . 16

    3.2 Analysis . . . . . 17

    3.3 *In situ* calibration . . . . . 17

    3.4 Discussion . . . . . 18

4 Sensor networks . . . . . 19

    4.1 Definition . . . . . 19

    4.2 Characteristics of interest in this work . . . . . 19

        4.2.1 Presence of references . . . . . 20

        4.2.2 Mobility . . . . . 20

    4.3 Conclusion . . . . . 21

5 Problem statement . . . . . 21

    5.1 Motivations . . . . . 21

    5.2 Objectives of the thesis . . . . . 21

## 812 Introduction

813 This chapter is aimed at recalling several concepts related to measuring instruments first. The  
814 question of the measurement of ambient quantities is detailed by defining low-cost instruments  
815 and by reviewing their performances in the context of air quality monitoring. Afterwards, the  
816 challenges for these instruments are defined. The threats to the data quality of measuring  
817 instruments are discussed before investigating how calibration can mitigate drift problems.  
818 Finally, the concept of *in situ* calibration is introduced, followed by the definition of a sensor  
819 network and its interesting characteristics for our problem. At the end of the chapter, the  
820 problem statement is detailed based on the previous developments.

## 821 1 Measurement of ambient quantities with low-cost instruments

822 In this section, the definition of a measuring instrument is recalled first. Then, a general  
823 description of how it makes measurements is provided, followed by the definition of low-cost  
824 instruments, their challenges and a review of their actual performances in metrological terms.

### 825 1.1 Definition of a measuring instrument

826 In environmental monitoring, we aim at tracking one or several quantities through measure-  
827 ments of their values over time. Such quantities are for instance temperature, relative humidity,  
828 pressure, concentration of chemical components in a gas or a liquid, noise... These quantities we  
829 aim at monitoring are called measurands.

830 **Definition 1** (Quantity [14]). A quantity is the "*property of a phenomenon, body, or substance,*  
831 *where the property has a magnitude that can be expressed as a number and a reference*".

832 **Definition 2** (Quantity value [14]). A quantity value is a "*number and reference together*  
833 *expressing the magnitude of a quantity*".

834 **Definition 3** (Measurement [14]). A measurement is the "*process of experimentally obtaining*  
835 *one or more quantity values that can reasonably be attributed to a quantity*".

836 **Definition 4** (Measurand [14]). A measurand is the "*quantity intended to be measured*".

837 To carry out measurements, one or several measuring instruments are used. Multiple devices  
838 may be required to perform a measurement and they can be embedded into a single device as a  
839 measuring system. Such a system may be used to track several measurands.

840 **Definition 5** (Measuring instrument [14]). A measuring instrument is a "*device used for making*  
841 *measurements, alone or in conjunction with one or more supplementary devices*".

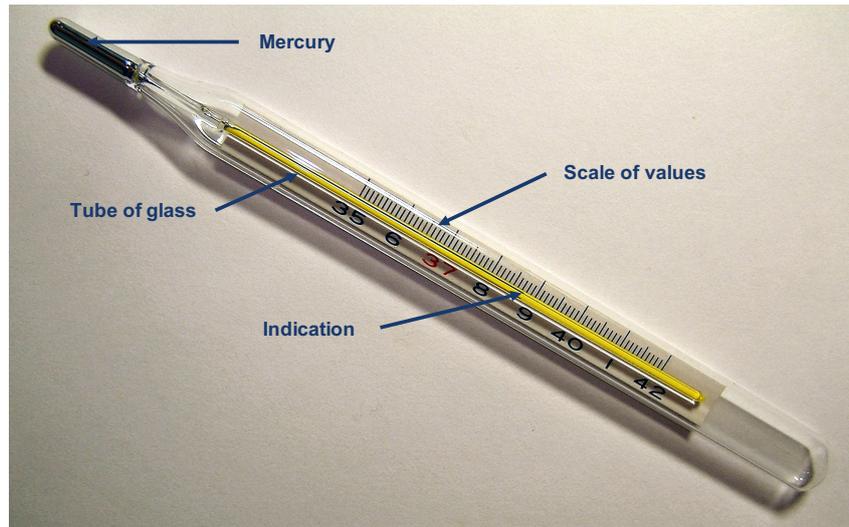
842 **Definition 6** (Measuring system [14]). A measuring system is a "*set of one or more measuring*  
843 *instruments and often other devices, including any reagent and supply, assembled and adapted*  
844 *to give information used to generate measured values within specified intervals for quantities of*  
845 *specified kinds*".

### 846 1.2 Measuring chain of an instrument

847 In a measuring instrument (or system), several components may be necessary to perform a  
848 measurement, forming a measuring chain. It usually starts by one or several sensors<sup>1</sup> and provides  
849 as an output one or several indications. For example, consider a mercury-in-glass thermometer  
850 (Figure 1.1.1). The sensor is the mercury. Depending on the variation of the volume of the liquid  
851 following temperature, its level in a tube of glass is the indication. With an appropriate scale of  
852 value, this level can be converted back to a temperature through direct reading.

---

<sup>1</sup>In ordinary use, the term sensor may refer to a measuring instrument. This is a slight misuse of language.



**Figure 1.1.1:** Description of a mercury-in-glass thermometer<sup>2</sup>

853 **Definition 7** (Measuring chain [14]). A measuring chain is a "series of elements of a measuring  
854 system constituting a single path of the signal from a sensor to an output element".

855 **Definition 8** (Sensor [14]). A sensor is the part of a measuring instrument "that is directly  
856 affected by a phenomenon, body, or substance carrying a quantity to be measured".

857 **Definition 9** (Indication [14]). An indication is a "quantity value provided by a measuring  
858 instrument or a measuring system".

859 As shown in Figure 1.1.1, the raw indication provided by a measuring instrument is not  
860 necessarily of the same kind as the measurand. There is often a quantity conversion that is  
861 performed at the end of the measuring chain so that the measurand and the indications provided  
862 are of the same kind to facilitate the readings. In this case, the indication provided by a measuring  
863 instrument is a measured value. For the conversion step, values from different instruments may  
864 be involved in the case of a measuring system.

865 **Definition 10** (Measured value [14]). A measured value is a "quantity value representing a  
866 measurement result".

867 The values provided by measuring instruments are not necessarily equal to the actual values  
868 of the measurand affecting the sensor. The actual values of measurand are called true quantity  
869 values.

870 **Definition 11** (True quantity value [14]). A true quantity value, or true value, is a "quantity  
871 value consistent with the definition of a quantity".

872 Thus, an uncertainty is usually associated with a measured value to give an idea of its  
873 precision. Its value can be determined through various means (see [72]), the components of the  
874 measuring chain having an influence.

875 **Definition 12** (Uncertainty<sup>3</sup> [14]). An uncertainty is a "non-negative parameter characterizing  
876 the dispersion of the quantity values being attributed to a measurand, based on the information  
877 used."

<sup>2</sup>Modified image. Original source: Menchi (Licence CC BY-SA), <https://commons.wikimedia.org/w/index.php?curid=51236>

<sup>3</sup>The concept of uncertainty is different from the concept of measurement error because even after evaluating the components of the error that are known or supposed, an uncertainty always remains [72]

878 This is notably why a measured value is a value representing a measurement result but not a  
 879 measurement result itself. Usually, a measurement result is expressed by a measured value and an  
 880 uncertainty. A measurement result composed of only a measured value does not allow to estimate  
 881 how close it is to the true value. In fact, any important information for the interpretation of  
 882 a measurement result should be added to it. It can concern the operating conditions, or data  
 883 helping to identify when or where a result was obtained, like a timestamp or GPS coordinates.

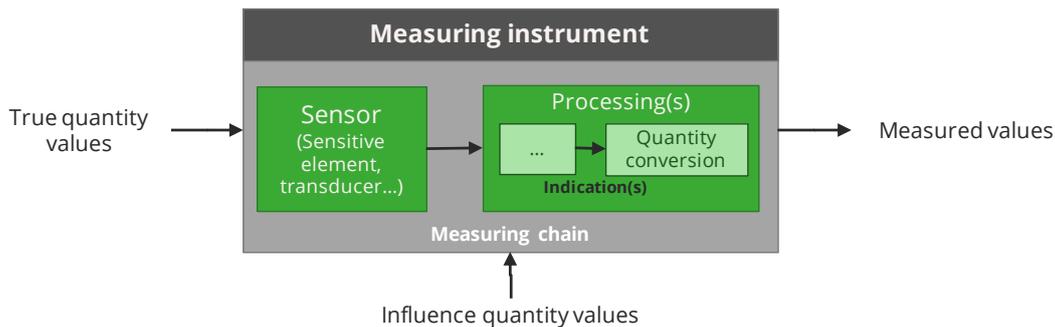
884 **Definition 13** (Measurement result [14]). A measurement result is a "set of quantity values  
 885 being attributed to a measurand together with any other available relevant information".

886 All along the measuring chain, influence quantities may act on the output of the instrument.

887 **Definition 14** (Influence quantity [14]). An influence quantity is a "quantity that, in a direct  
 888 measurement, does not affect the quantity that is actually measured, but affects the relation  
 889 between the indication and the measurement result".

890 The measuring chain of an instrument can be more or less long depending on the measurement  
 891 method. For a digital thermometer, based on a thermocouple for instance, the sensor provides  
 892 a voltage, as an indication of the temperature. Several processings are usually carried out  
 893 afterwards like amplification and filtering, which adds steps in the measuring chain compared to  
 894 the case of a mercury-in-glass thermometer.

895 The description of the measuring chain of an instrument is summarised in Figure 1.1.2.



**Figure 1.1.2:** Schematic description of the measuring chain of a measuring instrument

### 896 1.3 Low-cost instruments

#### 897 1.3.1 Definition

898 For any measurand, there is a wide range of measuring instruments notably in terms of cost.  
 899 Their price is related to the capabilities of instruments in terms of measurement range, sensitivity,  
 900 uncertainty and so on. In the General Introduction, we exposed that reference instruments were  
 901 sometimes too expensive to enable a dense environmental monitoring of different quantities,  
 902 taking as an example the context of air quality monitoring. Consequently, low-cost instruments  
 903 are necessary to help carry out this task.

904 **Definition 15** (Reference measuring instrument). A reference (or regulatory-grade) measuring  
 905 instrument is a measuring instrument implementing a reference measurement procedure or  
 906 equivalent. Such an instrument is usually expensive.

907 **Definition 16** (Low-cost measuring instrument). A low-cost measuring instrument is a measuring  
 908 instrument significantly affordable compared to reference instruments.

909 Depending on the application, what a low-cost measuring instrument is may be very different  
910 [27, 110, 138]. In the context of air pollution monitoring, we speak of devices that cost US\$ 100  
911 or less to few US\$ 1,000 [71, 85, 110] at most, considering that reference instruments can be  
912 worth from US\$ 10,000 to more than US\$ 100,000.

913 Low-cost instruments are often characterised by a small size, thanks to the use of micro  
914 and nano technologies [183]. They are also frequently autonomous in terms of energy: they are  
915 powered by batteries or small solar panels. Most of the time they have wireless communications  
916 capabilities. They can be deployed at fixed positions or be mobile and even wearable [66].

### 917 1.3.2 Challenges of environmental monitoring

918 Few years ago, Hart *et al.* [62] and Rundel *et al.* [127] exposed various challenges for  
919 measuring devices used in environmental monitoring:

920 **Energy supply and management** This subject concerns mainly autonomous devices. In this  
921 case, it aims notably at minimising the energy consumption so that size of the system of  
922 energy storage and charge can be reduced. In addition, the lifespan of the energy supply  
923 components has to be the longest as possible to limit the maintenance operations and its  
924 environmental impact.

925 **Hardware standardisation** To have upgradeable systems, interoperability of the hardware  
926 must be met, with user-friendly manageability.

927 **Software standardisation** In the same way that hardware has to be standardised, the software  
928 of measuring instruments should also be normalised to facilitate the usage of devices from  
929 different manufacturers in the same deployment.

930 **Data standardisation** To help data management and the use of values, data formats must be  
931 standardised.

932 **Data quality improvement** This challenge concerns different aspects. First, in conjunction  
933 with data standardisation, the data quality has to be ensured by providing for each mea-  
934 surement all the necessary and meaningful information regarding its realisation. Secondly,  
935 the accuracy of the measurements has to be improved through the upgrade of the hardware  
936 but also with the help of data processing algorithms.

937 **Data usage reinforcement** To make sure the most is taken out of the acquired data, it is  
938 necessary to facilitate its management and use. It would also be valuable to exploit the  
939 measurements in the largest number of possible applications.

940 **Security reinforcement** As systems are more and more connected, they become vulnerable to  
941 attacks that can affect data integrity. Protocols must be established to prevent tampering  
942 and protect privacy.

943 Among all these challenges, data quality is an important one for low-cost instruments in  
944 particular and is still a major issue [119]. Indeed, if a satisfying data quality cannot be achieved,  
945 the responses to any of the other challenges cited will have a limited impact.

### 946 1.3.3 Performances of low-cost instruments in the literature for air quality 947 monitoring

948 While the concepts introduced and methods developed in this work are generic, the examples  
949 and illustrations provided are taken from the field of air quality monitoring. Indeed, regulatory-  
950 grade instruments as used by air quality surveillance agencies are extremely expensive. Therefore,

951 the interest for low-cost devices is major in this field [27]. To evaluate the severity of data quality  
 952 issues for these devices, a review of their actual performances is conducted in this section.

953 In the literature, there are numerous works proposing an evaluation of the performances of  
 954 low-cost sensors for air quality monitoring. In a review from 2019 [73], the authors reported 105  
 955 references "that include quantitative comparison of sensor data with reference measurements"  
 956 from 2004 to April 2019, with 91 between January 2015 and December 2018. We consider here  
 957 the following works: [15, 16, 71, 73, 85, 121, 141], some of them being aggregates of previous  
 958 publications [73, 121]. They tackle the most common measurands regarding to air quality:  $PM_{10}$   
 959 [15, 16, 73, 121, 141],  $PM_{2.5}$  [15, 16, 71, 73, 85, 121, 141],  $PM_1$  [73, 141],  $O_3$  [15, 16, 71, 73, 85,  
 960 121, 141],  $NO_2$  [15, 16, 71, 73, 121, 141],  $NO$  [15, 16, 71, 73],  $CO$  [15, 16, 71, 73, 85, 121] and  
 961  $SO_2$  [15, 16, 71]. In these studies, the considered low-cost measuring instruments or systems  
 962 are co-located to reference, e.g. regulatory-grade, measuring instruments at a given location, for  
 963 instance an urban environment [15, 141], a suburban environment [71] or a mountain site [85].

964 These studies report a wide range of performances for the measuring instruments or systems  
 965 they consider: from poor to suitable at least for indicative measurements, according to the Data  
 966 Quality Objectives (DQO) set by the European Union [158], and this for all the measurands.  
 967 Minimum and maximal values of the coefficient of determination  $R^2$  reported in the cited  
 968 publications for  $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_1$ ,  $O_3$ ,  $NO_2$ ,  $NO$ ,  $CO$  and  $SO_2$  are listed in Table 1.1.1.  
 969 This metric indicates the quality of the linear regression between the values of an instrument  
 970 and a reference.

Publication	$PM_{10}$	$PM_{2.5}$	$PM_1$	$O_3$	$NO_2$	$NO$	$CO$	$SO_2$
	min – max	min – max	min – max	min – max	min – max	min – max	min – max	min – max
Borrego <i>et al.</i> [15]	0.13–0.36	0.07–0.27	n/a	0.12–0.77	0.02–0.89	0.34–0.80	0.53–0.87	0.09–0.20
Jiao <i>et al.</i> [71]	< 0.25–0.45		n/a	< 0.25–0.94	< 0.25–0.57	0.77–0.87	< 0.25–0.68	< 0.25
Li <i>et al.</i> [85]	n/a	0.82	n/a	0.62	n/a	n/a	0.68	n/a
Spinelle <i>et al.</i> [141]	0.01–0.81	0.00–0.86	0.26–0.95	0.04–0.96	0.35–0.88	n/a	n/a	n/a
Rai <i>et al.</i> [121]*		0.07–0.99		0.01–0.99	< 0.10–0.89	n/a	< 0.10–0.99	n/a
Karagulian <i>et al.</i> [73]*	0.00–1.00	0.00–1.00	< 0.60–0.95 <	0.00–1.00	0.00–1.00	< 0.10–1.00	0.25–1.00	n/a

**Table 1.1.1:** Minimum and maximal values of the coefficient of determination  $R^2$  reported in the cited publications for  $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_1$ ,  $O_3$ ,  $NO_2$ ,  $NO$ ,  $CO$  and  $SO_2$ . Publications marked with a "\*" are surveys aggregating results from other works. Sometimes the values were given for several types of  $PM$  instruments. The values of this table concern different types of devices and do not relate the variability that can be observed with several instruments of the same manufacturer [141].

971 Several authors report that their results should be considered with caution. On the first hand,  
 972 those who obtained results only at one site, under a certain type of operating conditions, explain  
 973 that their work should be repeated with different conditions [141]. On the other hand, those who  
 974 aggregated results point out the diversity of the metrics used [73] and of the conditions under  
 975 which the results were obtained [121], which makes the comparisons difficult. Karagulian *et al.*  
 976 [73] chose to rely on the coefficient of determination  $R^2$  to compare the highest number of studies,  
 977 but they explained that this metric is not the most appropriate. They call for standardised  
 978 protocols for the evaluation of low-cost measuring instruments. Such a protocol for the evaluation  
 979 for low-cost gas sensors for air quality monitoring has been proposed [138].

980 In all of these works, the calibration of the instruments is a main concern. In Section 1.2, we  
 981 mentioned that in general, measuring instruments provide indications not necessarily of the same  
 982 kind as the measurand and that a quantity conversion is often carried out to facilitate the reading  
 983 of measurement results. Determining the relationship to obtain measurements results from  
 984 indications is what calibration allows to achieve. This operation is carried out on a measuring  
 985 instrument for the first time right after its manufacturing or before its first use [67]. Generally,

986 it is performed in a dedicated facility. However, it is only mandatory to have an individual  
987 calibration for regulatory-grade instruments. In most cases, several instruments are calibrated  
988 altogether. Sometimes it also happens that instruments are not calibrated at all.

989 In the publications cited, out-of-the-box measurement results were quite inaccurate most  
990 of the time and some instruments had to be compensated following influence quantities like  
991 temperature, relative humidity or chemical components depending on the measurand and on the  
992 measuring principle at stake [16, 71, 73, 121]. Jiao *et al.* [71] showed for instance an improvement  
993 of the coefficient  $R^2$ :

- 994 • from 0.36 to 0.43, from 0.45 to 0.60 and from 0.43 to 0.50 for three different *PM* instruments
- 995 • from 0.88 to 0.94 for an  $O_3$  instrument
- 996 • from 0.57 to 0.82 for a  $NO_2$  instrument

997 In conclusion, all the authors report that the quality of measurement results can be strongly  
998 improved with the use of data treatment and processing tools. To better understand how data  
999 processing algorithms can improve the quality of the measurement results, the issues causing  
1000 their degradation must be identified.

## 1001 2 Threats to data quality for measuring instruments

### 1002 2.1 Introduction

1003 In Section 1.3.2, the challenge of data quality was briefly presented from different perspectives.  
1004 First, it can be related to the quality of the information composing a measurement result, e.g.  
1005 the measured value, its uncertainty, the conditions under which the measurement was made and  
1006 so on.

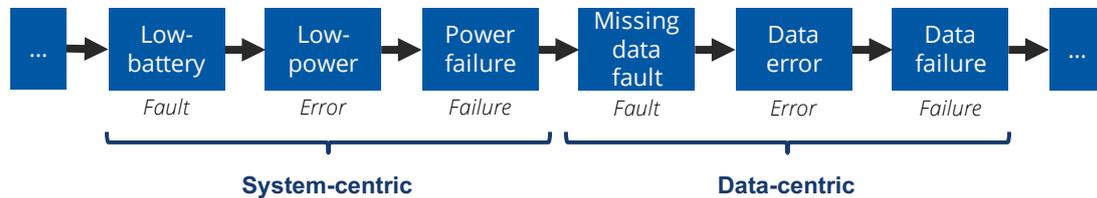
1007 In another way, it can be considered as a lack (or not) of measurement accuracy compared to  
1008 the one of reference instruments in terms of sensitivity or resolution for instance. Such issues  
1009 can be known at the design or prototyping stage of measuring instruments depending on their  
1010 components. Therefore, to improve the quality of the measurements, the performances of the  
1011 instruments have to be upgraded. This can be achieved along the advances during the conception  
1012 and manufacturing of its components.

1013 The data quality issue can also be observed as a challenge regarding the ability of a measuring  
1014 instrument to deliver results consistent with its specifications during its life. In the field of  
1015 dependability, this can be considered as a question of reliability. It is the continuity of correct  
1016 service delivered by a system [5]. For an instrument, the correct service is the achievement of  
1017 measurements consistent with its metrological specifications. It is often observed that the quality  
1018 of low-cost devices decays with time, even under regular operating conditions [33]. This is the  
1019 general issue we focus on in this work.

1020 The threats to the dependability of a system, and consequently the threats to its reliability,  
1021 are **failures**, **errors** and **faults** [5]. A failure is an event making the instrument no longer able  
1022 to deliver a correct service regarding its specifications or requirements. This event is triggered by  
1023 a deviation of the normal state of the system, called an error, and its potential cause is a fault.  
1024 The fault itself may be explained by the failure of another part of the system and so on. Thus,  
1025 there is a relationship of causality between these concepts [5].

1026 In our case, the failure we are interested in for a measuring system is the event when the  
1027 measurement results it delivers are not correct regarding the actual values of the measurand  
1028 and the results that should be provided by the device. To improve the quality of data provided  
1029 by low-cost measuring instruments, it is necessary to determine the faults that are causing this  
1030 failure. The faults of a measuring system can be studied from different perspectives depending  
1031 on the case [115]. Consider an instrument that is in failure mode from the point of view of data

1032 due to the absence of measurement results. With a data-centric point of view, the fault that  
 1033 generated this failure could be called "missing data". However, with a system-centric approach,  
 1034 this "missing data" fault can be explained in several ways, for instance a failure from a component  
 1035 of the measuring chain, a power failure or a communication failure, and these failures could be  
 1036 explained by a wide variety of faults depending on the system considered as illustrated in Figure  
 1037 1.2.1. Therefore, a point of view must be chosen.



**Figure 1.2.1:** Example of situation where a fault from a system-centric point of view generates a fault and then a failure from a data-centric perspective.

1038 We propose to take a data-centric point of view. Indeed, from the review of Section 1.3.3,  
 1039 data processing techniques are seen as promising means to improve the quality of the results of  
 1040 measuring instruments. Consequently, we choose to define the faults that we are interested in  
 1041 from what can be observed from the output of the measuring chain of an instrument, e.g. its  
 1042 indications or its measured values, because they are the inputs of data processing algorithms.  
 1043 The identification of the cause of the faults, whether they concern the measuring chain of the  
 1044 instruments or not, will not be studied here.

## 1045 2.2 Faults

1046 Several taxonomies have been proposed to differentiate faults of measuring instruments and  
 1047 sensor networks [98, 114, 115, 125].

1048 In this work, we propose to use the following taxonomy, inspired from a previous work  
 1049 conducted by Ni *et al.* [115] that focuses on the relationships between faults taken from different  
 1050 points of view.

1051 Eight different types of faults conducting to a possible data failure were identified. The list of  
 1052 possible faults is illustrated in Figure 1.2.2. In this figure, the fault-less signal is  $v_{true}(t) = \sin(\omega t)$   
 1053 with  $t \in [0; 100]$  and  $\omega = 2\pi \times 0.01$ .<sup>4</sup>

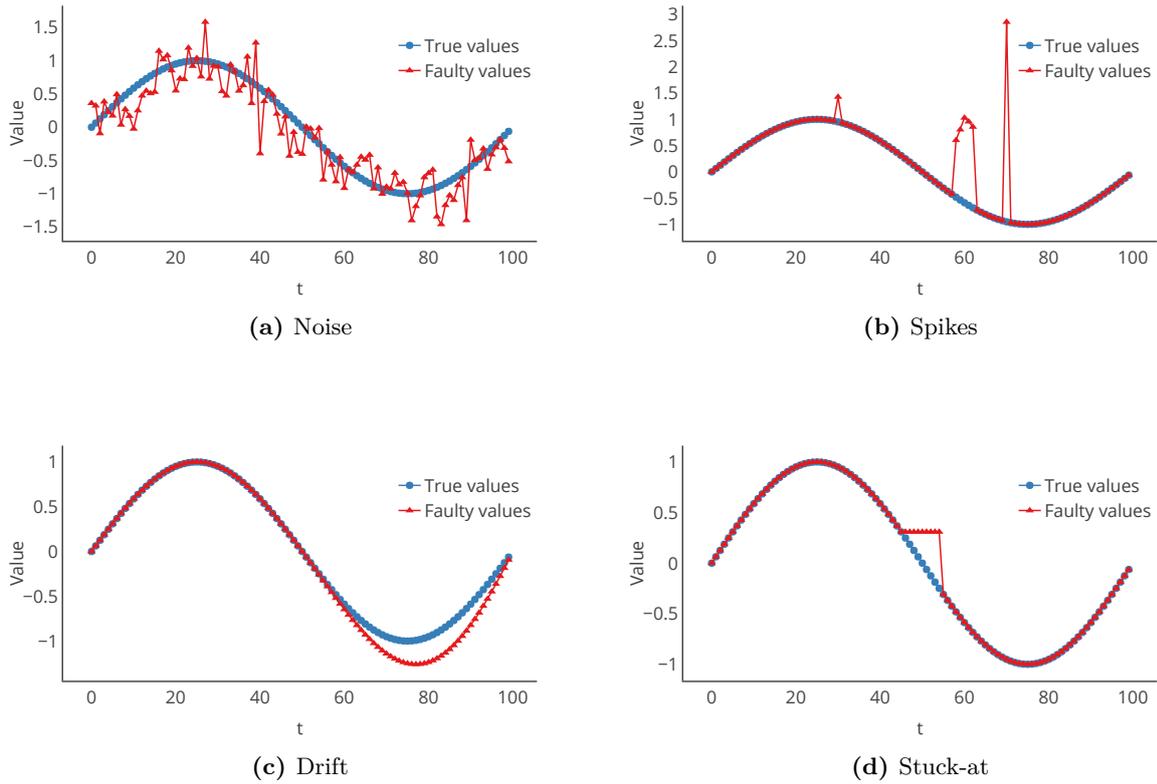
1054 **Noise** Disturbances present in the indications and which may be correlated or not to the true  
 1055 signal. Usually, it is modelled as a Gaussian white noise, e.g. a random variable following  
 1056 a normal law  $\mathcal{N}$  of null average and with a given standard deviation, that is added to the  
 1057 true values. This fault is usually permanent but is also intrinsic to a digital measuring  
 1058 instrument. In fact, noise is considered as a fault when it exceeds the specifications of the  
 1059 instrument. Its severity may vary over time.

1060 In Figure 1.2.2a, the measured values  $v(t)$  are equal to  $v(t) = v_{true}(t) + \varepsilon(t)$  with  $\varepsilon(t) \sim$   
 1061  $\mathcal{N}(0, 0.3)$ .

1062 **Spikes** The measured values at some instants are very different from the ones expected. This  
 1063 observation is based on the previous and following values, if there is no physical cause  
 1064 related to the measurand that could explain this variation. They can be of two types:  
 1065 single-point (only one point differs in the data set) or multi-points (a few points differ in  
 1066 the data set). The spike is characterised by its amplitude and, for multi-points, by its

---

<sup>4</sup>The unit of time is arbitrary.



**Figure 1.2.2:** Considered faults for measuring instruments in the manuscript (continued)

1067 shape, which both may vary. A probability law is associated to determine the frequency of  
 1068 spike occurrence. This fault is transient, but its frequency and severity may vary over time.

1069 In Figure 1.2.2b, two one-point spikes appear at  $t = 30$  and  $t = 70$  with a respective  
 1070 amplitude of 50% and 250% to the related values, and one multi-points spike appears at  
 1071  $t = 58$  for 5 time steps with a maximal amplitude of 75% of the absolute value of the signal  
 1072 and with a triangular shape.

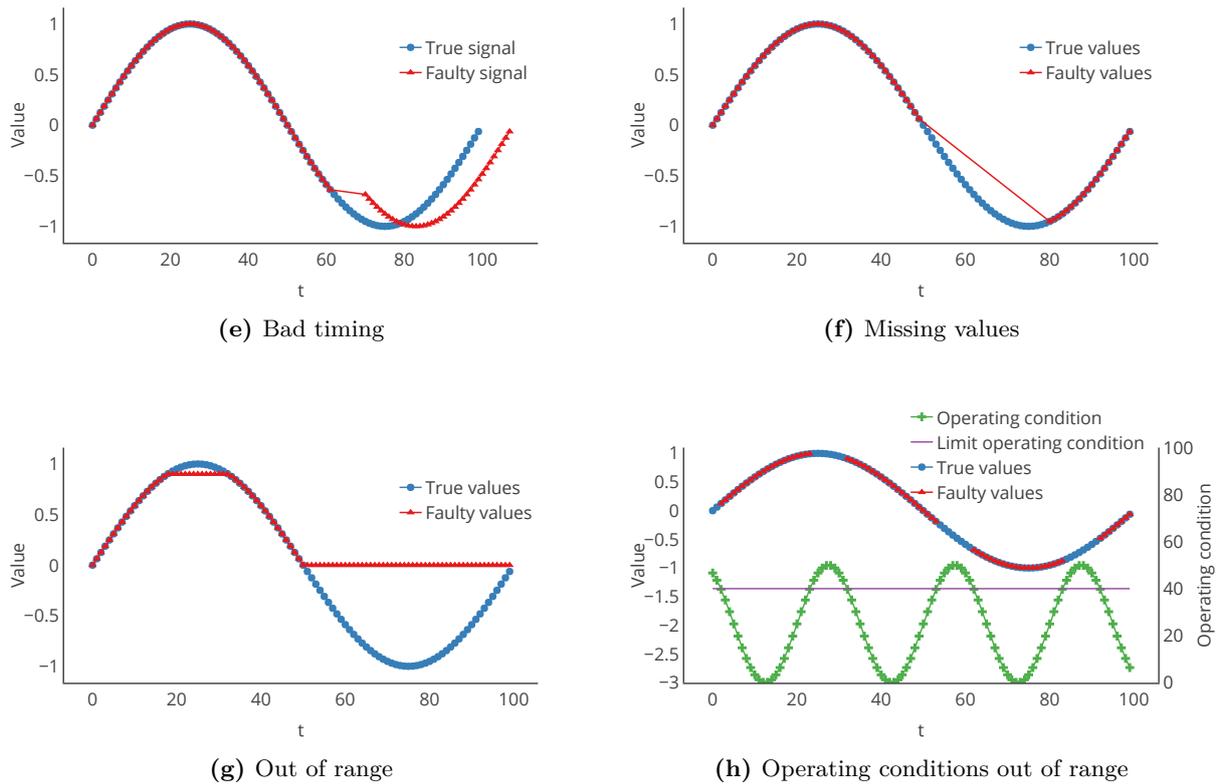
1073 **Drift** The drift fault is "*a continuous or incremental change over time in indication, due to*  
 1074 *changes in metrological properties of a measuring instrument*" [14]. Usually, this fault is  
 1075 permanent and irreversible, except if it is due to influence quantities which the effect is not  
 1076 compensated by the measuring system and is reversible.

1077 In Figure 1.2.2c, there is a drift of the gain  $G$  added, so that  $v(t) = (1 + G(t)) \cdot v_{true}(t)$  It  
 1078 is defined as:

$$G(t) = \begin{cases} 0 & \text{if } t < 50 \\ 0.01(t - 50) & \text{if } t \geq 50 \end{cases}$$

1079 **Stuck-at** Several consecutive values are identical when they should not. This fault can be  
 1080 transient or permanent, e.g. it may last for a certain amount of time or forever when it  
 1081 appears.

1082 In Figure 1.2.2d, this fault appears at  $t = 45$  and lasts for 10 time steps.



**Figure 1.2.2:** Considered faults for measuring instruments in the manuscript

1083 **Bad timing** A measurement result was obtained but has a timestamp differing from the one  
 1084 expected. It can be a transient fault, for instance due to a momentary latency of the  
 1085 instrument, or a permanent fault due to the drift of the clock of the device.

1086 In Figure 1.2.2e, values expected at  $t \in [60; 100]$  are recorded with a delay of 8 units of  
 1087 time.

1088 **Missing values** Measured values have been lost. As for the stuck-at fault, it can be a transient  
 1089 or a permanent fault. In Figure 1.2.2f, values at  $t \in [50; 80]$  are missing.

1090 **Measured values out of range** The recorded measured values are out of the measurement  
 1091 range of the sensor.

1092 The instrument may behave differently depending how it was designed to manage such  
 1093 fault. It may detect this fault and do not record the value, detect it and record it anyway,  
 1094 or it does not have a detection system and record the values normally. If recorded, the  
 1095 value may be equal to one of the bounds of the measurement range, simulating as it is  
 1096 stuck-at, or to another one, more or less related to the true value of the measurand.

1097 In Figure 1.2.2g, we considered the instrument having a range of  $[0; 0.90]$  and recording the  
 1098 values, behaving as it is stuck-at when values are out-of-range.

1099 **Operating conditions out of range** The environment of measurement is out of range of the  
 1100 sensor's normal operating conditions. Like for the fault of measured values out of range,  
 1101 the possibly wrong values may be managed differently depending on the instrument. They  
 1102 can be recorded normally, or recorded but marked as unreliable or not recorded at all.

1103 In Figure 1.2.2h, the instrument is not recording values anymore when the operation  
1104 condition signal exceeds 40.

### 1105 2.3 Discussion

1106 This fault taxonomy is a list of the most common data faults that can be observed in sets of  
1107 measurement results from measuring instruments. Although several faults may be related to  
1108 each other<sup>5</sup>, they provide a reasonable overview of the threats to the data quality of measuring  
1109 instruments, in terms of accuracy of the results. The way they appear and how they manifest in  
1110 the data can be discussed in detail for each type of device but is out of the scope of this work.

1111 These different faults can be differentiated in terms of how an instrument may be able to deal  
1112 with them. Indeed, low-cost environmental measuring instruments have computing capabilities,  
1113 though sometimes minimal. Thus, they can be self-aware of their characteristics and may be  
1114 able to pre-treat their measured values and potentially identify some faults.

1115 For instance, concerning the "out of range" fault, a device can be able to detect when it  
1116 produces a value higher than its maximal possible output if it knows its measurement range. The  
1117 instrument may then raise an error indicating it made an incorrect measurement, either because  
1118 the value of its own measurand was indeed too high compared to its properties or because the  
1119 measuring instrument is producing an erroneous response to the value of the measurand for  
1120 instance. This reasoning can be extended to the faults "bad timing", "missing value" or even  
1121 "operating conditions out of range" if a measuring instrument can access to information on its  
1122 operating conditions, with the help of other devices. Note that these three faults do not concern  
1123 directly the measured value but more how it was obtained. On the contrary, "noise", "stuck-at",  
1124 "spike" and "drift" faults concern directly the measured value like the "out of range" fault. To  
1125 detect them with a data-centric approach, an analysis of the measured values has to be carried  
1126 out. "Noise", "stuck-at" and "spike" are faults having a characteristic time shorter than a "drift"  
1127 fault which usually affects a measuring instrument on a longer time basis. For them it is possible  
1128 to consider studying the variance or the gradient of the measured signal in a first step [112],  
1129 although more elaborated techniques have been developed [114]. Be that as it may, self correcting  
1130 its faults is not straightforward for measuring instruments.

1131 In fact, drift is one of the most problematic faults because it is a continuous change over time  
1132 in the indications of the instruments. In this way, the relationship between the true values of  
1133 the measurand and the indications evolves. Consequently, the relationship determined through  
1134 a previous calibration to derive measured values from true values has to be determined again.  
1135 This is why calibration is an operation also conducted during the maintenance of an instrument  
1136 and it must be carried out periodically. To follow the traditional guidelines, it requires to take it  
1137 out of service and to bring it to a calibration facility. After this operation and the adjustment of  
1138 the instruments, it can be brought back to its deployment site and put back in service. However,  
1139 in the context of large deployments of measuring instruments, carrying out this maintenance  
1140 operation in this manner can have a significant cost, both economically and technically. It is all  
1141 the more important that low-cost devices are particularly prone to declining performances, often  
1142 sooner than expected [33].

## 1143 3 Calibration of measuring instruments

1144 In Section 1.3.3, we briefly introduced the concept of calibration. It was reported as a main  
1145 concern in studies on the performances of low-cost instruments. Thus, we provide here its formal  
1146 definition and develop why it is a challenging procedure for large deployments of measuring

---

<sup>5</sup>For instance, a "spike" fault may trigger a "measured value out of range" fault which itself may be represented by a "stuck-at". However, this is not always the case: a "measured value out of range" fault can happen due to the normal evolution of a measurand.

1147 instruments.

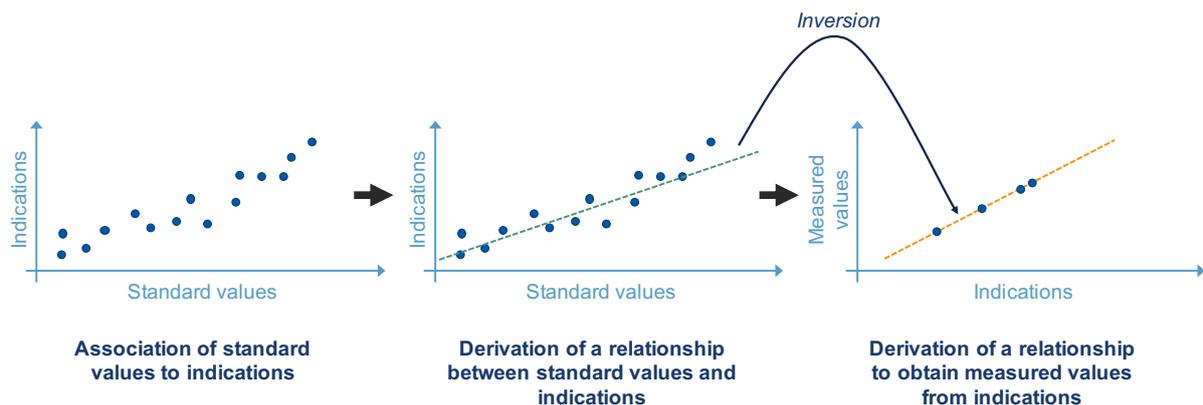
### 1148 3.1 Definition

1149 The formal definition of calibration is the following:

1150 **Definition 17** (Calibration [14]). Calibration is the "operation that, under specified conditions,  
 1151 in a first step, establishes a relation between the quantity values with measurement uncertainties  
 1152 provided by measurement standards and corresponding indications with associated measurement  
 1153 uncertainties and, in a second step, uses this information to establish a relation for obtaining a  
 1154 measurement result from an indication".

1155 **Definition 18** (Measurement standard [14]). A measurement standard is the "realisation of the  
 1156 definition of a given quantity, with stated quantity value and associated measurement uncertainty,  
 1157 used as a reference".

1158 In practice, it means in a first step that the calibration of a measuring instruments consists  
 1159 into observing its response to known values of the measurand. From these observations it is  
 1160 possible to determine a relationship giving the indication provided by a measuring instrument  
 1161 following the value of the measurand. To obtain the measured values from the indications, this  
 1162 relationship has to be inverted. This is summarised in Figure 1.3.1



**Figure 1.3.1:** Schematic description of the steps of a calibration operation

1163 As an example, consider the case of a mercury-in-class thermometer once again. Without  
 1164 its scale, we do not read a temperature value (measured value), we read a level of mercury  
 1165 (indication). Thus, the calibration of this instrument consists into putting the device into several  
 1166 environments for which the conditions are controlled, and the value of the temperature known  
 1167 with its uncertainty. For each value of temperature, the level of mercury is read. Then, it is  
 1168 possible to determine the level of mercury as a function of the temperature. Building the scale  
 1169 for the thermometer consists into inverting this relationship, which is a simple task in this case  
 1170 because it can be approximated as a linear relationship.

1171 Following the VIM [14], "calibration may be expressed by a statement, calibration function,  
 1172 calibration diagram, calibration curve, or calibration table. In some cases, it may consist of an  
 1173 additive or multiplicative correction of the indication with associated measurement uncertainty",  
 1174 a correction being a "compensation for an estimated systematic effect".

1175 The proper achievement of calibration for a measuring instrument is notably meant to ensure  
 1176 its trueness.

### 3.2 Analysis

Let us have a short analysis of calibration from a mathematical perspective. In this case, the result of calibration is expressed as a function. Consider that  $\zeta(s_i, t)$  is the indication provided by  $s_i$  at  $t$  and that the measuring chain of the measuring instrument is represented as a function  $\mathcal{F}$ .

Therefore,

$$\forall t, \zeta(s_i, t) = \mathcal{F}(v_{true}(s_i, t), \mathbf{q}(t))$$

where  $v_{true}$  are the true values of the measurand and  $\mathbf{q}$  is a vector of the values of the influence quantities of the considered instrument.<sup>6</sup>

The first step of calibration aims at determining  $\hat{\mathcal{F}}$  which is an estimate of  $\mathcal{F}$ . Then, the purpose is to derive  $\hat{\mathcal{F}}^{-1}$  such as:

$$v(s_i, t) = (\mathcal{F} \circ \hat{\mathcal{F}}^{-1})(v_{true}(s_i, t), \mathbf{q}(t))$$

where  $v$  are the measured values. Ideally,  $v(s_i, t) = v_{true}(s_i, t)$ .

However, deriving this function may be challenging because  $\hat{\mathcal{F}}$  depends on  $\mathbf{q}(t_{calibration})$ , where  $t_{calibration}$  is the instant of the calibration. If there is no influence quantity,  $\hat{\mathcal{F}}^{-1}$  can be determined without any problem. On the opposite case,  $\hat{\mathcal{F}}^{-1}$  should be determined on the entire space of  $\mathbf{q}$ , so that with  $\hat{\mathbf{q}}$ , an estimator of the values of the influence quantities, we can have:

$$v(s_i, t) = \hat{\mathcal{F}}^{-1}(\zeta(s_i, t), \hat{\mathbf{q}}(t)) = \hat{\mathcal{F}}^{-1}(\mathcal{F}(v_{true}(s_i, t), \mathbf{q}(t)), \hat{\mathbf{q}}(t)) \quad (1.1)$$

Obtaining this relationship over the entire space of  $\mathbf{q}$  is difficult, without considering that the uncertainties also have to be derived. This is why there are two steps in a calibration procedure, e.g. the determination of  $\hat{\mathcal{F}}$  and then of  $\hat{\mathcal{F}}^{-1}$ , the first step alone being often perceived as calibration.  $\hat{\mathcal{F}}$  and  $\hat{\mathcal{F}}^{-1}$  are often determined under a given finite set of operating conditions. The same reasoning can be extended to any type of expression chosen for the results of calibration (curve, diagram, table...).

### 3.3 *In situ* calibration

Calibration is an operation having a major impact on the quality of the measurements performed by an instrument. Indeed, the output of calibration is a relationship enabling to derive measurement results from indications, e.g. measured values and their associated uncertainties. Through this step, making the budget of the uncertainties on the indications is a prerequisite, and if there are influence quantities at stake, they must be taken into account. Thus, from its definition and all its constraints, it is justified to usually carry out calibration operations in dedicated facilities. This is manageable for few tens of instruments and corresponds to what is done by air quality monitoring agencies for instance [3] but it may be less the case in a deployment involving hundreds of nodes, particularly if they are low-cost. Indeed, these nodes are more prone to drift, requiring more frequent calibrations. Also, it was observed in some works [27, 124] that even if a laboratory calibration is performed, such instruments may behave differently once deployed in the field due to the uncontrolled operating conditions. In this context, maintaining a dense deployment of instruments for environmental sensing, with a significant number of low-cost devices is a tremendous task with classical approaches.

This is why research works emerged on *in situ* calibration algorithms, mostly in the past decade.

**Definition 19** (*In situ* calibration algorithms). *In situ* calibration algorithms aim at calibrating measuring instruments while leaving them in the field, preferably without any physical

<sup>6</sup>Quantities involved in the operating conditions of the calibration can be influence quantities too.

1210 intervention.

1211 They are also called *field* [64], *in place* [24] or *online* [150] calibration algorithms.<sup>7</sup>

1212 The literature on the subject studies under which hypotheses, in which manner and with  
 1213 which performances the measurement accuracy of one or several instruments can be improved  
 1214 with such algorithms. Notably in the context of a dense deployment of measuring instruments,  
 1215 comparisons of the measurement results of different instruments could be carried out for instance  
 1216 to perform this task and this is exactly what these algorithms aims at exploiting in practice.

Most of the time, *in situ* calibration strategies do not take indications as an input but measured values. Mathematically speaking, considering the developments presented in Section 3.2, it means it is not the function  $\widehat{\mathcal{F}}$  that is derived but a function  $\mathcal{H}$  mapping what are considered as standard values and potentially influence quantity values to measured values so that:

$$\mathcal{H} : (v_{standard}(s_i, t), \mathbf{q}(t)) \rightarrow v(s_i, t)$$

Then  $\mathcal{H}^{-1}$  is derived and, in the same way that we expressed concerns regarding the determination of  $\widehat{\mathcal{F}}^{-18}$ , the corrected values  $v_{corr}(s_i, t)$  that we obtain after an *in situ* calibration can be expressed as:

$$v_{corr}(s_i, t) = \widehat{\mathcal{H}}^{-1}(v(s_i, t), \widehat{\mathbf{q}}(t)) \quad (1.2)$$

1217 If the *in situ* calibration is perfect, then  $v_{corr}(s_i, t) = v_{true}(s_i, t)$ .

1218 In summary, *in situ* calibration strategies aim at giving relationships to obtain **corrected**  
 1219 **measured values** from measured values. To do so, it uses standards defined with the measured  
 1220 results of other measuring instruments. The measurement uncertainties associated with the  
 1221 measured value measured both by the instruments to calibrate and the ones used to provide  
 1222 standard values can be used to derive the uncertainties of the corrected values. Finally, the  
 1223 operating conditions may be known through instruments measuring other quantities. Thus, *in*  
 1224 *situ* calibration is a procedure conceptually close to the formal definition of calibration.

### 1225 3.4 Discussion

1226 Deploying a large number of instruments in an environment may offer opportunities to  
 1227 overcome the problem of calibration for low-cost measuring instruments and more generally the  
 1228 issue of data quality. Indeed, the devices deployed can be related in a network, allowing them to  
 1229 share information like their measurement results.

1230 However, regarding the formal definition of calibration and depending on the point of view,  
 1231 using results from instruments obtained in an uncontrolled environment to calibrate other devices  
 1232 may not be considered as equivalent to employing measurement standards or to being under  
 1233 controlled conditions. This is particularly true if results from an instrument of an equivalent  
 1234 accuracy class than another is used to calibrate another device.

1235 **Definition 20** (Accuracy class [14]). The accuracy class (or class) is the "*class of measuring*  
 1236 *instruments or measuring systems that meet stated metrological requirements that are intended to*  
 1237 *keep measurement errors or instrumental measurement uncertainties within specified limits under*  
 1238 *specified operating conditions*".

1239 Indeed, values in which a high confidence can be put are used to calibrate measuring  
 1240 instruments. Thus, the sources of the measurement standards and the conditions under which  
 1241 an algorithm is applied are critical for *in situ* calibration strategies.

---

<sup>7</sup>Calibration without any adjective is also used but it may be confusing regarding the definition of calibration in the VIM [14].

<sup>8</sup>e.g. it should be carried out on the entire space of  $\mathbf{q}$ , with an estimator  $\widehat{\mathbf{q}}$  of the values of the potential influence quantities.

1242 **4 Sensor networks**

1243 We define in this section what a sensor network is. Note that we do not discuss (wireless  
 1244 or not) sensor networks in general here. We focus on its characteristics of interest from the  
 1245 perspective of *in situ* calibration of measuring instruments.

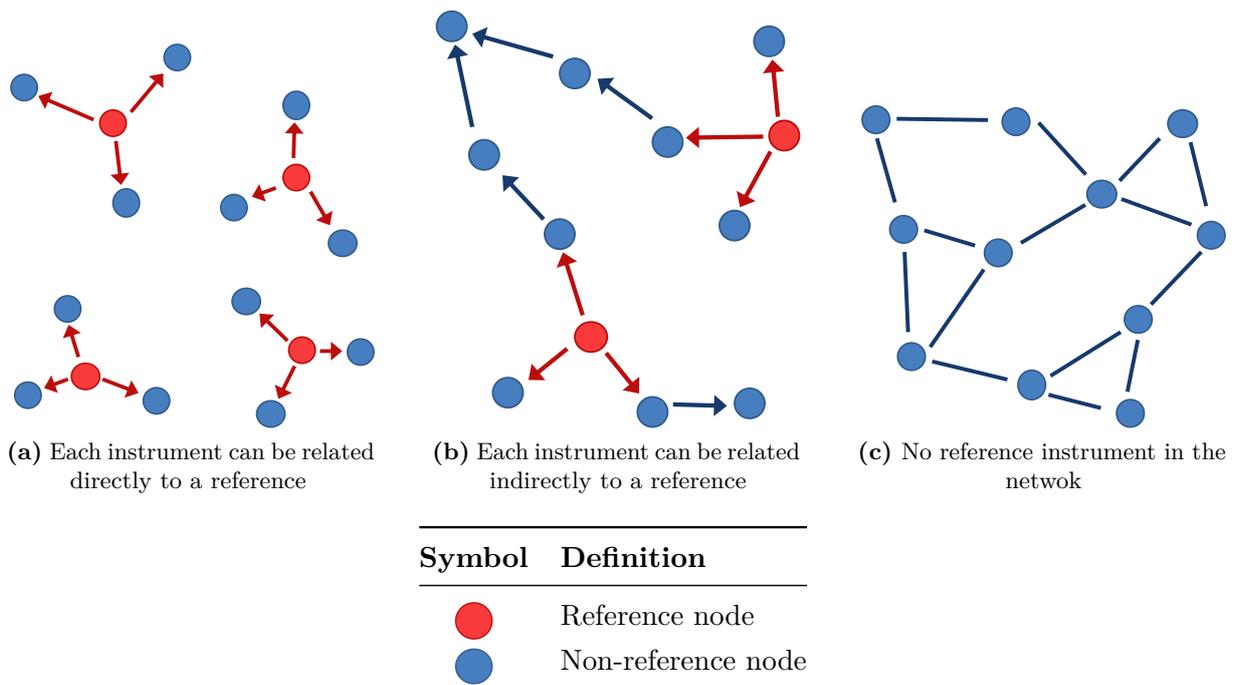
1246 **4.1 Definition**

1247 **Definition 21** (Sensor network [62]). A sensor network<sup>9</sup> is a set of measuring systems spatially  
 1248 deployed in order to periodically measure one or more quantities in an environment and that  
 1249 are exchanging information between them through a communication layer that can be wired or  
 1250 wireless.

1251 The measuring systems of the network are also called **nodes**. A node may be static or mobile.  
 1252 The network can be either meshed, with device to device communications, or a collection of stars  
 1253 centred on gateways for instance. A sensor network may target a single or multiple measurands.

1254 For a given measurand, instruments in a same network are not necessarily of the same  
 1255 accuracy class. In this work, those which are known to be more accurate than the others of the  
 1256 network are called **reference instruments**.

1257 It is also supposed that the nodes have sufficient computing and communication capabilities to  
 1258 carry out the operations proposed by algorithms to overcome the issue of their *in situ* calibration.



**Figure 1.4.1:** Examples of sensor networks with a different number of reference instruments and different relationships between the instruments.

1259 **4.2 Characteristics of interest in this work**

1260 Regarding the issue of data quality for low-cost instruments, there are two characteristics of  
 1261 sensor networks particularly interesting to tackle this issue: the presence of reference instruments

<sup>9</sup>Like sensor being used to refer to a measuring instrument, there is also a slight misuse of language here. However, this expression is widely used.

1262 for each measurand and the mobility of the nodes.

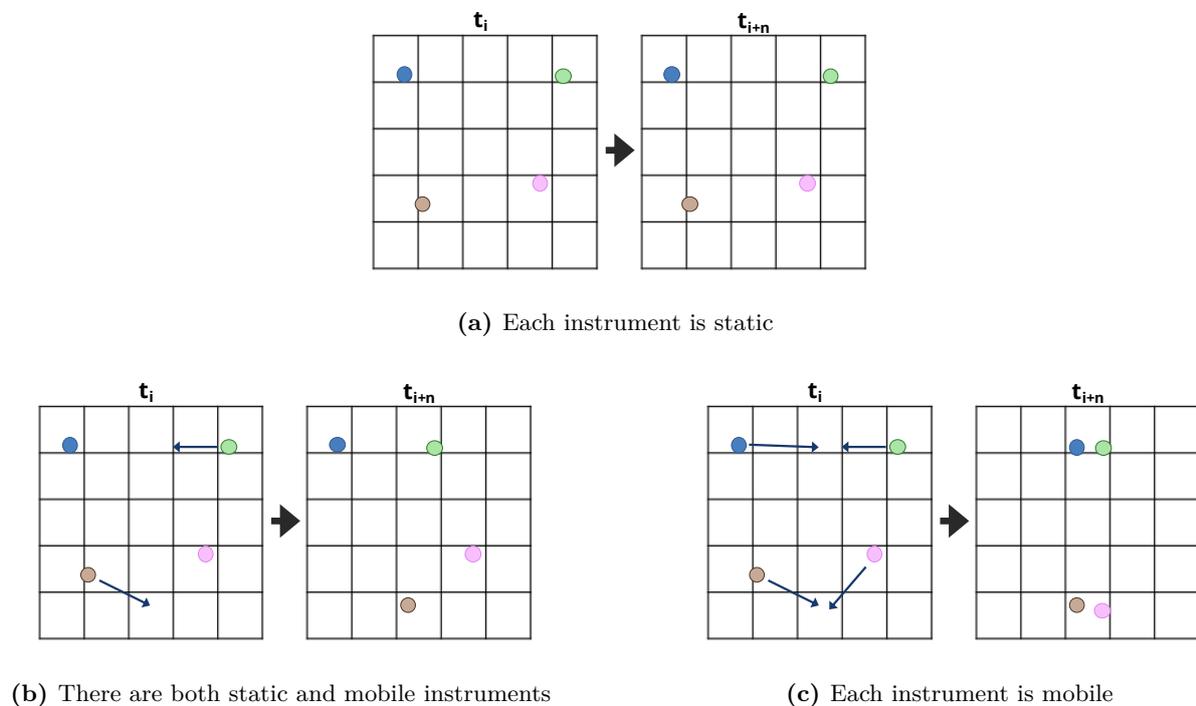
1263 **4.2.1 Presence of references**

1264 In sensor networks, reference instruments may be present in different manners. First of  
 1265 all, it is possible that each non-reference node can compare its measurement results directly to  
 1266 those of a reference one. It is illustrated in Figure 1.4.1a where the edges represent the possible  
 1267 comparisons. They can be defined following a criterion of distance between the nodes.

1268 Another possibility is that each instrument can be related to a reference node but in this  
 1269 case there may be intermediary nodes for some instruments as shown in Figure 1.4.1b.

1270 Finally, the last possible case is that there is no reference instrument in the network. This is  
 1271 the example of Figure 1.4.1c. Edges were drawn representing relationships between non-reference  
 1272 instruments.

1273 In the case of nodes that are measuring systems which do not measure the same quantities,  
 1274 it is possible that the graph of relationship between instruments is different depending on the  
 1275 measurand.



**Figure 1.4.2:** Illustration of the principle of static, static and mobile, and mobile sensor networks.

1276 **4.2.2 Mobility**

1277 We already mentioned that measuring instruments deployed for environmental monitoring  
 1278 can remain at a fixed position, e.g. they are static, or they can change of position over time, e.g.  
 1279 they are mobile.<sup>10</sup> Thus there are three different situation that can happen, regardless of the  
 1280 accuracy class of the instruments:

- 1281 • all the nodes of the network are static.
- 1282 • all the nodes are mobile.

<sup>10</sup>The case of wearable instruments is considered as a particular case of mobile instruments

- 1283     • there are both static and mobile nodes in the network.

1284     This is illustrated in Figure 1.4.2.

1285     Note that in the presence of mobile nodes, the relationships between reference and non-  
1286 reference instruments as exposed in the previous section may change overtime.

### 1287     **4.3 Conclusion**

1288     In this section, we defined what a sensor network is in a general manner. Two characteristics  
1289 of sensor network were presented respectively regarding the presence of reference instruments in  
1290 the network and the mobility of the nodes. The different possible cases for each property were  
1291 exposed. They are interesting for the problem we want to address because the configuration of  
1292 the sensor network can be exploited to drive the design of *in situ* calibration algorithms.

## 1293     **5 Problem statement**

1294     Based on the previous sections, we now detail the problem statement of this thesis briefly  
1295 exposed above in the General Introduction.

### 1296     **5.1 Motivations**

1297     To perform a dense measurement of ambient quantities in the context of environmental  
1298 monitoring activities such as air quality, the use of low-cost instruments is necessary to achieve  
1299 it at a reasonable cost economically speaking.

1300     Among all the challenges for the measurement of ambient quantities in the context of  
1301 environmental monitoring, the need for an improvement of data quality appears to be major,  
1302 particularly for low-cost devices. Indeed, they are prone to faults, one of them being the  
1303 instrumental drift.

1304     The traditional solution to mitigate this fault is to regularly perform calibration, as part  
1305 of maintenance operations. However, the cost of these tasks may limit the feasibility of dense  
1306 deployments of measuring instruments in practice. Indeed, it usually requires taking out the  
1307 node of the service, bringing it to a calibration facility, calibrating it and putting it back in  
1308 service. Therefore, there is a need for tools that allow recalibrating measuring instruments while  
1309 leaving them on the field.

1310     Because the aim of environmental monitoring with low-cost instruments is notably to have  
1311 a high density of instruments in the monitored area, it is tempting to use the measurement  
1312 results of other instruments to detect and potentially correct drift faults. The sensor network  
1313 thus constituted may offer opportunities to overcome this issue of data quality. In addition, the  
1314 presence of reference instruments in the network, or the mobility of the nodes are characteristics  
1315 that can facilitate the *in situ* calibration of measuring instruments.

1316     Fortunately, researchers have begun to tackle this issue and there are already methodologies  
1317 exploiting the properties of sensor networks to enable an *in situ* calibration of measuring  
1318 instruments.

### 1319     **5.2 Objectives of the thesis**

1320     As stated in the General Introduction, the objective of this thesis is to contribute to the  
1321 research effort on the improvement of data quality for low-cost measuring instruments through  
1322 their *in situ* calibration.

1323     In the first place, the goal is to identify the existing techniques in the literature enabling such  
1324 a calibration. In Section 4, we reported there are sensor networks with different characteristics in  
1325 terms of presence of reference instruments and of mobility of the nodes. Thus, the same *in situ*  
1326 calibration algorithm may not be applicable to each type of sensor network for instance. This is

1327 why a taxonomy is developed for the description of such algorithms, facilitating the identification  
1328 of the strategies that could be applied to a given sensor network depending on its properties and  
1329 the characteristics of its instruments.

1330 In the case where several *in situ* calibration algorithms are adapted to a sensor network, it  
1331 is possible that each strategy does not yield the same performances. In other words, they may  
1332 not all allow correcting instruments as well as possible. Being able to evaluate the performances  
1333 of different *in situ* calibration strategies on the same use case would be valuable. The second  
1334 objective of this thesis is to provide a means to achieve this task.

1335 Finally, before calibrating measuring instruments, it is necessary to determine if an instrument  
1336 needs to undergo such an operation. In publications introducing *in situ* calibration algorithms,  
1337 they mainly focus on how to perform the calibration itself and less on how to identify the faulty  
1338 instruments in a sensor network. Identifying these instruments is the last objective of this thesis.

Chapter  
**2**

**In Situ Calibration Algorithms for  
Environmental Sensor Networks**

---

**Contents**

---

1339  
  
1340  
  
1341  
  
1342  
  
  
  
1343  
  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364

Introduction . . . . .	<b>24</b>
1 Scope of the taxonomy . . . . .	<b>24</b>
2 Taxonomy for the classification of the algorithms . . . . .	<b>25</b>
2.1 Use of reference instruments . . . . .	25
2.2 Mobility of the instruments . . . . .	26
2.3 Calibration relationships . . . . .	26
2.4 Instrument grouping strategies . . . . .	27
3 Comparison to other taxonomies . . . . .	<b>28</b>
4 Review of the literature based on this classification . . . . .	<b>30</b>
4.1 Overview . . . . .	30
4.2 Mobile and static nodes . . . . .	32
4.3 Calibration relationships . . . . .	33
4.4 Pairwise strategies . . . . .	33
4.5 Blind macro calibration . . . . .	35
4.6 Group strategies . . . . .	36
4.7 Comment regarding other surveys . . . . .	37
5 Conclusion . . . . .	<b>37</b>

---

1365

## Introduction

1366

1367

1368

1369

1370

1371

1372

1373

1374

1375

1376

1377

1378

1379

1380

1381

1382

1383

1384

1385

1386

1387

1388

1389

1390

1391

1392

1393

1394

1395

1396

1397

1398

In Chapter 1, we recalled that calibration is a procedure which is part of the maintenance operations that must be carried out periodically on measuring instruments. To follow the traditional guidelines for a device deployed in the field, it requires to take it out of service and to bring it to a calibration facility. After this operation and the adjustment of the instruments, it can be brought back to its deployment site and put in service again. This is manageable for few tens of instruments and corresponds to what is done by air quality monitoring agencies for instance [3]. However, notably in the particular case of environmental sensing, the increase of the number of measuring instruments deployed to ensure a higher spatial coverage may not be feasible due to the cost of regulatory-grade instruments. This is why low-cost instruments are necessary to achieve this goal. In the literature, it is reported they are suffering from multiple issues, one of them being a faster drift than usual instruments. Regarding calibration, it forces to increase the frequency of this maintenance operation and thus its cost, economically and technically. Also, it was observed in some works [27, 124] that even if a laboratory calibration is performed, once deployed in the field, the instruments may behave differently due to the uncontrolled operating conditions. In this context, maintaining a sensor network for environmental sensing and composed of a significant number of low-cost instruments with classical approaches is a tremendous task.

This is why research works emerged on *in situ* calibration algorithms, mostly in the past decade. The literature on the subject is abundant. This is expected as the topic of sensor networks is a very hot research subject due to the trend for Smart Cities and the Internet of Things. More practically, it is also predictable due to the variety of sensor networks as described in Chapter 1, Section 4. Indeed, one can easily suspect that an algorithm may not be applicable to any type of sensor network, on top of the assumptions surrounding it.

A major gap is the absence of a unified terminology to describe the field of application of *in situ* calibration algorithms, regardless of the quantities measured by the sensor network. The large variety of different techniques are reported under different terms in the literature, such as "blind calibration", "multi-hop calibration", "macro calibration" and so on. The expression describing an algorithm, notably in the title or the abstract of publications, is sometimes not unique and is often incomplete. For instance, Balzano *et al.* [7] or Wang *et al.* [168] present their strategies as "blind calibration" algorithms but it does not indicate if it is applicable to static sensor networks, mobile ones or both. Consequently, it is not crystal-clear for an end user from the first lines of an article if its content matches his/her needs.

Thus, through the literature review carried out on *in situ* calibration algorithms in this thesis, a synthetic taxonomy for their classification is proposed in the following sections.

1399

### 1 Scope of the taxonomy

1400

1401

1402

1403

1404

1405

1406

1407

1408

1409

1410

1411

1412

To build the taxonomy, several guidelines were adopted. First of all, it should help to identify relevant algorithms for any type of sensor network.

Also, the terms must be independent from the kind of measurand of the sensor networks: the groups of categories described are relevant for any environmental phenomenon.

The attributes should also be limited to the properties involved in the definition of *in situ* calibration, e.g. the definition of the measurement standards, how the relationship between measured values and standard values is established and the relationship correcting the measured values.

There are other considerations that could matter and interact with calibration, such as data integrity, security or even privacy, in particular when crowdsensing platforms involving citizens are developed [102]. The way the calibration relationship is computed, e.g. in a centralised or decentralised manner is also a major subject. As a matter of fact, the care for the management of these questions at an early step of the design of the sensor networks is crucial. The way these

1413 concepts are implemented with respect to the system architecture may have a significant impact  
1414 on the effectiveness of some calibration methods.

1415 Nevertheless, they are not attributes that are directly driving the result of the calibration  
1416 procedure compared to the network architecture and the algorithmic principles underlying the  
1417 calibration method. Thus they are not in the scope of the presented taxonomy but defining  
1418 characteristics that should be clarified in a second step to describe an *in situ* calibration strategy  
1419 as part of a "datasheet" could be valuable for future works.

## 1420 2 Taxonomy for the classification of the algorithms

1421 The following subsections introduce the proposed taxonomy for the classification of *in situ*  
1422 calibration strategies for sensor networks.

1423 They can be divided into two categories: network architecture characteristics, namely the  
1424 nature of instruments and their potential mobility, and the algorithmic principles of the calibration  
1425 techniques, namely the mathematical structure of the calibration relationship and to which point  
1426 the algorithm can be distributed.

1427 Each subsection represents a primary level group of categories that may have others nested.  
1428 Categories for each group are in bold font. Figure 2.2.1 represents all the primary levels of the  
1429 taxonomy and the defined categories.

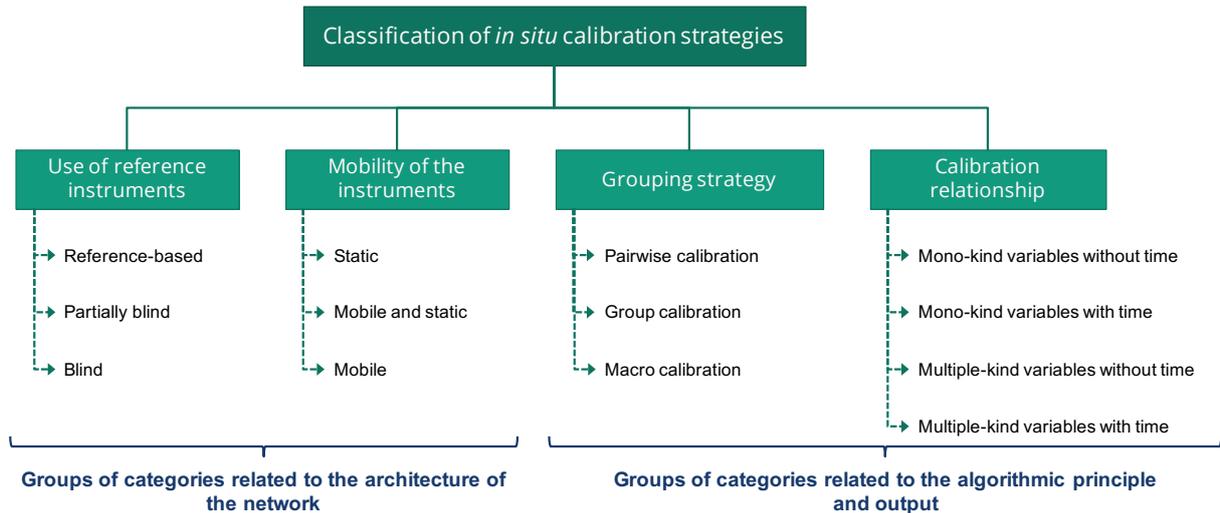


Figure 2.2.1: Proposed taxonomy for the classification of *in situ* calibration algorithms

### 1430 2.1 Use of reference instruments

1431 One of the first criteria of classification is how the calibration method assumes the presence  
1432 of reference instruments within the network. This is why this characteristic was introduced as a  
1433 major feature of sensor networks for our problem in Chapter 1 Section 4.2.1. Three categories of  
1434 *in situ* calibration strategies are derived from this property with the examples from Figure 1.4.1  
1435 in Chapter 1 to illustrate them.

1436 The calibration of measuring instruments using a sufficient number of reference measurement  
1437 instruments is called **reference-based calibration**. It means the network is composed of  
1438 both reference and non-reference instruments and that all the non-reference instruments can  
1439 be calibrated using at least one reference instrument. The approach postulates the existence  
1440 of a calibration relationship between each non-reference instrument and at least one reference  
1441 instrument because they are close enough for instance.

1442 The calibration of measuring instruments in the absence of reference values is called **blind**  
1443 **calibration**. It means the network is composed of only non-reference instruments. These various  
1444 methods may or may not assume the existence of a correlation between the instrument outputs.

1445 The hybrid situation is called **partially blind calibration**. In this setting, the network  
1446 may gather both reference and non-reference instruments, but a reference-based calibration is  
1447 not achievable, e.g. when some of the non-reference instruments can never be compared to a  
1448 reference instrument. It also captures the cases where a cluster of non-reference instruments is  
1449 considered as good enough to approximate a reference instrument.

## 1450 2.2 Mobility of the instruments

1451 The second significant aspect of the network architecture is the potential mobility of nodes.  
1452 As for the presence of references, this characteristic was also introduced as a major feature  
1453 of sensor networks in Chapter 1 Section 4.2.2. In the same way, three categories of *in situ*  
1454 calibration strategies are considered to illustrate how the methods exploit this feature. Figure  
1455 1.4.2 in Chapter 1 provides examples also illustrating these categories.

1456 A first category of methods addresses networks with **exclusively static** nodes. A second  
1457 one addresses networks with **exclusively mobile** nodes. The corresponding methods rely often  
1458 strongly on the mobility of the nodes to achieve calibration. A last group of methods addresses  
1459 heterogeneous networks with **both mobile and static** nodes. In such cases the mobility of the  
1460 nodes is not systematically exploited in the calibration strategy.

## 1461 2.3 Calibration relationships

1462 In Chapter 1 Section 3, the purpose of calibration was defined as the establishment of a  
1463 mathematical relationship between the indications of the instrument and the standard values of  
1464 the measurand and then the derivation of a relationship for obtaining measurement results from  
1465 indications. Based on the mathematical developments in Chapter 1 Section 3.2, it is possible to  
1466 have different types of relationships, for instance if there are influence quantities or not.

1467 The categories identified in this section are first based on the number of kinds of quantities as  
1468 input variables in the relationship: the measurand, the indications, the influence quantities, and  
1469 so on. In terms of algorithmic principles of the calibration methods, it implies the variety and  
1470 quantity of data to exchange as well as the computational effort necessary to achieve a target  
1471 accuracy.

1472 The most straightforward relationships are called **mono-kind variables without time**.  
1473 They only take a single quantity as input variable and do not depend on time.

1474 The second category of relationships gathers the ones that have **mono-kind variables with**  
1475 **time**. It accounts for a relationship with mono-kind variables which is influenced by time, for  
1476 instance in case of an instrument drifting due to ageing [160]. This is the reason why time is  
1477 considered as a particular influence quantity.

1478 The relationships with **multiple-kind variables without time** account for two or more  
1479 quantities as variables but remain independent from time. These models are mainly used  
1480 to include the effect of influence quantities (except time) in the calibration relationship. In  
1481 these cases, the networks include instruments measuring the influence quantities. They are not  
1482 systematically reference instruments and therefore their calibration may also be included in the  
1483 calibration strategy.

1484 Finally, this last approach may be extended into relationships with **multiple-kind variables**  
1485 **with time** when appropriate.

1486 For each of these categories, subcategories can be defined based on the kind of mathematical  
1487 expression used for the calibration relationship. Popular examples are polynomials with constant  
1488 coefficients [7], gain-phase [13], variable offset [168], neural networks [36].

These categories are particularly important regarding *in situ* calibration strategies. Indeed, considering Equations 1.1 and 1.2 in Chapter 1 Section 3, we have then:

$$\begin{aligned} v_{corr}(s_i, t) &= \widehat{\mathcal{H}}^{-1}(\widehat{\mathcal{F}}^{-1}(\zeta(s_i, t), \widehat{\mathbf{q}}(t)), \widehat{\mathbf{q}}(t)) \\ &= \widehat{\mathcal{H}}^{-1}(\widehat{\mathcal{F}}^{-1}(\mathcal{F}(v_{true}(s_i, t), \mathbf{q}(t)), \widehat{\mathbf{q}}(t)), \widehat{\mathbf{q}}(t)) \end{aligned} \quad (2.1)$$

1489 For example, and recalling that *in fine* the purpose of *in situ* calibration is to determine the  
1490 relationship  $\widehat{\mathcal{H}}^{-1}$ , assume that the calibration relationship  $\widehat{\mathcal{F}}^{-1}$  of an instrument is defined so  
1491 that:

$$v(s_i, t) = G \cdot \zeta(s_i, t) + O + \kappa \cdot \theta(t)$$

1492 where  $G$ ,  $O$  and  $\kappa$  are constant and  $G$  is a correction of the gain of the instrument,  $O$  a  
1493 correction of the offset, and  $\kappa \cdot \theta(t)$  a temperature compensation standing for  $\widehat{\mathbf{q}}(t)$  in the general  
1494 expression of  $\widehat{\mathcal{F}}^{-1}$ .

If the behaviour of the measuring instrument changes over time,  $\widehat{\mathcal{F}}^{-1}$  cannot ensure that  $v(s_i, t) = v_{true}(s_i, t)$ . If the calibration relationship  $\widehat{\mathcal{F}}^{-1}$  cannot be adjusted, the relationship  $\widehat{\mathcal{H}}^{-1}$  obtained with an *in situ* calibration strategy may compensate the change of behaviour. Consider  $\widehat{\mathcal{H}}^{-1}$  such as:

$$v_{corr}(s_i, t) = G_{corr} \cdot v(s_i, t) + O_{corr} + \kappa_{corr} \cdot \theta(t)$$

Therefore:

$$v_{corr}(s_i, t) = G_{corr}G \cdot \zeta(s_i, t) + (G_{corr}O + O_{corr}) + (G_{corr}\kappa + \kappa_{corr}) \cdot \theta(t)$$

1495 Here,  $\widehat{\mathcal{H}}^{-1}$  allows determining a calibration relationship equivalent to the type of  $\widehat{\mathcal{F}}^{-1}$ .  
However, if  $\widehat{\mathcal{H}}^{-1}$  is such as:

$$v_{corr}(s_i, t) = G_{corr} \cdot v(s_i, t) + O_{corr}$$

Then:

$$v_{corr}(s_i, t) = G_{corr}G \cdot \zeta(s_i, t) + (G_{corr}O + O_{corr}) + G_{corr}\kappa \cdot \theta(t)$$

1496 In this case, it may not be possible to correctly compensate the drift of the instrument  $s_i$   
1497 because  $G_{corr}G$  and  $G_{corr}\kappa$  are coupled: it is not possible to correct them independently.

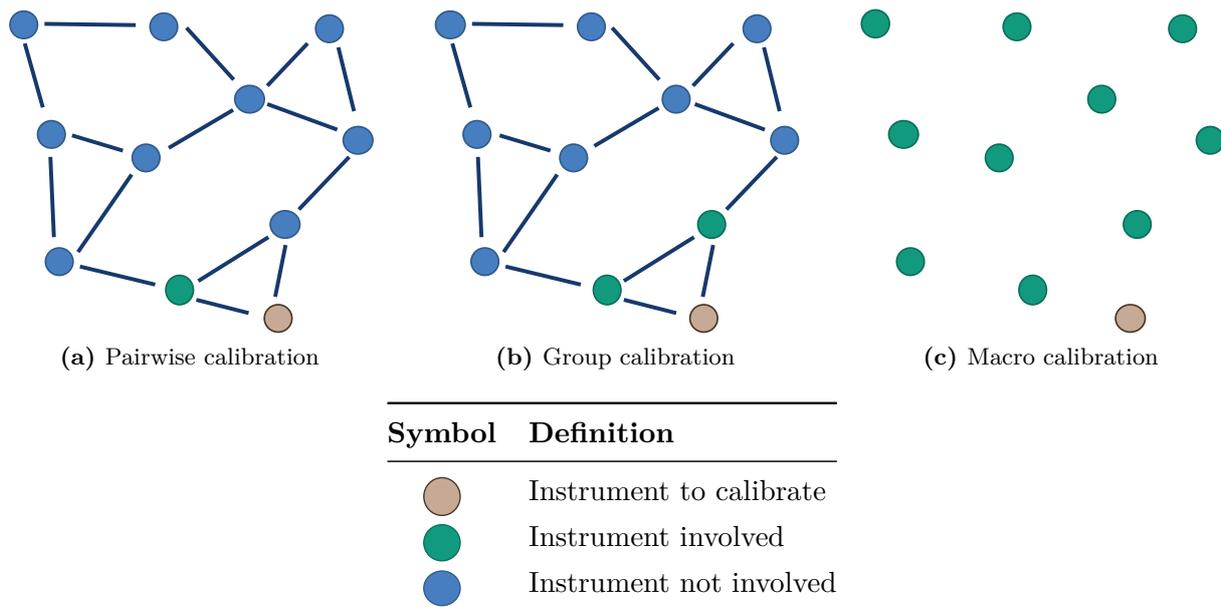
1498 This is why the type of relationship, both considering the quantities at stake and the kind of  
1499 relationship between them, is a key parameter of *in situ* calibration algorithms.

## 1500 2.4 Instrument grouping strategies

1501 While the previous categories are mostly driven by operational constraints (deployment  
1502 strategy, properties of the measurand and of the selected instruments), the present paragraph  
1503 considers the number of nodes involved in each calibration step and to which point the algorithm  
1504 can be distributed.

1505 A first approach is **pairwise calibration**. Two instruments are used, one providing standard  
1506 values for the other. It is classically applied between a reference instrument (or an approximation  
1507 of reference) and each of the nodes related to it. It can be a distributed or even localised  
1508 algorithm. This case is illustrated in Figure 2.2.2a.

1509 A **macro calibration** strategy consists in calibrating the network as a whole, e.g. the  
1510 values of all the instruments are used to calibrate each node of the network. Even if they exist,  
1511 node-to-node relationships may not be exploited directly. A centralised algorithm might be  
1512 necessary with this grouping strategy. This case is illustrated in Figure 2.2.2c.



**Figure 2.2.2:** Examples of cases for grouping strategies from the point of view of an instrument to calibrate. It shows which instruments are used to calibrate it. In these figures, the sensor networks are equivalent in terms of the number of instruments and relationships between the instruments. The relationships between the instruments in Figure 2.2c are not represented because they may not be exploited directly.

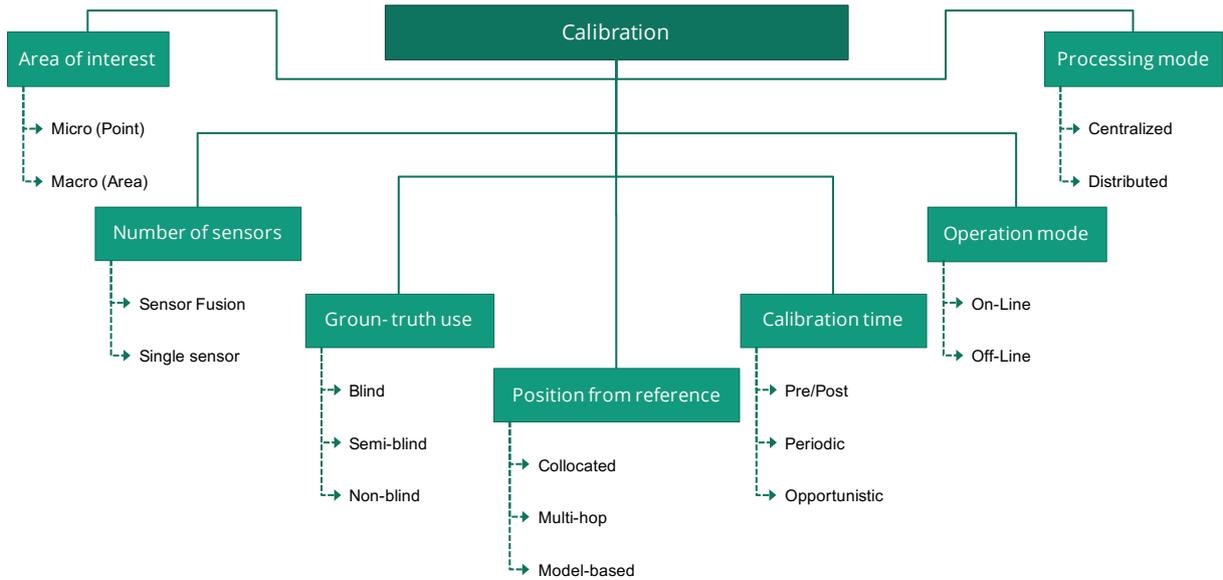
1513 **Group calibration** is an intermediate approach consisting in carrying calibration operation  
 1514 among groups of measuring instruments among the whole network. In this case, the criteria  
 1515 defining these groups become essential. This approach may be used when pairwise calibration  
 1516 induces significant errors, while macro calibration approaches are not fine adjusted enough.  
 1517 This category includes notably strategies where groups are composed of instruments measuring  
 1518 additional quantities besides the main target quantity. These additional quantities are often  
 1519 included as influence quantities in the calibration relationship. These algorithms can be at least  
 1520 partially distributed, e.g. the computation is concentrated on an elected group leader, or fully  
 1521 distributed at the cost of messages broadcasting. This case is illustrated in Figure 2.2.2b.

### 1522 3 Comparison to other taxonomies

1523 During the development of the presented taxonomy, two notable works reviewing calibration  
 1524 algorithms for sensor networks were conducted and published within a few months.

1525 The first publication is from Maag *et al.* [95] and focus on sensor networks for air pollution  
 1526 monitoring, addressing operational concerns regarding to calibration. Maag *et al.* build their  
 1527 classification of the algorithms around the terms "blind calibration", "collaborative calibration",  
 1528 "blind and collaborative calibration" and "transfer calibration". They add precision regarding  
 1529 the mobility of the instruments in the publications they cite. Without going into details of  
 1530 the definitions of the terms, it is a taxonomy with few attributes: blind group calibration and  
 1531 blind macro calibration strategies cannot be distinguished with this taxonomy for instance. This  
 1532 information on how the instruments are grouped to perform a calibration is important, notably  
 1533 if a distributed algorithm is targeted. This is why this feature was added in the presented  
 1534 taxonomy.

1535 The second publication is from Barcelo-Ordinas *et al.* [8]. The taxonomy they developed



**Figure 2.2.3:** Taxonomy for the classification of self-calibration algorithms according Barcelo-Ordinas *et al.* [8]

1536 is more detailed. All the categories of attributes they proposed are represented in Figure 2.2.3.  
 1537 Some of them are common with the presented taxonomy, like the use of reference instruments  
 1538 which is called the "ground-truth use". There are also differences. First, two groups of categories  
 1539 are close to what was defined here as the grouping strategy in the presented taxonomy: the "area  
 1540 of interest" and the "number of instruments". However, the associated definitions and the number  
 1541 of categories in these groups do not strictly match the ones given in Section 2. Also the idea of  
 1542 using reference instruments is refined in this taxonomy with a "position from reference" group  
 1543 of attributes, meaning the authors do not consider a relationship between the use of reference  
 1544 instruments and the category "position from reference". Barcelo-Ordinas *et al.* also provided  
 1545 other groups of attributes like the "calibration time", the "operation mode" or the "processing  
 1546 mode". The processing mode was stated as not in the scope of the presented taxonomy in  
 1547 Section 1. The "calibration time" (pre/post, periodic, opportunistic) and the "operation mode"  
 1548 (on-line, off-line) can have an influence on the data quality of the instruments. Indeed, two  
 1549 algorithms with the same attributes except one that performs a periodic calibration and the  
 1550 other an opportunistic one may not give the same results. However, we estimate that such  
 1551 features should be considered in a second phase while identifying relevant algorithms considering  
 1552 a given sensor network. For this reason, these features were not added to the presented taxonomy,  
 1553 contrary to Barcelo-Ordinas *et al.* Finally, contrary to the work of Maag *et al.* and the work  
 1554 presented here, Barcelo-Ordinas *et al.* did not consider the mobility of the instruments of the  
 1555 network as an attribute of *in situ* calibration algorithms.

1556 In both of these works, the mathematical expression of the calibration relationship is not  
 1557 directly an attribute of the algorithms. This topic is discussed but separately from others like  
 1558 the presence of reference in the network. In addition, calibration relationships are considered  
 1559 from the point of view of the mathematical expression only in both publications. The question  
 1560 of the calibration relationship is more developed in the presented taxonomy. The classification of  
 1561 calibration relationships is decomposed in two levels in our work: first the kinds of quantities  
 1562 involved in the relationships and then their mathematical expression. In this way, the proposed  
 1563 taxonomy provides a description capturing more information on the calibration relationships in  
 1564 *in situ* calibration algorithms.

1565 To conclude, the presented taxonomy is halfway between the one of Maag *et al.* which is  
1566 based on few groups of attributes, and the one of Barcelo-Ordinas *et al.* that is more extensive  
1567 but with information excluded from the scope of this work. Our taxonomy is limited to what  
1568 is seen as the essential attributes of an *in situ* calibration algorithm and designed with an  
1569 effort to avoid any redundancy or misunderstanding from the terms used, the taxonomy of  
1570 Barcelo-Ordinas *et al.* being sometimes confusing due to some groups of categories that are  
1571 semantically close. Nevertheless, all these works are interesting basis towards the definition and  
1572 use a shared terminology.

## 1573 4 Review of the literature based on this classification

1574 An application of the classification is provided here with highlights on the existing literature.  
1575 A large number of *in situ* calibration studies according to this classification are sorted in Table  
1576 2.4.1. Some rows refer to multiple papers as they are related somehow (same technique or same  
1577 authors) and consist in developments of the same initial paper. The current section focuses on a  
1578 description of the methods. The topic of performance comparison between methods is addressed  
1579 in the next section. The addressed measurands cover a wide range of environmental quantities:  
1580 temperature [7], pressure [182], noise [129], air pollutants [63], light [163]... Most of the reported  
1581 studies have generic approaches that can be transposed to other measurands.

### 1582 4.1 Overview

1583 Regarding pairwise strategies, relatively few papers address methodological issues related to  
1584 reference-based pairwise strategies, as this approach is the closest to a "traditional" calibration  
1585 approach with measurement standards and features fewer challenges. Partially blind and blind  
1586 pairwise calibration methods (often focusing on mobile nodes) are more complex as they require  
1587 to define calibration relationships not only between reference and non-reference nodes, but also  
1588 between non-reference nodes only. This translates into error propagation issues.

1589 Macro calibration approaches were initially developed to address the absence of reference  
1590 instruments in a network and thus are mostly blind or partially blind. In the absence of reference,  
1591 there is a strong challenge in defining valid calibration relationships based on values from  
1592 non-reference instruments, which explains the major consideration for these methods.

1593 Group strategies have been generating strong interests as they appear to outperform both  
1594 pairwise and macro strategies with or without reference instruments.

1595 Most methods are based on relationships with mono-kind variables without time and with  
1596 a linear expression, but more complex models are progressively appearing to better address  
1597 the complexity of environmental sensing. Moreover, a particular attention is given in most  
1598 publications to the calibration relationship used compared to the other attributes of *in situ*  
1599 calibration that we identified. This can be explained by the significance of this information for  
1600 readers, e.g. it allows determining if the algorithm is able to provide a correction corresponding  
1601 to the way an instrument drifts.

1602 Likewise, while most works initially focused on static networks, there are now many interests  
1603 for mobile nodes as they allow for physical rendez-vous between nodes. Henceforth, calibration  
1604 methods are less impacted by the physical variability of the phenomena.

1605 Finally, an underlying question addressed is the ability to distribute the computation of  
1606 calibration relationships [23, 55, 107, 129, 143, 144, 177]. The topic is of strong interest when  
1607 considering privacy preservation issues [102]. The capability to decentralise is linked to the  
1608 grouping strategy: pairwise and group strategies foster more naturally decentralised computation,  
1609 under the condition that the nodes are capable of individual procession and of bidirectional  
1610 communication. On the contrary, macro-calibration strategies tend to be centralised, except  
1611 when the characteristics of the parameter identification methods allow for partially or fully

**Table 2.4.1:** *In situ* calibration strategies for sensor networks  
R: Reference-based, PB: Partially blind, B: Blind, P: Pairwise, Gr: Group, Ma: Macro,  
1V: Mono-kind variables without time, 1VT: Mono-kind variables with time, MV: Multiple-kind  
variables without time, MVT: Multiple-kind variables with time, M: Exclusively mobile, S:  
Exclusively static, MS: Mobile and static

Papers	Availability of reference instruments			Instruments grouping strategies			Kinds of variables in calibration relationships				Mobility of the instruments		
	R	PB	B	P	Gr	Ma	1V	1VT	MV	MVT	M	S	MS
Ramanathan <i>et al.</i> [123]	✓	-	-	✓	-	-	✓	-	-	-	-	✓	-
Miluzzo <i>et al.</i> [107]	✓	-	-	-	✓	-	✓	-	-	-	✓	-	✓
Deshmukh <i>et al.</i> [39]	✓	-	-	✓	-	-	✓	-	-	-	-	✓	-
Spinelle <i>et al.</i> [139, 140]	✓	-	-	✓	✓	-	✓	-	✓	-	-	✓	-
Moltchanov <i>et al.</i> [109]	✓	-	-	✓	-	-	✓	-	-	-	-	✓	-
Gao <i>et al.</i> [58]	✓	-	-	✓	-	-	✓	-	✓	-	-	✓	-
Lin <i>et al.</i> [88]	✓	-	-	✓	-	-	✓	-	-	-	-	✓	-
Fang <i>et al.</i> [50, 51]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Martin <i>et al.</i> [103]	✓	✓	-	-	✓	-	-	-	✓	-	-	✓	-
Sun <i>et al.</i> [147]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Zimmerman <i>et al.</i> [186]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Weissert <i>et al.</i> [108, 171]	✓	-	-	-	✓	-	✓	-	-	-	-	✓	-
Yu <i>et al.</i> [184]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Topalovic <i>et al.</i> [155]	✓	-	-	✓	✓	-	✓	-	✓	-	-	-	-
Wei <i>et al.</i> [169, 170]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Mahajan <i>et al.</i> [96]	✓	-	-	✓	✓	-	✓	-	✓	-	-	✓	-
Malings <i>et al.</i> [100]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Qin <i>et al.</i> [120]	✓	-	-	-	✓	-	-	-	✓	-	-	-	✓
Barcelo-Ordinas <i>et al.</i> [9, 10, 53]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Wang <i>et al.</i> [165]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Loh <i>et al.</i> [91]	✓	-	-	✓	-	-	✓	-	✓	-	-	✓	-
Badura <i>et al.</i> [6]	✓	-	-	✓	✓	-	✓	-	✓	-	-	✓	-
Cordero <i>et al.</i> [34]	✓	-	-	-	✓	-	-	-	✓	-	-	✓	-
Tsujita <i>et al.</i> [156]	-	✓	✓	✓	-	-	✓	-	-	-	✓	-	✓
Tsujita <i>et al.</i> [157]	-	✓	✓	✓	✓	-	-	-	✓	-	-	✓	-
Y. Xiang <i>et al.</i> [177]	-	✓	✓	✓	-	-	✓	-	-	-	✓	-	✓
Hasenfratz <i>et al.</i> [63], Saukh <i>et al.</i> [131, 132]	-	✓	✓	✓	-	-	✓	-	-	-	✓	-	✓
Fu <i>et al.</i> [57]	-	✓	✓	✓	-	-	✓	-	-	-	✓	-	✓
Maag <i>et al.</i> [93, 94]	-	✓	✓	-	✓	-	-	-	✓	-	✓	-	✓
Arfire <i>et al.</i> [4]	-	✓	✓	-	✓	-	✓	✓	✓	✓	✓	-	✓
Markert <i>et al.</i> [102]	-	✓	✓	✓	✓	-	✓	-	-	-	✓	-	✓
Kizel <i>et al.</i> [75]	-	✓	-	✓	-	-	✓	-	-	-	-	✓	-
Sailhan <i>et al.</i> [129]	-	-	✓	✓	✓	-	✓	-	-	-	✓	-	-
Fonollosa <i>et al.</i> [56]	-	✓	✓	✓	-	-	✓	-	-	-	-	✓	-
Whitehouse <i>et al.</i> [172]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Ihler <i>et al.</i> [65]	-	✓	✓	-	✓	-	✓	-	-	-	-	✓	-
Taylor <i>et al.</i> [152]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-

continued on next page

*continued from previous page*

Papers	Availability of reference instruments			Instruments grouping strategies			Kinds of variables in calibration relationships				Mobility of the instruments		
	R	PB	B	P	Gr	Ma	1V	1VT	MV	MVT	M	S	MS
Tan <i>et al.</i> [151]	-	-	✓	-	✓	-	✓	-	-	-	-	✓	-
Bychkovskiy <i>et al.</i> [24]	-	✓	✓	✓	-	-	✓	-	-	-	-	✓	-
Balzano <i>et al.</i> [7], Lipor <i>et al.</i> [89], Dorffer <i>et al.</i> [44]	-	✓	✓	-	-	✓	✓	-	-	-	-	✓	-
Takruri <i>et al.</i> [148–150]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Kumar <i>et al.</i> [76, 77]	-	-	✓	-	✓	-	✓	-	-	-	-	✓	-
Ramakrishnan <i>et al.</i> [122]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Buadhachain <i>et al.</i> [23]	-	-	✓	-	✓	✓	✓	-	-	-	✓	-	✓
Bilen <i>et al.</i> [13]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Cambareri <i>et al.</i> [25]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
C. Wang <i>et al.</i> [161–163]	-	-	✓	-	-	✓	✓	-	-	-	✓	-	-
C. Xiang <i>et al.</i> [176]	-	-	✓	-	-	✓	✓	-	-	-	✓	-	-
Dorffer <i>et al.</i> [41–43, 45]	-	✓	✓	-	-	✓	✓	-	-	-	✓	-	-
Y. Wang <i>et al.</i> [166–168] Li [87]	-	-	✓	-	-	✓	-	✓	-	-	-	✓	-
Ye <i>et al.</i> [182]	-	-	✓	✓	-	-	✓	-	-	-	✓	-	-
De Vito <i>et al.</i> [36, 47]	-	-	✓	-	✓	-	-	-	✓	-	-	✓	-
Son <i>et al.</i> [137], Lee [82]	-	-	✓	-	-	✓	✓	-	-	-	✓	-	-
Stankovic <i>et al.</i> [143, 144]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Fishbain <i>et al.</i> [55]	-	✓	✓	-	✓	-	✓	-	-	-	-	✓	-
Popoola <i>et al.</i> [118]	-	-	✓	✓	-	-	-	-	✓	-	-	✓	-
Yan <i>et al.</i> [180]	-	-	✓	-	✓	-	-	-	✓	-	-	✓	-
Yang <i>et al.</i> [181]	-	-	✓	-	-	✓	✓	-	-	-	-	✓	-
Mueller <i>et al.</i> [113]	-	-	✓	-	✓	-	-	-	✓	-	-	✓	-
Kim <i>et al.</i> [74]	-	-	✓	-	✓	-	-	-	✓	-	-	✓	-
Chen <i>et al.</i> [30]	-	-	✓	-	-	✓	-	-	✓	-	-	✓	-
Wu <i>et al.</i> [175]	-	-	✓	-	✓	-	✓	-	-	-	-	✓	-
Becnel <i>et al.</i> [11]	-	✓	-	✓	✓	-	✓	-	-	-	-	✓	-
Cheng <i>et al.</i> [31]	-	✓	-	✓	-	-	✓	-	-	-	-	✓	-
Sun <i>et al.</i> [146]	-	✓	-	-	✓	-	-	-	✓	-	✓	-	-

1612 decentralised computation. However, while distributed computing impacts the computational  
1613 performance of algorithms, there is no report on how it affects calibration performances so far.

## 1614 4.2 Mobile and static nodes

1615 Static networks are more frequently studied than mobile ones. A wide range of solutions  
1616 is now available to calibrate them. However, these calibration methods usually require a high  
1617 spatial density of nodes to overcome the spatial variability of the phenomena, which is not always  
1618 viable technically or economically. The availability of mobile nodes could alleviate this constraint,  
1619 as calibration operations exploit physical rendez-vous between nodes. In turn, the methods based  
1620 on this principle are challenged when the rendez-vous frequency is too low compared to the speed  
1621 of degradation of the measurement accuracy [132]. In such cases, the addition of a few reference  
1622 nodes seems to yield satisfying results [45, 63, 146]. Moreover, a challenge of mobile instruments  
1623 is that they face rapid transients. To address this, methods initially developed for static networks  
1624 appear promising, such as the work of De Vito *et al.* [36, 47] which uses dynamic and nonlinear

1625 supervised machine learning tools.

### 1626 **4.3 Calibration relationships**

1627 Most reported relationships are of mono-kind variables without time type and based on  
1628 linear expressions. Nevertheless, there is a rising interest for models with multiple-kind variables,  
1629 which stems from the observation that there are indeed significant influence quantities for various  
1630 environmental measurands, notably air pollutant concentrations or temperature and humidity.  
1631 It often depends on the technology of the instruments used [50, 68, 71, 136, 169, 170]. Such  
1632 relationships gave very interesting results compared to simpler relationship models:

- 1633 • for reference-based group calibration in [6, 9, 10, 34, 53, 120, 139, 140, 147, 155]
- 1634 • for partially blind group calibration strategies in [4, 50, 93, 103], including with time-  
1635 sensitive models in [4, 23]
- 1636 • for blind strategies, either pairwise or group based, in [47, 113, 118, 180].

1637 On the contrary, relationships with multiple-kind variables were shown to be unnecessary in  
1638 [22, 64] where the control of the operating temperature of the device was sufficient to perform a  
1639 pairwise calibration without being influenced by this quantity. Malings *et al.* [100] reached a  
1640 similar conclusion while also indicating that generalised models, e.g. calibration relationships  
1641 built on all the values of the different measurands of a measuring system for all the instruments of  
1642 the device, have several advantages: they reduce the effort required to calibrate the instruments  
1643 because only one generalised relationship is derived instead of individual ones.

1644 In general, time-dependent approaches are used to address drift issues. Drift is often modelled  
1645 as an additive random variable with a given probability distribution [167, 177], so that drift-  
1646 compensation translates as an offset correction.

### 1647 **4.4 Pairwise strategies**

#### 1648 *Reference-based pairwise*

1649 Relatively few papers address methodological issues related to reference-based pairwise  
1650 strategies, as this approach is the closest to a "traditional" calibration approach with measurement  
1651 standards. Primarily, reference instruments may be directly co-located in the field with non-  
1652 reference instruments to achieve their calibration [6, 50, 88, 96, 123, 139, 140, 155].

1653 However, more automated strategies are expected, requiring less the co-location of instruments.  
1654 Nevertheless, even in the simple case of a relatively dense sensor network, the measurand may  
1655 spatially vary too much in general to relate a reference instrument at a given location to an  
1656 instrument at another location for calibration purposes. As an elementary solution to this,  
1657 Moltchanov *et al.* [109] proposed to carry out calibration against the reference node using only  
1658 the data collected during a specific time span based on the postulate that the phenomenon varies  
1659 less during this time span. This was an idea previously developed by Tsujita *et al.* [157] including  
1660 weather conditions used to correct the measured values but not with a reference-based approach.  
1661 In the context of the Internet of Things, Loh *et al.* proposed a web-query based framework using  
1662 machine learning to calibrate portable particulate matter measuring instruments. Their reference  
1663 values are obtained from governmental monitoring stations.

#### 1664 *Partially blind pairwise*

1665 Partially blind pairwise calibration focuses mostly on mobile nodes. Tsujita *et al.* [156]  
1666 tackled it first for mobile nodes by proposing that the device to calibrate should display either the  
1667 value of a reference node that is close enough, or the average measurements between co-located

1668 nodes if no reference node is available. A calibration parameter is adjusted with these values to  
1669 correct measurements between rendez-vous.

1670 Y. Xiang *et al.* [177] later proposed another method. They also distinguish calibration  
1671 based on the values of a reference instrument and on the values of a non-reference instrument.  
1672 Their originality relies in the correction of the values that is performed with an estimator of the  
1673 drift error of the node. This error is recalculated at each calibration by minimising its variance  
1674 according to a linear combination of the values of the instruments involved in the calibration  
1675 process.

1676 Hasenfrazz *et al.* [63] addressed by various methods the case of calibration for mobile devices  
1677 against reference instruments or not. They notably provided dedicated extensions for the case  
1678 where some devices rarely encounter reference instruments. They also demonstrated a linear  
1679 dependency between the measurement error and the number of intermediary calibrations between  
1680 a given node and the reference node it is calibrated against. In [132] and [131], Saukh *et al.*  
1681 proposed solutions to this issue of error accumulation by working on the occurrence of rendez-vous  
1682 between nodes, in view of maximising the opportunities of calibration. An alternative idea  
1683 was developed by Fu *et al.* [57] who proposed the optimisation of the deployment of reference  
1684 instruments to ensure that all nodes can be calibrated against one of the references with a path  
1685 no longer than  $k$  hops. Then Maag *et al.* [93, 94] and Arfire *et al.* [4] extended this work to  
1686 models with multiple-kind variables, with and without time dependency. They showed that the  
1687 complexity of the model should be adjusted based on the frequency of rendez-vous.

1688 In a similar way, Markert *et al.* [102] introduced a calibration strategy based on rendez-vous  
1689 but with a particular focus on privacy aspects for the exchange of data.

1690 Kizel *et al.* [75] also proposed a multi-hop calibration method consisting into co-locating two  
1691 devices for a certain time, one being the reference to the other, and then moving the freshly  
1692 calibrated device close to another non-calibrated, a reference instrument being introduced in the  
1693 loop to reset the error that accumulates. The advantage is the error is related to the number of  
1694 hops that took place like in [63].

1695 Sailhan *et al.* [129] developed a multi-hop, multiparty calibration scheme with the addition  
1696 of an assessment protocol for the relevancy of the calibration, based on a weighted directed  
1697 hypergraph of the network, the weights indicating the quality of the calibration. The presented  
1698 strategy was applied to blind networks but as in [63], it could be extended to partially blind  
1699 networks.

1700 In the case where the instruments cannot be mobile, Weissert *et al.* [108, 171] proposed to  
1701 use a proxy model in their framework for the calibration of sensor networks. It consists into  
1702 using the mean and standard deviation of the data obtained at a given location that is equivalent  
1703 in terms of land use compared to the location of the instrument to calibrate.

1704 Fonollosa *et al.* [56] used various models for calibration of chemical measuring instruments  
1705 with an approach called "calibration transfer" which is a kind of multi-hop calibration. Indeed,  
1706 the principle is to calibrate one of the instruments and then apply the same model on other  
1707 instruments to calibrate, eventually with a transformation to properly map the measurement  
1708 spaces. This approach was also used by Laref *et al.* [80]. Such strategies are widely used in  
1709 the field of spectroscopy [173]. It, however, requires the measuring instruments are more or  
1710 less behaving the same under identical varying conditions, which is not always the case when  
1711 dealing with low-cost instruments. Moreover, in the context of environmental sensing with  
1712 static instruments over a large area, it is very unlikely that the instruments measure the same  
1713 quantities at the same time. Thus, deriving the correct transformation to transfer the calibration  
1714 relationship is challenging. Cheng *et al.* [31] proposed an in-field calibration transfer method to  
1715 learn this transformation when the distributions of the reference values at the source and target  
1716 locations are similar, and when this transformation can be assumed as linear.

---

## 4.5 Blind macro calibration

Blind calibration strategies for sensor networks were first developed as a way to locate spatially the nodes in a static network. Whitehouse *et al.* [172] proposed to solve an optimisation problem ensuring that consistency and geometrical constraints were respected. Ihler *et al.* [65] proposed nonparametric belief propagation instead. Taylor *et al.* [152] developed an inference technique using a Bayesian filter. However, as they targeted spatial localisation, most of these methods based their algorithms on electromagnetic or acoustic [151] propagation (time delay, intensity loss...), and cannot be applied directly to other calibration problems.

Bychkovskiy *et al.* [24] proposed a first solution that could be used for any measurand, provided that the existence of relations between instruments of the network is known. It demands first to estimate the parameters of each existing relationship between instruments in the network. Then, the consistency of the derived relationships must be maximised to be resilient to circular dependencies. This technique was applied to a dense static sensor network and has not been yet extended to a mobile sensor network. In theory, it could also be applied to a sensor network with reference nodes but there is no report on the topic.

Later, Balzano *et al.* [7] developed a blind and partially blind calibration strategy for static sensor network suitable for any measurands, without any prior knowledge on existing relationships between the values of instruments. They tested it notably on temperature, light intensity or CO<sub>2</sub> level measurements. The key postulate is the sensor network is dense enough to oversample the signal of interest. They proposed that the true signal lies in a subspace of the space formed by the measured values. Considering the prior choice of the subspace, the parameters of the calibration relationships for all nodes are then estimated using singular value decomposition or by solving a system of equations using a least square estimator. This method was extended later in [89] to provide a total least square formulation and also in [44] to take into account outliers and separate them from the measurement matrix.

Alternately, Takruri *et al.* [148–150] addressed calibration as a drift compensation problem. They proposed to proceed recursively: measured values at step  $n$  are first corrected with predicted drifts obtained at step  $n-1$ , then the next measurements are predicted using support vector regression (SVR). Finally, the predicted values are used to estimate the drifts using a Kalman filter. Kumar *et al.* [76, 77] replaced SVR by kriging, which is a method of interpolation originally from geostatistics, as a prediction method for next values. In the same vein of using spatial relationship between the measured values, Chen *et al.* [31] introduced a spatial correlation model for drift correction targeting sensor networks deployed in buildings.

Ramakrishnan *et al.* [122] proposed another blind calibration strategy based on a gossip-based distributed consensus strategy, with SAGE algorithm used for parameter estimation. Later Buadhachain *et al.* [23] used an expectation-maximisation algorithm instead. The consensus-based approach is interesting as it can reduce the communication bandwidth.

Bilen *et al.* [13] extended the problem of blind calibration to the case of sparse input signals, which are measurements with missing/useless information, and exploited compressive sensing to estimate the instruments' corrective gains.

Cambareri *et al.* [25] proposed a non-convex formulation of the problem and gave a formal criterion of convergence of the calibration method enabling to estimate corrective gains for the values.

For mobile nodes, C. Wang *et al.* [162] proposed a method which exploits the moments of the measurements, here the average and the variance. They formulate calibration as an optimisation problem minimising the difference between the moments of the true signal and the measured one. The method was extended in [161] and [163] for different expressions of the calibration relationship. The approach was later adjusted by C. Xiang *et al.* [176] to address specifically mobile crowdsensing, with instruments embedded in mobile phones for instance. Dorffer *et al.*

1766 [41, 42, 45] considered mobile sensor network calibration with mono-kind variables and linear  
1767 relationships for spatially sparse signals by using matrix factoring, with an extension regarding  
1768 non-linear calibration [43].

1769 Wang *et al.* [168] proposed an extension of [7] relaxing some hypothesis and based their  
1770 estimation on a Kalman filter with the help of a drift detection strategy. The approach was  
1771 improved in [87] in terms of drift detection, number of instruments allowed to drift at the same  
1772 time and pre-processing of the input signal with wavelet denoising. Wu *et al.* [175] applied the  
1773 same idea recently but in the case of clustered sensor networks, e.g. when the instruments of  
1774 the networks can be grouped into several clusters. In this case, it is more a group calibration  
1775 strategy than a macro calibration.

1776 Wang *et al.* then extended their work for sparse signals with either a Bayesian approach  
1777 [167] or a deep learning approach [166].

1778 Yang *et al.* [181] also based their work on the idea of [7]. They prove that, if the underlying  
1779 signals follow a first-order auto-regressive process, then the parameters of the linear calibration  
1780 model are recoverable. They use a nonparametric Bayesian model to do so.

1781 Overall, macro calibration methods do not suffer from error propagation issues unlike pairwise  
1782 approaches. However, because of the absence of references, they usually require large amount of  
1783 data, which is typically available in the case of highly dense static networks or mobile networks  
1784 with high frequency of rendez-vous.

#### 1785 4.6 Group strategies

1786 On the first hand, group calibration strategies are used to calibrate multiple instruments  
1787 located at the same place and measuring different quantities that could influence the calibration  
1788 relationship. It was shown that their corrected values were more accurate when exploiting  
1789 multiple-kind variables in calibration relationships [9, 10, 47, 53, 100, 120, 139, 140]. Therefore,  
1790 strategies [63] that were developed for pairwise and mono-kind variables calibration in the first  
1791 place were extended to group calibration [4, 93, 94]. New methods were also developed.

1792 Based on results like [71, 139, 140], Zimmerman *et al.* [186] introduced a calibration strategy  
1793 using random forests with multiple-kind variable relationships for measuring systems when the  
1794 latter are co-located to reference instruments.

1795 Kim *et al.* [74] recently presented an approach of blind group calibration with prior information  
1796 on the cross-sensitivities of instruments, known from laboratory experiments, to build calibration  
1797 relationships with multiple-kind variables.

1798 On the other hand, other works proposed group calibration strategies to reinforce the  
1799 confidence in the values used as standards.

1800 Miluzzo *et al.* [107] proposed a calibration technique with multiple reference nodes in a sensor  
1801 network by formulating a distributed average consensus problem estimating the offset of each  
1802 non-reference node. The concept of sensing factor was introduced. It refers to the area within  
1803 which the measurand value can be assumed to be identical for the reference instruments and the  
1804 instruments to calibrate.

1805 Lee *et al.* [82, 137] proposed a blind group approach for mobile nodes. The area of deployment  
1806 is divided into several non-overlapping regions. It is assumed that calibration relationships exist  
1807 in each of these areas. This is used to formulate the parameter estimation problem as a Laplacian  
1808 linear equation relating the drift, the measured values and noise.

1809 Stankovic *et al.* [144] developed a novel methodology based on exploiting the neighbours  
1810 of each node. Relations in the network are expressed in a matrix and, as there are groups of  
1811 neighbours, the matrix can be decomposed into blocks. Each block represents a group of relations  
1812 with parameters to estimate, reducing the problem compared to a macro blind calibration. This  
1813 approach was extended in [143] to better deal with the case of additive measurement noise.

1814 Fishbain *et al.* [55] proposed a method of aggregation for non-calibrated sensor network  
 1815 relying on a group-consensus strategy.

1816 Fang *et al.* [50] later introduced reference instruments that are measuring different quantities,  
 1817 in view of analysing the influence quantities in the calibration relationship. More recently [51],  
 1818 they also added an outlier detection and removal strategy before the calibration stage to improve  
 1819 the quality of the input data.

1820 Yu *et al.* [184] also proposed to derive features from the time series of the measurands to  
 1821 reduce the effect of cross-interference phenomenon between two quantities and then to use a deep  
 1822 calibration model to correct the measuring instruments. The same idea of feature extraction was  
 1823 also developed by Wang *et al.* [165] in a work extending the idea of piecewise linear calibration of  
 1824 [163] to a group calibration with different types of calibration relationship computed piecewise.

1825 Becnel *et al.* [11] used the concepts developed in [63, 131, 132] notably to propose a group  
 1826 calibration strategy with a weighted neighbourhood approach. On a same dataset, their approach  
 1827 seems to improve the results by 20 % compared to the previous works they considered.

#### 1828 4.7 Comment regarding other surveys

1829 By comparing the common references in Table 2.4.1 to the references listed in Table 5 of  
 1830 the survey of Barcelo-Ordinas *et al.* [8], we can observe that some works are not classified in  
 1831 the same way like for instance the work of Balzano *et al.* [7] that is classified as "micro" for the  
 1832 attribute "area of interest" and "single sensor" for the attribute "number of sensors" whereas it was  
 1833 classified as a macro calibration technique here. This was expected as differences were pointed  
 1834 out between the two taxonomies in Section 2. As the groups "area of interest" and "number of  
 1835 sensors" of Barcelo-Ordinas *et al.* are similar to the grouping strategy in the presented taxonomy,  
 1836 it appears we did not understand the algorithms in the same way. This underlines the use of a  
 1837 shared terminology is critical to ensure a good comprehension of the algorithms.

## 1838 5 Conclusion

1839 In this chapter, a taxonomy for the classification of *in situ* calibration algorithms for sensor  
 1840 networks has been proposed. It is based on four groups of categories capturing both the different  
 1841 network architectures and algorithmic principles: the availability of reference instruments in  
 1842 the network, the mobility of the instruments, the kind of input variables in the calibration  
 1843 relationships, and the instruments grouping strategy (pairwise, macro or by group) used for a  
 1844 calibration procedure.

1845 The review shows that relatively few papers address methodological issues related to reference-  
 1846 based pairwise strategies, as this approach is the closest to a "traditional" calibration approach  
 1847 and features relatively little challenges. Partially blind and blind pairwise calibration methods,  
 1848 which often focus on mobile nodes, are more complex as they require to define calibration  
 1849 relationships between non-reference nodes, which translates into error propagation issues. Macro  
 1850 calibration approaches are mostly used to deal with reference-less situation. Their challenge lies  
 1851 in defining valid calibration relationships between non-reference instruments. Group strategies  
 1852 appear to improve on performance of both pairwise and macro strategies with or without reference  
 1853 instruments. Most methods are based on calibration relationships with mono-kind variables  
 1854 and a linear expression, but more complex models are progressively appearing to better address  
 1855 the complexity of environmental sensing. Likewise, while most work initially focused on static  
 1856 networks, there is now a strong interest for mobile nodes as they allow for physical rendez-vous  
 1857 between nodes, which reduces the impact of the physical variability of the phenomena between  
 1858 static distant nodes.

1859 In a general manner, the main situations and issues for the *in situ* calibration of environmental  
 1860 sensor networks have been addressed multiple times in the literature. To identify the improvements

1861 that could be brought, the question of the performances of the existing algorithms must be  
1862 tackled, which is the subject of the next chapter.

---

1863 The content of this chapter is based on the following publication:

1864 F. Delaine, B. Lebental and H. Rivano, "In Situ Calibration Algorithms for Environ-  
1865 mental Sensor Networks: A Review," in *IEEE Sensors Journal*, vol. 19, no. 15, pp.  
1866 5968-5978, 1 Aug.1, 2019, DOI: 10.1109/JSEN.2019.2910317.

1867 The review presented in this thesis was extended to the papers published after this survey.

---

Chapter  
**3**

**Framework for the Simulation of  
Sensor Networks Aimed at Evaluating  
In Situ Calibration Algorithms**

---

**Contents**

---

1874	1	Challenges for the comparison of <i>in situ</i> calibration algorithms . . . . .	41
1875			
1876	2	Description of the framework . . . . .	43
1877	2.1	Simulation-based strategy . . . . .	43
1878	2.2	Functional decomposition . . . . .	44
1879	2.2.1	Build ideal measured values . . . . .	45
1880	2.2.2	Build measured values from the ideal measured values . . . . .	46
1881	2.2.3	Perform calibration and build corrected values . . . . .	46
1882	2.2.4	Evaluate . . . . .	46
1883	3	Comparison of <i>in situ</i> calibration strategies for blind static sensor networks . .	46
1884	3.1	Frame of the study . . . . .	46
1885	3.2	Application of the framework . . . . .	47
1886	3.2.1	Simulation of the ideal measured values . . . . .	47
1887	3.2.2	Building of the measured values . . . . .	50
1888	3.2.3	Configuration of the strategies . . . . .	50
1889	3.2.4	Evaluation . . . . .	51
1890	3.3	Results . . . . .	51
1891	3.4	Conclusions . . . . .	53
1892	4	Evaluation of measurements after correction . . . . .	55
1893	4.1	Problem statement . . . . .	55
1894	4.2	Evaluation with an error model . . . . .	56
1895	4.3	Means of visualisation . . . . .	60
1896	4.4	Conclusion . . . . .	61
1897	5	Sensitivity of the calibration algorithms to the specificities of the case study .	62
1898	5.1	Using a more realistic model of the true values . . . . .	62
1899	5.2	Density of the sensor network . . . . .	67
1900	5.3	Instrument modelling . . . . .	71
1901	5.3.1	Drift model . . . . .	71

1902	5.3.2	Other faults . . . . .	75
1903	5.4	Parameters of calibration strategies . . . . .	77
1904	5.5	Summary of the results . . . . .	79
1905	6	Discussion and conclusion . . . . .	<b>80</b>

---

## 1 Challenges for the comparison of *in situ* calibration algorithms

In Chapter 2, after proposing a taxonomy for the classification of *in situ* calibration algorithms, a review of the existing works was conducted. The literature on the subject is abundant and numerous contributions were reported, covering all the possible types of sensor networks. Nevertheless, no comparison of the *in situ* calibration strategies in terms of performances was proposed in this survey. Indeed, though the presented algorithms in publications are operational from a mathematical perspective, a formal quantification of the improvement is never provided. Instead, one or several case studies are systematically provided to demonstrate how the algorithm is put in practice and its efficiency on these specific cases. This step is either performed through simulations [45], laboratory experiments or field experiments [109], or both [7].

The principle of such studies consists in comparing the measured values, both before and after calibration, to the values that should have been measured by the instruments if they were ideal. The latter values are obtained with the help of simulations or with higher quality instruments co-located to the instruments to recalibrate.

Regarding approaches by simulation, different ones are proposed to compute numerically the values of the measurand. It can be based on 2D Gaussian fields [41] or ARMA processes [166] for instance. Disturbances like noise or drifts are added to the reference values in order to generate the measured values. However, the models can be more or less realistic. Thus, the model used may raise the question of the relevance of the results obtained, notably with respect to experiments.

Concerning experimental approaches, there are lots of datasets produced and used in various studies (for instance the Intel Lab Data [79], a deployment at James Reserve used in [7], the OpenSense Zurich Dataset [86]). They may be representative of particular types of deployments, for instance indoor (Intel Lab Data) or outdoor (James Reserve, OpenSense) measurements, with static and dense sensor networks (Intel Lab Data, James Reserve) or mobile sensor networks (OpenSense). Moreover, whereas repeating a simulation with different network settings (number of instruments, types of instruments, positions...) but identical phenomenon is feasible, it is almost impossible with experiments except under laboratory conditions. Although there are existing facilities that could allow it [38], they are not adapted for the study of sensor networks over large areas. A solution could be the deployment of as many instruments as necessary to produce the desired configurations, but this would drastically increase the cost of the experiment. Moreover, this would have to be reproduced in multiple environments in order to study its influence on the performance of *in situ* calibration strategies.

More generally, whether we focus on simulation-based or experimental approaches, there is no single case study that is reused a significant number of times in multiple publications by different groups of authors according to the review conducted in Chapter 2. It prevents the comparison of existing works. Such a study is conducted in [35] but as for the work of Karagulian *et al.* [73] reported in Chapter 1 Section 1.3.3, the conclusions must be considered with precautions. In this case, the performances after calibration of measuring systems using instruments from the same manufacturer are compared based on the results published by different groups of researchers. However, the results were not obtained on the same case study. In this way, conclusions may be incorrect.

As an explanation to the absence of one or several main case studies in the community, they are not always easily reproducible. For simulation-based ones, some of their parameters may have not been provided or even mentioned. This is why an identical approach may be used in multiple papers but rarely from different authors. Regarding experiments, datasets are often shared with the community, though they sometimes become unavailable over time or require post-processing that is left to the user. Nevertheless, as for simulation, a dataset is rarely used in publications from authors who did not produce it. This may be due again to the fact that

1958 experiments are tailored for specific cases and calibration strategies.

1959 Another difficulty is the multiplicity of metrics that can be used to assess the performances  
 1960 of the strategies [17]. Such metrics are for instance the mean absolute error, the mean absolute  
 1961 percentage error, the root-mean-square error, the Pearson’s correlation coefficient and so on.  
 1962 Original metrics developed in works on the evaluation of the performance of instruments [54]  
 1963 or ones related to official recommendations such as the data quality objective (DQO) in the  
 1964 context of air quality measuring instruments [16, 27, 95] are also used. They are mostly based on  
 1965 prior information concerning the expected corrected values. As shown in the comparison of the  
 1966 existing works carried out by Cross *et al.*[35], authors do not all use the same metrics in their  
 1967 case studies.

1968 Also, in the perspective of comparative studies, while a few authors have shared the code  
 1969 associated to their publications as in [89] or [44], most of the codes are not open-sourced.  
 1970 Therefore, algorithms have to be fully reimplemented in most cases to achieve comparison [168]  
 1971 and due to the efforts it requires, it is carried out in comparative studies only for strategies with  
 1972 features similar to the one they propose—for example the use of machine learning techniques [26,  
 1973 48, 49] or the use of Kalman filter [168].

1974 Overall, while a new strategy can be overperforming previous ones on a given case study,  
 1975 one rarely knows whether the new strategy is better in general or only for the specific case (Fig.  
 1976 3.1.1), regardless of whether it is based on simulation or on experiment.

<b>Publication</b>	 $P_a$	 $P_b$	 $P_c$
<b>Algorithm</b>	 $A_a$	 $A_b$	 $A_c$
<b>Case study</b>	 $C_a$	 $C_b$	 $C_c$
<b>Conclusion</b>	$A_a$ works on $C_a$	$A_b > A_a$ on $C_b$	$A_c > A_b$ on $C_c$
<b>Not proved</b>	-	$A_b > A_a$ on $C_a$	$A_c > A_b$ on $C_{a,b}$ $A_c > A_a$ on $C_{a,b,c}$

**Figure 3.1.1:** Usual evaluation process performed across papers: consider three publications,  $P_a$ ,  $P_b$  and  $P_c$  that are published successively.  $P_a$  introduces an algorithm  $A_a$  which is evaluated on case study  $C_a$ . It concludes that the algorithm is working. Later,  $P_b$  presents  $A_b$  and evaluate it with  $A_a$  on case study  $C_b$ , concluding that  $A_b$  has better performances than  $A_a$ . Finally,  $P_c$  brings forward  $A_c$  and show it is better than  $A_b$  on case study  $C_c$ . In the absence of unified validation case study,  $A_b$  may still be better than  $A_c$  on  $C_{a,b}$ , and  $A_a$  better than  $A_b$  and  $A_c$  on  $C_a$  and  $C_{a,b,c}$  respectively.

1977 Thus, there is a strong need for systematic tools and protocols [84, 95] enabling to compare  
 1978 across studies the performances of *in situ* calibration methodologies.

1979 This chapter is addressing this question. The contribution consists in a generic framework  
 1980 to design case studies that can be used for quantitative comparisons of *in situ* calibration  
 1981 strategies. Its objective is to get a better understanding of the factors driving the performances  
 1982 of calibration strategies and the quality of sensor networks. For reproducibility and scaling issues,  
 1983 it is based on numerical simulations of the environmental phenomenon and of the measuring  
 1984 instruments, their metrological properties as well as the faults introduced in the measurements.  
 1985 It is henceforth not a new calibration methodology, but the first methodology to carry out a  
 1986 systematic inter-comparison of the performances of existing methodologies.

1987 We apply this framework on seven calibration algorithms. In a first step, we consider an

1988 elementary case study on a blind static network that is sufficient to get first insights by analysing  
1989 the performances with commonly used metrics at different scales. Then, we evaluate the impact  
1990 of the realism of the model of pollution emission and dispersion, the density of the deployment  
1991 of the measuring instruments, and the error model of the devices. Finally, several engineering  
1992 insights are deduced from the investigations carried out, validating the relevance of such a  
1993 framework in the design phase of a practical network of low-cost environmental sensors.

## 1994 **2 Description of the framework**

1995 A generic framework to design case studies is proposed in this section for quantitative  
1996 comparisons of *in situ* calibration strategies. It aims at yielding a better understanding of the  
1997 factors driving the performances of calibration strategies, and at providing a protocol as rigorous  
1998 as possible to conduct such studies. The different steps are described in a generic way to ensure  
1999 their applicability for most cases.

2000 Like for faults in Chapter 1 Section 2, we adopt a data-centric point of view. We do not target  
2001 the modelling of the entire measuring chain of an instrument described in Chapter 1 Section 1.2  
2002 and illustrated in Figure 1.1.2 in the same chapter. We assume that instruments are grey boxes  
2003 which provide measured values based on true values of a measurand and possibly also on true  
2004 values of influence quantities. The grey boxes representing instruments consist into algorithms  
2005 mimicking the features of real instruments.

2006 In this work, we focus on the metrological performances of *in situ* calibration strategies.  
2007 Other subjects such as the communication costs, the energy consumption or the computational  
2008 efforts for instance are not in the frame of this study.

### 2009 **2.1 Simulation-based strategy**

2010 The methodology is based on simulation because it enables the following properties that are  
2011 difficult to get with field or lab experiments

- 2012 • Ability to perform a study with different operating conditions for a same sensor network.
- 2013 • Reproducibility: same operating conditions on different sensor networks.
- 2014 • Knowledge of true values: in experimental operating conditions, true values are not  
2015 known perfectly—there is always an uncertainty— and having very accurate values requires  
2016 high-quality instruments which are usually expensive.
- 2017 • Scalability: the only limit to the density, number and diversity of measuring instruments is  
2018 the computing time.

2019 To conduct the evaluation of *in situ* calibration strategies, the framework combines the  
2020 simulations of the environmental phenomena to produce the true values of the measurand, and  
2021 of the potential influence quantities, at any position and time, the simulations of the mobility –if  
2022 any– of the nodes of the network to know where and when measurements are performed, and  
2023 the simulations of the measuring chain of the instruments to produce the measured values. The  
2024 realism of the results will depend on the complexity and accuracy of each simulation model. On  
2025 the other hand, being able to analyse results on a simplistic model can also help to highlight  
2026 fundamental properties before confronting them to more complex situations. In particular, even  
2027 if we consider networks of sensors, we neglect to simulate the system and networking aspects.  
2028 We consider they have a lesser influence on the metrological performances of *in situ* calibration  
2029 strategies. One could, however, argue that system and network issues could challenge the  
2030 robustness of a given implementation of a calibration protocol because of packet loss. It could  
2031 also be interesting to evaluate the energetic cost or time to converge of such implementation. It

2032 is still possible to extend the framework to these issues that could be covered by tools such as  
 2033 WSN<sup>11</sup>, NS-3<sup>12</sup> or other<sup>13</sup>.

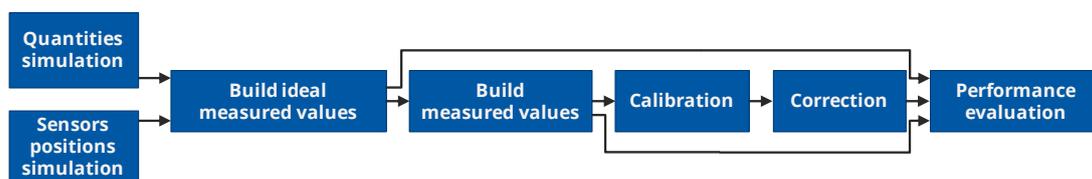
2034 The objective of this framework is to help end users in the design of their sensor networks by:

- 2035 • showing which strategies are applicable to a given use case. Currently only a few strategies  
 2036 are supplemented with strong formal criteria enabling to determine whether they can be  
 2037 used for a given network and in a given situation, for instance in [45] or [7]. Such criterion  
 2038 may, however, be not easy to define and therefore simulation is an interesting solution.
- 2039 • showing which strategies should give the best performances regarding the assumptions that  
 2040 can be made in practice.
- 2041 • allowing the optimisation of the settings of the calibration strategies –as in [109] where the  
 2042 time period of calibration is studied. As there are rarely formal applicability criteria, there  
 2043 is also rarely protocols defined for the adjustment of the parameters.

## 2044 2.2 Functional decomposition

2045 Our methodology, represented in Figure 3.2.1, can be described schematically as follows:

- 2046 1. Build a dataset of ideal measured values.
  - 2047 (a) Simulate the quantities involved with a sufficiently high resolution (spatially and  
 2048 temporally).
  - 2049 (b) Simulate the positions of the measuring instruments.
  - 2050 (c) Combine the two simulations.
- 2051 2. Build measured values from the ideal measured values, e.g. add defects to these values.
- 2052 3. Perform calibration, e.g. determine the correction to apply to the measured values.
- 2053 4. Derive corrected measured values.
- 2054 5. Compare measured and corrected measured values with ideal measured values. If the  
 2055 algorithm performs well, the corrected values should be closer to the ideal values than the  
 2056 measured ones.



**Figure 3.2.1:** Schematic diagram of the methodology proposed for *in situ* calibration strategies evaluation

2057 To apply this methodology, input data are the number of instruments involved, the quantities  
 2058 measured, the *in situ* calibration strategies to apply and the metrics considered for the evaluation  
 2059 of the results but also:

<sup>11</sup>WSN<sup>et</sup>: <http://wsnet.gforge.inria.fr/>

<sup>12</sup>NS-3: <https://www.nsnam.org/>

<sup>13</sup>OMNet++, OpNet, LoRaSim...

- 2060 • Information to simulate the quantities involved. Depending on the model, it can be as  
2061 simple as some parameters of an equation (e.g. a Gaussian model) or a complex description  
2062 of an urban scene (e.g. <http://air.ec-lyon.fr/SIRANE/>)
- 2063 • Specifications on the deployment and possible mobility of the instruments. Since the  
2064 positions of the instruments is critical for the performances of the sensor network, the  
2065 number of instruments deployed, and their positions may be optimised [19].
- 2066 • Descriptions of the measuring chain of the instruments involved, in terms of drift and of  
2067 their potential influence quantities.
- 2068 • Configuration of the considered *in situ* calibration strategies.

2069 As outputs, the framework provides time series of the true values, of the measured values  
2070 and of the corrected values for each instrument. Values of metrics computed for the comparison  
2071 of these times series may also be given.

2072 The following subsections give details on each step.

### 2073 2.2.1 Build ideal measured values

2074 **Ideal measured values** are measured values that should be measured by instruments if  
2075 they were perfect. It means for an instrument  $s_i$  that its measured values  $v(s_i, t)$  are equal to  
2076 the true value of the measurand, noted  $v_{true}(s_i, t)$ , for all  $t$ .

2077 Generating these ideal measured values for all the instruments of a sensor network can be  
2078 carried out as follows.

2079 To build ideal measured values, true values of the measurand must be computed. True values  
2080 of other quantities may also be derived for the simulation such as influence quantities if the  
2081 instruments to model undergo such effects.

2082 To be able to repeat the study with different sensor networks, the quantities can be simulated  
2083 independently of the positions of the instruments so that one may build different sets of ideal  
2084 measured values from a same simulation of the quantities depending on the sensor network.

2085 Therefore, the simulation is performed on the targeted domain of study, e.g. a specific  
2086 geometry, at a sufficient spatial and temporal resolution, regarding the potential positions of the  
2087 instruments and their frequency of measurement. The model used for the simulation depends on  
2088 the application and on the targeted accuracy. Ideally, the simulation should perfectly represent  
2089 the studied environmental phenomena and the quantities affecting instruments operating under  
2090 real conditions. This is not always possible and thus the models used may not reflect the real  
2091 temporal and spatial variability of environmental phenomena. This is particularly true for air  
2092 pollution monitoring. While high model accuracy can be achieved with advanced computational  
2093 fluid dynamics models [12], such models require time and resources that are not always acceptable  
2094 or available for users. In multiple works, case studies were conducted with abstract models [7],  
2095 enabling to demonstrate that the proposed calibration strategy is functional. However, such  
2096 models do not enable determining if an algorithm is applicable on real cases.

2097 In addition to the true values of the considered quantities, the sets of the positions where  
2098 measurements are performed must be built with respect to the position and possible mobility of  
2099 each instrument. Based on the positions of the instruments over time, true values are extracted  
2100 and stored from the simulation of the studied quantities at the locations of each instrument at  
2101 each time step according to their sampling periods.

2102 Afterwards, a time series of ideal measured values is available for each measuring instrument.<sup>14</sup>

---

<sup>14</sup>In this section, we considered that the simulation of the quantities and the simulation of the positions of the instruments are carried out separately and then combined. However, it might be possible to determine first the positions of the instruments and then to derive the ideal measured values only at the positions for instance.

2103 After this step, these time series must be validated against the hypotheses of the targeted  
2104 calibration strategies.

### 2105           **2.2.2 Build measured values from the ideal measured values**

2106       Until this step, measuring instruments are assumed ideal. Thus, to obtain realistic measured  
2107 values, e.g. values that include the real behaviour of measuring instruments, faults must be  
2108 introduced. The taxonomy presented in Chapter 1 Section 2.2 can be used here to model the  
2109 measured values. As we study *in situ* calibration strategies, drift faults are introduced principally.  
2110 The influence on the performances of *in situ* calibration strategies of other faults (e.g. noise,  
2111 spikes, missing data...) can also be studied by introducing them in the measured values (see  
2112 Section 5.3.2).

### 2113           **2.2.3 Perform calibration and build corrected values**

2114       Once measured values are obtained, calibration may then be performed low-cost to correct  
2115 them. It must be based on algorithms applicable to the specific case study addressed.

2116       Calibration and correction are separated in the methodology for the sake of genericity. Some  
2117 calibration strategies directly determine the correction to apply to each value [149], while others  
2118 provide parameters of a mathematical relationship to apply to each measured value [7].

### 2119           **2.2.4 Evaluate**

2120       Finally, corrected values, measured values and true values can be compared to assess the  
2121 performances of the algorithms. Most of the time, usual metrics are employed such as root-mean-  
2122 square error, mean absolute or Pearson's correlation coefficient for each measuring instrument in  
2123 the network. Results of case studies may also be presented as the mean of these metrics over the  
2124 whole network [150]. We discuss the most suitable metrics to use in Section 4.

2125       In the following sections, the framework is applied to the comparison of several strategies for  
2126 blind static networks on a case study that is simple enough to focus on fundamental issues.

## 2127       **3 Comparison of *in situ* calibration strategies for blind static sensor net-** 2128       **works**

2129       The purpose of this section is to give a simple example of how to apply the framework.

### 2130           **3.1 Frame of the study**

2131       In this section, the case of blind static sensor networks is considered. All the instruments  
2132 remain at the same position and no reference instrument is present. This type of sensor network  
2133 offers a particular challenge in terms of calibration. By contrast, when reference instruments are  
2134 in the network, trustworthy values are available. Secondly, when instruments are mobile, it is  
2135 possible that two or more instruments are measuring the same true quantity value when they are  
2136 in a spatio-temporal vicinity. For blind static sensor networks, the availability of standard values  
2137 may not be assumed, and only the instruments that are deployed close enough can compare their  
2138 measurements.

2139       We consider seven calibration strategies, from four different groups of authors, that were  
2140 identified in across the review of Chapter 2. These strategies have never been compared all  
2141 together on a common case study.

- 2142       • Balzano et al. [7]: To apply this algorithm, the sensor network must be dense enough to  
2143       oversample the signal of interest. We refer to this assumption by the term "oversampling"  
2144       afterwards. With such a sensor network, the true values lie in a subspace of the space  
2145       formed by the measured values. To calibrate the instruments, this subspace must be

known. The gains and offsets for all nodes are then estimated with the chosen subspace and the measured values using singular value decomposition (SVD) or by solving a system of equations using a least square (LS) estimator. These strategies are respectively called **SM-SVD** and **SM-LS**, SM standing for subspace matching. Note that gains are estimated up to a multiplicative coefficient. Consequently, the gain of one instrument should be known to complete the calibration. Likewise, the offset computation requires either to know some of the offsets, or that an additional hypothesis on the signal is valid.<sup>15</sup> As such, the network is not perfectly blind in truth.

- Lipor et al. [89]: This work is an extent of [7]. It shows first that the solution obtained with SM-SVD is equivalent to the one obtained by solving the system of equations expressed as a total least squares problem (TLS). The latter strategy is called **SM-TLS**.<sup>16</sup>
- Takruri et al. [148–150]: Three contributions from these authors are considered. The first one is based on the idea that the average of the values of neighbouring measuring instruments gives a good estimation of the correct value of an instrument to calibrate. If the absolute difference between the measured value and the average is greater than a given threshold, this difference is added to the instrument's value. We call this strategy **AB-DT** (average-based estimation for difference-based threshold). Two variations are also proposed, a first one in which the difference-based threshold mechanism is replaced by a Kalman filter that estimates the error affecting the instruments (**AB-KF**), and a second one where the average-based part is replaced by support vector regression (**SVR-KF**).
- Kumar et al. [76, 77]: This work, inspired from the previous one, also proposes the use of a Kalman filter to estimate the error of each instrument but instead of performing an average over neighbouring nodes or using SVR for true value estimation, kriging is used (**K-KF**).

Note that not every strategy that could have been applied to our case study is considered here. For instance, the work of Dorffer et al. [44] extends [7] and [89] for sparse signals, but this is not in the scope of the study.

## 3.2 Application of the framework

### 3.2.1 Simulation of the ideal measured values

A space of  $1000 \times 1000\text{m}$  is considered. It is discretised with a step of 10m. At the centre of this area, a  $NO_2$  source is considered. The concentration of  $NO_2$ ,  $C$ , at the instant  $t$  and position  $(x, y)$ , is modelled as:

$$C(x, y, t) = A(t) \exp\left(-\frac{x^2 + y^2}{\sigma(t)^2}\right)$$

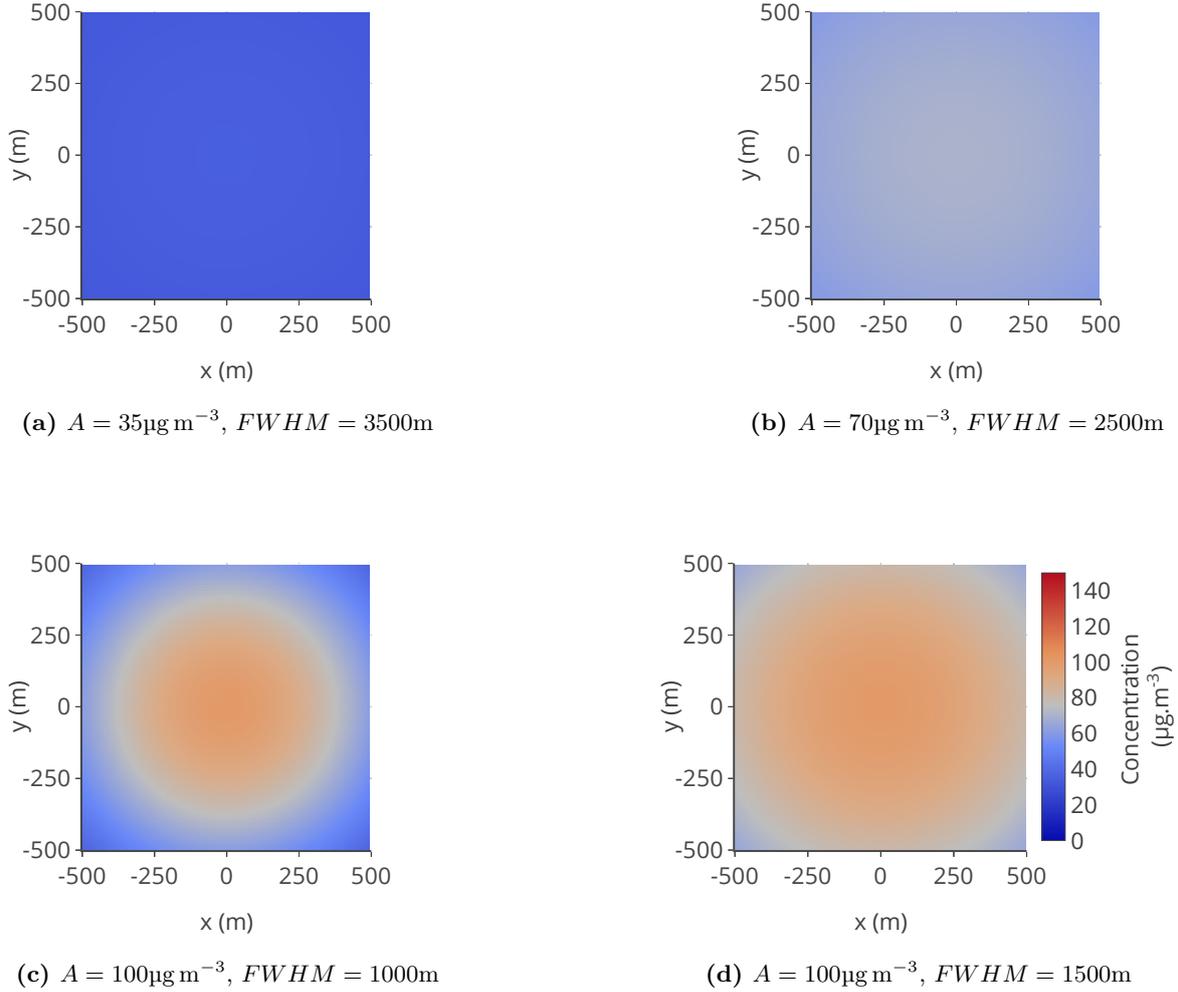
It is a 2D Gaussian function with an equal spread  $\sigma$  for  $x$  and  $y$ . This model is not very realistic but has been used in other papers [45] for its simplicity of implementation. Examples of pollution maps are represented in Figure 3.3.1.

To facilitate the interpretation of  $\sigma$ , it is expressed as a function of the full width at half maximum (FWHM) of the Gaussian curve:

$$\sigma = \frac{FWHM}{2 \ln(2)}$$

<sup>15</sup>A usual assumption is that "the signal has a null (or known) average".

<sup>16</sup>Note that additional formulations given in the publication, in the case where some gains are known for a subset of instruments, are not considered here as it would be equivalent to have a partially blind sensor network.



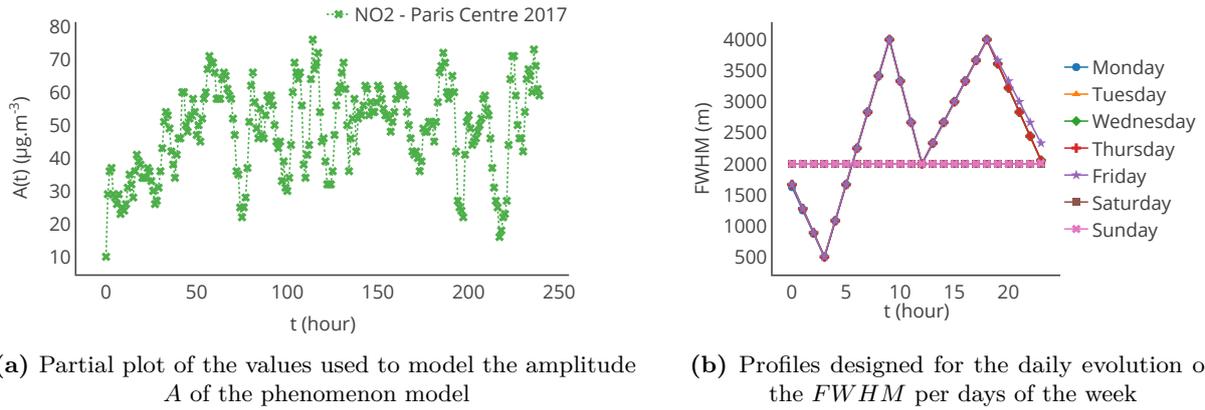
**Figure 3.3.1:** Examples of maps of concentration  $C$  used following  $C(x, y) = A \exp\left(-\frac{(2\ln(2))^2}{FWHM^2}(x^2 + y^2)\right)$  for given  $A$  and  $FWHM$

2177  $A$  and  $FWHM$  are functions representing respectively the temporal evolution of the amplitude  
 2178 of the 2D Gaussian function and of its FWHM.

2179 To be representative of actually measured concentrations, the function  $A$  is based on values  
 2180 measured by a  $NO_2$  monitoring station in Paris between 2017/01/01 00:00 and 2017/12/31 23:00,  
 2181 with an hourly time step (Paris centre station [2]).<sup>17</sup> We consider it represents our pollution  
 2182 source at the coordinates  $(0, 0)$ . Even if the values of the station are not the ones of a real source,  
 2183 we assume it gives a reasonable order of magnitude of variation over time.

2184 The FWHM represents the spread of the pollutant around the source. Two linear piecewise  
 2185 functions are defined to represent the daily evolution of the FWHM: one for weekdays and one  
 2186 for weekends. They are represented in Figure 3.3.2b. Their shape is inspired from the dataset

<sup>17</sup>There were missing entries in the time series. Values were first interpolated with a limit of 3 consecutive values to fill, e.g. 3 hours. Then, in the case where more than 3 consecutive values are missing, the interpolation is made based on the values at the same hour of the previous and next day. The values are interpolated with a linear function for both cases.



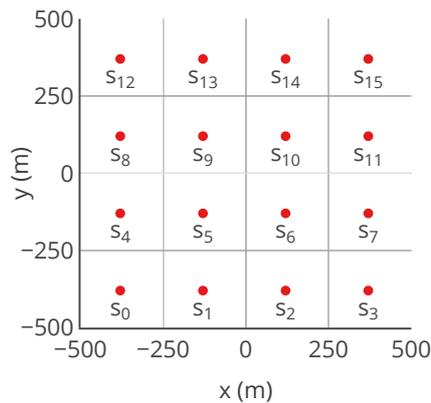
(a) Partial plot of the values used to model the amplitude  $A$  of the phenomenon model (b) Profiles designed for the daily evolution of the  $FWHM$  per days of the week

**Figure 3.3.2:** Evolution of  $A$  and  $FWHM$  for the modelling of the concentration of pollutant

2187 used for the amplitude  $A$ . This is not fully realistic, but it provides an order of magnitude of  
 2188 spatial and temporal variations occurring in urban context.

2189 The concentration of pollutant  $C$  is simulated for a year at an hourly time-step over the  
 2190 area of study. To avoid the computation of  $C$ , at each time step, for the current value of  $A$  and  
 2191  $FWHM$ , a catalogue of maps is generated from sets of possible values for the amplitude and  
 2192 the  $FWHM$ : respectively from 0 to  $150\mu\text{g m}^{-3}$  with a step of  $5\mu\text{g m}^{-3}$  and from 0 to 4000m  
 2193 with a step of 500m. In our case, it required the simulation of 157 maps. To derive the time  
 2194 series, the map generated with the closest allowed values of  $A$  and  $FWHM$  is picked from the  
 2195 catalogue at each time step. Finally, a time series of 8760 maps is obtained.

2196 A sensor network  $S$  of  $|S| = 16$  static measuring instruments is considered. It is uniformly  
 2197 deployed spatially. The positions are represented in Figure 3.3.3. Then, the time series of  
 2198 concentration maps and the positions of the instruments are combined to obtain the time series  
 2199 of ideal measured values for all the instruments, noted  $v_{true}$ . In this case, we consider that the  
 2200 instruments measure one value per hour.



**Figure 3.3.3:** Positions of the 16 measuring instruments considered in the case study, deployed uniformly in the field

2201 The validity of this set of ideal measured values must be verified for **SM-(SVD,LS,TLS)**

2202 strategies. The network should satisfy the hypothesis of "oversampling". To verify it, a principal  
 2203 component analysis (PCA) of the matrix formed by the concatenation of the ideal measured  
 2204 values time series of each node is carried out. We assume that the sensor network is satisfying the  
 2205 oversampling hypothesis if the explained variance ratio of the  $\nu$  first components is greater than  
 2206 0.999 and  $\nu < |S|$ . In our case, the condition is met for  $\nu = 2 < |S|$ . Therefore, the oversampling  
 2207 hypothesis is satisfied.

### 2208 3.2.2 Building of the measured values

2209 The network is defined as "blind", which can also be expressed by "all the instruments have the  
 2210 same sensing reliability". In a first step, instruments are only drifting. Drift of each instrument is  
 2211 only function of time. We assume all the instruments undergo a linear drift increasing up to 5%  
 2212 of the gain every 7 days. The actual drift is randomly drawn following a uniform law. The value  
 2213 of 5% is set according to what is reported in typical measuring instrument datasheets [105].

2214 The instruments are assumed to be initially calibrated and to remain faultless for eight weeks,  
 2215 which gives a training dataset for the strategies needing it, notably to learn the subspace of the  
 2216 signal with PCA for **SM-(SVD,LS,TLS)**. For four additional weeks, the instruments are kept  
 2217 without drift to study how calibrations strategies behave if the devices are not drifting. Finally,  
 2218 the instruments start drifting after twelve weeks. This instant is noted  $t_{\text{start drift}}$ .

2219 Therefore, the gain  $G(s_i, t)$  of the instrument  $s_i$  at  $t$  is :

$$G(s_i, t) = \begin{cases} 1 & \text{if } t < t_{\text{start drift}} \\ G(s_i, t - 1) & \text{if } t \geq t_{\text{start drift}} \text{ and } (t - t_{\text{start drift}}) \bmod 7 \text{ days} \neq 0 \\ G(s_i, t - 1) + \delta G(s_i, t) & \text{if } t \geq t_{\text{start drift}} \text{ and } (t - t_{\text{start drift}}) \bmod 7 \text{ days} = 0 \end{cases}$$

$$\text{with } \forall t, \delta G(s_i, t) \sim \mathcal{U}(0, 0.05)$$

2220 This drift model is called "Weekly Gain Linear Increase" (**WGLI**).

The measured value  $v(s_i, t)$  of the instrument  $s_i$  at  $t$  is expressed from the true value  $v_{\text{true}}(s_i, t)$   
 using the relationship:

$$v(s_i, t) = G(s_i, t) \cdot v_{\text{true}}(s_i, t)$$

### 2221 3.2.3 Configuration of the strategies

2222 Each strategy has parameters to set up. They are defined as follows:

- 2223 • **SM-(SVD,LS,TLS)**: All these strategies share the same parameters, namely the period-  
 2224 icity of calibration and the time interval on which the linear system is solved. Note that  
 2225 the adjustment of the parameters was not discussed by the original authors. We chose to  
 2226 apply the strategy each week and to solve the linear system over the past  $w = 7$  days.
- 2227 • Kalman filter of **(AB, SVR, K)-KF**: Variables  $R$  and  $Q$  are set to 0.0001 and 1 respec-  
 2228 tively.
- 2229 • Definition of neighbours for **AB-(DT,KF)**: Instruments are defined as neighbours if they  
 2230 are distant of less than 250m.
- 2231 • Parameters of the SVR for **SVR-KF**: Kernel used is 'rbf', the kernel coefficient gamma is  
 2232 set to  $10^{-8}$  and the penalty parameter  $C$  is set to  $10^3$ .
- 2233 • Parameters for the weighting for **K-KF**:  $a$ ,  $c_0$  and  $c_1$  were defined as in [77], therefore  
 2234  $a = 12$ ,  $c_0 = 0.25$ ,  $c_1 = 0.85$ .

2235 We invite the reader to refer to the original publications for a more in-depth understanding  
2236 of the parameters of these calibration strategies.

### 2237 3.2.4 Evaluation

2238 To compare the corrected values of an instrument, obtained with each strategy, and its  
2239 measured values against its ideal measured values, metrics are needed. We consider the most  
2240 commonly used metrics in the publications reported in Chapter 2 and described as follows. They  
2241 can be computed for each instrument or averaged over the network. We discuss the relevance  
2242 of these statistics in Section 4. In the following, the set  $x$  stands for the  $k$  measured values  
2243 of an instrument  $s_i$  over  $[t - \Delta t; t]$ ,  $V(s_i, (t, \Delta t))$  or the corresponding set of corrected values  
2244  $V_{corr}(s_i, (t, \Delta t))$  and the set  $y$  is the associated  $k$  true values  $V_{true}(s_i, (t, \Delta t))$ .

- Mean absolute error (**MAE**):

$$MAE(x, y) = \frac{1}{k} \sum_{i=0}^k |x_i - y_i|$$

- Root-mean-square error (**RMSE**):

$$RMSE(x, y) = \sqrt{\frac{1}{k} \sum_{i=0}^k (x_i - y_i)^2}$$

- Mean absolute percentage error (**MAPE**):

$$MAPE(x, y) = \frac{100}{k} \sum_{i=0}^k \frac{|x_i - y_i|}{|y_i|}$$

2245 Remark: the values of  $y$  must all be different from zero.

- Pearson correlation coefficient ( $\rho$ ):

$$\rho(x, y) = \frac{E[(x - E[x])(y - E[y])]}{\sqrt{E[(x - E[x])^2]E[(y - E[y])^2]}}$$

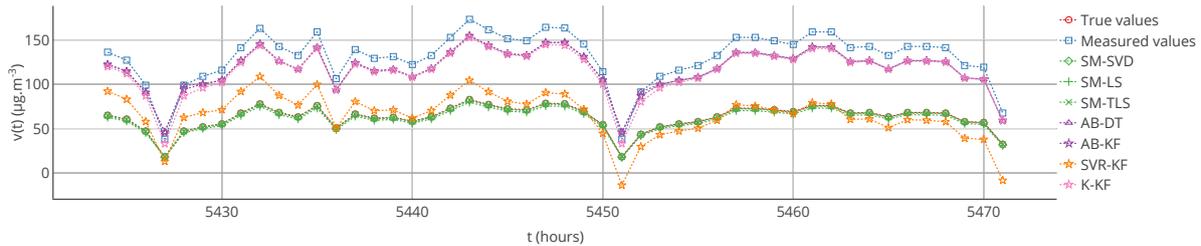
2246 with  $E[.]$  being the expectation.

2247 For perfect instruments before correction (or after with  $v_{corr}$  instead of  $v$ ):

- 2248 •  $MAE(V(s_i, (t, \Delta t)), V_{true}(s_i, (t, \Delta t))) = 0$ .
- 2249 •  $RMSE(V(s_i, (t, \Delta t)), V_{true}(s_i, (t, \Delta t))) = 0$ .
- 2250 •  $MAPE(V(s_i, (t, \Delta t)), V_{true}(s_i, (t, \Delta t))) = 0\%$ .
- 2251 •  $\rho(V(s_i, (t, \Delta t)), V_{true}(s_i, (t, \Delta t))) = 1$ .

### 2252 3.3 Results

2253 In Figure 3.3.4, a partial plot of true values, measured values and corrected values obtained  
2254 with each strategy is displayed for a particular instrument. For the considered device, a visual  
2255 observation indicates that strategies SM-(SVD, LS, TLS) and SVR-KF provide better results  
2256 than AB-DT, AB-KF and K-KF.



**Figure 3.3.4:** True values, measured values and corrected values with the strategies considered for a particular instrument  $s_1 = 2$  between  $t = 5424$ h and  $t = 5471$ h. SM-(SVD, LS, TLS) and SVR-KF seem to provide better results than AB-DT, AB-KF and K-KF

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	18	2	53	6	25	2	0.935	0.007
<b>SM-SVD</b>	1	1	5	4	2	2	0.994	0.006
<b>SM-LS</b>	1	1	3	2	2	1	0.995	0.005
<b>SM-TLS</b>	1	1	5	4	2	2	0.994	0.006
<b>AB-DT</b>	18	1	55	6	25	1	0.932	0.004
<b>AB-KF</b>	18	1	58	11	25	1	0.933	0.003
<b>SVR-KF</b>	16	1	83	11	18	1	0.825	0.012
<b>K-KF</b>	18	1	53	10	24	1	0.927	0.005

**Table 3.3.1:** Mean and standard deviation of each metric, computed on the entire time interval of drift, over the 16 nodes of the network. SM-(SVD, LS, TLS) strategies have the best results overall whatever the metric considered. SVR-KF provides corrected values only slightly better than before calibration according to MAE and RMSE but not according to MAPE and Pearson correlation coefficient. AB-(DT, KF) and K-KF do not improve the measured values significantly.

2257 Despite its advantage to easily visualise the results, this representation is not representative  
 2258 of all the instruments of the network.

2259 The computation of the mean and standard deviation of each metric over the whole sensor  
 2260 network, on the entire time interval of study and with the results of each strategy, plus without  
 2261 calibration, is given in Table 3.3.1. With it, we observe that:

- 2262 • SM-(SVD, LS, TLS) strategies have the best results overall with a small mean error and  
 2263 standard deviation whatever the metric considered
- 2264 • SVR-KF which seemed to give interesting results in Figure 3.3.4 provides corrected values  
 2265 only slightly better than before calibration according to MAE and RMSE but not according  
 2266 to MAPE and Pearson correlation coefficient. This could be explained by two reasons. The  
 2267 strategy may correct well for high measured values but correct poorly for low ones. It could  
 2268 also be due to errors in the corrected values of particular instruments introduced by the  
 2269 calibration algorithm. The aggregation by averaging makes it impossible to discriminate  
 2270 between these two possible explanations.
- 2271 • AB-(DT, KF) and K-KF do not improve the measured values

	MAE		MAPE		RMSE		Pearson	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	22	18	66	51	30	24	0.924	0.942
<b>SM-SVD</b>	1	3	4	10	2	5	0.996	0.983
<b>SM-LS</b>	1	2	3	7	2	3	0.997	0.986
<b>SM-TLS</b>	1	3	4	10	2	5	0.996	0.983
<b>AB-DT</b>	20	17	61	49	26	23	0.933	0.932
<b>AB-KF</b>	18	17	56	45	25	23	0.935	0.927
<b>SVR-KF</b>	16	16	80	71	19	19	0.828	0.805
<b>K-KF</b>	17	16	51	43	24	23	0.930	0.918

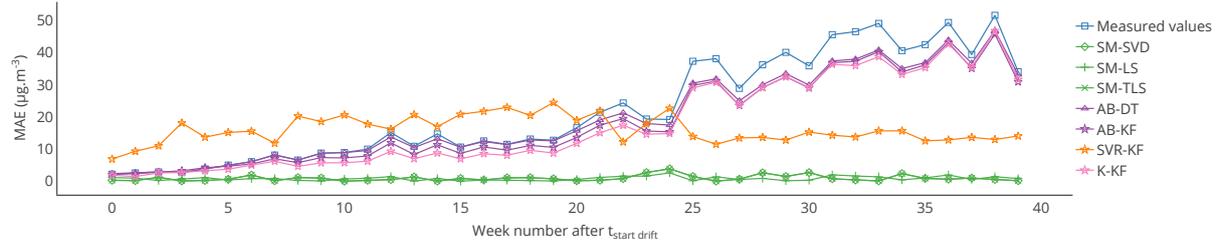
**Table 3.3.2:** Values of the metrics, computed on the entire time interval of drift, for two particular instruments of the network  $s_1 = 2$  and  $s_2 = 10$ . Strategies AB-(DT, KF) and (SVR, K)-KF are quite equivalent for these two instruments. The improvements are rather small for both instruments. For SM-(SVD, LS, TLS), results are consistent with the observations of Table 3.3.1.

2272 These last two observations invited us to look at results for particular instruments. Table 3.3.2  
2273 gathers the results for two instruments selected randomly. We observe that strategies AB-(DT,  
2274 KF) and (SVR, K)-KF are quite equivalent for these two instruments. The improvements are  
2275 rather small for both instruments. For SM-(SVD, LS, TLS), results are consistent with our  
2276 previous observation based on the average and the standard deviation in Table 3.3.1. The standard  
2277 deviation does give the information of the existence but not the identity of the instruments that  
2278 are degraded or improved by calibration, which is needed for practical deployments to drive the  
2279 maintenance operations.

2280 Furthermore, note that the results of Table 3.3.1 were computed over the 12 weeks of drift.  
2281 The results may be different if computed over a different time range, for instance over each week.  
2282 Figure 3.3.5 shows the evolution over time of the MAE computed each week for a particular  
2283 instrument and each strategy. Table 3.3.3 provides statistics on the MAE computed each week  
2284 but also for the other metrics that were computed in the same way. From the standard deviations  
2285 of MAE, MAPE and RMSE, it shows that the observations made previously could be locally  
2286 false, e.g. a strategy is better than others considering a computation of the metrics over the 12  
2287 weeks of drift but not always considering a computation of the metrics over each week. This  
2288 is shown in Figure 3.3.5 where results with MAE for SVR-KF are nearly always worse than  
2289 those for AB-DT, AB-KF and K-KF until week 24 but are better afterwards. This figure also  
2290 shows that the performances of SVR-KF could be explained by the presence of a bias, at least  
2291 according to MAE because it is quite constant in Figure 3.3.5.

### 2292 3.4 Conclusions

2293 Through this evaluation, we have shown that existing *in situ* calibration strategies could  
2294 improve the measured values of a blind static sensor network in the considered case study. Overall,  
2295 the strategies SM-(SVD, LS, TLS) have the best performances. This can be explained by the  
2296 fact that the gain of at least one of the instruments in the network has to be known for these  
2297 methods. In a way, the sensor network is only partially blind, but acquiring this information may  
2298 not be a major issue in practice for large sensor networks. The other strategies appear to be able  
2299 to mitigate the drift by few per cents at most only in the best cases, depending on the metric



**Figure 3.3.5:** Evolution of the MAE computed for each week of the drift period between the drifted values and the true values, and between the corrected values for each strategy and the true values for a particular instrument, after the start of drift. MAE for SVR-KF is nearly always worse than those for AB-DT, AB-KF and K-KF until week 24 but are better afterwards. The performances of SVR-KF could be explained by the presence of a bias, at least according to this metrics, as its evolution is quite flat.

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	22	16	66	36	25	17	1.000	0.000
<b>SM-SVD</b>	1	1	3	3	1	1	1.000	0.000
<b>SM-LS</b>	1	1	3	2	1	1	1.000	0.000
<b>SM-TLS</b>	1	1	3	3	1	1	1.000	0.000
<b>AB-DT</b>	20	13	61	32	22	14	0.999	0.001
<b>AB-KF</b>	18	13	56	31	21	14	0.999	0.000
<b>SVR-KF</b>	16	4	80	39	18	4	0.935	0.034
<b>K-KF</b>	17	14	51	32	20	15	1.000	0.000

**Table 3.3.3:** Mean and standard deviation of the weekly mean of each metric for  $s_1$  during the drift period. From the standard deviations of MAE, MAPE and RMSE, the observations made based on Table 3.3.1 could be locally false, e.g. a strategy is better than others considering a computation of the metrics over the 12 weeks of drift but not always considering a computation of the metrics over each week.

2300 considered. The comparison between metrics enables understanding better how they operate.  
 2301 For all algorithms, studying the average values of the metrics over all the network hides strong  
 2302 disparities from instrument to instrument.

2303 However, these results could be challenged based on the following considerations:

- 2304 • The results may be different with another network (different positions and increased  
 2305 number of instruments, notably for AB-DT, AB-KF and K-KF strategies that use values  
 2306 of neighbour instruments for the prediction of correct values). However, the network is  
 2307 often defined by the use case. Thus, it is not always a criterion that can be optimised.
- 2308 • The model of the considered phenomenon is not realistic and therefore results may not be  
 2309 practical. It is, however, sufficient to highlight issues of the algorithms under test. A more  
 2310 realistic model may not yield better result but would certainly make the analysis more  
 2311 complex.

2312 • Only a drift of the gain was added which does not fully relate the behaviour of a real  
 2313 instrument. Adopting a finer-grained model of the behaviour also increases the complexity  
 2314 of the case study.

2315 • Other parameters of the calibration strategies may change the results. We empirically  
 2316 explored the parameters of the algorithms, keeping in this study the values that give the  
 2317 best results for each strategy. There is, however, no claim for optimality here, but the  
 2318 framework can be used to optimise the parameters of a given strategy.

2319 In Section 5, the impact of these four separate aspects on the algorithms is tested. Before  
 2320 moving on to this, we discuss in the next section the metrics used and how additional ones, or  
 2321 other ways to visualise results, can give an enhanced comprehension of the behaviour of the  
 2322 calibration algorithms on a sensor network.

## 2323 4 Evaluation of measurements after correction

### 2324 4.1 Problem statement

2325 To evaluate the performances of *in situ* calibration strategies, various metrics are commonly  
 2326 used like MAE, MAPE, RMSE and Pearson’s correlation coefficient notably. We presented the  
 2327 results as statistics of these metrics over the entire network or values for particular instruments.  
 2328 The metrics were computed on the entire time interval of study or on subintervals. Results were  
 2329 displayed as tables and plots. These choices enabled carrying out an analysis at multiple scales  
 2330 (entire network versus individual instruments). This approach underlines disparities of efficiency  
 2331 of the algorithms on particular instruments for instance.

2332 Multiple surveys exist on metrics [17]. They aim at providing a clear reading on the differences  
 2333 and similarities among the wide spectra of metrics used. While metrics are numerous, groups of  
 2334 them are based on the same idea (difference, absolute difference, square difference...). In some  
 2335 works, such as in [54], an integrated performance index was developed to combine the values  
 2336 of several metrics into a single one. However, some metrics are interdependent [154], which  
 2337 can exaggerate or hide information when they are combined. In the literature regarding *in situ*  
 2338 calibration strategies, the choice of metrics is not always clearly justified, nor is the choice of  
 2339 the values presented and how they are computed (values for particular instruments, statistics  
 2340 of the metric over the network and so on). Moreover, Tian *et al.* [154] showed it is possible to  
 2341 have equal values of multiple metrics for two sets of values very different from each other. They  
 2342 justified it by the interdependence, underdetermination and incompleteness of usual metrics, in  
 2343 addition to the fact that they assume the error to be linear. This issue is major as metrics are  
 2344 used in official recommendations such as the data quality objectives (DQO) in the context of air  
 2345 quality measuring instruments [16, 27, 95]. In this way, we suggest in the next section to use an  
 2346 error model as introduced by Tian *et al.* [154]. It is considered as a satisfying means to deal  
 2347 with these issues.

2348 Another question lies in the way a metric is computed or in how statistics are derived from  
 2349 it. As shown in the previous section, looking at values of metrics computed over the full-time  
 2350 interval of drift or computed over each week may yield different interpretations. The choice of  
 2351 the time interval on which it is derived can be guided by requirements on the minimal duration  
 2352 for which the mean error of an instrument must be under a given threshold for instance.

2353 Regarding the use of statistics, it concerns particularly networks with hundreds of nodes,  
 2354 for which interpreting the values of metrics for each node may not be easily feasible. However,  
 2355 statistics derived over the entire network do not enable discriminating easily between instruments  
 2356 differently affected by an algorithm. The use of statistics is perfectly appropriate if the quality  
 2357 of the calibration of the instruments belonging to a sensor network should be evaluated at the  
 2358 level of the network, e.g. when it is the collective performance that matter. Otherwise, if each

instrument is expected to be perfectly calibrated after an *in situ* calibration, statistics may only give a first idea of the performances of the algorithm. At the same time, evaluating the performances of an algorithm on a very large sensor network by looking at each instrument individually may not be feasible. Thus, the scale at which the performances of the instruments are evaluated must be specified in the description of a study, to make a proper use of statistics and to determine if the analysis is manageable in practice.

In addition, the way the results are displayed is important. We discuss this topic in Section 4.3 and propose the use of an easily readable representation, compatible with the error model we use hereafter.

## 4.2 Evaluation with an error model

We investigate in this section an approach which captures better individual instrument behaviours. It applied to the previous case study for the sake of comparison. We follow the error model approach proposed by Tian *et al.*[154]. It consists in finding a function  $\mathcal{F}$  associating each measured value of an instrument  $s_i$ , noted  $v(s_i, t)$ , to the true values of the measurand  $v_{true}(s_i, t)$  so that  $v(s_i, t) = \mathcal{F}(v_{true}(s_i, t))$ .<sup>18</sup> They illustrated this with a linear additive error model expressed as:

$$v(s_i, t) = a \cdot v_{true}(s_i, t) + b + \varepsilon(t)$$

with  $a$  and  $b$  being respectively a constant gain and a constant offset, and  $\varepsilon$  a random error, following a given probability law depending on the case. If the actual distribution of  $\varepsilon$  is not determined, it is equivalent to the fitting a linear trend model between  $v$  and  $v_{true}$ . A linear regression can be used to determine the slope  $a$  and the intercept  $b$  in this case.  $a = 1$  and  $b = 0$  are the ideal results, indicating a perfect calibration. The linear regression between  $v$  and  $v_{true}$  is appropriate to describe the performances before calibration. Comparing the slope and intercept of this regression to the ones obtained with the linear regression between the corrected values after calibration  $v_{corr}$  and the true values  $v_{true}$  enables determining the improvement brought by the calibration algorithm. This approach with linear regression is used in the literature on calibration of individual instruments [139]. In our case, this error model is perfectly appropriate to evaluate the remaining error after calibration, considering the drift model defined in Section 3.

The case study of Section 3 is analysed here again with the error model. The results associated with all the tables and figures presented are provided. First, the overall results obtained with this approach for the slope and intercept of the linear model, plus the score of the regression, e.g. the coefficient of determination of the regression, are in Table 3.4.1. Then, results are reported in Table 3.4.2 for the same particular instruments considered in Section 3. From these tables, the observations that can be made are consistent with those made with Table 3.3.1 and Table 3.3.2.

These tables also seem to confirm a bias for the strategy SVR-KF: the standard deviation of the slope and intercept among instruments is small, the mean slope is close to 1 but the mean intercept is equal to 10. However, the poor mean score of the regression invites to be careful.

Table 3.4.3 shows the mean and standard deviation of  $a$ ,  $b$  and of the score computed weekly for a particular instrument. In this case, the means are different from those of Table 3.4.2 for  $s_1$  (a linear regression is not a linear operation). We observe that the slope for SVR-KF is varying significantly. More importantly, the mean intercept is equal to zero with a standard deviation of 17, which is not the case for the other strategies. Therefore, the observation made with Table 3.4.2 about SVR-KF is not valid locally and can explain why the mean score is poor compared to the ones obtained with the other strategies.

In conclusion, the use of an error model allows to make the same conclusions as in Section 3. However, clearer information can be derived such as the remaining error after calibration,

<sup>18</sup>This idea of error model is close to the concept of a calibration relationship. The inverse function of the error model is the relationship to apply to an instrument so that calibration is perfect.

2398 expressed in terms of gain and offset.

	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.62	0.07	-2	0	0.87	0.01
<b>SM-SVD</b>	1.02	0.02	1	0	0.99	0.01
<b>SM-LS</b>	0.97	0.02	1	0	0.99	0.01
<b>SM-TLS</b>	1.02	0.02	1	0	0.99	0.01
<b>AB-DT</b>	1.62	0.05	-2	1	0.87	0.01
<b>AB-KF</b>	1.62	0.04	-1	1	0.87	0.01
<b>SVR-KF</b>	1.05	0.03	10	1	0.68	0.02
<b>K-KF</b>	1.62	0.04	-2	1	0.86	0.01

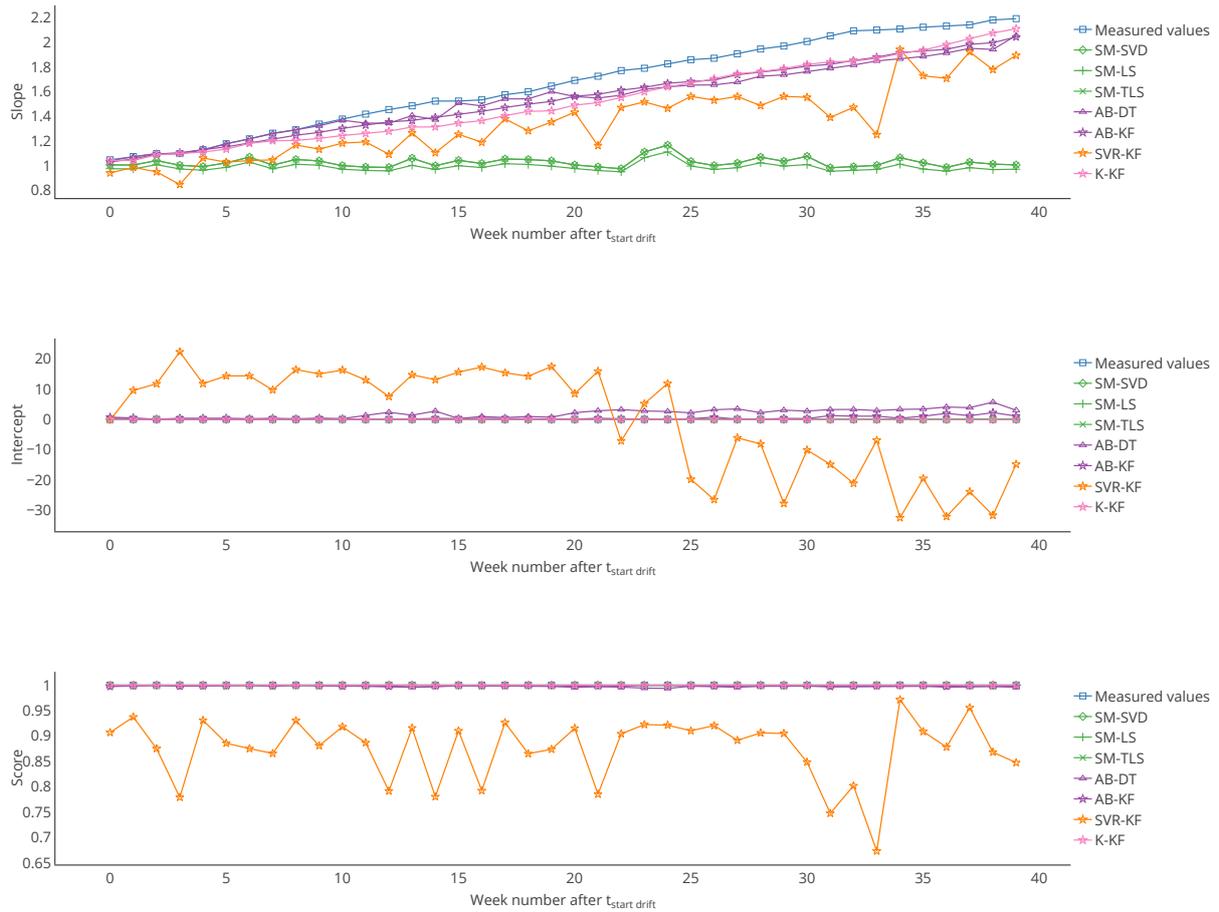
**Table 3.4.1:** Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the parameters of the error model and the regression score, computed on the entire time interval of drift, over the 16 nodes of the network. The observations that can be made with these values are identical to those of Table 3.3.1. This table also seems to confirm that there is a bias for SVR-KF as the standard deviation of the slope and intercept among instruments is small, with mean slope close to 1 but a mean intercept equal to 10. However, the poor mean score of the regression invites to be careful with this statement.

	Slope		Intercept		Score	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	1.77	1.58	-2	-1	0.85	0.89
<b>SM-SVD</b>	1.01	1.05	0	1	0.99	0.97
<b>SM-LS</b>	0.97	0.94	0	1	0.99	0.97
<b>SM-TLS</b>	1.01	1.05	0	1	0.99	0.97
<b>AB-DT</b>	1.62	1.55	0	-1	0.87	0.87
<b>AB-KF</b>	1.63	1.56	-1	-3	0.87	0.86
<b>SVR-KF</b>	1.09	1.01	9	12	0.69	0.65
<b>K-KF</b>	1.62	1.56	-2	-4	0.87	0.84

**Table 3.4.2:** Values of the parameters of the error model and the regression score, computed on the entire time interval of drift, for two particular instruments of the network  $s_1 = 2$  and  $s_2 = 10$ . The observations that can be made with these values are identical to those of Table 3.3.2.

	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.66	0.36	0	0	1.00	0.00
<b>SM-SVD</b>	1.03	0.04	0	0	1.00	0.00
<b>SM-LS</b>	0.99	0.03	0	0	1.00	0.00
<b>SM-TLS</b>	1.03	0.04	0	0	1.00	0.00
<b>AB-DT</b>	1.53	0.28	2	1	1.00	0.00
<b>AB-KF</b>	1.54	0.30	0	1	1.00	0.00
<b>SVR-KF</b>	1.35	0.28	0	17	0.88	0.06
<b>K-KF</b>	1.51	0.32	0	0	1.00	0.00

**Table 3.4.3:** Mean and standard deviation of the parameters of the error model and the regression score for  $s_1$ , computed on each week of the drift period. The slope for SVR-KF varies significantly and the mean intercept is equal to zero with a standard deviation of 17. This is not the case for the other strategies. Thus, the observation made with Table 3.4.2 on SVR-KF is not valid locally and can explain why the mean score is poor compared to the ones obtained with the other strategies.



**Figure 3.4.1:** Evolution of the slope, intercept and score of the error model of  $s_1$ , computed on each week of the drift period. Both the slope and intercept are locally poor for SVR-KF, although this is not visible when computing the error model on the entire time interval of study. This explains the poor values of the score compared to the other strategies.

2399

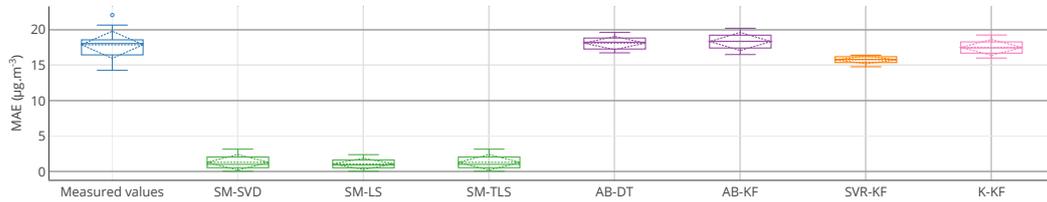
### 4.3 Means of visualisation

2400

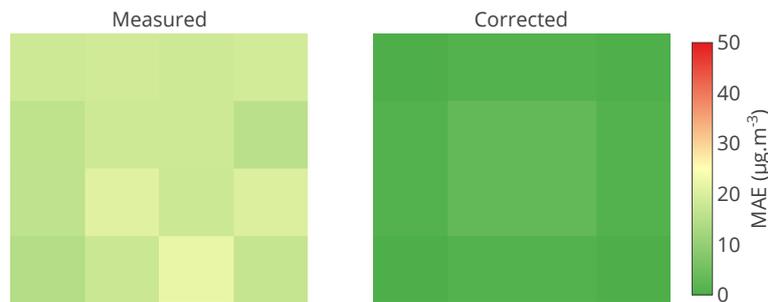
Choosing other ways of visualising results may help to better understand what happens to each instrument. Various representations (Figures 3.4.2, 3.4.3 and 3.4.4) are discussed in terms of suitability for the comparison of methods.

2401

2402



**Figure 3.4.2:** Box plots of the MAE, computed on the entire time interval of drift, of the 16 nodes of the network without calibration and with **SM-SVD**. It is a graphical representation of the information displayed in Table 3.3.1 for MAE. Supplementary information is provided compared to the table: quartiles, minimum and maximal values.



**Figure 3.4.3:** Matrix of the MAE, computed on the entire time interval of drift, of the 16 nodes of the network without calibration and with **SM-SVD**. It shows exactly the information for each instrument. The colour scale can help to identify problematic instruments.

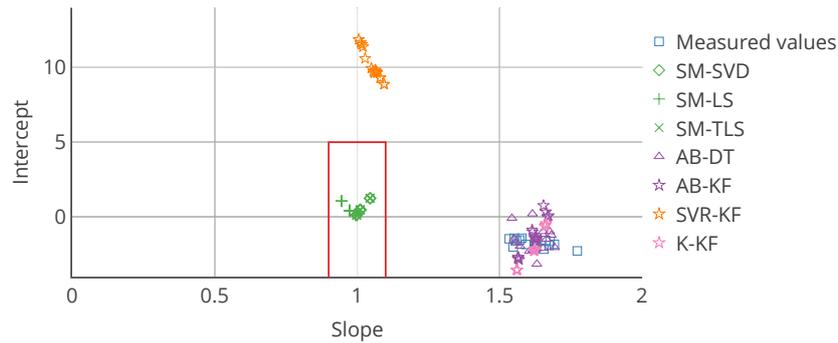
2403

Figure 3.4.2 is a boxplot-based representation of the values of MAE computed on the entire time interval of study. A boxplot is provided for each strategy. This representation is a graphical illustration of the information displayed in Table 3.3.1 for MAE. In addition, it provides supplementary information compared to the table: quartiles, minimum and maximal values. However, this representation does not show exactly the results for each instrument. One graph is needed for each metric. The plot can also show statistics on the evolution of a metric for a particular instrument like Table 3.3.3 but again it can only display information about one metric.

2409

To display values regarding each instrument of a network, a matrix of values can be a solution. A colour scale can make the values more easily readable. Figure 3.4.3 is a matrix of the mean of MAE on the entire time interval of study. This figure shows exactly the information for each instrument. The colour scale helps to identify problematic instruments. It is appropriate for real-time display. Nevertheless, if multiple metrics (or variables if an error model is considered) are used, one matrix must be plotted for each one. The study of multiple calibration strategies is not possible on a single plot.

2416



**Figure 3.4.4:** Target plot of the 16 nodes as a function of their slope and intercept in the error model for each calibration strategy. The slope and intercept are computed on the entire time interval of drift. In this case, it allows to locate an instrument according to the slope and intercept of its associated error model. The red rectangle is an example of area that can be defined to quickly identify instruments that are not satisfying a requirement, here a slope in  $[0.9; 1.1]$  and an intercept in  $[-5; 5]$ .

2417 To overcome the challenge of displaying information for multiple instrument on a single chart,  
 2418 we propose to use target plots. They enable locating a point along multiple axes (two for a 2D  
 2419 diagram, three for a 3D diagram). Boundaries within which the point should be ideally can be  
 2420 depicted. Each axis can be used for a different metrics. A colour scale can be used to depict  
 2421 a third one. In the previous section, we proposed to use an error model with two variables.  
 2422 Therefore, a target plot is perfectly adapted. Figure 3.4.4 is a target plot of the parameters of  
 2423 the error model, each point's coordinates being the values of the slope and the intercept of the  
 2424 error model for each instrument, computed on the entire time interval of study. Indeed, this  
 2425 diagram shows exactly the information for each instrument. Moreover, statistics like standard  
 2426 deviation can be added with error bars. The target diagram is also appropriate for real-time  
 2427 display and results for multiple calibration strategies can be plotted. Thus, it is a very powerful  
 2428 representation, although it is limited to two or three variables, respectively for a 2D plot and  
 2429 a 3D plot (potentially three or four with a colour scale on the markers) to allow a graphical  
 2430 representation, and only one statistic per axis can be added with error bars.

#### 2431 4.4 Conclusion

2432 In this section, we investigated further issues regarding the evaluation of performances of *in*  
 2433 *situ* calibration strategies in terms of metrics and display. The conclusions are as follows:

- 2434 • In the case of an evaluation based on usual metrics, the study with multiple metrics is  
 2435 strongly recommended, with a careful regard on the purpose of each metric. Possible  
 2436 interdependence and underdetermination between them<sup>19</sup> may mislead interpretations.  
 2437 We recommend using an error model-based approach, which allows to better capture the  
 2438 correction brought by a calibration algorithm, for instance in terms of gain and offset.
- 2439 • The way metrics (or the parameters when an error model is used) are computed (time  
 2440 range on which the calculation is performed, periodicity of the computation) should be  
 2441 consistent with the details of the case study.

<sup>19</sup>Note that the set of metrics that is used in this work (or a subset) is not optimal itself regarding these questions. See [154]

- 2442 • Computing statistics of metrics over the whole network is advantageous for large sensor  
2443 networks, but may hide important information regarding to individual instrument (for  
2444 instance, a specific device degrading the overall performances of a given method due to  
2445 outliers)
- 2446 • Graphical visualisations of metrics (or parameters of an error model) can enable a better  
2447 analysis, notably with a target plot of the model's parameters when an error model with  
2448 two or three parameters is used.

## 2449 5 Sensitivity of the calibration algorithms to the specificities of the case 2450 study

2451 In this section, we discuss the influence of the solutions chosen for the implementation of  
2452 each step of the framework on the performances of the algorithms studied. First the function  
2453 used to model the concentration of pollutant is modified. Then, the number of instruments is  
2454 changed. Afterwards, the influence of the drift model is investigated. Finally, the impacts of  
2455 a change of parameters for the calibration algorithms are studied. For comparison purposes,  
2456 the metrics chosen in Section 3 are used, completed with the error model and the target plot  
2457 proposed in Section 4 when appropriate.

### 2458 5.1 Using a more realistic model of the true values

2459 The model used in Section 3 is simplistic and can legitimately raise questions on the perfor-  
2460 mances of each calibration strategy with a more representative model. In this section, consider  
2461 the Gaussian Plume model [61, 145], a refined model with meteorological parameters, taken  
2462 from a real-world data set. Otherwise, the space of study and its discretisation, the pollutant  
2463 source location and duration of study, the sensor network, the instrument drift model, and the  
2464 parameters of the calibration strategies are kept similar to the previous case study.

2465 We suppose that at each time step, the pollutant dispersion to be in steady state. Con-  
2466 centration  $C$  at instant  $t$  and position  $(x, y, z)$  for a wind direction following  $x$  is expressed  
2467 as:

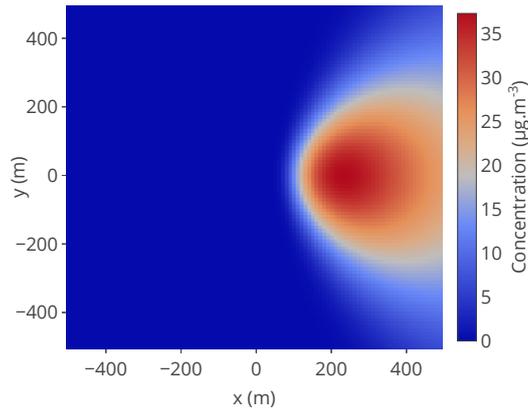
$$C(x, y, z, t) = \frac{Q}{4\pi V_w(t) \sigma_y \sigma_z} e^{-\frac{y^2}{4\sigma_y^2}} \left( e^{-\frac{(z-H)^2}{4\sigma_z^2}} + e^{-\frac{(z+H)^2}{4\sigma_z^2}} \right)$$

2468 with

- 2469 •  $\sigma_y$  and  $\sigma_z$ : respectively horizontal and vertical dispersion coefficients
- 2470 •  $Q$ : Emission rate at the source
- 2471 •  $V_w$ : Wind speed
- $H$ : Pollutant effective release.  $H = h_s + \Delta h(t)$  where

$$\Delta h(t) = \frac{1.6F^{\frac{1}{3}}x^{\frac{2}{3}}}{V_w} \text{ with } F = \frac{g}{\pi} D \left( \frac{T_s - T(t)}{T_s} \right)$$

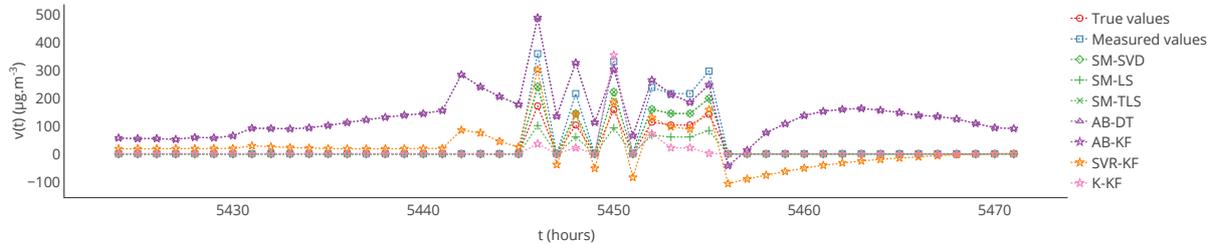
- 2472 –  $h_s$ : pollutant source height
- 2473 –  $g$ : gravity constant
- 2474 –  $D$ : volumetric flow
- 2475 –  $T_s$ : source temperature
- 2476 –  $T$ : ambient temperature



**Figure 3.5.1:** Example of concentration map used and based on the Gaussian plume model. The pollutant source is at coordinates  $(x_s, y_s) = (0, 0)$ ,  $h_s = 25\text{m}$ , the map is observed at  $z = 2\text{m}$ . Other parameters are :  $T = 25^\circ\text{C}$ ,  $V_w = 10\text{m s}^{-1}$ ,  $T_s = 30^\circ\text{C}$ ,  $g = 9.8\text{m s}^{-2}$ ,  $D = 1.9 \times 10^{-9}\text{m}^3\text{ s}^{-1}$ ,  $Q = 5 \times 10^{-3}\text{kg s}^{-1}$ ,  $\sigma_y = 1.36|x - x_s|^{0.82}$ ,  $\sigma_z = 0.275|x - x_s|^{0.69}$ . Wind direction is equal to  $0^\circ$  here.

2477 To allow for wind direction changes, the cylindrical coordinate system is used to rotate the  
2478 plume.

2479 Emission rate, temperature of the source and dispersion coefficients are supposed constant  
2480 over the simulation range. Ambient temperature, wind speed and wind direction are extracted  
2481 from METAR weather reports collected over a year. An example of pollution map obtained for a  
2482 given altitude  $z$  is represented in Figure 3.5.1.



**Figure 3.5.2:** True values, measured values and corrected values with the strategies considered for a particular instrument between  $t = 5424\text{h}$  and  $t = 5471\text{h}$  with the Gaussian plume model. Note that the profiles of the curves are very different from those of Figure 3.3.4 as instruments are not necessarily exposed to the plume of pollutant due to the wind direction. In this case it is exposed to the pollutant between  $t = 5445\text{h}$  and  $t = 5456\text{h}$ .

2483 Figure 3.5.2 is a temporal plot for a particular instrument like in Section 3. From this plot  
2484 we observe that the true signal appears to vary more quickly from high to low values. This is  
2485 due to the fact that only one source of pollutant is considered in the area, and that the wind  
2486 varies quickly from one direction to another in this dataset. Moreover, as we assumed that  
2487 the concentration at each time step could be represented following the equation of a stationary  
2488 Gaussian Plume, the pollutant quickly spreads at some points if the wind direction changes a lot.  
2489 In practice the pollutant may not disperse as fast.

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	13	6	17	6	38	21	0.98	0.00
<b>SM-SVD</b>	$2.30 \times 10^3$	$6.12 \times 10^3$	$2.28 \times 10^5$	$6.13 \times 10^5$	$1.39 \times 10^4$	$3.73 \times 10^4$	0.55	0.43
<b>SM-LS</b>	26	24	952	$2.42 \times 10^3$	93	138	0.85	0.34
<b>SM-TLS</b>	$2.30 \times 10^3$	$6.12 \times 10^3$	$2.28 \times 10^5$	$6.13 \times 10^5$	$1.39 \times 10^4$	$3.73 \times 10^4$	0.55	0.43
<b>AB-DT</b>	36	3	$2.11 \times 10^3$	529	54	7	0.62	0.19
<b>AB-KF</b>	36	3	$2.11 \times 10^3$	529	54	7	0.62	0.19
<b>SVR-KF</b>	249	357	$1.51 \times 10^4$	$2.15 \times 10^4$	270	380	0.54	0.18
<b>K-KF</b>	47	22	$2.18 \times 10^3$	$1.31 \times 10^3$	85	36	0.11	0.43

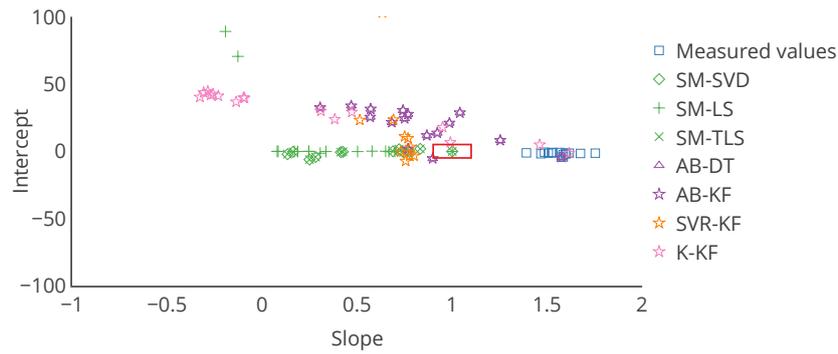
	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.56	0.08	-2	1	0.97	0.01
<b>SM-SVD</b>	-21.11	60.40	$2.49 \times 10^3$	$6.99 \times 10^3$	0.48	0.42
<b>SM-LS</b>	0.34	0.32	10	28	0.83	0.33
<b>SM-TLS</b>	-21.11	60.40	$2.49 \times 10^3$	$6.99 \times 10^3$	0.48	0.42
<b>AB-DT</b>	0.82	0.31	19	13	0.42	0.24
<b>AB-KF</b>	0.82	0.31	19	13	0.42	0.24
<b>SVR-KF</b>	0.83	0.19	234	357	0.32	0.19
<b>K-KF</b>	0.26	0.66	30	15	0.18	0.27

**Table 3.5.1:** Mean and standard deviation of each metric, computed on the entire time interval of drift, over the 16 nodes of the network with the Gaussian plume model. No calibration strategy allows to obtain less error than without calibration, even for SM-(SVD, LS, TLS) strategies that had the best performances in Section 3.

2490 Results with usual metrics and statistics on the parameters of the linear error model are  
 2491 presented in Table 3.5.1 on the entire network. From this table, we observe first that no calibration  
 2492 strategy provides less error than without calibration, even for SM-(SVD, LS, TLS), the strategies  
 2493 that had the best performances in Section 3. This is also the case in Table 3.5.2 for the same  
 2494 particular instruments considered in Section 3, despite apparently satisfying results according to  
 2495 MAE but not according to the slope and intercept notably. The results for each instrument are  
 2496 depicted in detail in Figure 3.5.3 with the target plot of the parameters of the error model for all  
 2497 the instruments.

2498 Using a more realistic model of dispersion obviously produces a more complex signal that  
 2499 may require a higher density of instruments to capture the phenomenon. Indeed, the subspace  
 2500 size for SM-(SVD, LS, TLS) is equal to 11 (it was 2 in Section 3.2.1). This is still fitting the  
 2501 oversampling condition required for these algorithms, but it could explain the actual results:  
 2502 there may not be enough measuring instruments to capture all the features of the measurand.  
 2503 The same reasoning applies for SVR-KF which is also model based. For the other strategies, for  
 2504 which the prediction of true values is based on neighbouring nodes, the quality of the results  
 2505 which were already poor in Section 3 is even worse.

2506 In conclusion, the influence of the model used to derive the true values of the measuring  
 2507 instruments is important. Considering cases on urban air pollution monitoring, the dispersion of



**Figure 3.5.3:** Target plot of the 16 nodes as a function of their slope and intercept in the error model, computed on the entire time interval of drift, for each calibration strategy and with the Gaussian plume model. The instruments are not all present, notably for those after the correction by SVR-KF. Axis were truncated to keep the plot readable. The red rectangle depicts the same requirement as in Figure 3.4.4 e.g. a slope in  $[0.9; 1.1]$  and an intercept in  $[-5; 5]$ . Few instruments are inside this area whatever the *in situ* calibration strategy considered.

2508 pollutant for cities with different street layout may lead to different concentration fields despite  
 2509 identical pollutant sources and experimental conditions. Therefore, even with an identical sensor  
 2510 network with instruments drifting in the same way, *in situ* calibration strategies could produce  
 2511 corrections of a variable quality.

	MAE		MAPE		RMSE		Pearson	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	12	12	19	13	36	35	0.98	0.98
<b>SM-SVD</b>	6	33	108	201	24	269	0.83	0.06
<b>SM-LS</b>	8	25	55	86	20	57	0.97	0.97
<b>SM-TLS</b>	6	33	108	201	24	269	0.83	0.06
<b>AB-DT</b>	37	32	$2.59 \times 10^3$	$1.61 \times 10^3$	51	45	0.57	0.79
<b>AB-KF</b>	37	32	$2.60 \times 10^3$	$1.61 \times 10^3$	51	45	0.57	0.79
<b>SVR-KF</b>	24	387	$1.31 \times 10^3$	$2.60 \times 10^4$	34	415	0.69	0.34
<b>K-KF</b>	27	63	818	$3.41 \times 10^3$	55	100	0.60	-0.25

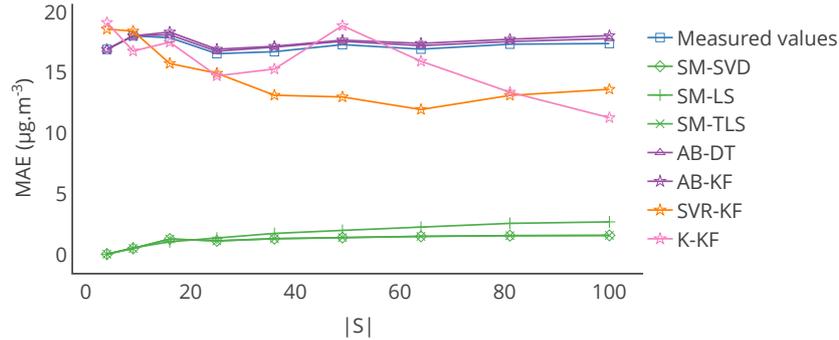
	Slope		Intercept		Score	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	1.75	1.47	-1	-2	0.96	0.97
<b>SM-SVD</b>	0.83	0.25	2	-6	0.68	0.00
<b>SM-LS</b>	0.58	0.17	0	-0	0.94	0.93
<b>SM-TLS</b>	0.83	0.25	2	-6	0.68	0.00
<b>AB-DT</b>	0.77	0.92	27	14	0.33	0.63
<b>AB-KF</b>	0.77	0.92	28	14	0.33	0.63
<b>SVR-KF</b>	0.78	0.98	3	382	0.48	0.12
<b>K-KF</b>	0.95	-0.26	18	42	0.36	0.06

**Table 3.5.2:** Values of the metrics, computed on the entire time interval of drift, for two particular instruments of the network  $s_1 = 2$  and  $s_2 = 10$ , with the Gaussian plume model. No calibration strategy allows to obtain less error than without calibration, even for SM-(SVD, LS, TLS) strategies that had the best performances in Section 3, despite apparently satisfying results according to MAE but not according to the slope and intercept notably.

## 5.2 Density of the sensor network

The density of the network of sensors is crucial to capture fine spatial features of the phenomenon. In this section, we investigate the impact on the performances of the calibration strategies.

First, we replay the initial case study (2D Gauss dispersion model) with regular grid networks with one instrument every  $\frac{1000}{n}$  m,  $n \in [2..10]$ . The size of the network,  $|S| = n^2$ , grows from 4 to 100. It was equal to 16 Section 3. All other parameters are kept.



**Figure 3.5.4:** Evolution of the mean of the MAE, computed on the entire time interval of study, of all the nodes of the network, as a function of the number of nodes  $|S|$ , for the 2D Gauss model

Evolution of the mean of the MAE of the network with  $|S|$  is shown in Figure 3.5.4. From it, we observe that increasing the number of instruments improves the results with strategies K-KF and SVR-KF. However, the results are not significantly impacted by  $|S|$  for strategies AB-(DT, KF) and SM-(SVD, LS, TLS) compared to the baseline case ( $|S| = 16$ ). It can be explained by the simplicity of the model used to derive the true values. Moreover, for these strategies, slightly better results are obtained with a smaller number of instruments than with  $|S| = 16$ . It seems that such a number (4 or 9 nodes) already enables capturing enough information to calibrate instruments for SM-(SVD, LS, TLS). For AB-(DT, KF), the prediction based on averaging values of neighbour instruments cannot be improved by increasing the number of instruments. Indeed, the network is blind, and instruments are all drifting at the same time, following the same model, with different values of drift. For SVR-KF, issues identified in Sections 3 and 4 seem to prevail despite an improvement of the performances with more instruments (but not with a lower number of nodes). However, in a general manner for K-KF, the more instruments are deployed, the more the kriging seems to provide accurate predictions. Results still have to be improved for  $|S| = 100$  though.

Thus, the number of instruments has an influence, though limited, on the results of the *in situ* calibration strategies. We also conjecture that the way instruments are deployed can have an influence, as it is crucial for an efficient reconstruction of the signal [20].

We also replay Gaussian Plume case study with a regular grid network of 100 instruments. We conjectured in Section 5.1 that increasing the density of instruments could improve on the poor calibration performances witnessed. Unexpectedly, our results contradict this intuition: the increased number of instruments does not improve the performances of the considered strategies.

Results with usual metrics and statistics on the parameters of the linear error model are presented in Table 3.5.3 for the entire network. From this table, we indeed observe that no calibration strategy allows obtaining less error than without calibration. In Table 3.5.4 results are quite equivalent to the previous ones in orders of magnitude. SM-(SVD, LS, TLS) strategies

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	13	7	17	6	38	25	0.98	0.00
<b>SM-SVD</b>	$1.04 \times 10^{12}$	$1.04 \times 10^{13}$	$1.04 \times 10^{14}$	$1.04 \times 10^{15}$	$6.32 \times 10^{12}$	$6.32 \times 10^{13}$	0.55	0.37
<b>SM-LS</b>	$4.77 \times 10^8$	$4.77 \times 10^9$	$4.77 \times 10^{10}$	$4.77 \times 10^{11}$	$2.91 \times 10^9$	$2.91 \times 10^{10}$	0.91	0.26
<b>SM-TLS</b>	$8.67 \times 10^{11}$	$8.67 \times 10^{12}$	$8.67 \times 10^{13}$	$8.67 \times 10^{14}$	$5.29 \times 10^{12}$	$5.29 \times 10^{13}$	0.54	0.37
<b>AB-DT</b>	37	3	$2.09 \times 10^3$	580	57	8	0.63	0.20
<b>AB-KF</b>	37	3	$2.10 \times 10^3$	578	57	8	0.63	0.20
<b>SVR-KF</b>	412	208	$2.68 \times 10^4$	$1.33 \times 10^4$	432	232	0.51	0.18
<b>K-KF</b>	39	21	$1.66 \times 10^3$	973	71	41	0.10	0.44

	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.55	0.07	-2	1	0.97	0.01
<b>SM-SVD</b>	$-1.01 \times 10^{11}$	$1.01 \times 10^{12}$	$1.22 \times 10^{12}$	$1.22 \times 10^{13}$	0.43	0.38
<b>SM-LS</b>	$-4.63 \times 10^7$	$4.63 \times 10^8$	$5.63 \times 10^8$	$5.63 \times 10^9$	0.89	0.26
<b>SM-TLS</b>	$-8.43 \times 10^{10}$	$8.43 \times 10^{11}$	$1.02 \times 10^{12}$	$1.02 \times 10^{13}$	0.42	0.38
<b>AB-DT</b>	1.02	0.71	17	14	0.43	0.25
<b>AB-KF</b>	1.02	0.71	17	14	0.43	0.25
<b>SVR-KF</b>	0.82	0.14	415	208	0.29	0.16
<b>K-KF</b>	0.19	1.38	23	11	0.20	0.30

**Table 3.5.3:** Mean and standard deviation of each metric, computed on the entire time interval of study, over the 100 nodes of the network with the Gaussian plume model. Again, no calibration strategy allows to obtain less error than without calibration.

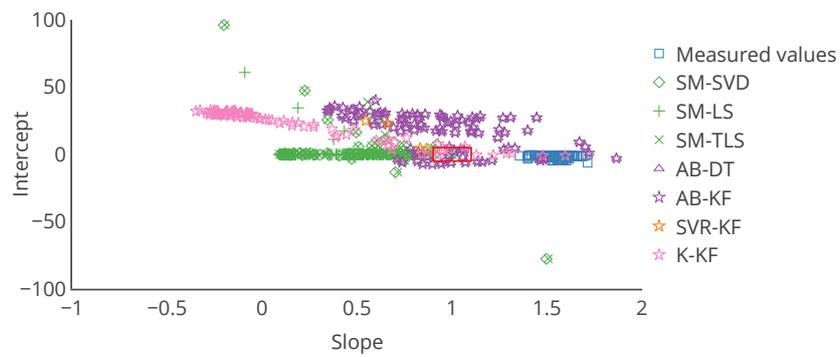
2545 have abnormal performances, notably due to outliers in the results for some instruments. Figure  
 2546 3.5.5 depicts the target plot of the parameters of the error model for all the instruments.

2547 For the strategies AB-DT, AB-KF and K-KF, which perform the prediction of true values  
 2548 based on neighbouring nodes, the fact that instruments are drifting altogether may have a more  
 2549 important effect than the density of the network. For SM-(SVD, LS, TLS), the subspace of the  
 2550 signal, which is at the heart of these algorithms, was chosen considering the same criteria on the  
 2551 sum of the explained variance ratios. In this case, 13 components of the PCA with the highest  
 2552 variance ratios were kept (11 in Section 5.1). Thus, setting that a subspace is satisfying if the  
 2553 sum of the explained variance ratios of its components is greater than a threshold may not be  
 2554 sufficient to define it properly. The same reasoning applies for SVR-KF which is also model  
 2555 based.

	MAE		MAPE		RMSE		Pearson	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	10	10	22	22	23	21	0.98	0.99
<b>SM-SVD</b>	5	3	36	25	13	6	0.92	0.99
<b>SM-LS</b>	6	4	42	30	11	6	0.99	1.00
<b>SM-TLS</b>	5	3	36	25	13	6	0.92	0.99
<b>AB-DT</b>	37	37	$2.16 \times 10^3$	$2.08 \times 10^3$	53	54	0.59	0.63
<b>AB-KF</b>	38	37	$2.17 \times 10^3$	$2.09 \times 10^3$	53	54	0.59	0.63
<b>SVR-KF</b>	373	309	$2.22 \times 10^4$	$1.75 \times 10^4$	374	311	0.63	0.68
<b>K-KF</b>	9	10	44	62	18	22	0.95	0.90

	Slope		Intercept		Score	
	$s_1$	$s_2$	$s_1$	$s_2$	$s_1$	$s_2$
<b>No calibration</b>	1.59	1.54	-1	-1	0.97	0.97
<b>SM-SVD</b>	0.84	0.87	-0	-0	0.85	0.98
<b>SM-LS</b>	0.71	0.84	-0	-0	0.99	0.99
<b>SM-TLS</b>	0.84	0.87	-0	-0	0.85	0.98
<b>AB-DT</b>	1.11	1.28	20	16	0.35	0.39
<b>AB-KF</b>	1.11	1.28	20	16	0.34	0.39
<b>SVR-KF</b>	0.89	1.02	375	309	0.39	0.46
<b>K-KF</b>	1.31	1.29	1	1	0.91	0.81

**Table 3.5.4:** Values of the metrics, computed on the entire time interval of study, for two particular instruments of the network  $s_1 = 2$  and  $s_2 = 10$ , with the Gaussian plume model. For these instruments, the results are quite equivalent to those of Table 3.5.2 in orders of magnitude



**Figure 3.5.5:** Target plot of the 16 nodes as a function of their slope and intercept in the error model, computed on the entire time interval of study, for each calibration strategy and with the Gaussian plume model. The instruments are not all present: axes were truncated to keep the plot readable. The red rectangle depicts the same requirement as in Figure 3.4.4 and 3.5.3 e.g. a slope in  $[0.9; 1.1]$  and an intercept in  $[-5; 5]$ . More instruments are inside this area compared to Figure 3.5.3 but most of them are outside whatever the *in situ* calibration strategy considered.

### 5.3 Instrument modelling

The modelling of the instruments used combines a simple drift model and the assumption that there are no other faults occurring to instruments than drifting. This hypothesis is not realistic, in particular when considering low-cost instruments. In this section, we introduce a more realistic drift model and evaluate the robustness of calibration strategies when the measurements are corrupted by other faults.

The model for the simulation of the true values, the sensor network and the parameters of the calibration strategies used in Section 3 are considered.

#### 5.3.1 Drift model

To determine the influence of the drift complexity, two models are considered, with  $t_{\text{start drift}}$  being the instant when the instrument begins to drift:

- Random gain and offset increase (**RGOI**): Gain  $G(s_i, t)$  and offset  $O(s_i, t)$  drift of instrument  $s_i$  are computed at each time step following:

$$G(s_i, t) = \begin{cases} 1 & \text{if } t < t_{\text{start drift}} \\ G(s_i, t-1) + \delta G(s_i, t) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\forall t, \delta G(s_i, t)$  drawn following  $\mathcal{U}(0, \delta G_{\text{max}})$

$$O(s_i, t) = \begin{cases} 0 & \text{if } t < t_{\text{start drift}} \\ O(s_i, t-1) + \delta O(s_i, t) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\forall t, \delta O(s_i, t)$  drawn following  $\mathcal{U}(0, \delta O_{\text{max}})$

with  $\delta G_{\text{max}}$  and  $\delta O_{\text{max}}$  being respectively the maximal gain and offset possible increase per time step.

- Continuous gain and offset increase (**CGOI**): A constant gain  $\delta G(s_i)$  and offset  $\delta O(s_i)$  increase of instrument  $s_i$  are drawn from uniform laws at the beginning of the drift and are added to respectively to the gain and offset of the instrument at each time step.

$$G(s_i, t) = \begin{cases} 1 & \text{if } t < t_{\text{start drift}} \\ G(s_i, t-1) + \delta G(s_i) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\delta G(s_i)$  drawn following  $\mathcal{U}(0, \delta G_{\text{max}})$

$$O(s_i, t) = \begin{cases} 0 & \text{if } t < t_{\text{start drift}} \\ O(s_i, t-1) + \delta O(s_i) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\delta O(s_i)$  drawn following  $\mathcal{U}(0, \delta O_{\text{max}})$

with  $\delta G_{\text{max}}$  and  $\delta O_{\text{max}}$  being respectively the maximal gain and offset possible increase per time step.

For both models, measured values are expressed following:

$$v(s_i, t) = G(s_i, t) \cdot v_{\text{true}}(s_i, t) + O(s_i, t)$$

In this case,  $\delta G_{\text{max}} = 6 \times 10^{-5}$ , and  $\delta O_{\text{max}} = 0.03$ .

2577 There is an offset in RGOI and CGOI drift models. The estimation of the offset could not be  
2578 carried out with SM-(SVD, LS, TLS) because the phenomenon does not have a null average on  
2579 the give time span of calibration. Hence, the offsets are expected to be wrong for these strategies.

2580 Results with usual metrics and statistics on the parameters of the linear error model are  
2581 presented in Table 3.5.5 and 3.5.6. From these tables and compared to Tables 3.3.1 and 3.4.1, we  
2582 observe that very different results can be obtained depending on the drift model and notably  
2583 for SM-(SVD, LS, TLS). For these strategies, the measured values with RGOI or CGOI drift  
2584 models are less improved by the calibration strategies even if we take into account that we  
2585 expected to have incorrect offsets. Indeed, the average slope is worse than the one in Table 3.4.1,  
2586 with an equivalent standard deviation. For both drift models, SVR-KF seems to improve the  
2587 measured values according to MAE, MAPE, RMSE and the Pearson correlation coefficient but  
2588 not according to the error model with a poor slope (0.76) and a high intercept (20) in average.  
2589 AB-(DT,KF) and K-KF give corrected values equivalent to the measured values with the RGOI  
2590 model. They are slightly improved if the measured values are built with the CGOI model and  
2591 more notably for K-KF.

2592 Thus, the drift model used to derive the measured values can drastically change the corrected  
2593 values obtained after calibration depending on the algorithm used.

	<b>MAE</b>		<b>MAPE</b>		<b>RMSE</b>		<b>Pearson</b>	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	56	0	273	46	65	0	0.65	0.00
<b>SM-SVD</b>	50	3	248	32	56	3	0.67	0.01
<b>SM-LS</b>	46	1	232	43	52	1	0.68	0.00
<b>SM-TLS</b>	50	3	248	32	56	3	0.67	0.01
<b>AB-DT</b>	56	0	274	48	65	0	0.65	0.00
<b>AB-KF</b>	56	1	276	54	65	1	0.64	0.00
<b>SVR-KF</b>	14	0	86	13	16	1	0.84	0.01
<b>K-KF</b>	56	1	276	54	65	1	0.64	0.00

	<b>Slope</b>		<b>Intercept</b>		<b>Score</b>	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.41	0.00	43	1	0.42	0.01
<b>SM-SVD</b>	1.24	0.01	42	2	0.45	0.01
<b>SM-LS</b>	1.18	0.04	40	0	0.46	0.00
<b>SM-TLS</b>	1.24	0.01	42	2	0.45	0.01
<b>AB-DT</b>	1.41	0.01	43	1	0.42	0.00
<b>AB-KF</b>	1.41	0.03	43	1	0.42	0.00
<b>SVR-KF</b>	0.76	0.02	20	1	0.71	0.02
<b>K-KF</b>	1.41	0.03	43	1	0.41	0.00

**Table 3.5.5:** Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the RGOI drift model. Results are very different with this model compared to Tables 3.3.1 and 3.4.1, notably for SM-(SVD, LS, TLS) even if we take into account that we expected to have incorrect offsets.

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	44	29	210	131	51	34	0.75	0.15
<b>SM-SVD</b>	33	21	162	99	38	24	0.78	0.14
<b>SM-LS</b>	26	16	133	89	31	19	0.77	0.13
<b>SM-TLS</b>	33	21	162	99	38	24	0.78	0.14
<b>AB-DT</b>	46	1	227	44	54	1	0.70	0.01
<b>AB-KF</b>	47	1	228	46	54	1	0.70	0.00
<b>SVR-KF</b>	14	0	85	13	16	0	0.84	0.01
<b>K-KF</b>	20	1	95	22	23	1	0.89	0.00

	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.34	0.20	33	23	0.58	0.23
<b>SM-SVD</b>	1.18	0.12	27	17	0.63	0.21
<b>SM-LS</b>	1.07	0.13	21	14	0.61	0.20
<b>SM-TLS</b>	1.18	0.12	27	17	0.63	0.21
<b>AB-DT</b>	1.35	0.03	35	1	0.49	0.01
<b>AB-KF</b>	1.35	0.03	36	1	0.49	0.00
<b>SVR-KF</b>	0.76	0.02	20	1	0.71	0.02
<b>K-KF</b>	1.18	0.02	14	1	0.79	0.01

**Table 3.5.6:** Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the CGOI drift model. Results are very different with this model compared to Tables 3.3.1 and 3.4.1, notably for SM-(SVD, LS, TLS) even if we take into account that we expected to have incorrect offsets.

### 5.3.2 Other faults

As stated in Chapter 1 Section 2.2, measuring instruments are subject to other faults but drift. Even if calibration algorithms are not capturing these faults, the quality of input data impacts the performances of calibration. In [50], it was shown that the outliers contained in the time series of measured values have a more important influence on the calibration results than the average quality of the data. As an early work on the topic, Ramanathan *et al.* [123] observed that the measured values were sometimes corrupted, even with an individual field calibration for each instrument performed with a mobile chemistry laboratory at the time of deployment. In this section, we investigate the influence of other faults on the calibration results.

Let us assume that in addition to a WGLI drift, all the instruments are undergoing after  $t_{\text{start drift}}$ :

**Noise** We model it as a random variable following a normal law. For each instrument  $s_i$ , the value of the noise  $\varepsilon(s_i, t)$  is drawn from  $\mathcal{N}(0, \varepsilon_{\text{max}})$  at each time step, with  $\varepsilon_{\text{max}} = 20$  here

**Spikes** In this study, a spike occurs depending on the random variable  $p_\psi(s_i, t)$  that follows  $\mathcal{U}(0, 1)$ . If  $p_\psi(s_i, t) < 0.05$ , a spike is added to the measured value  $v(s_i, t)$ . The value of the spike is equal to  $(\psi \cdot v)(s_i, t)$ , with  $\psi(s_i, t)$  following  $\mathcal{U}(-1, 1)$ .

Thus, the measured values of an instrument  $s_i$  are equal to:

$$v(s_i, t) = \begin{cases} G(s_i, t) \cdot v_{\text{true}}(s_i, t) & \text{if } t < t_{\text{start drift}} \\ G(s_i, t) \cdot v_{\text{true}}(s_i, t) + \varepsilon(s_i, t) & \text{if } t \geq t_{\text{start drift}} \text{ and } p_\psi(s_i, t) \geq 0.05 \\ (G(s_i, t) + \psi(s_i, t)) \cdot v_{\text{true}}(s_i, t) + \varepsilon(s_i, t) & \text{if } t \geq t_{\text{start drift}} \text{ and } p_\psi(s_i, t) < 0.05 \end{cases}$$

Results with usual metrics and statistics on the parameters of the linear error model are presented in Table 3.5.7. While SM-(SVD, LS, TLS) strategies worked very well for instruments suffering from drift only, their performances degrade significantly after introducing these faults in the measured values. For the other strategies, noise and spikes do not seem to degrade the results although the performances were fairly unsatisfying from the start like in Section 3. It is expected since this kind of fault is not considered by the calibration strategies. It is henceforth less a matter of performance than an issue of robustness of each method against abnormal values. In practice, one should try to identify and correct as many faults as possible, e.g. by filtering spikes and averaging noise, before applying an *in situ* calibration algorithm.

	MAE		MAPE		RMSE		Pearson	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	25	2	106	14	33	2	0.79	0.01
<b>SM-SVD</b>	60	50	198	144	82	70	0.23	0.33
<b>SM-LS</b>	28	6	93	10	33	7	0.35	0.57
<b>SM-TLS</b>	60	50	198	144	82	70	0.23	0.33
<b>AB-DT</b>	21	1	78	17	28	2	0.88	0.01
<b>AB-KF</b>	21	1	78	17	28	2	0.88	0.01
<b>SVR-KF</b>	18	0	92	15	22	0	0.75	0.01
<b>K-KF</b>	25	1	107	18	32	1	0.79	0.00

	Slope		Intercept		Score	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.62	0.07	-2	0	0.62	0.01
<b>SM-SVD</b>	0.18	0.69	-13	14	0.16	0.25
<b>SM-LS</b>	0.15	0.26	-0	1	0.43	0.08
<b>SM-TLS</b>	0.18	0.69	-13	14	0.16	0.25
<b>AB-DT</b>	1.63	0.04	-2	1	0.77	0.02
<b>AB-KF</b>	1.63	0.04	-1	1	0.78	0.02
<b>SVR-KF</b>	1.06	0.03	10	1	0.56	0.01
<b>K-KF</b>	1.61	0.04	-2	1	0.62	0.01

**Table 3.5.7:** Mean and standard deviation of each metric, computed on the entire time interval of study, over the 16 nodes of the network with the WGLI drift model, spikes and noise. The results with SM-(SVD, LS, TLS) are significantly influenced by the spikes and noise. The other strategies do not seem to provide more degraded results due to the noise and spikes although the performances were fairly unsatisfying from the start in Section 3.

#### 5.4 Parameters of calibration strategies

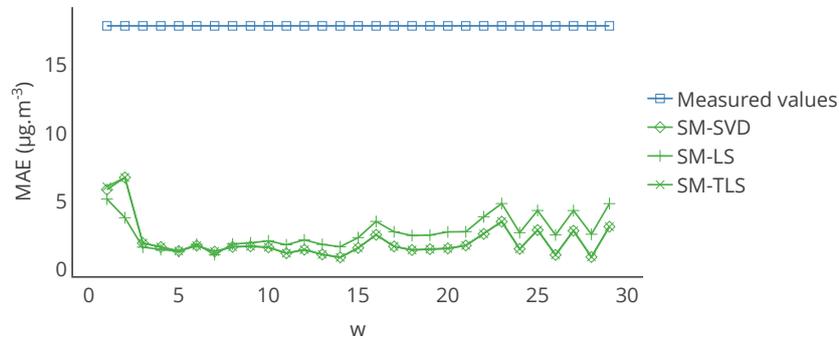
2620  
2621 Finally, the last step that can influence the results is the configuration of the calibration  
2622 strategies. Indeed, the strategies considered in Section 3 have multiple parameters.

2623 Multiple studies can be performed for each parameter of each strategy we considered. As an  
2624 example, we only investigate the case of SM-(SVD, LS, TLS). The influence of the frequency  
2625 of computation and of the duration of the time range used to select the measured values used  
2626 is studied. Note that this aspect was not treated in the original publication [7]. We consider  
2627 here that the period of computation and the duration of the time range for the selection of the  
2628 measured values are equal and denoted  $w$ . The same environment, sensor network and drift  
2629 model as in Section 3 are considered.

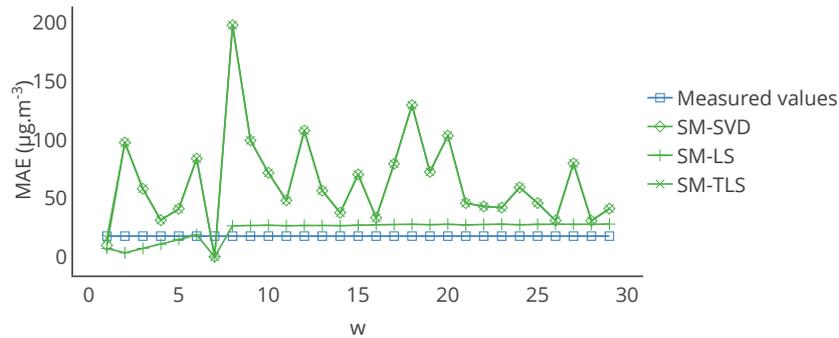
2630 Figure 3.5.6 represents the evolution of the mean of MAE over the network for the considered  
2631 strategies as a function of  $w$ . The best result is obtained with  $w = 7$ , which was our original  
2632 value in Section 3, considering the average of the error for the three strategies with the same  
2633  $w$ . Nevertheless, changing the value of  $w$  may lower the improvement brought by the *in situ*  
2634 calibration for low and high  $w$  but the corrected values are still closer to the true values compared  
2635 to the measured values.

2636 We also changed the criteria to determine the signal subspace. The sum of the explained  
2637 variance ratio of the components obtained by PCA must be greater  $1 - 1 \times 10^{-30}$  instead of  
2638  $1 - 1 \times 10^{-3}$  previously. This resulted in a change of the subspace with four components considered  
2639 instead of two. The same study with varying  $w$  was conducted and results are shown in Figure  
2640 3.5.7.

2641 We still observe that the best results are obtained for  $w = 7$ : for most of the values of  $w$ , the  
2642 algorithms degrade the quality of the measurements. This is even more significant for SM-SVD  
2643 and SM-TLS. We, however, do not claim that  $w = 7$  is an optimal parameter since we only  
2644 showcase this feature of the framework: it enables a full parametric optimisation of the methods.  
2645 Calibration methods are indeed very sensitive to the adjustment of several parameters and a  
2646 comparative study should include a parametric optimisation.



**Figure 3.5.6:** Evolution of the mean of MAE, computed on the entire time interval of study, over the network for the strategies SM-SVD, SM-LS and SM-TLS as a function of  $w$ . The best result is obtained for  $w = 7$ , which was our original value in Section 3, considering the average error of the three strategies for the same  $w$ .



**Figure 3.5.7:** Evolution of the mean of MAE, computed on the entire time interval of study, over the network for the strategies SM-SVD, SM-LS and SM-TLS as a function of  $w$  with a different signal subspace than in Figure 3.5.6. The best results are still obtained for  $w = 7$  but for most the values of  $w$ , the algorithms degrade the quality of the measurements and more significantly for SM-SVD and SM-TLS.

## 5.5 Summary of the results

2647

2648

2649

2650

In this section, we changed several parameters independently : the model of the phenomenon, the number of instruments, the faults introduced, notable the drift model, and the parameters of the calibration strategies for SM-(SVD, LS, TLS) that gave the best results previously.

Method	Baseline case	With Gaussian plume model	Varying number of nodes	Drift RGOI	Drift CGOI	Drift WGLI + faults	Varying calibration parameters
SM-SVD	Values improved	Values degraded	Values improved	Values not significantly improved or degraded	Values improved	Values degraded	Values not significantly improved or degraded
SM-LS	Values improved	Values degraded	Values improved	Values not significantly improved or degraded	Values improved	Values degraded	Values not significantly improved or degraded
SM-TLS	Values improved	Values degraded	Values improved	Values not significantly improved or degraded	Values improved	Values degraded	Values not significantly improved or degraded
AB-DT	Values not significantly improved or degraded	Values degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Case not studied
AB-KF	Values not significantly improved or degraded	Values degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Values not significantly improved or degraded	Case not studied
SVR-KF	Values not significantly improved or degraded	Values degraded	Values improved	Values improved	Values improved	Values not significantly improved or degraded	Case not studied
K-KF	Values not significantly improved or degraded	Values degraded	Values improved	Values not significantly improved or degraded	Values improved	Values not significantly improved or degraded	Case not studied

■ Values improved   
■ Values not significantly improved or degraded   
■ Values degraded   
■ Case not studied

**Figure 3.5.8:** Summary of the results observed by changing the parameters of the baseline case study which was built with a 2D Gaussian function to model the measurand, with the WGLI drift model applied to the measured values of the instruments, with 16 nodes in the network and the values indicated in Section 3 for the parameters of the *in situ* calibration algorithms. The improvements are judged relatively to results of the baseline case, recalled in the first column. Thus, this table does not indicate that the corrected values are very close to the true values when the measured values are improved.

2651

2652

The results obtained are reported in Figure 3.5.8. From a general perspective, we observed that:

2653

2654

- increasing the complexity of the model of the phenomenon drastically worsen the results for all the strategies

2655

2656

- changing the number of instruments between  $|S| = 4$  and  $|S| = 100$  in our case does not have a determining effect.

2657

- the model of the drift strongly impacts the results

2658

- the presence of other faults in the measured values can degrade the results

2659

- the (correct) adjustment of parameters of calibration strategies may have a significant effect

2660

2661

2662

Therefore, comparing *in situ* calibration strategies requires multiple precautions on how the corrected values were obtained and on what the measured values were based to provide fair and correct conclusions.

## 6 Discussion and conclusion

In this chapter, a framework for the simulation of sensor networks, which enables a systematic comparison of *in situ* calibration strategies with reproducibility, and scalability, was presented. We have showcased this methodology by comparing several calibration strategies for blind and static sensor networks. Our results provide several engineering insights on the design of a sensor network.

Based on the methodology described in Section 2, we have proposed a case study in Section 3 concerning blind and static sensor networks. Although we have shown that several strategies have better performances than the others, these results have to be tempered as the conclusions depends on how the evaluation is conducted. We have proposed solutions to conduct a balanced assessment of the performances in terms of metrics used in Section 4. It is still challenging for sensor networks with a high number of instruments that require multiple scale analysis. In addition, we have shown in Section 5 that results may not be the same depending on choices made during the design of the case study regardless of how the evaluation is carried out. We have developed variations of our initial case study by changing the model computing the true values of the measurand, the number of instruments, the fault model of the instruments and parameters of calibration strategies.

Our results highlight the dependence of the performances of *in situ* calibration algorithms to the case study. Besides, these performances seem to be often limited, even on the relatively simple environmental models discussed here. The latter explains why more complex and accurate environmental models were not tested here. Finding the good trade-off between simplicity and realism of the simulation models is important. At first sight, the goal could be to consider models as realistic as possible everywhere. However, the accuracy of simulation-based environmental models remains challenging to establish, and costly to run, while simplified models such as those used in the paper may be sufficient when the goal is to compare methods to single out the most promising ones among several for a predefined case study. This is also true for the models to simulate the measuring chain and the deployment of the instruments.

It allows to point out an additional advantage to the framework proposed in this paper: its suitability as a design tool for sensor networks. For instance, after having determined the number of instruments (and their positions) required to cover a given area with a static sensor network, the best *in situ* calibration strategy and its parameters could be identified among a set of algorithms applicable to this particular network. In such cases, however, a stronger focus on the accuracy of the environmental model should be put whenever possible. Beyond the quality of the models, the adjustment of the strategies is crucial for an objective evaluation. The reproducibility provided by the framework is an asset to conduct a thorough parametric optimisation and compare the best results of each calibration.

To conclude, we can point out the following engineering insights on the design of a sensor network and the evaluation of *in situ* calibration strategies.

- There is actually no method that is universally outperforming the other. The best calibration strategy to apply depends on the deployment of the instruments and the exploitation of the measurements.
- Even on the same scenario, two strategies can outperform each other depending on which metrics of performance is considered and the way they are computed. The metrics and their means of visualisation have to be chosen accordingly to the target exploitation of the sensor network, to focus on the relevant features. We also advise using an error model in order to get meaningful details on the performances.
- Increasing the density of the deployment of instruments does not always lead to better performances of a calibration strategy.

- 2711 • In some case, the quality of data can be degraded by a calibration strategy. Besides, other  
2712 faults than drift will happen, in particular when considering low-cost instruments with  
2713 low-quality electronics. To cope with robustness issues of the calibration strategy, it seems  
2714 relevant to process these errors upstream.

---

2715 The content of this chapter is based on the following publication:

2716 F. Delaine, B. Lebental and H. Rivano, "Framework for the Simulation of Sensor  
2717 Networks Aimed at Evaluating In Situ Calibration Algorithms," in *Sensors*, vol. 20,  
2718 no. 16, 2020, DOI: [10.3390/s20164577](https://doi.org/10.3390/s20164577).

2719 Codes and data files used to produce the results of Sections 3, 4 and 5 are available online  
2720 under licence AGPL 3.0 and ODbL 1.0 respectively in [37]



Chapter  
**4**

**Diagnosis of Drift Faults in Sensor Networks**

**Contents**

2726	1	Motivations . . . . .	85
2727			
2728	2	State-of-the-art of fault diagnosis applied to sensor networks . . . . .	86
2729	2.1	Overview of fault diagnosis for sensor networks through existing reviews	86
2730	2.2	Existing methods of diagnosis compatible with the concept of rendez-	
2731		vous addressing any fault . . . . .	87
2732	2.2.1	Introduction . . . . .	87
2733	2.2.2	Related works . . . . .	87
2734	2.2.3	Remarks . . . . .	89
2735	2.3	Positioning of the contribution . . . . .	89
2736	3	Definition of concepts for a drift diagnosis algorithm based on rendez-vous . .	89
2737	3.1	Validity of measurement results . . . . .	90
2738	3.2	Compatibility of measurement results . . . . .	90
2739	3.3	Rendez-vous . . . . .	91
2740	3.3.1	Formal and practical definitions . . . . .	91
2741	3.3.2	Compatible rendez-vous . . . . .	92
2742	4	Algorithm for the diagnosis of calibration issues in a sensor network . . . . .	92
2743	4.1	General idea . . . . .	92
2744	4.2	Procedure for the diagnosis of all the instruments in a sensor network	94
2745	4.3	Improvements and extensions of the presented algorithm . . . . .	95
2746	4.4	Conclusion . . . . .	95
2747	5	Application of the algorithm to a first case study . . . . .	97
2748	5.1	Definition of the case study . . . . .	97
2749	5.1.1	Sensor network . . . . .	97
2750	5.1.2	Instruments . . . . .	97
2751	5.1.3	True values . . . . .	98
2752	5.2	Configuration of the diagnosis algorithm . . . . .	99
2753	5.3	Definition of the true status of an instrument . . . . .	99
2754	5.4	Metrics for the evaluation of performances of the diagnosis algorithm .	99
2755	5.5	Results . . . . .	101

2756	5.5.1	Observations from the results of different instruments . . . . .	101
2757	5.5.2	Overall appreciation . . . . .	101
2758	5.5.3	Evolution over time of the metrics . . . . .	102
2759	5.6	Explanations of false results . . . . .	107
2760	5.7	On the parameters of the diagnosis algorithm . . . . .	108
2761	5.8	Conclusion . . . . .	110
2762	6	On the assumption regarding the top-class instruments being always predicted	
2763		as non-faulty . . . . .	<b>111</b>
2764	6.1	Theoretical discussion . . . . .	111
2765	6.2	Case study with the instrument of class $c_{max}$ drifting . . . . .	111
2766	6.3	Conclusion . . . . .	113
2767	7	Means to reduce false results . . . . .	<b>113</b>
2768	7.1	Keep the predicted status of instruments unchanged once they are	
2769		predicted as faulty . . . . .	113
2770	7.2	Alternate definition for rates used to decide of the status of the instruments	113
2771	7.3	Adjustment of the maximal tolerated values for the different rates used	
2772		to determine the statuses of the instruments . . . . .	115
2773	7.4	Conclusion . . . . .	116
2774	8	Adjustment of the minimal size required for a set of valid rendez-vous to allow	
2775		a prediction between the statuses faulty and non-faulty . . . . .	<b>117</b>
2776	8.1	Algorithm determining an upper boundary for $ \Phi_v _{min}$ . . . . .	117
2777	8.2	Application to the case study . . . . .	120
2778	8.3	Conclusion . . . . .	122
2779	9	Sensitivity of the algorithm to changes in the case study . . . . .	<b>122</b>
2780	9.1	Influence of the values of drift . . . . .	122
2781	9.2	Influence of the true values . . . . .	123
2782	9.3	Influence of the model used for the true values . . . . .	124
2783	9.4	Influence of the density of instruments . . . . .	125
2784	9.5	Influence of other faults . . . . .	131
2785	9.6	Conclusion . . . . .	131
2786	10	Combination with a simple calibration approach . . . . .	<b>133</b>
2787	11	Conclusion . . . . .	<b>138</b>
2788			
2790			

---

---

## 1 Motivations

Across the study of all the *in situ* calibration strategies in Chapter 2, and particularly regarding those considered in Chapter 3, most strategies perform the calibration of all instruments without evaluating first whether they actually need it. Moreover, we observed the algorithms may not improve, or may even degrade, the quality of the measurements. Two guidelines for *in situ* calibration algorithms can be established based on that. First, instruments providing measurement results agreeing with their specifications, should not be affected by an *in situ* calibration strategy. Secondly, the need for a calibration operation must be explicitly determined. This is particularly important if the execution of the algorithm is expensive, e.g. in terms of energy consumption in the case of autonomous nodes. It would add more confidence in the algorithms, namely it could inform it recalibrated an instrument because a criterion on its measurement results was met.

Some calibration algorithms are actually providing the identification of instruments needing a calibration, notably in [87, 168]. In these two publications, the algorithm determines first whether an instrument has drifted since the last time step. If it is the case, it gives the identity of the instrument that has drifted and correct it with the help of all the other instruments. However, if more than one instrument is drifting at a time, the algorithm raises an error. This is problematic as such a case is very likely in practice.

In the same vein of assumptions that are difficult to meet in practice, multiple strategies are designed based on the hypothesis that the sensor network is "dense". It means that each instrument is within few tens of meters to one or several nodes of the network. The need for this assumption to be verified can be explained as follows. To correct drifting instruments, their correct values—or the values of the drift—must be estimated. To do so, mainly two approaches are considered: the estimation from values measured by neighbouring instruments or the estimation from a model of the measurand. This is the case whether the network is static or not and blind or not.

Regarding the first approach, instruments should be physically close enough to have comparable values between them. This is something that can be easily achieved in practice and the reason why the assumption of a dense sensor network is often proposed. However, if the distance between the instrument to recalibrate and the instruments used to estimate its value is too important compared to the characteristic spatial length of variation of the measurand, and if it cannot be reduced, it may not be possible to have an accurate estimation with the desired uncertainty. Thus, as complex quantities are targeted in environmental monitoring, comparisons may be feasible for mobile sensor networks, but it is less likely to be the case for static sensor networks. Indeed, with mobile instruments can sometimes meet whatever the number of nodes in the network. For static ones, the number of instruments may have to be significantly increased depending on the application so that values of different devices may be compared but with an associated cost that may be significant.<sup>20</sup>

Therefore, a model-based approach for estimating the correct values of instruments may be preferred when meeting the required density of instruments is not possible, e.g. with static instruments close enough or with mobile nodes meeting often. With a model-based approach, it is assumed possible to model satisfyingly the measurand to estimate its value at several locations, with the help of *a priori* information and sometimes with measured values, preferably from

---

<sup>20</sup>The case where neighbouring instruments are used to evaluate the values of an instrument could be extended to the case where the nodes of the network embed several instruments with the same characteristics and measuring the same quantity. In this case, the set of instruments of a node measuring the same quantity can be seen as a measuring system, which would be equivalent to having a single instrument of a higher quality. Indeed, it can be legitimately expected that the average of the values of these instruments is more accurate than their values considered individually. This comes down to the case described in this paragraph. For the sake of clarity, this particular situation is not discussed in detail in this manuscript.

reference instruments. Acquiring such information may not be easy or may require a long training period to capture enough features of the phenomena. The latter may not be compatible with the lifespan of the instruments, especially if they are low cost. The deployment of expensive reference instruments during the deployment phase is a solution but has a major cost impact, notably if a high number of instruments is required. Additionally, using a model for the measurand raises the question of the interest of deploying the sensor network itself. If the model is able to derive at the required spatiotemporal resolution the values of the considered phenomenon, even with the help of a few high-quality instruments, this may be sufficient for most applications.

Overall, these elements point out that most approaches for *in situ* calibration can be challenged by the fact that the reference values used for calibration, e.g. standard values, need to meet certain conditions. They are not necessarily met in *in situ* calibration algorithms. The same precautions should be taken regarding the identification of instruments that have drifted, whatever the approach chosen to determine them.

In this chapter, we investigate the feasibility of drift detection for measuring instruments with the help of measurement results from other devices. The concept of **rendez-vous** is used to achieve this goal and is formally defined in Section 3, based on existing works on the subject [40, 132]. The study of these rendez-vous is expected to help to determine if an instrument needs to be recalibrated as long as one of the instruments involved in the rendez-vous can be considered as a standard, e.g. an instrument of which results can legitimately be trusted. To remain the most generic as possible, **no assumption is made for this algorithm on the type of sensor network, on the density of instruments and on the measurand** such as *a priori* known information regarding it.<sup>21</sup>

Thus, a formal definition of the concept of rendez-vous is proposed first based on existing works. Which values of instruments may be compared to each other in practice is briefly discussed. Then, a diagnosis algorithm, that is a way to identify the instruments that need a correction of their calibration relationship, is provided. The approach is based on rendez-vous. Afterwards, a case study is presented to illustrate the application of the diagnosis algorithm. The means to improve the results, in terms of false results, are discussed followed by guidelines on the adjustment of the parameters of the algorithm. The sensitivity of the algorithm to the parameters of the case study is studied. Finally, the algorithm is combined to a simple calibration strategy showing promising results. Before moving to these contributions, a literature review is conducted regarding fault diagnosis applied to sensor networks and in particular regarding the detection of drift.

## 2 State-of-the-art of fault diagnosis applied to sensor networks

### 2.1 Overview of fault diagnosis for sensor networks through existing reviews

In the last decade, while sensor networks gained in popularity, the question of their dependability became a major subject. In particular, multiple contributions were made on fault diagnosis as reported in various surveys [98, 106, 114, 185].

In these publications, the research works are classified by the way the diagnosis is performed. First, the place where the decision is computed, e.g. if it is a centralised, decentralised or hybrid approach, is considered. Then, the general mathematical approach used for the diagnosis (testing, comparisons, majority voting, statistics, probabilities, machine learning algorithms, fuzzy logic, clustering...) is studied.

Faults are generally characterised by "hard" or "soft" and as "permanent", "transient" or "temporary"[98, 106, 185]. Only Muhammed *et al.* [114] used a taxonomy similar to the one of Ni *et al.* [115] that inspired the one used in this manuscript. In particular, these authors report

---

<sup>21</sup>It could be for instance an actual correlation between the true values at two very distant points that could be exploited in a specific situation.

2880 that the existing diagnosis algorithms are often targeting multiple faults. For those addressing  
2881 the drift fault<sup>22</sup>, stuck-at or out-of-range faults are also detected by the same algorithm in the  
2882 15 associated references they cited.

2883 In the following surveys [98, 114, 185], the diagnosis approaches reported concern mainly  
2884 static sensor networks. Mahapatro *et al.* [98] listed eight algorithms that could be applied  
2885 to static and mobile sensor networks.<sup>23</sup> Zhang *et al.* [185] reported two references explicitly  
2886 dealing with mobile sensor networks [1, 28]. They also stated that approaches designed for static  
2887 networks behave poorly when applied to mobile ones. This indicates that algorithms exploiting  
2888 specificities related to networks with mobile nodes in the diagnosis approach have not been  
2889 deeply investigated so far.

2890 In addition, other subjects such as energy consumption, communication range needed,  
2891 communication failures, or the amount of required neighbour nodes are taken into account in the  
2892 surveys on fault diagnosis for sensor networks [98, 106, 114, 185]. Zhang *et al.* [185] underline  
2893 it is challenging to design a diagnosis approach having satisfying performances or properties  
2894 regarding all the features of interest for such an algorithm, e.g. for instance diagnosing faults  
2895 well with low energy consumption, few communications between nodes and so on.

## 2896 **2.2 Existing methods of diagnosis compatible with the concept of rendez-vous** 2897 **addressing any fault**

### 2898 **2.2.1 Introduction**

2899 The methods used to carry out the diagnosis of any fault in sensor networks are diverse in  
2900 terms of concepts and tools on which they are based on. According to the surveys cited [98,  
2901 106, 114, 185], different main ideas come out like the use of comparisons between instruments,  
2902 the use of statistics and probabilities, as well as the use of machine learning models (regressors,  
2903 classifiers...). As the goal in this chapter is to propose a diagnosis algorithm for sensor networks  
2904 based on rendez-vous between the nodes, we focus here on approaches compatible with this idea,  
2905 e.g. it consists at least in comparison of values measured by different instruments.

### 2906 **2.2.2 Related works**

2907 Chen *et al.* [29] introduced a faulty sensor detection algorithm consisting of four steps of  
2908 evaluation. First, each instrument compares its measured value at the instant of the diagnosis  
2909 with the measured values of its neighbours. If the difference between two instruments is higher  
2910 than a first threshold, the deviation of the difference since the last comparison is computed. If it  
2911 is higher than a second threshold, the test between the two instruments is marked as positive.  
2912 In a second step, each instrument determines if it is likely good or faulty depending on the  
2913 number of positive tests and its number of neighbours. This information is shared between the  
2914 instruments so that based on its neighbours that are likely good, an instrument determines if  
2915 it is actually good and share this result. Then, for the remaining undetermined instruments,  
2916 their statuses are decided if all their neighbours are non-faulty and if all the initial tests gave  
2917 the same results. If ambiguities remains, they are removed based on the likely statuses of the  
2918 instruments. Xu *et al.* [179] proposed an extension of this work dealing with the particular case  
2919 of tree-like networks to reduce the number of communications. To avoid intermittent faults, they  
2920 also proposed to compute the initial test result based on multiple comparisons of values between  
2921 two instruments, instead of a unique comparison. In the same way, Saha *et al.* [128] developed  
2922 a similar algorithm but with comparisons for multiple quantities (measurand and remaining

---

<sup>22</sup>In the publication of Muhammed *et al.* [114], calibration, gain, and offset faults correspond to what is notably a drift fault in this thesis.

<sup>23</sup>Note that the subject of mobile sensor networks was new at the time when Mahapatro *et al.* [98] published their survey.

2923 energy).

2924 Ssu *et al.* [142] proposed an approach to diagnosis faults between sources and sinks. The  
2925 main idea is to send a request through two paths and to compare the results obtained at the  
2926 sink node. If the results are different, then at least one faulty path exists. The algorithm tries to  
2927 identify it with the help of a third path and a majority voting procedure, but the faulty paths  
2928 cannot always be identified. Such a method is particularly relevant to diagnose communication  
2929 or hardware-related faults.

2930 Lee *et al.* [83] presented a distributed algorithm to isolate faulty nodes. Initially, the nodes  
2931 are all assumed as faulty. A comparison is made between the measured values of neighbour nodes  
2932 and if the result is higher than a threshold, then the test is positive. If less than a predefined  
2933 number of positive tests has been obtained, or if the test with a non-faulty node is negative, then  
2934 the diagnosed instrument is non-faulty. The algorithm works either based on single comparisons  
2935 between neighbours or with comparisons repeated multiple times.

2936 Mahapatro *et al.* [97, 99] introduced a clustering-based diagnosis algorithm. The clustering  
2937 part is used for the definition of the neighbours around cluster heads, the cluster heads being  
2938 the instruments with the highest residual energy levels. They also compare the measured values  
2939 between instruments of a cluster and a majority voting strategy is used to determine the state of  
2940 the nodes.

2941 Chanak *et al.* [28] developed a comparison-based scheme using a reference mobile sink node  
2942 moving between the static nodes of the network. When it is close to a node, several diagnoses are  
2943 performed to detect hardware and software faults. This also enables a low consumption of energy  
2944 for the transmission of the measurement results as it is no longer necessary to communication  
2945 with a distant gateway. The core of this contribution lies in the determination of an optimal  
2946 path to meet with each node.

2947 Luo *et al.* [92] proposed an approach using the concept of average consensus. Each instrument  
2948 estimates first its status regarding the average consensus measured value build with the values of  
2949 its neighbours. Each instrument also estimates the statuses of its neighbours. Then, the decisions  
2950 of all the instruments are merged to make the final decision.

2951 In reaction to contributions computing the average value of the neighbours of an instrument  
2952 by weighting their values with the inverse of the distance between them, Xiao *et al.* [178] argued  
2953 that the distance does not control alone the relationship between the values of two instruments,  
2954 notably if a closer one is actually faulty. Thus, they propose to take into account an indicator of  
2955 trustworthiness computed for each node of the network. This confidence value is used in voting  
2956 procedures that are used to determine the status of instruments.

2957 Ji *et al.* [69] also developed their algorithm around a weighted average of the values measured  
2958 by an instrument. The weights are representing a confidence level associated to each instrument.  
2959 The difference between the measured value of an instrument and the average is compared to a  
2960 threshold. If it is greater than the threshold, the confidence level of the instrument is decreased  
2961 and once it reaches zero, the instrument is reported as faulty.

2962 This idea of trust between instruments has been extensively studied and extended, as did  
2963 for instance Jiang *et al.* [70]. Their robust trust model is based on direct comparisons but also  
2964 on third parties and recommendations, with concerns about the security of the communications  
2965 between the instruments. Their trust framework is also adapted for mobile sensor networks.  
2966 Wang *et al.* [164] exploited Petri nets to introduce a trust-based formal model aimed at detecting  
2967 faults in sensor networks.

2968 Sharma *et al.* [134] also exploited the idea of confidence in a method similar to the one of  
2969 Chen *et al.* [29]. The main difference with Chen *et al.* is that in a first step each instrument  
2970 analyses its own behaviour and that a confidence level is associated to the determined statuses.

2971 Feiyue *et al.* [52] proposed to combine the information resulting of a self-evaluation and  
2972 of comparisons with the neighbours of each instrument. In a first step, the reliable nodes are

2973 determined through a majority voting procedure. Then, various metrics are computed and  
2974 combined to make the decisions for the instruments that were not considered reliable in the first  
2975 step.

### 2976 **2.2.3 Remarks**

2977 In this section, the way the computation is conducted (centralised, decentralised, hybrid) was  
2978 not discussed on purpose. Indeed, this topic impacts the number of communications required and  
2979 the energy consumption of the method more than the actual result of the diagnosis procedure.

2980 Another aspect is how the faults are managed after detection in a diagnosis algorithm. There  
2981 are different visions. Considering the diagnosis of outliers for instance, Ottosen *et al.* [116]  
2982 adopted a gap-filling strategy. Fang *et al.* [51] preferred a fault removal approach because this  
2983 procedure is a data processing carried out before the calibration of the instruments of the network.  
2984 As shown for the calibration of measuring instruments in Chapter 3, it may also be preferable to  
2985 correct measurement results prior to the diagnosis of drift if any other fault may be present.

2986 Finally, in the works presented here, the diagnosis is performed in most cases based on single  
2987 values from a given set of instruments measured at the same time. Taking a decision on a short  
2988 time range may not be correct, especially if drift is targeted. Indeed, drift is a fault increasing  
2989 little by little and thus a diagnosis on a long sequence of measured values over time could be  
2990 more appropriate.

### 2991 **2.3 Positioning of the contribution**

2992 From this review, we observed that the literature on fault diagnosis for sensor networks is rich  
2993 and abundant. Static sensor networks have been widely studied while mobile sensor networks  
2994 have been partially covered. Multiple types of faults are addressed, concerning various aspects  
2995 of measuring systems, but drift is not specifically targeted in existing works. Various methods  
2996 have been used to tackle the challenge of fault diagnosis, notably based on comparisons between  
2997 instruments.

2998 However, the validity of the comparisons is rarely discussed from a metrological perspective.  
2999 It would be valuable to develop a new approach with considerations on the quality of the  
3000 measurement results. Moreover, exploiting the concept of rendez-vous allows to take into account  
3001 that it happens in a spatiotemporal vicinity. This concept is more sophisticated than the simple  
3002 idea of comparison of measurement results.

3003 In addition, a diagnosis algorithm not requiring any assumptions on the type of sensor network  
3004 and its structure would be interesting. Indeed, the existing approaches are often targeting static  
3005 sensor networks, which may imply a dense network in the case of a comparison-based strategy,  
3006 and rarely networks with mobile nodes. Regarding rendez-vous, mobility is an important feature  
3007 that can facilitate their occurrence. Because mobile nodes can meet spatially with other mobile  
3008 or static nodes, a diagnosis approach based on these meeting points would allow more meaningful  
3009 comparisons of values if a very high density of measuring instruments is not feasible. In this  
3010 way, if the algorithm works whether the sensor network is static or has mobile node leaves to the  
3011 users the choice of the most appropriate type of sensor network for their application.

## 3012 **3 Definition of concepts for a drift diagnosis algorithm based on rendez-** 3013 **vous**

3014 Before the presentation of an algorithm for the diagnosis of drift in sensor networks, this section  
3015 introduces the definitions of the concepts of validity and compatibility of measurement results,  
3016 and of rendez-vous. A sensor network composed of a set of instruments  $S$  is considered here and  
3017 in the following sections without any other assumption on its type and on the characteristics of  
3018 its nodes.

### 3.1 Validity of measurement results

In Chapter 1 Section 1.2, we recalled that measuring instruments cannot exactly provide true values. In addition, we indicated they may be different accuracy class, e.g. of different qualities, in Chapter 1 Section 3.

On top of these ideas, we introduce here the notion of the **validity of measurement results**. For instance, a measuring instrument cannot measure all the values of a quantity. It has a bounded measuring interval which is the set of values of a quantity "*that can be measured by a given measuring instrument or measuring system with specified instrumental measurement uncertainty, under defined conditions*" [14]. In addition, it cannot work properly under any operating conditions. Therefore, a measurement result with a value outside an instrument's measuring interval or which has been obtained outside of the normal conditions of operation of a device, may not be **valid**.

The conditions under which measurement results are valid are specific to each type of instrument. In practice, they are usually provided in the technical documentation or it is possible to determine them experimentally. In this chapter, the conditions under which results are considered as valid are explained in the case study.

The notation  $M^*(s_i, (t, \Delta t))$  refers to the set of measurement results for  $s_i$ , over the time range  $[t - \Delta t; t]$  that are metrologically valid, e.g.  $v(s_i, t)$  in the measuring interval and so on depending on the definition of what a valid result is metrologically speaking.

### 3.2 Compatibility of measurement results

**Definition 22.** Consider  $s_i \in S$ . A measurement result  $m(s_i, t)$  is **compatible with true value** if  $v_{true}(s_i, t) \in [v(s_i, t) - \Delta v(s_i, t); v(s_i, t) + \Delta v(s_i, t)]$ .

**Definition 23.** Consider  $s_i \in S$ . A measurement result  $m(s_i, t)$  is **upper non-compatible with true value** if  $v_{true}(s_i, t) < v(s_i, t) - \Delta v(s_i, t)$ .

**Definition 24.** Consider  $s_i \in S$ . A measurement result  $m(s_i, t)$  is **lower non-compatible with true value** if  $v_{true}(s_i, t) > v(s_i, t) + \Delta v(s_i, t)$ .

The set of measurement results compatible with true values of  $s_i$  obtained during  $[t - \Delta t; t]$  is noted  $M^{\approx}(s_i, (t, \Delta t))$ .  $M^+(s_i, (t, \Delta t))$  and  $M^-(s_i, (t, \Delta t))$  are defined for upper and lower non-compatible measurement results respectively. It can be noticed that

$$M(s_i, (t, \Delta t)) = M^{\approx}(s_i, (t, \Delta t)) \cup M^+(s_i, (t, \Delta t)) \cup M^-(s_i, (t, \Delta t))$$

These definitions can be extended for the comparison of measurement results between different instruments.

**Definition 25.** Consider  $s_i$  and  $s_j \in S$ .  $m(s_i, t)$  is **compatible with**  $m(s_j, t')$  if  $v(s_j, t') - \Delta v(s_j, t') \leq v(s_i, t) - \Delta v(s_i, t) \leq v(s_j, t') + \Delta v(s_j, t')$  or  $v(s_j, t') - \Delta v(s_j, t') \leq v(s_i, t) + \Delta v(s_i, t) \leq v(s_j, t') + \Delta v(s_j, t')$ . It is noted  $m(s_i, t) \approx m(s_j, t')$ .

**Definition 26.** Consider  $s_i$  and  $s_j \in S$ .  $m(s_i, t)$  is **upper non-compatible with**  $m(s_j, t')$  if  $v(s_i, t) - \Delta v(s_i, t) \geq v(s_j, t') + \Delta v(s_j, t')$ . It is noted  $m(s_i, t) > m(s_j, t')$ .

**Definition 27.** Consider  $s_i$  and  $s_j \in S$ .  $m(s_i, t)$  is **lower non-compatible with**  $m(s_j, t')$  if  $v(s_i, t) + \Delta v(s_i, t) \leq v(s_j, t') - \Delta v(s_j, t')$ . It is noted  $m(s_i, t) < m(s_j, t')$ .

The sets of measurement results of  $s_i$  obtained during  $[t - \Delta t; t]$  associated to these definitions are respectively noted  $M^{\approx}(s_i \rightarrow s_j, (t, \Delta t))$ ,  $M^+(s_i \rightarrow s_j, (t, \Delta t))$  and  $M^-(s_i \rightarrow s_j, (t, \Delta t))$ .

More restrictive criteria can be set to determine if measurement results are compatible either with their associated true values or with results of other instruments. For instance, it

3058 could be a minimal level of overlapping between  $[v(s_i, t) - \Delta v(s_i, t); v(s_i, t) + \Delta v(s_i, t)]$  and  
3059  $[v(s_j, t') - \Delta v(s_j, t'); v(s_j, t') + \Delta v(s_j, t')]$ . However, the influence of these definitions are not  
3060 studied afterwards as it is more a matter of specification for the quality of measurements than a  
3061 matter of diagnosis.

### 3062 3.3 Rendez-vous

#### 3063 3.3.1 Formal and practical definitions

3064 A rendez-vous between two instruments is defined as follows.

3065 **Definition 28** (Rendez-vous). Two instruments  $s_i$  and  $s_j \in S$  are considered in rendez-vous  
3066 when they are in a spatiotemporal vicinity so that their measurement results can be compared. In  
3067 practice, it means for both their measurement results  $m(s_i, t)$  and  $m(s_j, t')$  that the instruments  
3068 were spatially close enough and the difference between the instants of measurement  $t$  and  $t'$  was  
3069 small enough to actually measure the same quantity value.

3070  $s_i$  being in a rendez-vous at  $t$  with  $s_j$  is noted  $\varphi(s_i \rightarrow s_j, t)$ . Respectively  $s_j$  being in a  
3071 rendez-vous at  $t'$  with  $s_i$  is noted  $\varphi(s_j \rightarrow s_i, t')$ .

3072 The concept of "close enough" instruments will be explained later in the present section.  
3073 Regarding the instants of measurement, two different ones are defined,  $t$  and  $t'$ , because  $s_i$   
3074 and  $s_j$  may not have the same measurement frequency or may not have synchronous clocks.  
3075 Consequently, in general, the measurement results of two instruments in rendez-vous are not  
3076 obtained exactly at the same time.

3077 The set of the rendez-vous encountered by  $s_i$  with  $s_j$  during  $[t - \Delta t; t]$  is noted  $\Phi(s_i \rightarrow$   
3078  $s_j, (t, \Delta t))$ . By extension,  $\Phi(s_i \rightarrow S, (t, \Delta t))$  is the set of rendez-vous between  $s_i$  and any other  
3079 instrument of  $S$  during  $[t - \Delta t; t]$ .

3080 First of all, with this definition of a rendez-vous, there is no reason that any of the instrument  
3081 should be mobile or static. As long as the conditions regarding the spatiotemporal vicinity are  
3082 respected, rendez-vous can happen.<sup>24</sup> Thus, the use of the concept of rendez-vous in practice  
3083 requires the definition of what an acceptable spatiotemporal vicinity is for a rendez-vous.

A first possible definition can be based on a maximal distance  $\Delta l_\varphi$  allowed between the  
instruments and on a maximal acceptable time difference  $\Delta t_\varphi$  between the measurement results  
of the two instruments. Therefore,  $s_i$  is in rendez-vous with  $s_j$  at  $t$  if

$$\exists t' \in \mathbb{R}^+ \text{ such as } \exists m(s_j, t') \text{ and } |t - t'| \leq \Delta t_\varphi \text{ and } |l(s_i, t) - l(s_j, t')| \leq \Delta l_\varphi$$

3084 with  $l(s_i, t)$  and  $l(s_j, t')$  being the positions of  $s_i$  at  $t$  and  $s_j$  at  $t'$ . The temporal condition  
3085 allows multiple rendez-vous with a single instrument for a same measurement result if  $t'$  is  
3086 not unique. In this case,  $t'$  can be chosen as the value reaching the minimum of  $|t - t'|$  or of  
3087  $|l(s_i, t) - l(s_j, t')|$ . This choice of a unique measurement result for  $s_j$  is necessary to ensure that  
3088 the measurement result of  $s_i$  is compared to the result the most probable to have been obtained  
3089 from the same quantity value.

3090 In practice,  $\Delta t_\varphi$  can be related to the characteristic time of variation of the measurand  
3091 whereas  $\Delta l_\varphi$  may be related to the spatial homogeneity of this quantity.

Another definition of the spatial closeness can be designed based on the concept of repre-  
sentative area of measurement results [126], noted  $a(s_i, t)$  for  $s_i$  at  $t$ . It is the area around the  
location of a measurement result in which the measurement result is representative of any other  
result that would have been obtained at another position within the area. This definition may  
be more usable in practice than the one based on a norm as it considers the geometry of the

---

<sup>24</sup>If no instrument is mobile, it may be necessary to have less restrictive requirements on the spatiotemporal vicinity to observe rendez-vous.

area in which the instruments are. There are buildings and crossings for instance in an urban environment. Thus, with this definition,  $s_i$  is in rendez-vous with  $s_j$  at  $t$  if for instance:

$$\exists t' \in \mathbb{R}^+ \text{ such as } \exists m(s_j, t') \text{ and } |t - t'| \leq \Delta t_\varphi \text{ and } a(s_i, t) \cap a(s_j, t') \neq \emptyset$$

An alternate expression could be the following:

$$\exists t' \in \mathbb{R}^+ \text{ such as } \exists m(s_j, t') \text{ and } |t - t'| \leq \Delta t_\varphi \text{ and } |a(s_i, t) \cap a(s_j, t')| \geq \Delta a_\varphi$$

3092 where  $\Delta a_\varphi$  is the minimal size of the intersection of the representativity area of both measurement  
3093 results.

3094 In conclusion, the concept of rendez-vous can be applied according to different definitions in  
3095 practice. In the following sections, the diagnosis algorithm is designed without any assumption  
3096 on the practical definition of a rendez-vous.

### 3097 3.3.2 Compatible rendez-vous

3098 Consider a rendez-vous  $\varphi(s_i \rightarrow s_j, t)$ . By associating the definitions of compatible measure-  
3099 ment results, a rendez-vous of  $s_i$  with  $s_j$  at  $t$  is stated as **compatible** if  $m(s_i, t) \approx m(s_j, t')$ . Oth-  
3100 erwise it is stated as **upper non-compatible** if  $m(s_i, t) > m(s_j, t')$  or **lower non-compatible**  
3101 if  $m(s_i, t) < m(s_j, t')$ .<sup>25</sup> Consequently,  $\Phi^\approx(s_i \rightarrow s_j, (t, \Delta t))$ ,  $\Phi^+(s_i \rightarrow s_j, (t, \Delta t))$  and  $\Phi^-(s_i \rightarrow$   
3102  $s_j, (t, \Delta t))$  are the respective associated sets of such rendez-vous.

## 3103 4 Algorithm for the diagnosis of calibration issues in a sensor network

### 3104 4.1 General idea

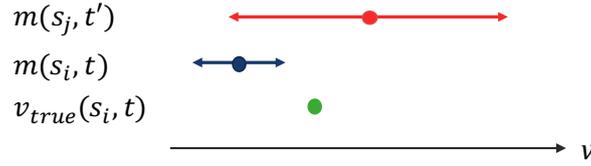
3105 The general purpose of a diagnosis algorithm is to determine whether if a system is **faulty**  
3106 (F) or **non-faulty** (NF). In the present case, the predicted status of an instrument  $s_i$  at  $t$  is  
3107 noted  $\hat{\Omega}(s_i, t)$ . Ideally,  $\hat{\Omega}(s_i, t)$  is equal to the true status of the instrument,  $\Omega(s_i, t)$ , which is  
3108 unknown in practice.

3109 To determine if instruments are correctly calibrated, the proposed approach consists in using  
3110 the concept of rendez-vous introduced in Section 3.3. When two instruments  $s_i$  and  $s_j$  are  
3111 in rendez-vous, it means that, according to chosen spatiotemporal conditions, they were both  
3112 measuring the same quantity value at the same time and the same place. Suppose that one of the  
3113 instruments,  $s_j$  for instance, was diagnosed as non-faulty and consider  $m(s_i, t)$  and  $m(s_j, t')$ , the  
3114 measurement results involved in a rendez-vous  $\varphi(s_i \rightarrow s_j, t)$ .<sup>26</sup> If both results are metrologically  
3115 valid, it means that  $m(s_j, t')$  can be seen as a standard value for  $s_i$ : it is a known value with an  
3116 associated uncertainty. Thus, if  $m(s_i, t) \approx m(s_j, t')$ , it denotes that  $s_i$  is correctly calibrated, at  
3117 least according to  $s_j$ , and its status can be predicted as non-faulty. Otherwise, it is predicted as  
3118 faulty. As  $s_j$  helps to predict the status of  $s_i$ , it is called a **diagnoser** of  $s_i$ .

Predicting the status of an instrument based on only one value may easily lead to false predictions because drift is a fault usually having a longer characteristic time than others like spike faults. For instance, if spikes are not correctly removed before carrying out a diagnosis, it is possible that  $m(s_i, t) > m(s_j, t')$  or that  $m(s_i, t) < m(s_j, t')$ , e.g. the measurement results  $m(s_i, t)$  and  $m(s_j, t')$  are not compatible, whereas the true status of  $s_i$ ,  $\Omega(s_i, t)$  is actually equal to non-faulty from a calibration perspective. Hence, the use of multiple rendez-vous with  $s_j$  and also with other instruments to predict the status of  $s_i$  can reduce the impact of these particular

<sup>25</sup>We recall that more restrictive criterion can be set to determine if measurement results are compatible with the results of other instruments but it is not studied in this work. See Section 3.2 for more details.

<sup>26</sup>We recall that the measurement results  $m(s_i, t)$  and  $m(s_j, t')$  have two different timestamps  $t$  and  $t'$  respectively because they may not have been obtained exactly at the same time (see Section 3.3).



**Figure 4.4.1:** Example where the measurement results  $m(s_i, t)$  and  $m(s_j, t')$  are compatible with each other but where  $m(s_i, t)$  is not compatible with its true value

cases. Moreover, if the class<sup>27</sup> of  $s_i$  is such as  $c(s_i) \gg c(s_j)$ , predicting the status of  $s_i$  based on  $s_j$ , even if  $\hat{\Omega}(s_j, t') = NF$ , may not be advisable. Indeed, it is possible that  $m(s_i, t) \approx m(s_j, t')$  but with  $m(s_i, t)$  not compatible with its true value as shown in Figure 4.4.1. A solution to that is to allow only instruments of a higher class than the one of  $s_i$  to take part in the prediction of its status. This minimal class is noted  $c_{min}^{\mathcal{D}}(s_i)$  and is defined by :

$$c_{min}^{\mathcal{D}}(s_i) = c(s_i) + \Delta c_{min}^{\mathcal{D}}(c(s_i))$$

3119 where  $\Delta c_{min}^{\mathcal{D}}(k)$  is the relative difference of class required between instruments of class  $k$  and  
 3120 the instruments allowed to be their diagnosers.<sup>28</sup>

Thus, to determine the status of an instrument  $s_i$  at  $t_d$ , the following set of **valid rendez-vous**  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  is used:

$$\begin{aligned} \Phi_v(s_i \rightarrow S, (t_d, \Delta t)) = & \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (t_d, \Delta t)), \text{ such as} \\ & s_j \notin S^{\mathcal{D}}, // s_j \text{ is not an instrument to diagnose} \\ & \hat{\Omega}(s_j, t') = NF, // s_j \text{ is non-faulty} \\ & c(s_j) \geq c_{min}^{\mathcal{D}}(s_i), // \text{ The class of } s_j \text{ is higher or equal to the minimal} \\ & \text{ class allowed to diagnose } s_i \\ & m(s_i, t) \in M^*(s_i, (t_d, \Delta t)), // \text{ The measurement result of } s_i \text{ is valid} \\ & m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\} // \text{ The measurement result of } s_j \text{ is valid} \end{aligned} \quad (4.1)$$

3121 It is also possible that the input information of a diagnosis algorithm is not sufficient to  
 3122 choose between F and NF statuses. Thus, a third option for the predicted status of a system is  
 3123 **ambiguous** (A).<sup>29</sup> In the present case, it is when  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  does not contain enough  
 3124 rendez-vous to allow a prediction with enough confidence. This minimal size for a set of valid  
 3125 rendez-vous for any instrument is noted  $|\Phi_v|_{min}$ .<sup>30</sup>

3126 Based on the definition of  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ , an algorithm can be designed to determine  
 3127 the statuses of all the instruments in a sensor network.

<sup>27</sup>See Definition 20 in Chapter 1 Section 3, page 18.

<sup>28</sup> $\Delta c_{min}^{\mathcal{D}}(k)$  can be identical for each class e.g.  $\Delta c_{min}^{\mathcal{D}}(k) = \Delta c_{min}^{\mathcal{D}}(S) \forall k \in [0..c_{max}]$ .

<sup>29</sup>The true status of an instrument cannot be ambiguous. It is either faulty or non-faulty strictly.

<sup>30</sup>It is possible to define different  $|\Phi_v|_{min}$  for each class of instrument like for  $\Delta c_{min}^{\mathcal{D}}(k)$ .  $|\Phi_v|_{min}$  was introduced as a common constant for the diagnosis of all the instruments because the amount of valid information to perform a binary diagnosis has no major reason to change from a class to another and in the worst case,  $|\Phi_v|_{min}$  can be set as:

$$|\Phi_v|_{min} = \min_{k \in [0..c_{max}]} |\Phi_v|_{min}(k)$$

## 4.2 Procedure for the diagnosis of all the instruments in a sensor network

Consider a diagnosis procedure  $d$  that occurs at  $t_d$ , and  $\Delta t$  such as  $[t_d - \Delta t; t_d]$  is the time range on which the diagnosis procedure is carried out. The measurement results obtained and the rendez-vous that occurred during  $[t_d - \Delta t; t_d]$  are the data that will be used to predict the statuses of the instruments of  $S$ . This forms a scene as defined by [40].

All the instruments in  $S^{c_{max}}$  are assumed as non-faulty.<sup>31</sup> Thus, at the beginning of a diagnosis procedure, **the set of instruments to diagnose**  $S^{\mathcal{D}}$  is equal to  $S \setminus S^{c_{max}}$  where  $S^{c_{max}}$  is the set of instruments of class  $c_{max}$ .

Then, the predicted statuses  $\hat{\Omega}(s_i, t_d)$  for each  $s_i \in S$  are initialised. The predicted statuses of the instruments in  $S^{c_{max}}$  are set to non-faulty and those of the instruments in  $S^{\mathcal{D}}$  are initially set to ambiguous.

Afterwards the predicted statuses of all the instruments, noted  $\hat{\Omega}(S, t_d)$ , are actualised. These actualised statuses are noted  $\tilde{\Omega}(S, t_d)$ .

They are determined as follows for each instrument  $s_i \in S^{\mathcal{D}}$ . First, if  $|\Phi(s_i \rightarrow S, (t_d, \Delta t))| < |\Phi_v|_{min}$ , it means that during  $[t_d - \Delta t; t_d]$ ,  $s_i$  did not meet other instruments enough times to be able to diagnose its status, whatever the predicted statuses of the other instruments. Therefore, it is not possible to actualise its predicted status with another value than ambiguous, which is already the value of  $\tilde{\Omega}(s_i, t_d)$ , and  $s_i$  is removed from  $S^{\mathcal{D}}$ .

Otherwise,  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  is determined. If its size is lower than  $|\Phi_v|_{min}$ , the actualised predicted status of  $s_i$  stays equal to ambiguous and  $s_i$  remains in  $S^{\mathcal{D}}$  as the size of  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  may change in a future iteration.

If  $|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))| \geq |\Phi_v|_{min}$ , the actualised predicted status can be determined between non-faulty and faulty.

To do so, the rates  $r_{\Phi_v}^{\sim}(s_i \rightarrow S, (t_d, \Delta t))$ ,  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t))$  and  $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t))$  are computed. They are defined as follows :

$r_{\Phi_v}^{\sim}(s_i, (t_d, \Delta t))$  is the rate of compatible rendez-vous in the set of valid rendez-vous of  $s_i$  over the time range  $[t_d - \Delta t; t_d]$ . It is equal to

$$r_{\Phi_v}^{\sim}(s_i, (t_d, \Delta t)) = \frac{|\Phi_v^{\sim}(s_i, (t_d, \Delta t))|}{|\Phi_v(s_i, (t_d, \Delta t))|}$$

$r_{\Phi_v}^+(s_i, (t_d, \Delta t))$  is the rate of upper non-compatible rendez-vous in the set of valid rendez-vous of  $s_i$  over the time range  $[t_d - \Delta t; t_d]$ . It is equal to

$$r_{\Phi_v}^+(s_i, (t_d, \Delta t)) = \frac{|\Phi_v^+(s_i, (t_d, \Delta t))|}{|\Phi_v(s_i, (t_d, \Delta t))|}$$

$r_{\Phi_v}^-(s_i, (t_d, \Delta t))$  is the rate of lower non-compatible rendez-vous in the set of valid rendez-vous of  $s_i$  over the time range  $[t_d - \Delta t; t_d]$ . It is equal to

$$r_{\Phi_v}^-(s_i, (t_d, \Delta t)) = \frac{|\Phi_v^-(s_i, (t_d, \Delta t))|}{|\Phi_v(s_i, (t_d, \Delta t))|}$$

Based on  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$ , which are the maximal tolerated values associated to the rates  $r_{\Phi_v}^{\sim}(s_i, (t_d, \Delta t))$ ,  $r_{\Phi_v}^+(s_i, (t_d, \Delta t))$  and  $r_{\Phi_v}^-(s_i, (t_d, \Delta t))$ , if one of the following conditions is true :

- $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^+)_{max}$

<sup>31</sup>This assumption is discussed in Section 6.

- 3157     •  $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^-)_{max}$
- 3158     •  $(1 - r_{\Phi_v}^{\sim}(s_i \rightarrow S, (t_d, \Delta t))) > (r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$

3159     then  $\tilde{\Omega}(S, t_d)$  is equal to faulty. Otherwise  $\tilde{\Omega}(s_i, t_d)$  is set to non-faulty. In the end,  $s_i$  is  
3160 removed from  $S^{\mathcal{D}}$ .<sup>32</sup>

3161     The use of three different rates,  $r_{\Phi_v}^{\sim}(s_i \rightarrow S, (t_d, \Delta t))$ ,  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t))$  and  $r_{\Phi_v}^-(s_i \rightarrow$   
3162  $S, (t_d, \Delta t))$ , is useful because it brings an additional information on the status of the instrument if  
3163 it is diagnosed as faulty. If this prediction was triggered by the condition on  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t))$ ,  
3164 it indicates that the instrument is overestimating its measured values for instance.

3165     After all the instruments in  $S^{\mathcal{D}}$  are treated, if  $\hat{\Omega}(S, t_d) = \tilde{\Omega}(S, t_d)$ , it means that no statuses  
3166 of the instruments in  $S^{\mathcal{D}}$  changed. Consequently, the statuses of the instruments of the network  
3167 at  $t_d$ ,  $\hat{\Omega}(S, t_d)$ , are determined and the diagnosis procedure ends. Otherwise  $\hat{\Omega}(S, t_d)$  takes the  
3168 values of  $\tilde{\Omega}(S, t_d)$  and the actualised statuses are determined again for each instrument  $s_i \in S^{\mathcal{D}}$ .

3169     The pseudo-code of this algorithm is provided in Algorithm 1.

### 3170     4.3 Improvements and extensions of the presented algorithm

3171     The considered algorithm can be easily improved in terms of efficiency. Appendix A Section  
3172 1 provides a version with several changes that may reduce the number of iterations of the main  
3173 loop of Algorithm 1.

3174     Moreover, it is possible to add more confidence into the predicted statuses of the instruments  
3175 by restricting the set of the diagnosers of an instrument  $s_i$  to the one of the diagnosers of  
3176 class  $k > c_{min}^{\mathcal{D}}(s_i)$ , with  $k$  being the highest class such as the predicted status  $\hat{\Omega}(s_i, t_d)$  can be  
3177 determined as non-faulty. This formulation is provided in Appendix A Section 2.

3178     In addition, the formulation of Algorithm 1 invites to compute it in a centralised manner.  
3179 Appendix A Section 3 introduces elements allowing an application of the principles of the  
3180 diagnosis algorithm in a decentralised manner.

3181     Finally, Appendix A Section 4 gives insights on how the diagnosis procedure could be extended  
3182 to the case of a sensor network measuring different quantities and for which some of its measurands  
3183 are influence quantities for several instruments in the network.

3184     For the sake of clarity, Algorithm 1 is the version used in the following sections of this chapter.

### 3185     4.4 Conclusion

3186     In this section, the principle of a diagnosis algorithm for sensor networks aiming at detecting  
3187 drift faults of measuring instruments was introduced. It is strongly based on the concepts of  
3188 rendez-vous between measuring instruments and of class of a measuring instrument. As a major  
3189 assumption, the instruments with the highest class must be assumed as non-faulty. Several  
3190 parameters drive the algorithm and are studied in the following sections. In a first step, the next  
3191 section gives an example of case study applying the presented algorithm.

---

<sup>32</sup> Alternate manners to determine the predicted status of an instrument according to compatible and non-compatible valid rendez-vous could be chosen. For instance, the value of the predicted status could be chosen according to the number of compatible and non-compatible valid rendez-vous. Thus  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  would be replaced by maximal numbers of tolerated non-compatible rendez-vous regarding  $|\Phi_v^+(s_i \rightarrow S, (t_d, \Delta t))|$ ,  $|\Phi_v^-(s_i \rightarrow S, (t_d, \Delta t))|$  and their sum.

---

**Algorithm 1:** Algorithm of the diagnosis procedure proposed for the detection of drift in sensor networks

---

**Data:**  $S, \Phi(S \rightarrow S, (t_d, \Delta t)), |\Phi_v|_{min}, \Delta c_{min}^D$   
**Result:**  $\hat{\Omega}(S, t_d)$

/\* Initiate the set of instruments to diagnose and the predicted statuses of the instruments \*/  
 $S^D \leftarrow S \setminus S^{cmax}$   
 $\hat{\Omega}(S^{cmax}, t_d) \leftarrow NF$   
 $\hat{\Omega}(S^D, t_d) \leftarrow A$   
 $\tilde{\Omega}(S^D, t_d) \leftarrow \hat{\Omega}(S^D, t_d)$  /\* Initiate the actualised statuses \*/  
/\* Ignore the instruments that cannot have enough valid rendez-vous \*/  
**for**  $s_i \in S^D$  **do**  
  **if**  $|\Phi(s_i \rightarrow S, (t_d, \Delta t))| < |\Phi_v|_{min}$  **then**  
     $S^D \leftarrow S^D \setminus \{s_i\}$   
  **end**  
/\* Predict the status of each instrument to diagnose \*/  
**repeat**  
   $\hat{\Omega}(S, t_d) \leftarrow \tilde{\Omega}(S, t_d)$  /\* The actualised statuses are now the predicted statuses \*/  
  **for**  $s_i \in S^D$  **do**  
    /\* Build the current set of valid rendez-vous for  $s_i$  \*/  
     $c_{min}^D(s_i) \leftarrow c(s_i) + \Delta c_{min}^D(c(s_i))$   
     $\Phi_v(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (t_d, \Delta t)), \text{ such as}$   
       $s_j \notin S^D, \hat{\Omega}(s_j, t) = NF, c(s_j) \geq c_{min}^D(s_i),$   
       $m(s_i, t) \in M^*(s_i, (t_d, \Delta t)) \text{ and } m(s_j, t) \in M^*(s_j, (t_d, \Delta t))\}$   
    /\* If  $s_i$  have enough valid rendez-vous, then compute the different rates to actualize its status \*/  
    **if**  $|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))| \geq |\Phi_v|_{min}$  **then**  
       $r_{\Phi_v}^{\approx}(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^{\approx}(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$   
       $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^+(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$   
       $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^-(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$   
      /\* If a condition on the different rates is met, then the actualised status of  $s_i$  is set to faulty, otherwise, it is set to non-faulty \*/  
      **if**  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^+)_{max}$  **or**  $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^-)_{max}$  **or**  
       $(1 - r_{\Phi_v}^{\approx}(s_i \rightarrow S, (t_d, \Delta t))) > (r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  **then**  
         $\tilde{\Omega}(s_i, t_d) \leftarrow F$   
      **else**  
         $\tilde{\Omega}(s_i, t_d) \leftarrow NF$   
      **end**  
       $S^D \leftarrow S^D \setminus \{s_i\}$  /\*  $s_i$  is diagnosed so it can be removed from  $S^D$  \*/  
    **end**  
  **end**  
**end**  
**until**  $\hat{\Omega}(S, t_d) = \tilde{\Omega}(S, t_d)$  /\* Repeat until there is no difference between the predicted and actualised statuses \*/

---

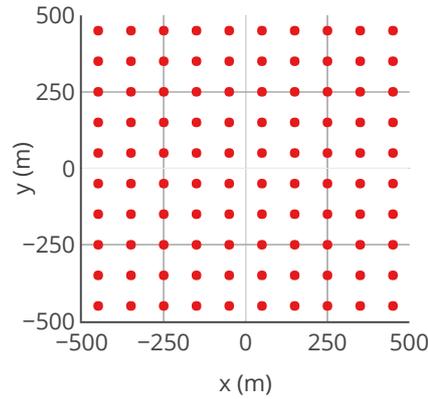
## 3192 5 Application of the algorithm to a first case study

3193 To illustrate the presentation of the diagnosis algorithm, this section provides a first case  
3194 study based on simulation to appreciate its performances.

### 3195 5.1 Definition of the case study

#### 3196 5.1.1 Sensor network

3197 A sensor network of 10 instruments is considered. The class of one of them is equal to 1, the  
3198 others being of class zero. Thus,  $c_{max} = 1$  in this case.



**Figure 4.5.1:** Map of the 100 positions available for the instruments in the case study.

3199 The instruments of class  $k = 0$  move randomly in a discrete space of 100 positions at a  
3200 discrete time. The positions of the instruments are geometrically defined according to a grid of  
3201  $10 \times 10$  with a step that represents 100m, and centred on  $(x, y) = (0, 0)$  as shown in Figure 4.5.1.  
3202 At each time step a new position is chosen randomly for each instrument following a uniform  
3203 law. Instruments may remain in position and multiple instruments may share a position. Two  
3204 instruments are in rendez-vous when they are at the same position at the same time. In addition,  
3205 the instrument of class  $c_{max}$  is static. Its position is randomly drawn among the 100 positions.

3206 As the case study is based on simulation, the time step has no actual physical meaning. To  
3207 ease the comprehension of the study, the time step of the simulation represents 10 min and the  
3208 case study lasts 265 days.

#### 3209 5.1.2 Instruments

3210 Instruments are assumed to be initially calibrated. In the first place, the instrument of class  
3211  $c_{max}$  is assumed as perfect, e.g. it does not drift. The instruments of class zero all follow the  
3212 same drift model. The model chosen corresponds to the RGOI model defined in Chapter 5  
3213 Section 5.3.1, page 71. As a reminder, gain  $G(s_i, t)$  and offset  $O(s_i, t)$  drift of instrument  $s_i$  are  
3214 computed for this model at each time step following:

$$G(s_i, t) = \begin{cases} 1 & \text{if } t < t_{\text{start drift}} \\ G(s_i, t-1) + \delta G(s_i, t) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\forall t, \delta G(s_i, t)$  drawn following  $\mathcal{U}(0, \delta G_{max})$

$$O(s_i, t) = \begin{cases} 0 & \text{if } t < t_{\text{start drift}} \\ O(s_i, t-1) + \delta O(s_i, t) & \text{if } t \geq t_{\text{start drift}} \end{cases}$$

with  $\forall t, \delta O(s_i, t)$  drawn following  $\mathcal{U}(0, \delta O_{max})$

3215 with  $\delta G_{max}$  and  $\delta O_{max}$  being respectively the maximal gain and offset possible increase per  
3216 time step.

Measured values are expressed following:

$$v(s_i, t) = G(s_i, t) \cdot v_{true}(s_i, t) + O(s_i, t)$$

3217 The instruments start to drift at  $t = 0$

Each measured value  $v(s_i, t)$  is associated to a constant relative uncertainty  $\Delta_r v(c_i)$  depending on the class of the instrument such as

$$\frac{\Delta v(s_i, t)}{v(s_i, t)} = \Delta_r v(c_i)$$

3218 A detection limit<sup>33</sup>  $v_{min}(c_i)$  is also defined to balance the effect of high uncertainties for low  
3219 measured values. This detection limit is used to determine if a measurement result  $m(s_i, t)$  is  
3220 metrologically valid, e.g. all the values below this detection limit are not considered valid.

3221 All the values of the parameters are listed in Table 4.5.1.

Parameter	Value	Unit
Class 1		
$\Delta_r v$	1	%
$v_{min}$	0.752	$\mu\text{g m}^{-3}$
Class 0		
$\delta G_{max}$	2/(24*6*30)	%/10min
$\delta O_{max}$	18.8/(24*6*30)	$\mu\text{g/m}^3/10\text{min}$
$\Delta_r v$	30	%
$v_{min}$	37.6	$\mu\text{g m}^{-3}$
True values model		
$A_{max}$	200	$\mu\text{g m}^{-3}$
$\sigma_{max}$	3826	m

**Table 4.5.1:** Values of the parameters of the case study. The values of  $\delta G_{max}$  and  $\delta O_{max}$  are displayed as a fraction of  $(24 * 6 * 30)$  min for the sake of clarity, e.g. the numerator is the maximal drift of the gain and of the offset per 30 days.

### 5.1.3 True values

3222  
3223 To model the true values of the instruments, the model used in Chapter 3 Section 3.2.1 is  
3224 considered. However in this case, instead of using the data that was considered to determine  
3225 the values of  $A(t)$  and  $\sigma(t)$ ,  $A(t)$  and  $\sigma(t)$  are drawn randomly at each time step following the

<sup>33</sup>The detection limit is the value "obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component is  $\beta$ , given a probability  $\alpha$  of falsely claiming its presence" [14]. In practice, it is recommended that  $\beta = \alpha = 0.05$  for information.

uniform laws  $\mathcal{U}(0, A_{max})$  and  $\mathcal{U}(0, \sigma_{max})$  respectively. The values of  $A_{max}$  and  $\sigma_{max}$  are reported in Table 4.5.1.

The impact of this choice on the results is discussed in Section 5.6 and 9.3.

## 5.2 Configuration of the diagnosis algorithm

We aim at performing a diagnosis **every 15 days**. Thus, for  $d \in D$ ,  $t_d = (d + 1) \times 15$  days, with  $D = [0..16]$ .

The goal of this diagnosis is to detect when an instrument has provided **at least 25% of measurement results non-compatible with their true values over the past 15 days**. According to this specification, we choose  $\Delta t = 15$  days and the rate thresholds  $(r_{\Phi_v^+})_{max}$ ,  $(r_{\Phi_v^-})_{max}$  and  $(r_{\Phi_v^+} + r_{\Phi_v^-})_{max}$  are all set to 25%.

The minimal number of valid rendez-vous to conclude with a predicted status different from ambiguous  $|\Phi_v|_{min}$  is set initially to 15.

## 5.3 Definition of the true status of an instrument

As it is a simulation, it is possible to know the true status of an instrument. Indeed its measured values and the true values are accessible, and therefore, it can be determined if 25% of measurement results non-compatible with their true values were obtained or not during a period of 15 days.

$r_{true}(s_i, (t, \Delta t))$  is the rate of measurement results compatible with true values of  $s_i$  over  $[t - \Delta t; t]$ . It is equal to

$$r_{true}(s_i, (t, \Delta t)) = \frac{|M^{\approx}(s_i, (t, \Delta t))|}{|M(s_i, (t, \Delta t))|}$$

In this case, if  $r_{true}(s_i, (t, \Delta t)) < 0.75$ , then  $\Omega(s_i, t) = F$ . Otherwise,  $\Omega(s_i, t) = NF$ .

Note that the true status of an instrument is based on **all its measurement results against the true values** on the considered time range whereas its predicted status is based on **the measurement results of its valid rendez-vous with other devices**. Thus, it is expected that instruments which are actually non-faulty may be predicted as faulty and *vice versa* because the sets of measurement results used to compute the true and predicted status of an instrument are different.

## 5.4 Metrics for the evaluation of performances of the diagnosis algorithm

To estimate the performance of the algorithm, vocabulary and metrics that are used to evaluate binary classifiers [130, 153] are appropriate but must be adapted as the predicted statuses can take three values instead of two.

Regarding the possible true statuses and predicted statuses, there are six cases:

- A non-faulty status predicted for an instrument which true status is non-faulty is a **true negative (TN)**
- A faulty status predicted for an instrument which true status is non-faulty is a **false negative (FN)**
- An ambiguous status predicted for an instrument which true status is non-faulty is a **non-determined negative (NDN)**
- A non-faulty status predicted for an instrument which true status is faulty is a **false positive (FP)**
- A faulty status predicted for an instrument which true status is faulty is a **true positive (TP)**

- 3265 • An ambiguous status predicted for an instrument which true status is faulty is a **non-**  
3266 **determined positive (NDP)**

3267 To these primary metrics,  $P$  and  $N$  are added, which are respectively **the number of**  
3268 **positives** (faulty true status) and **the number of negatives** (non-faulty true status).

The relationships between these metrics are summed up in Table 4.5.2

		True status	
		Non-faulty (N)	Faulty (P)
Predicted status	Non-faulty	True negative (TN)	False negative (FN)
	Ambiguous	Non-determined negative (NDN)	Non-determined positive (NDP)
	Faulty	False positive (FP)	True positive (TP)

**Table 4.5.2:** Contingency table of the different primary metrics

3269 For these metrics, only the instruments in  $S \setminus S^{c_{max}}$  are considered. The instruments of class  
3270  $c_{max}$  being always assumed as non-faulty and being not drifting in the following sections, taking  
3271 them into account would bias the metrics defined afterwards in Table 4.5.3.  
3272

**Table 4.5.3:** Metrics derived from the metrics  $P$ ,  $N$ ,  $TP$ ,  $TN$ ,  $FP$ ,  $FN$ ,  $NDP$  and  $NDN$

Name	Expression	Role
Prevalence	$Prev = \frac{P}{P+N}$	Indicates the proportion of positive cases among all the cases
True positive rate	$TPR = \frac{TP}{P}$	Indicates the proportion of positive cases correctly detected
True negative rate	$TNR = \frac{TN}{N}$	Indicates the proportion of negative cases correctly detected
False positive rate	$FPR = \frac{FP}{N}$	Indicates the proportion of positive cases incorrectly detected
False negative rate	$FNR = \frac{FN}{P}$	Indicates the proportion of negative cases incorrectly detected
Non-determined positive rate	$NDPR = \frac{NDP}{P}$	Indicates the proportion of positive cases detected as ambiguous
Non-determined negative rate	$NDNR = \frac{NDN}{N}$	Indicates the proportion of negative cases detected as ambiguous
Non-determined rate	$NDR = \frac{NDP+NDN}{P+N}$	Indicates the proportion of all cases detected as ambiguous
Positive predictive value	$PPV = \frac{TP}{TP+FP}$	Indicates the proportion of correctly detected positive cases per positive call
False discovery rate	$FDR = \frac{FP}{TP+FP}$	Indicates the proportion of incorrectly detected positive cases per positive call
Negative predictive value	$NPV = \frac{TN}{TN+FN}$	Indicates the proportion of correctly detected negative cases per negative call
False omission rate	$FOR = \frac{FN}{TN+FN}$	Indicates the proportion of incorrectly detected negative cases per negative call

*continued on next page*

---

continued from previous page

---

Name	Expression	Role
Accuracy	$ACC = \frac{TP+TN}{P+N}$	Indicates the proportion of correct detection

---

Another metric added is the **delay of first positive detection for an instrument**  $s_i$ , noted  $\Delta\mathcal{D}(s_i)$ . It is the difference between the index of the diagnosis procedure  $d$  where  $\Omega(s_i, t_d)$  changes from  $NF$  to  $F$  and the index of the diagnosis procedure  $d'$  where  $\widehat{\Omega}(s_i, t_{d'}) = F$  for the first time. Thus:

$$\Delta\mathcal{D}(s_i) = d - d'$$

3273 This value can be positive or negative as there can be early or late positive detection. In  
 3274 the following results, the average and the standard deviation of this metric over  $S \setminus S^{c_{max}}$  are  
 3275 considered.

## 3276 5.5 Results

### 3277 5.5.1 Observations from the results of different instruments

3278 In Figure 4.5.2, the evolution of the true and predicted statuses over time are represented for  
 3279 different instruments. It shows that the diagnosis algorithm can produce different results over  
 3280 the instruments:

- 3281 • the status of the instrument at each diagnosis procedure is often correctly predicted, except  
 3282 for a delay right after it becomes faulty (Figures 4.5.2a and 4.5.2c)
- 3283 • there are predicted statuses equal to ambiguous, once only in Figure 4.5.2a, but it can  
 3284 happen frequently considering Figure 4.5.2b
- 3285 • Changing decisions happen, e.g. an instrument predicted as faulty may be predicted as  
 3286 non-faulty later (Figure 4.5.2b)

3287 In all these particular cases, false results are observed, e.g. for instance an instrument is  
 3288 predicted as non-faulty when it is faulty and so on. In Section 5.3, we stated that these behaviours  
 3289 were expected. We investigate the reasons behind these incorrect predictions in Section 5.6.

### 3290 5.5.2 Overall appreciation

3291 Globally, the values of the metrics computed over all the diagnosis procedures are listed in  
 3292 Table 4.5.4.<sup>34</sup> In this study, the prevalence of positive cases, e.g. the proportion of cases where  
 3293 the instruments are actually faulty over all the diagnosis procedures, is equal to 76%, for 153  
 3294 cases (9 instruments diagnosed 17 times each).

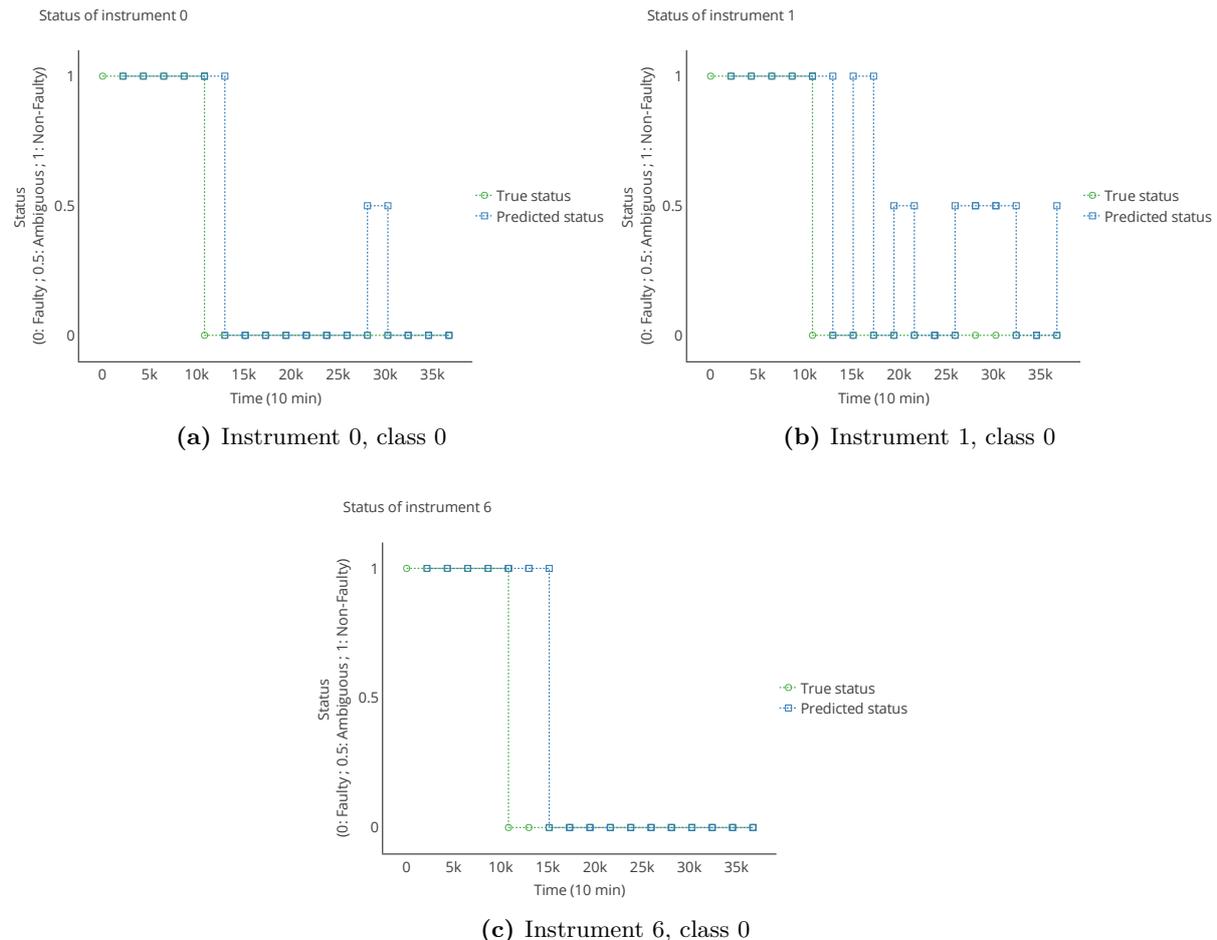
3295 From the number of  $TP$ ,  $FP$ ,  $TN$ ,  $FN$ ,  $NDP$  and  $NDN$ , it shows that:

- 3296 • Most of the predicted statuses are true negatives (36 cases) and true positives (78 cases)
- 3297 • Few predicted statuses are false negatives (20 cases) and non-determined positives (19  
 3298 cases)
- 3299 • No non-determined negative and false positive are predicted

3300 Consequently, the  $TNR$  is equal to 1 and the  $FPR$  is equal to zero whereas the  $TPR$  is  
 3301 important (0.67) and the  $FNR$  is moderate (0.17). Moreover, the  $NDNR$  is equal to zero and

---

<sup>34</sup>We remind that the metrics do not take into account the instruments of class  $c_{max}$  in this Section.



**Figure 4.5.2:** Evolution of the true status and of the predicted status for several instruments

3302 the  $NDPR$  is equal to 0.16, thus  $NDR = 16\%$  of ambiguous predictions were made which is  
 3303 quite important.

3304 That being said, it appears that the algorithm is correct when it predicts that instruments  
 3305 are faulty ( $PPV = 1$ ,  $FDR = 0$ ). It is often not correct when it predicts that instruments are  
 3306 non-faulty, about one third of the time ( $NPV = 0.64$ ,  $FOR = 0.36$ ). According to Figure 4.5.2,  
 3307 there is a delay when an instrument actually becomes faulty and the first prediction as faulty.  
 3308 It confirmed by the statistics of the delay of positive detection  $\Delta D$  in Table 4.5.5: the average  
 3309 value is negative and in fact it is always lower or equal to zero according to the maximal value of  
 3310 this metric.

3311 Nevertheless, the accuracy is equal to 75% which is acceptable for a first case study without  
 3312 a fine adjustment of the diagnosis algorithm's parameters.

### 3313 5.5.3 Evolution over time of the metrics

3314 The values of the metrics given in the previous subsection are representative of the results  
 3315 for all the diagnosis 17 procedures. They do not allow to understand how this overall result was  
 3316 build. Therefore, we study here their evolution.

3317 Figure 4.5.3 shows the evolution of the metrics as a function of the index of the diagnosis  
 3318 procedure. Each curve gives the values obtained for the corresponding metric at the diagnosis  $d$   
 3319 with the results of the nine instruments, where  $d$  is the index of the diagnosis procedure.

		True status		Prevalence	Accuracy
		Non-faulty $N = 36$	Faulty $P = 117$		
Predicted status	Non-faulty	$TN$ 36	$FN$ 20	$NPV$ 0.64	$FOR$ 0.36
	Ambiguous	$NDN$ 0	$NDP$ 19		
	Faulty	$FP$ 0	$TP$ 78	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.17		
		$NDNR$ 0.00	$NDPR$ 0.16	$NDR$ 0.16	
		$FPR$ 0.00	$TPR$ 0.67		

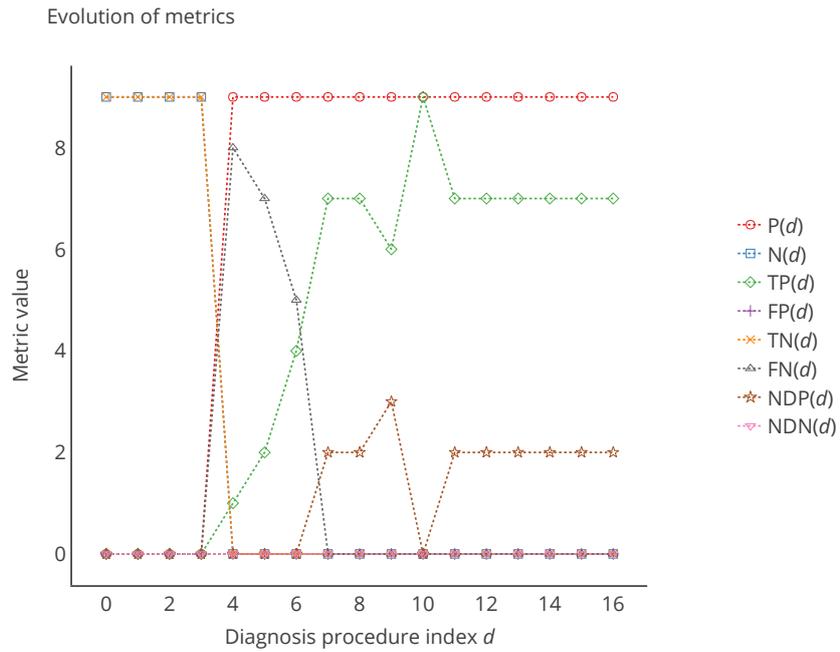
**Table 4.5.4:** Confusion matrix of the case study (metrics are defined in Section 5.4).

Delay of positive detection			
$\mu$	$\sigma$	min	max
-2.1	1.2	-4	0

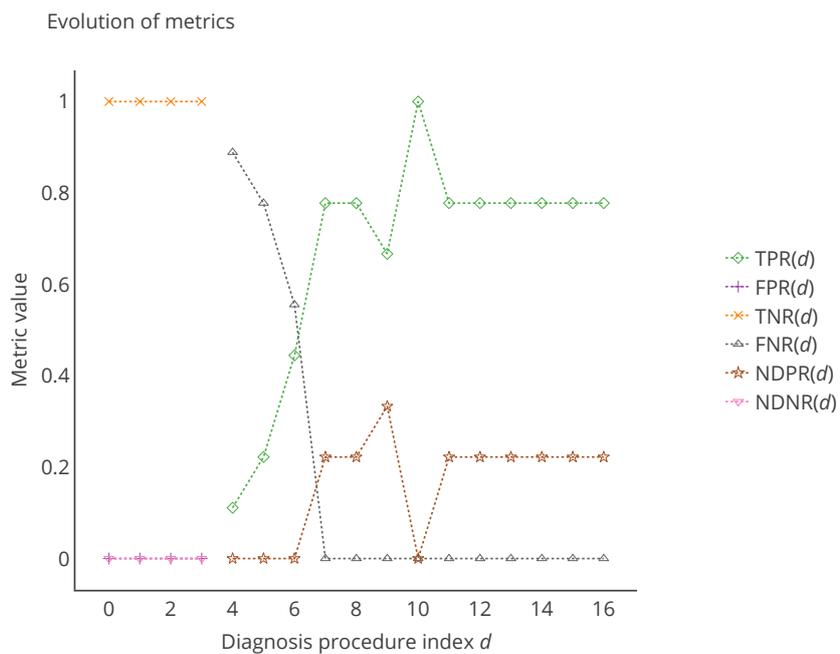
**Table 4.5.5:** Statistics of the delay of positive detection  $\Delta\mathcal{D}$  for the case study

3320 Figure 4.5.3a shows that first there are only non-faulty instruments (curve of  $N$ ), and faulty  
3321 ones appear at the 5<sup>th</sup> diagnosis ( $d = 4$ ). After the 6<sup>th</sup> diagnosis, all the instruments of class  
3322 zero are faulty (curve of  $P$ ). The transition is quite abrupt but this was expected because all  
3323 the instruments of class zero follow the same law of drift. The curves of the true positives  $TP$   
3324 and true negatives  $TN$  follow mostly the ones of  $P$  and  $N$  respectively. The false positives or  
3325 negatives are in fact occurring around the transition where instruments of class zero become  
3326 faulty. This behaviour is more clearly represented in Figures 4.5.3b and 4.5.3c with the curves of  
3327 the  $TPR$  and  $FNR$ , and the curves of the  $NPV$  and  $FOR$  respectively. Finally, with Figure  
3328 4.5.3d, we observe that the global accuracy, initially equals to one, decreases until the transition  
3329 phase and then increases again, converging to 0.8, which is explained by the 20% of  $NDP$  cases  
3330 in the last diagnosis procedures (Figure 4.5.3c).

3331 To conclude, this particular study of the evolution of the metrics allowed to understand  
3332 the global results over the 17 diagnosis procedures: the false results occur mainly when the  
3333 instruments actually become faulty and the number of predicted statuses equal to ambiguous  
3334 increases as more as there are instruments actually predicted as faulty. Thus, worse overall results  
3335 could have been displayed by considering the results of few diagnosis procedures around the  
3336 transition phase. We conjecture that better results would be obtained by increasing the number  
3337 of diagnosis procedures in the case study, without changing any parameter of the algorithm, e.g.



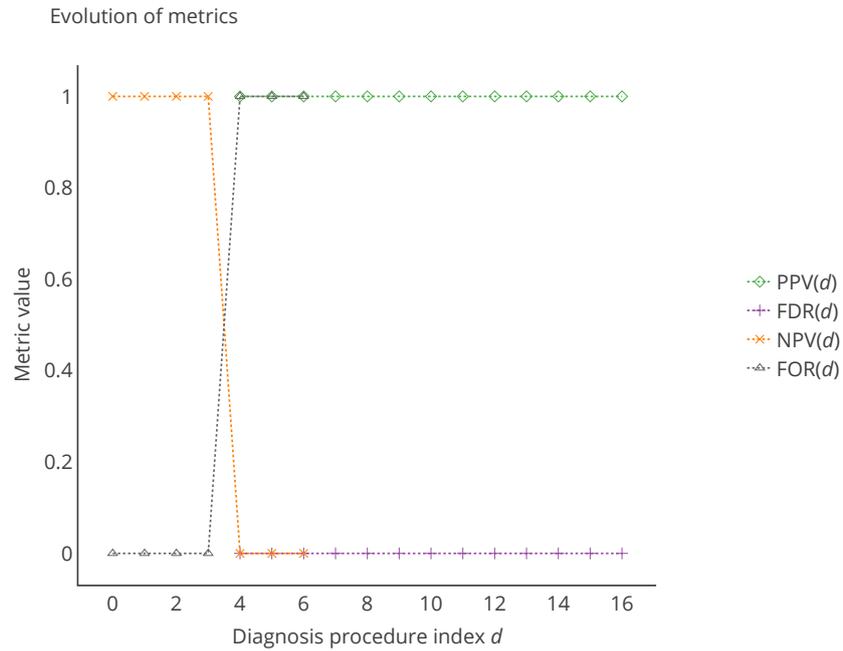
(a)



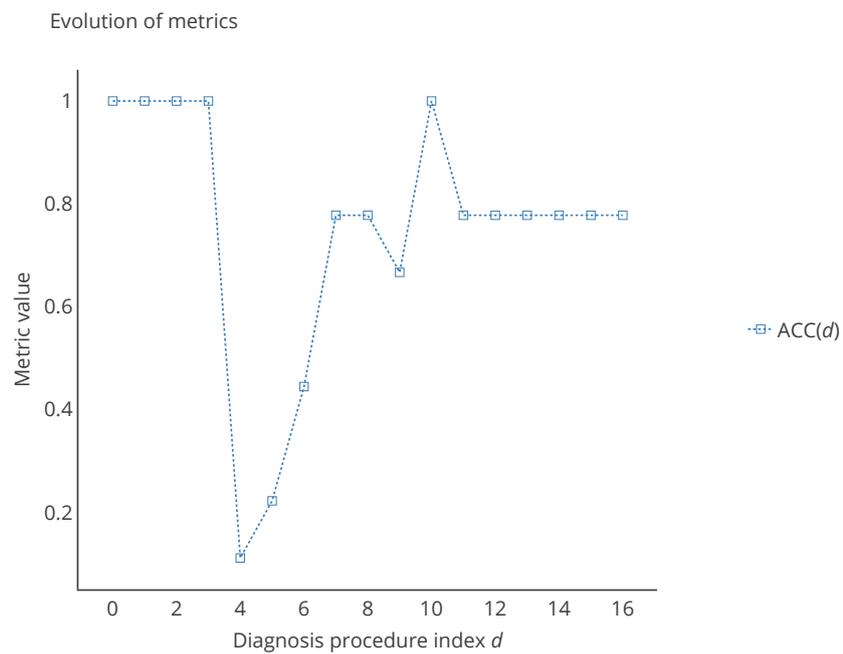
(b)

**Figure 4.5.3:** Evolution of the metrics computed for each diagnosis procedure as a function of the current diagnosis procedure id (continued) (metrics are defined in Section 5.4).

3338 by increasing the duration of the case study. However, this would not reduce the number of false  
 3339 that may happen in absolute terms. The reasons why false results happen are explained in the  
 3340 following section.

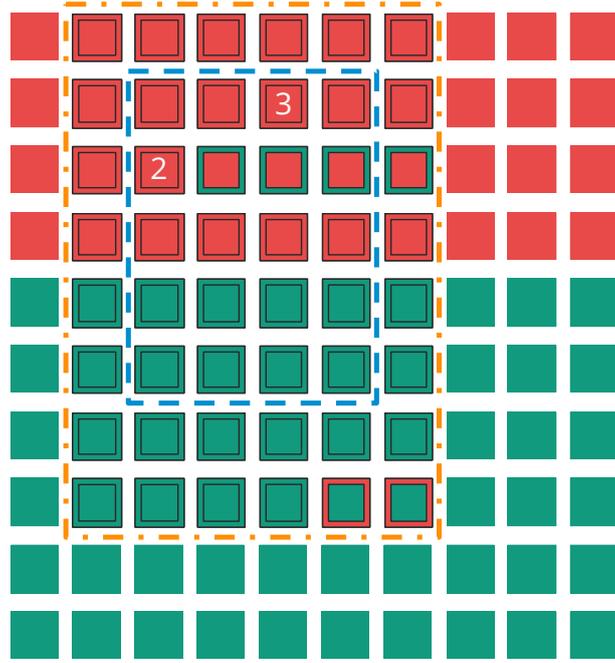


(c)



(d)

**Figure 4.5.3:** Evolution of the metrics computed for each diagnosis procedure as a function of the current diagnosis procedure id (metrics are defined in Section 5.4).



Symbol	$m(s_i, t)$ compatible with true value?	$m(s_i, t)$ involved in a rendez-vous?	$m(s_i, t) \approx m(s_j, t')$ ?
	Yes	No	-
	No	No	-
	Yes	Yes	Yes
	Yes	Yes	No
	No	Yes	Yes
	No	Yes	No
	The number $i$ indicates that the measurement result $m(s_i, t)$ is involved in $i$ rendez-vous.		
	Set of the measurements results involved in rendez-vous $M(s_i, (t_d, \Delta t)) _{\Phi}$		
	Set of the measurements results involved in valid rendez-vous $M(s_i, (t_d, \Delta t)) _{\Phi_v}$		

**Figure 4.5.4:** Representation of an example set of measurement results  $M(s_i, (t_d, \Delta t))$ . First, we observe that the restrictions of the set of measurement results  $M(s_i, (t_d, \Delta t))$  to the ones where they are respectively involved in rendez-vous  $M(s_i, (t_d, \Delta t))|_{\Phi}$  and involved in valid rendez-vous  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$  do not have the same proportion of measurement results compatible with their true values. Thus, the rates of measurement results compatible with true values  $r_{true}(s_i, (t_d, \Delta t))$ ,  $r_{true}(s_i, (t_d, \Delta t))|_{\Phi}$  and  $r_{true}(s_i, (t_d, \Delta t))|_{\Phi_v}$  are not equal. Therefore, considering that the predicted statuses are determined with  $r_{\tilde{\Phi}_v}(s_i, (t_d, \Delta t))$  for instance, which computation is based on the set  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ , it is possible that  $r_{\tilde{\Phi}_v}(s_i, (t_d, \Delta t))$  from  $r_{true}(s_i, (t_d, \Delta t))$ . Here  $r_{true}(s_i, (t_d, \Delta t)) = 0.6$  and  $r_{\tilde{\Phi}_v}(s_i, (t_d, \Delta t)) = 0.48$ . If a threshold of 0.5 is considered to differentiate faulty instruments from non-faulty ones, then the true status of  $s_i$  is  $\Omega(s_i, t_d) = NF$  here but its predicted status is  $\tilde{\Omega}(s_i, t_d) = F$ : it is a false result.

## 3341 5.6 Explanations of false results

3342 In this case study, multiple false results, e.g. false positive or false negative, were observed.  
3343 The reasons why they happen can easily be explained.

3344 Consider as an example a set of measurement results  $M(s_i, (t_d, \Delta t))$ . This set is composed of  
3345 100 measurement results, represented as squares in Figure 4.5.4, such as the rate of compatible  
3346 measurement results with their true values  $r_{true}(s_i, (t_d, \Delta t))$  is equal to  $\frac{60}{100} = 0.6$ . If a threshold  
3347 equal to 0.5 for  $r_{true}$  is considered to differentiate faulty instruments from non-faulty ones<sup>35</sup>,  
3348 then  $\Omega(s_i, t_d) = NF$  here.

3349 In the diagnosis algorithm, the predicted status of an instrument is determined based on  
3350 the measurement results associated to the rendez-vous in the set  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ . These  
3351 measurement results are the ones in the blue rectangle dashed line in Figure 4.5.4. The question is  
3352 whether these measurement results are representative of  $M(s_i, (t_d, \Delta t))$  and thus allow obtaining  
3353 a  $r_{\Phi_v}^{\approx}(s_i, (t_d, \Delta t))$  equal to  $r_{true}(s_i, (t_d, \Delta t))$  or not.

3354 Consider first  $M(s_i, (t_d, \Delta t))|_{\Phi}$  which is the restriction of  $M(s_i, (t_d, \Delta t))$  to the measurement  
3355 results associated to rendez-vous that are in the set  $\Phi(s_i \rightarrow S, (t_d, \Delta t))$ . In this case, not all the  
3356 measurement results of  $s_i$  are involved in a rendez-vous, only those with a border in Figure 4.5.4  
3357 (grouped in the orange rectangle dashed line in Figure 4.5.4). Computing  $r_{true}(s_i, (t_d, \Delta t))|_{\Phi}$ ,  
3358 which is defined similarly to  $r_{true}(s_i, (t_d, \Delta t))$ , gives a value of  $\frac{24}{48} = 0.50$  which is worse than  
3359  $r_{true}(s_i, (t_d, \Delta t))$  but the result is still correct: it allows to say that  $s_i$  is non-faulty because  
3360  $r_{true}(s_i, (t_d, \Delta t))|_{\Phi} \geq 0.5$  according to the threshold previously chosen.

3361 However, the rate used for the prediction of the status of an instrument,  $r_{\Phi_v}^{\approx}(s_i, (t_d, \Delta t))$ , is  
3362 not based on comparisons with true values but with other instruments. Different situations can  
3363 be observed as shown in the table of Figure 4.5.4:

- 3364 • a measurement result compatible with its true values can be compatible or not with the  
3365 measurement result of the other instrument involved in a rendez-vous
- 3366 • a measurement result non-compatible with its true values can be compatible or not with  
3367 the measurement result of the other instrument involved in a rendez-vous
- 3368 • measurement results of an instrument may be involved in several rendez-vous.

3369 From the perspective of  $\Phi(s_i \rightarrow S, (t_d, \Delta t))$  first before moving to  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ ,  
3370  $r_{true}(s_i, (t_d, \Delta t))|_{\Phi}$  may not be equal to  $r_{\Phi}^{\approx}(s_i, (t_d, \Delta t))$  and it is the case here as  $r_{\Phi}^{\approx}(s_i, (t_d, \Delta t)) =$   
3371  $\frac{24-2+4}{48+3} = 0.51 \neq r_{true}(s_i, (t_d, \Delta t))|_{\Phi}$ . The difference is small but the reason why false results  
3372 happen begins to appear. On top of the fact that the compatibility between measurement results  
3373 of two instruments may not correctly report the true status of an instrument<sup>36</sup>, the sets used  
3374 for the computations of the rates  $r_{true}$  and  $r_{\Phi}$  do not necessarily have the same composition of  
3375 measurement results.

Considering then  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$ , we have:

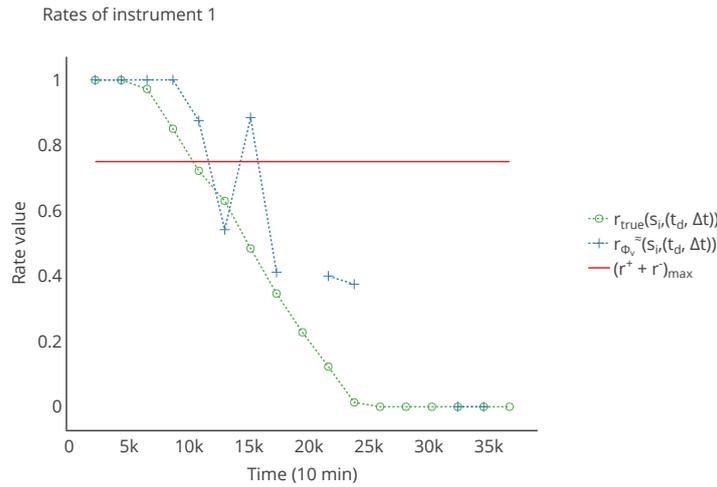
$$r_{true}(s_i, (t_d, \Delta t))|_{\Phi_v} = \frac{8}{20} = 0.40 \text{ and } r_{\Phi_v}^{\approx}(s_i, (t_d, \Delta t)) = \frac{8+3}{20+3} = 0.48$$

3376 Thus, with a threshold  $(r_{\Phi}^+ + r_{\Phi}^-)_{max} = 0.5$ , the predicted status of  $s_i$ ,  $\widehat{\Omega}(s_i, t_d)$ , is equal to faulty  
3377 (because  $r_{\Phi_v}^{\approx}(s_i, (t_d, \Delta t)) < (r_{\Phi}^+ + r_{\Phi}^-)_{max}$ ) which is different from  $\Omega(s_i, t_d)$ : it is a false result.

3378 Figure 4.5.5 shows the evolution of the different rates  $r_{true}$  and  $r_{\Phi_v}^{\approx}$  computed for each block  
3379 of 15 days for the instrument that was considered in Figure 4.5.2b. With the trigger level in  
3380 red, the evolution of the green and blue curves in Figure 4.5.2b can be explained following the

<sup>35</sup>e.g. if  $r_{true}(s_i, (t_d, \Delta t)) \geq 0.5$  then  $\Omega(s_i, t_d) = NF$

<sup>36</sup>See Section 4.1



**Figure 4.5.5:** Evolution of the rates  $r_{true}$  and  $r_{\Phi_v}$  for  $s_1$  in the case study of Section 5. The missing points for  $r_{\Phi_v}$  are due to the ambiguous cases where  $r_{\Phi_v}$  could not be computed.

3381 reasoning developed for the presented theoretical example. In this Figure, the diagnosis curve is  
 3382 particularly interesting as the instrument is alternatively diagnosed as faulty and non-faulty, and  
 3383 several times.

3384 Through this development, the reason why false results appear has been explained. It is  
 3385 because the properties of sets of measurement results like  $M(s_i, (t_d, \Delta t))$  are not necessarily  
 3386 conserved while building  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  and it is this latter set that is used to compute the  
 3387 rates  $r_{\Phi_v}$ . In fact, there is a sort of "sampling" of  $M(s_i, (t_d, \Delta t))$  that is performed through the  
 3388 rendez-vous and resulting in  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$  in the end. This explains mainly the false results  
 3389 of the diagnosis algorithm. Solving this issue is quite challenging. It might be possible in some  
 3390 situations where the characteristics of the sets of measurement results could be conserved by the  
 3391 "sampling" at stake for the restriction of the sets of measurement results but it would generally  
 3392 require strong assumptions on the measurement results and on the occurrence of rendez-vous,  
 3393 without taking into account the notion of validity of the rendez-vous. However, by adjusting  
 3394 the parameters of the diagnosis algorithm, it may also be possible to reduce the number of false  
 3395 results. Therefore, they are reviewed in the following section.

### 3396 5.7 On the parameters of the diagnosis algorithm

3397 There are several parameters that drive the proposed algorithm:

- 3398 • the instant of diagnosis  $t_d$  and its periodicity
- 3399 • the time difference  $\Delta t$  defining the length of the time range on which measurement results  
 3400 selected for the input data of the algorithm
- 3401 • the minimal relative differences of class required for instruments of class  $k$  with their  
 3402 diagnoser instruments  $\Delta c_{min}^{\mathcal{D}}(k)$
- 3403 • the minimal required number of valid rendez-vous  $|\Phi_v|_{min}$
- 3404 • the conditions for a measurement result to be considered as valid
- 3405 • the conditions of spatiotemporal vicinity for two instruments to be in rendez-vous

- 3406 • the maximal tolerated values  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  for the rates of upper  
3407 and lower non-compatible rendez-vous in sets of valid rendez-vous.

3408 It may be possible to improve the results of case study with different rules or values for them.  
3409 Indeed, values were chosen but may not be the ones allowing obtaining the best results. They  
3410 may explain the reason why there are many false results in particular. The following paragraphs  
3411 provide a discussion regarding the choice of the parameters.

### 3412 Instant of diagnosis and its periodicity

3413  $t_d$  is the instant of the  $d^{\text{th}}$  diagnosis. In this case study, we chose to have a periodic diagnosis,  
3414 e.g. diagnosis procedures are carried out on a regular basis (every 15 days).<sup>37</sup> With a different  
3415 frequency of diagnosis, for instance every week or even every day in the case study, it may  
3416 be possible to detect sooner that an instrument has become faulty in exchange of a higher  
3417 computational cost. Consequently, the overall results would be different but this would not  
3418 change the general behaviour of the algorithm. The choice of the instant of diagnosis and its  
3419 periodicity is not discussed in detail in this work despite the fact that it may have a significance  
3420 in practice because it strongly depends on the instruments used and on the context of deployment.  
3421 Indeed, all the instruments do not have the same characteristic time of drift for instance. This  
3422 information is often provided as an order of magnitude of drift of the gain and offset over a week  
3423 or a month in percentage in datasheets [104]. In this way, it would influence the periodicity  
3424 of the diagnosis. Moreover, the context of deployment can justify to perform the diagnosis at  
3425 specific times of the day.<sup>38</sup> Thus, the characteristics of the instruments used and the context of  
3426 deployment can help in the choice of the moment of diagnosis and its periodicity.<sup>39</sup>

### 3427 Length of the time range on which the algorithm is applied

3428 The length of the time range  $\Delta t$  has an important influence on the results. Indeed, the higher  
3429  $\Delta t$  is, the more rendez-vous may be in  $\Phi(s_i \rightarrow S, (t_d, \Delta t))$  and consequently in  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ ,  
3430 and vice versa. Thus, it influences the contents of  $M(s_i, (t_d, \Delta t))|_{\Phi}$  and  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$ .  
3431 However, a high value of  $\Delta t$  implies that measurement results obtained at very different instants  
3432 would be used by the diagnosis algorithm. Considering that old measurement results are more  
3433 accurate than recent ones in the case of an irreversible drift, the higher  $\Delta t$  is, the less the  
3434 measurement results at the actual level of the drift will be preponderant in  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$ .  
3435 Therefore, choosing a different value for  $\Delta t$  while trying to bring closer the distributions of  
3436 measured values in  $M(s_i, (t_d, \Delta t))$  and  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$  could also induce false results for  
3437 another reason because as shown in Section 5.6, the content of the sets  $M(s_i, (t_d, \Delta t))$  and  
3438  $M(s_i, (t_d, \Delta t))|_{\Phi_v}$  are not necessarily equivalent, regardless of  $\Delta t$ . Moreover, reducing  $\Delta t$  in  
3439 particular could also have another drawback: on a short time range there may likely be less  
3440 rendez-vous than on a longer one. Depending on the minimal required number of valid rendez-  
3441 vous  $|\Phi_v|_{min}$ , the value of  $\Delta t$  can make the algorithm not functional, e.g. all the instruments

<sup>37</sup>An actual day and hour is not associated to  $t_d$  because it does not have a real signification here.

<sup>38</sup>As an example, in [109] for an *in situ* calibration strategy, Moltchanov *et al.* chose to apply their *in situ* calibration algorithm every day after 4:00a.m. and to use the measured values between 1:00a.m. and 4:00a.m.. They proceeded like this because the measuring instruments to calibrate were distant from the reference and at this moment, the measurand was supposed to be homogenous in the measurement area, enabling to compare between them the values measured by the instruments. Therefore, in the same context with our diagnosis algorithm, it may be relevant to apply it at a similar time of the day that was used in [109].

<sup>39</sup>Note that too high a frequency of diagnosis may be pointless if no rendez-vous occurs between two instants of diagnosis. It would be more relevant to have in this case a formulation of the algorithm so that the diagnosis procedure is carried out when new rendez-vous happen. In this way, Appendix A Section 5 provides a discussion regarding first how to reformulate the algorithm so that it is possible to trigger a diagnosis procedure on an event, e.g. a rendez-vous, and how to use an event-based range on which the algorithm is applied. Then, elements regarding a real-time version of the diagnosis algorithm are given in Appendix A Section 6.

3442 different from the class  $c_{max}$  would be predicted as ambiguous.

3443 In fact, as for the instant of diagnosis and its periodicity, the adjustment of  $\Delta t$  could be  
3444 defined based on the properties of the instruments used like their characteristic time of drift.  
3445 Indeed, it can give an order of magnitude of the relevant length of the time range on which the  
3446 diagnosis algorithm should be applied.

#### 3447 Minimal relative differences of class required

3448 The minimal relative differences of class required  $\Delta c_{min}^{\mathcal{D}}(k)$  would usually be set to +1 or 0.  
3449 Indeed, it is rarely advisable to determine the statuses of instruments with ones of a lower class  
3450 as shown in Section 4.1. Predicting the status of an instrument with diagnoser instruments of a  
3451 strictly higher class can give a higher confidence in the diagnosis as their measurement results  
3452 are standards of a higher quality than the ones provided by diagnoser instruments of an identical  
3453 class.

#### 3454 Minimal required number of valid rendez-vous

3455 The minimal required number of valid rendez-vous  $|\Phi_v|_{min}$  can be a chosen integer but to  
3456 add up more confidence in the predicted status, it is advisable to make it the highest as possible.  
3457 However, too high a value can make the diagnosis impossible: all the instruments of class  $c_{max} - 1$   
3458 and lower would be diagnosed as ambiguous. If a training dataset of rendez-vous is available, it  
3459 is possible to give an upper boundary. This is explained in Section 8.

#### 3460 Conditions for a measurement result to be considered as valid

3461 The conditions to satisfy to state a measurement result of an instrument as valid can be set  
3462 according to the datasheets of the instruments. Measurement range, detection limit and so on  
3463 are values usually given. Thus, it is not really a parameter driving the diagnosis algorithm, it is  
3464 a constraint driven by the choice of the measuring instruments.

#### 3465 Conditions of spatiotemporal vicinity for two instruments to be in rendez- 3466 vous

3467 This subject was already discussed in Section 3.3. We introduced different definitions for  
3468 these conditions. In summary, they should be determined based on expert specifications or on  
3469 characteristics of the measurand and of the measuring instruments.

#### 3470 Maximal tolerated values for the rates of upper and lower non-compatible 3471 rendez-vous in sets of valid rendez-vous

3472 Like  $t_d$ , its periodicity and  $\Delta t$ , the maximal tolerated values  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ +$   
3473  $r_{\Phi_v}^-)_{max}$  for the rates of upper and lower non-compatible rendez-vous in sets of valid rendez-vous  
3474 can be derived according to a requirement. This is what was done in Section 5.2. Nevertheless,  
3475 these parameters have a major role in the way false results are produced because they are used  
3476 to make the final decision between a status predicted as faulty or non-faulty. Therefore, their  
3477 adjustment is discussed in Section 7 where the reduction of false results is tackled.

### 3478 5.8 Conclusion

3479 In this section, a case study was conducted to put in practice the algorithm of diagnosis  
3480 presented in Section 4.2. It provides satisfying results according to the initial specification.  
3481 However, the results are not excellent. There are false results and, in this particular case, a  
3482 significant number of false negatives. There are also several cases of predicted statuses equal to  
3483 ambiguous. The reasons behind the false results were explained theoretically.

3484 To improve the results, the adjustment of the parameters of the algorithm was discussed.  
 3485 The brief overview of each one highlighted that the adjustment of some of them can be driven  
 3486 by expert specifications and the characteristics of the measuring instruments composing the  
 3487 sensor network. For  $|\Phi_v|_{min}$ ,  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$ , guidelines are provided in  
 3488 Sections 7 and 8 to set their values.

3489 Before that, we discuss in the following section the assumption for the instruments of class  
 3490  $c_{max}$  that are considered as non-faulty—they were even not drifting in the present case study.

## 3491 6 On the assumption regarding the top-class instruments being always 3492 predicted as non-faulty

### 3493 6.1 Theoretical discussion

3494 The assumption that instruments of class  $c_{max}$  are always predicted as non-faulty is a strong  
 3495 one. It is, however, necessary for the proper operation of the presented algorithm. If there is no  
 3496 assumption regarding the status of any instrument of class  $c_{max}$ , they are initially considered as  
 3497 ambiguous and thus the sets  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  are always empty. In this case, the algorithm  
 3498 predicts a status equal to ambiguous for each instrument, which makes it useless.

3499 Actually, the assumption on the instruments of class  $c_{max}$  can be relaxed: **one single**  
 3500 **instrument of class  $c_{max}$  supposed as non-faulty may be sufficient**. More practically,  
 3501 depending on the value of  $|\Phi_v|_{min}$  and on the configuration of the sensor network, only a subset  
 3502 of  $S^{c_{max}}$  may be sufficient to ensure the lowest number of instruments systematically diagnosed  
 3503 as ambiguous.<sup>40</sup> Note that this formulation is also more appropriate in the case of a blind sensor  
 3504 network. Indeed, for this type of sensor network, all the instruments are considered as equivalent  
 3505 from a metrological perspective. Consequently, there is no subset of instruments of a higher class  
 3506 than others.<sup>41</sup>

3507 Following on from this new formulation of the assumption regarding the statuses of the  
 3508 instruments of class  $c_{max}$ , we might be tempted to try to relax it to point where the status of  
 3509 only one instrument of any class has to be known. Consider such an instrument  $s_i$ . Again, all  
 3510 the instruments  $s_j \in S$  such as  $c_{min}^D(s_j) > c(s_i)$  cannot be diagnosed differently than ambiguous  
 3511 if the minimal relative differences of class required  $\Delta c_{min}^D(k)$  is set to +1 or 0 as recommended  
 3512 in Section 5.7. In particular, if the class of  $s_i$  is the lowest class in the network, noted  $c_{min}$ ,  
 3513 and if  $\Delta c_{min}^D(c_{min}) = +1$ , then all the instruments of the network, except  $s_i$  are systematically  
 3514 diagnosed as ambiguous. In fact, the lower the class of  $s_i$  is, the more instruments will always be  
 3515 predicted as ambiguous. Thus, it may be simpler to come back to assuming some instruments of  
 3516 class  $c_{max}$  as non-faulty. It is acceptable to guarantee such a hypothesis in real situations.

### 3517 6.2 Case study with the instrument of class $c_{max}$ drifting

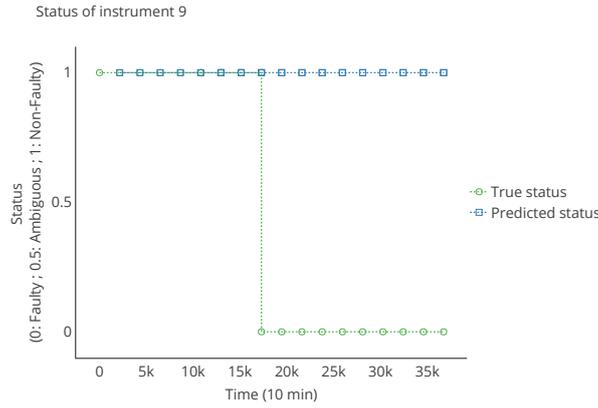
3518 In the case study of Section 5, the instrument of class  $c_{max}$ , noted here  $s_9$ , was first assumed  
 3519 as perfect, e.g. not drifting, for the sake of clarity.

3520 Consider here that this instrument is also undergoing drift following the RGOI model but  
 3521 with  $\delta G_{max}(s_9) = 0.5/(24 * 6 * 30)\%/10\text{min}$  and  $\delta O_{max}(s_9) = 1.88/(24 * 6 * 30)\mu\text{g}/\text{m}^3/10\text{min}$ .<sup>42</sup>  
 3522 With the previous definition of  $r_{true}$  and the same criteria to determine the true status of the  
 3523 instrument  $\Omega(s_9, t)$  like the other instruments, the evolution of the true and predicted statuses

<sup>40</sup>Note that if  $\Delta c_{min}^D(c_{max}) > 0$ , the instruments of class  $c_{max}$  which statuses are not assumed as non-faulty, cannot be diagnosed differently than ambiguous. Hence, it may be more relevant to define a sub-network without these instruments on which the diagnosis algorithm is applied so that the results are not biased by the behaviour of the algorithm with these instruments.

<sup>41</sup>In the case of a blind sensor network, necessarily  $\Delta c_{min}^D = 0$  for all the instruments, otherwise the algorithm cannot work.

<sup>42</sup>See Table 4.5.1 for the explanation regarding  $G_{max}(s_9)$  and  $O_{max}(s_9)$  expressed as fractions.



**Figure 4.6.1:** Evolution of the true status and of the predicted status of  $s_9$ , the instrument of class 1, when it is drifting

		True status		Prevalence	Accuracy
		Non-faulty $N = 36$	Faulty $P = 117$		
Predicted status	Non-faulty	$TN$ 36	$FN$ 28	$NPV$ 0.56	$FOR$ 0.44
	Ambiguous	$NDN$ 0	$NDP$ 15		
	Faulty	$FP$ 0	$TP$ 74	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.24		
		$NDNR$ 0.00	$NDPR$ 0.13	$NDR$ 0.10	
		$FPR$ 0.00	$TPR$ 0.63		

**Table 4.6.1:** Confusion matrix of the case study when the instrument of class 1 is drifting (metrics are defined in Section 5.4).

3524 over time for this particular instrument is given in Figure 4.6.1, the predicted status being again  
 3525 assumed as always non-faulty. We observe that there are consequently false negative if the  
 3526 assumption is kept. However, still without taking into account  $s_9$  in the metrics to evaluate the  
 3527 results, the diagnosis algorithm gives the results of Tables 4.6.1 and 4.6.2. We observe that the  
 3528 values of the metrics do not change significantly from Tables 4.5.4 and 4.5.5. They are slightly  
 3529 worse sometimes: more false negatives (28 instead of 20), less true positives (74 instead of 78),  
 3530 thus  $NPV$  is lower and the  $FOR$  is higher but the  $PPV$  is higher.

3531 Therefore, it appears that even with  $s_9$  drifting, the diagnosis algorithm still provides satisfying  
 3532 results. In this way, assuming that the instruments of class  $c_{max}$  are not drifting is acceptable.

Delay of positive detection			
$\mu$	$\sigma$	min	max
-2.3	0.9	-4	-1

**Table 4.6.2:** Statistics of the delay of positive detection  $\Delta\mathcal{D}$  for the case study when the instrument of class 1 is drifting

3533 More detailed explanations are provided in Appendix B to show that even when drifting, this  
 3534 instrument is still more accurate than the instruments of a lower class. In fact, the metric  
 3535 used to determine if an instrument is truly faulty or not is not fairly appreciating the status  
 3536 of an instrument when the goal is to compare its measurement results to the ones of another  
 3537 instrument, potentially of a worse quality.

### 3538 6.3 Conclusion

3539 In this section, the need for an assumption regarding the instruments of class  $c_{max}$  to be  
 3540 considered as non-faulty was explained and relaxed to only require this assumption for one of  
 3541 them at least. The case study of Section 5 was extended by making the instrument of class  $c_{max}$   
 3542 drifting. It showed that the results were only slightly worse in this configuration. In conclusion,  
 3543 assuming that instruments of class  $c_{max}$  are always non-faulty is a reasonable hypothesis.

## 3544 7 Means to reduce false results

3545 In this section, we discuss several means to reduce the number of false results that happen  
 3546 for the reasons explained in Section 5.6.

### 3547 7.1 Keep the predicted status of instruments unchanged once they are pre- 3548 dicted as faulty

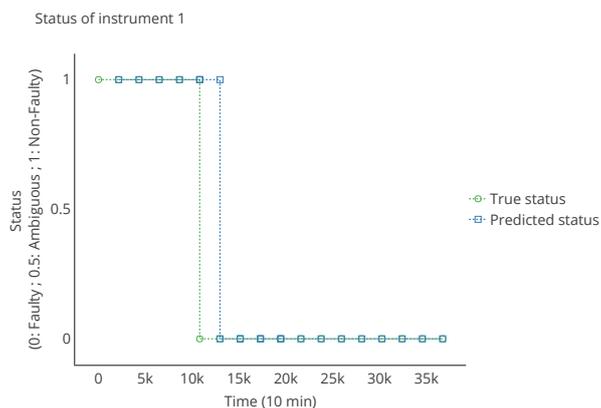
3549 A first idea to overcome faults results could be to keep the predicted status of an instrument  
 3550 equal to faulty for the next diagnosis procedures once it is first diagnosed as so. Figure 4.7.1  
 3551 shows the effect of such choice on the same instrument as in Figure 4.5.2b: here the instrument  
 3552 changes of status only once.

3553 The global performances of the diagnosis algorithm on the case study with this choice of  
 3554 behaviour are shown in Table 4.7.1. The results obtained are improved: there are less false  
 3555 negatives, less non-determined positives and much more true positives. Thus, all the associated  
 3556 metrics based on these metrics are improved. An accuracy of 88% is obtained, while the *PPV*  
 3557 goes up to 1 for instance. Only the *FNR* and *FOR* are still important but this is explained by  
 3558 the late detection of the change of status of the instruments in average (-2.1) as shown in Table  
 3559 4.7.2.

3560 Therefore, this solution of keeping the predicted status of an instrument equal faulty, once it  
 3561 is predicted as so, avoids observing changes of predicted statuses after being predicted as faulty  
 3562 a first time but it does not solve entirely the problem of false results.

### 3563 7.2 Alternate definition for rates used to decide of the status of the instru- 3564 ments

In Section 5.6, we observed that measurement results may be considered multiple times in the  
 computation of the rates  $r_{\Phi_v}^+$ ,  $r_{\Phi_v}^-$  and  $r_{\Phi_v}^{\approx}$  if an instrument encounters multiple other instruments  
 at the same time. It could be preferable to have each measurement result counted only once,



**Figure 4.7.1:** Evolution of the true status and of the predicted status of  $s_1$  while keeping  $\hat{\Omega}(s_1, t) = F$  for  $t \geq t_d$  such as  $\hat{\Omega}(s_1, t_d) = F$  for the first time.

		True status		Prevalence	Accuracy
		Non-faulty $N = 36$	Faulty $P = 117$		
Predicted status	Non-faulty	$TN$ 36	$FN$ 18	$NPV$ 0.67	$FOR$ 0.33
	Ambiguous	$NDN$ 0	$NDP$ 1		
	Faulty	$FP$ 0	$TP$ 98	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.15		
	$NDNR$ 0.00	$NDPR$ 0.01	$NDR$ 0.01		
	$FPR$ 0.00	$TPR$ 0.84			

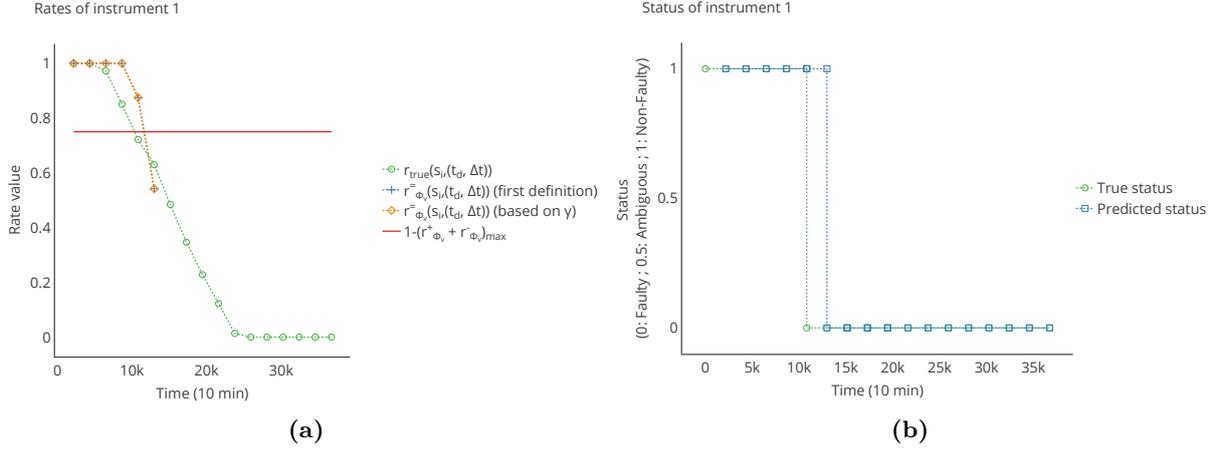
**Table 4.7.1:** Confusion matrix of the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time (metrics are defined in Section 5.4).

notably if  $M(s_i, (t_d, \Delta t))|_{\Phi_v} \sim M(s_i, (t_d, \Delta t))$  in terms of distribution of the measured values. To achieve that, we define  $\gamma(s_i \rightarrow s_j, t)$ , the weight of the rendez-vous  $\varphi(s_i \rightarrow s_j, t)$  :

$$\gamma(\varphi(s_i \rightarrow s_j, t)) = \frac{1}{|\Phi(s_i \rightarrow S, t)|}$$

Delay of positive detection			
$\mu$	$\sigma$	min	max
-2.1	1.2	-4	0

**Table 4.7.2:** Statistics of the delay of positive detection  $\Delta\mathcal{D}$  for the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time



**Figure 4.7.2:** Evolution for  $s_1$  in the case study of Section 5 (a) of the rates  $r_{true}$  and  $r_{\Phi_v}^{\approx}$  (both definitions) ; (b) of its true status and of its predicted status with the alternate definition for  $r_{\Phi_v}^{\approx}$  ; while keeping  $\hat{\Omega}(s_1, t) = F$  for  $t \geq t_d$  such as  $\hat{\Omega}(s_1, t_d) = F$  for the first time.

$\gamma_v$  can be defined for  $\Phi_v$  too. With this weight, we have notably:

$$\sum_{\varphi(s_i \rightarrow s_j, t') \in \Phi_v(s_i, (t, \Delta t))} \gamma_v(\varphi(s_i \rightarrow s_j, t')) = |M(s_i, (t, \Delta t))|_{\Phi_v}|$$

Thus, an alternate definition of  $r_{\Phi_v}^{\approx}(s_i, (t, \Delta t))$  is for instance:

$$r_{\Phi_v}^{\approx}(s_i, (t, \Delta t)) = \frac{\sum_{\varphi(s_i \rightarrow s_j, t') \in \Phi_v^{\approx}(s_i, (t, \Delta t))} \gamma_v(\varphi(s_i \rightarrow s_j, t'))}{\sum_{\varphi(s_i \rightarrow s_j, t') \in \Phi_v(s_i, (t, \Delta t))} \gamma_v(\varphi(s_i \rightarrow s_j, t'))}$$

3566 However, Figure 4.7.2a shows that using this definition in this case study does not change the  
 3567 values of  $r_{\Phi_v}^{\approx}(s_1, (t, \Delta t))$  compared to the results obtained with the previous definition. Hence,  
 3568 there is no change in the predicted status of the instrument that was considered in Figure 4.7.1  
 3569 as shown in Figure 4.7.2b. This is also globally the case according to Table 4.7.3 compared to  
 3570 Table 4.7.1.

### 3571 7.3 Adjustment of the maximal tolerated values for the different rates used 3572 to determine the statuses of the instruments

3573 The issue of false results can be seen from another perspective. In fact, the most problematic  
 3574 false results are the false negatives. Indeed, false positives will only trigger an early maintenance

		True status			
		Non-faulty $N = 36$	Faulty $P = 117$	Prevalence 0.76	Accuracy 0.88
Predicted status	Non-faulty	$TN$ 36	$FN$ 18	$NPV$ 0.67	$FOR$ 0.33
	Ambiguous	$NDN$ 0	$NDP$ 1		
	Faulty	$FP$ 0	$TP$ 98	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.15		
		$NDNR$ 0.00	$NDPR$ 0.01	$NDR$ 0.01	
		$FPR$ 0.00	$TPR$ 0.84		

**Table 4.7.3:** Confusion matrix of the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time and with the alternate definition for  $r_{\Phi_v}^{\approx}$  (metrics are defined in Section 5.4).

3575 of the instruments which is preferable in terms of quality of the measurements. The cost of the  
3576 maintenance has to be taken into account but regarding calibration issues, an *in situ* calibration  
3577 could be performed.

3578 The parameters that can be adjusted to obtain hopefully more false negatives and less false  
3579 positives are the maximal tolerated values  $(r_{\Phi_v}^+)^{max}$ ,  $(r_{\Phi_v}^-)^{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)^{max}$  for the rates  
3580 of upper and lower non-compatible rendez-vous in sets of valid rendez-vous. The lower these  
3581 values are, the less non-compatible result will be accepted.

3582 It is difficult to give a formal procedure to adjust these values and would require additional  
3583 developments. By the way of a possible solution, an empirical approach could be considered if a  
3584 training database or test cases are available. While setting first values for  $(r_{\Phi_v}^+)^{max}$ ,  $(r_{\Phi_v}^-)^{max}$   
3585 and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)^{max}$  arbitrarily, instruments may be diagnosed and the results studied to see if  
3586 the predictions were correct. If it is the case, it means that the maximal tolerated values for the  
3587 rates could be increased if there were only true positives, and decreased if there were only true  
3588 negatives. In the opposite cases, e.g. if there were only false positives or only false negatives, it  
3589 means that these values should be increased and decreased respectively. However, if there are  
3590 both true positives, true negatives, false positives and false negatives, the values for  $(r_{\Phi_v}^+)^{max}$ ,  
3591  $(r_{\Phi_v}^-)^{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)^{max}$  should be adjusted so that the true positives and true negatives are  
3592 maximised and the false positives and false negatives minimised. It may be impossible to find  
3593 values so that there are no false results.

#### 3594 7.4 Conclusion

3595 In this section, two means to reduce false results were presented, one having an interesting  
3596 effect (keeping the predicted status of instruments unchanged once they are predicted as faulty)

3597 and the other one having no influence on the results (an alternate definition for rates used to  
3598 decide of the status of the instruments). Then, we discussed the adjustment of the maximal  
3599 tolerated values for the different rates used to determine the statuses of the instruments because  
3600 it can help to have more false positives and less false negatives. We explained notably why false  
3601 positives were less prejudiciable than false negatives.

3602 In conclusion, we can say that despite a careful design of the sensor network and adjustment  
3603 of the parameters of the diagnosis algorithm, false results may still happen. Therefore, a careful  
3604 investigation must be carried out if a high rate of false results is observed to understand their  
3605 causes.

## 3606 8 Adjustment of the minimal size required for a set of valid rendez-vous 3607 to allow a prediction between the statuses faulty and non-faulty

3608 The choice of the value of  $|\Phi_v|_{min}$  can make the algorithm non-functional: in the worst  
3609 situation, all the instruments, except those of class  $c_{max}$ , are always diagnosed as ambiguous if  
3610 the condition  $|\Phi_v(s_i, (t, \Delta t))| \geq |\Phi_v|_{min}$  is never satisfied for any  $s_i$  to diagnose. Thus, having  
3611 a method to determine first if the value of  $|\Phi_v|_{min}$  is compatible with the considered sensor  
3612 network and then to give an upper boundary for its value would be interesting.

3613 Note that in the case where the condition  $|\Phi_v(s_i, (t, \Delta t))| \geq |\Phi_v|_{min}$  is never satisfied, it is  
3614 also possible to increase  $\Delta t$  so that more rendez-vous could be included in the set  $\Phi_v(s_i, (t, \Delta t))$ .  
3615 We assume here that  $\Delta t$  was defined first to determine a relevant set of rendez-vous based on  
3616 expert knowledge (Section 5.7) and cannot be changed.

### 3617 8.1 Algorithm determining an upper boundary for $|\Phi_v|_{min}$

3618 Consider a training database giving the possible rendez-vous between the instruments of a  
3619 sensor network over a given time range, such as a set  $D$  of diagnosis procedures could be carried  
3620 out on it.

3621 For each  $d \in D$ , it is possible to determine for each instrument the set of its rendez-vous  
3622  $|\Phi(s_i \rightarrow s_j, (t_d, \Delta t))|$ .

Consider the matrix  $\mathbf{Mat}_\Phi(D, \Delta t) \in \mathbb{N}^{|S| \times |S|}$  where its values are defined by:

$$\phi_{i,j} = \min_{d \in D} |\Phi(s_i \rightarrow s_j, (t_d, \Delta t))|$$

3623 Each line of this matrix gives the worst-case scenario of rendez-vous occurring on a range of  
3624 diagnosis according to the training database.

3625 We will now apply a procedure similar to the diagnosis algorithm to obtain a maximum value  
3626 for  $|\Phi_v|_{min}$  such as all the instruments are diagnosable as non-faulty, noted  $\max |\Phi_v|_{min}$ , under a  
3627 number of assumptions.

3628 Consider first the set  $S^{NF} = S^{c_{max}}$  because the instruments of class  $c_{max}$  are assumed as  
3629 non-faulty by the diagnosis algorithm.<sup>43</sup> and the set of instruments to diagnose  $S^D = S \setminus S^{c_{max}}$ .

For each  $s_i \in S^D$ , it is possible to compute:

$$|\Phi_v|_{min}(s_i) = \sum_{\substack{j=1 \\ c(s_j) \geq c_{min}^D(s_i) \\ s_j \in S^{NF}}}^{|S|} \phi_{i,j}$$

3630 The maximum of these  $|\Phi_v|_{min}(s_i)$  gives a first value for the upper boundary  $\max |\Phi_v|_{min}$ . If

---

<sup>43</sup>In the case where only a subset of  $S^{c_{max}}$  is assumed as non-faulty (see Section 6), only the instruments of this subset should be in  $S^{NF}$ .

we suppose that all the rendez-vous are likely to be valid in the worst case scenario and that all the instruments are likely to be non-faulty, then with  $|\Phi_v|_{min} = \max |\Phi_v|_{min}$ , all the  $s_i$  such as  $|\Phi_v|_{min}(s_i) = \max |\Phi_v|_{min}$  would be predicted as non-faulty. Representing this set of instruments by  $\widehat{S}^{NF}$ , we would have in a following iteration of the diagnosis algorithm  $S^{NF} = S^{NF} \cup \widehat{S}^{NF}$  and  $S^D = S^D \setminus \widehat{S}^{NF}$ .

$|\Phi_v|_{min}(s_i)$  can be computed again for the remaining instruments in  $S^D$ . If the maximum of these new values is lower than the previous value obtained for  $\max |\Phi_v|_{min}$ , then  $\max |\Phi_v|_{min}$  is set to this new maximum so that some of the remaining instruments in  $S^D$  can be diagnosed in the worst case scenario of rendez-vous during a second iteration of the algorithm. This procedure is repeated until  $S^D$  is empty. The value of  $\max |\Phi_v|_{min}$  obtained at the end is the maximal value for  $|\Phi_v|_{min}$  ensuring that all the instruments can be diagnosed as non-faulty when they are likely to be so in the worst case scenario of rendez-vous.

However, some instruments can be diagnosed as faulty in practice. In this case, the value  $\max |\Phi_v|_{min}$  obtained previously may not be reachable for some instruments.

To overcome this issue, we can compute  $\max |\Phi_v|_{min}$  as a function of the number  $\lambda$  of instruments predicted as faulty in the network.

Consider  $\mathcal{C}^F(S^{(c_{max}-1)^-}, \lambda)$  that is the set of all the possible combinations of  $\lambda$  faulty instruments in  $S^{(c_{max}-1)^-}$ . For each set of instruments that are faulty  $S^F \in \mathcal{C}^F(S^{(c_{max}-1)^-}, \lambda)$ , it is possible to determine  $\max |\Phi_v|_{min}(S^F)$  following the procedure described previously by initialising  $S^D$  to  $S \setminus (S^{c_{max}} \cup S^F)$ . For  $\lambda > 0$ , it is possible that some instruments have no rendez-vous with the non-faulty instruments depending on  $S^F$ . Thus, the end of the procedure described previously is reached when  $S^D$  is empty or when  $S^D$  is identical after an iteration.

Then:

$$\max |\Phi_v|_{min}(\lambda) = \min_{S^F \in \mathcal{C}^F(S^{(c_{max}-1)^-}, \lambda)} (\max |\Phi_v|_{min}(S^F))$$

is the maximal value allowed for  $|\Phi_v|_{min}$  considering  $\lambda$  instruments as faulty.

However, it is possible that  $\max |\Phi_v|_{min}(\lambda) > \max |\Phi_v|_{min}(\lambda')$  for  $\lambda' < \lambda$ .

Thus:

$$\max |\Phi_v|_{min}(\lambda) = \min_{\lambda' \in [0.. \lambda]} (\max |\Phi_v|_{min}(\lambda'))$$

Algorithm 2 gives a pseudo-code to determine the values of  $\max |\Phi_v|_{min}$  as a function of the number of instruments considered as faulty  $\lambda$ .

---

**Algorithm 2:** Algorithm to determine the maximal value of the minimal size required for a set of valid rendez-vous to allow prediction between the statuses faulty and non-faulty for an instrument, as a function of the number of faulty instruments in the sensor network

---

**Data:**  $S, \{\Phi(S, (t_d, \Delta t)), d \in D\}$

**Result:**  $\max |\Phi_v|_{min}$

/\* Compute the matrix of the minimal numbers of rendez-vous encountered on the diagnosis ranges between each couple of instruments \*/

**for**  $i \in [1..|S|]$  **do**

**for**  $j \in [1..|S|]$  **do**  
    |  $\phi_{i,j} \leftarrow \min_{d \in D} |\Phi(s_i \rightarrow s_j, (t_d, \Delta t))|$   
    **end**

**end**

/\* Determine the maximal value allowed for  $|\Phi_v|_{min}$  as a function of the number of faulty instruments \*/

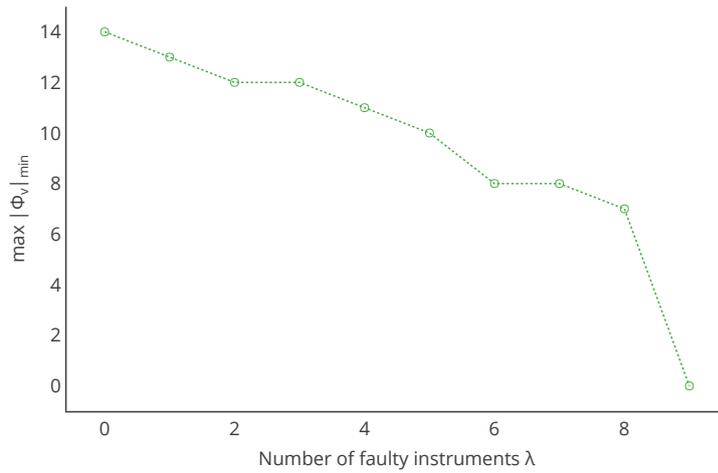
---

```

for  $\lambda \in [0..|S^{(c_{max}-1)^-}|]$  do /* For each number of instruments that can be
faulty                                                                    */
     $\max |\Phi_v|_{min}(\lambda) \leftarrow +\infty$ 
    /* For each combination of  $\lambda$  instruments that can be faulty          */
    for  $S^F \in \mathcal{C}^F(S^{(c_{max}-1)^-}, \lambda)$  do
         $S^{NF} \leftarrow S^{c_{max}}$ 
         $S^D \leftarrow S \setminus (S^{c_{max}} \cup S^F)$ 
         $\max |\Phi_v|_{min}(S^F) \leftarrow +\infty$ 
        /* Determine the maximal value for  $|\Phi_v|_{min}$  according to  $S^F$  so that
           all the instruments in  $S^D$  can be predicted as non-faulty      */
        repeat
             $\tilde{S}^D \leftarrow S^D$ 
             $\max |\Phi_v|_{min}(S^F)' \leftarrow 0$ 
            for  $s_i \in S^D$  do
                 $\max |\Phi_v|_{min}(s_i) \leftarrow \sum_{\substack{j=1 \\ c(s_j) \geq c_{min}^D(s_i) \\ s_j \in S^{NF}}}^{|S|} \phi_{i,j}$ 
                if  $\max |\Phi_v|_{min}(s_i) > \max |\Phi_v|_{min}(S^F)'$  then
                     $\max |\Phi_v|_{min}(S^F)' \leftarrow \max |\Phi_v|_{min}(s_i)$ 
                     $\hat{S}^{NF} \leftarrow \{s_i\}$ 
                else if  $\max |\Phi_v|_{min}(s_i) = \max |\Phi_v|_{min}(S^F)'$  then
                     $\hat{S}^{NF} \leftarrow \hat{S}^{NF} \cup \{s_i\}$ 
                end
            if  $\max |\Phi_v|_{min}(S^F)' < \max |\Phi_v|_{min}(S^F)$  then
                 $\max |\Phi_v|_{min}(S^F) \leftarrow \max |\Phi_v|_{min}(S^F)'$ 
            end
             $S^{NF} \leftarrow S^{NF} \cup \hat{S}^{NF}$ 
             $S^D \leftarrow S^D \setminus \hat{S}^{NF}$ 
        until  $S^D = \emptyset$  or  $S^D = \tilde{S}^D$ 
        /* If this new maximal value for  $|\Phi_v|_{min}$  is lower than the current
           one of  $\max |\Phi_v|_{min}(\lambda)$ , then it is the new maximal value allowed
           for  $|\Phi_v|_{min}$  with  $\lambda$  faulty instruments                    */
        if  $\max |\Phi_v|_{min}(S^F) < \max |\Phi_v|_{min}(\lambda)$  then
             $\max |\Phi_v|_{min}(\lambda) \leftarrow \max |\Phi_v|_{min}(S^F)$ 
        end
        /* If the maximal value allowed for  $|\Phi_v|_{min}$  with  $\lambda$  faulty
           instruments is higher than the maximal value allowed for a lower
           number of instruments, then it is this value that is the maximal
           value allowed for  $|\Phi_v|_{min}$  with  $\lambda$  faulty instruments      */
        if  $\lambda > 0$  and  $\max |\Phi_v|_{min}(\lambda - 1) < \max |\Phi_v|_{min}(\lambda)$  then
             $\max |\Phi_v|_{min}(\lambda) \leftarrow \max |\Phi_v|_{min}(\lambda - 1)$ 
        end
    end
end

```

---



**Figure 4.8.1:** Evolution of  $\max |\Phi_v|_{\min}$  as a function of the number of faulty instruments  $\lambda$  in the network, based on the matrix  $\mathbf{Mat}_\Phi(D, \Delta t)$  derived from the database used for the case study of Section 5

3657

## 8.2 Application to the case study

Consider the following matrix  $\mathbf{Mat}_\Phi(D, \Delta t)$  obtained with the database used in the case study of Section 5:

$$\mathbf{Mat}_\Phi(D, \Delta t) = \begin{pmatrix} 0 & 13 & 11 & 8 & 4 & 9 & 10 & 8 & 9 & 13 \\ 13 & 0 & 7 & 12 & 7 & 11 & 12 & 10 & 9 & 12 \\ 11 & 7 & 0 & 9 & 10 & 9 & 11 & 11 & 12 & 10 \\ 8 & 12 & 9 & 0 & 9 & 9 & 12 & 9 & 13 & 8 \\ 4 & 7 & 10 & 9 & 0 & 8 & 13 & 11 & 13 & 7 \\ 9 & 11 & 9 & 9 & 8 & 0 & 10 & 8 & 13 & 11 \\ 10 & 12 & 11 & 12 & 13 & 10 & 0 & 14 & 9 & 12 \\ 8 & 10 & 11 & 9 & 11 & 8 & 14 & 0 & 13 & 14 \\ 9 & 9 & 12 & 13 & 13 & 13 & 9 & 13 & 0 & 8 \\ 13 & 12 & 10 & 8 & 7 & 11 & 12 & 14 & 8 & 0 \end{pmatrix}$$

3658

The last row corresponds to the instrument of class  $c_{\max} = 1$ .

3659

Applying the considered algorithm to determine  $\max |\Phi_v|_{\min}$  as a function of  $\lambda$  gives the Figure 4.8.1.

3660

3661

First of all, note that the overall upper boundary for  $|\Phi_v|_{\min}$  given by this algorithm is equal to 14, which is lower than the value used in the case study in Section 5: this value was set to 15. This could explain some of the predictions that were observed of instruments diagnosed as ambiguous. Hence, the algorithm we provided allow to identify if a value of  $|\Phi_v|_{\min}$  is likely to lead to instruments predicted as ambiguous.

3662

3663

3664

3665

Suppose that we want to be still able to diagnose the sensor network without any ambiguous predicted statuses due to a lack of valid rendez-vous when five instruments are predicted as faulty for instance. With Figure 4.8.1, then  $|\Phi_v|_{\min}$  must be lower or equal to 10 according to the training database used.

3666

		True status		Prevalence	Accuracy
		Non-faulty $N = 36$	Faulty $P = 117$		
Predicted status	Non-faulty	$TN$ 36	$FN$ 14	$NPV$ 0.72	$FOR$ 0.28
	Ambiguous	$NDN$ 0	$NDP$ 0		
	Faulty	$FP$ 0	$TP$ 103	$FDR$ 0.00	$PPV$ 1.00
			$TNR$ 1.00	$FNR$ 0.12	
		$NDNR$ 0.00	$NDPR$ 0.00	$NDR$ 0.00	
		$FPR$ 0.00	$TPR$ 0.88		

**Table 4.8.1:** Confusion matrix of the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time and  $|\Phi_v|_{min} = 10$  (metrics are defined in Section 5.4).

Delay of positive detection			
$\mu$	$\sigma$	min	max
-1.6	1.2	-3	0

**Table 4.8.2:** Statistics of the delay of positive detection  $\Delta D$  for the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time,  $|\Phi_v|_{min} = 10$

3670 Consider now  $|\Phi_v|_{min} = 10$  in our case study with the predicted status of the instrument kept  
3671 unchanged once they are equal to faulty (Section 7). The execution of the diagnosis algorithm  
3672 with this value for  $|\Phi_v|_{min}$  gives the results of Tables 4.8.1 and 4.8.2. We observe that the results  
3673 are significantly improved from what was obtained previously (Tables 4.7.1 and 4.7.2). There are  
3674 less false negatives, no more non-determined positives, and more true positives. The  $PPV$  is  
3675 now equal to 1 and the  $NPV$  increased from 0.67 to 0.72. The accuracy is now equal to 0.91  
3676 instead of 0.88. Moreover, the average delay of positive detection has been improved.

3677 Therefore, changing the value of  $|\Phi_v|_{min}$  is necessary if the algorithm is not functional at  
3678 all (e.g. when  $\hat{\Omega}(s_i, t) = A, \forall s_i$  and  $t$ ) but it can also be useful to improve the performances of  
3679 the algorithm. In our case, a lower value of  $|\Phi_v|_{min}$  is interesting.<sup>44</sup> The algorithm provided to  
3680 determine  $\max |\Phi_v|_{min}$  as a function of the number of faulty instruments in the sensor network  
3681 is adapted to help to find a suitable  $|\Phi_v|_{min}$ .

<sup>44</sup>Note that a higher (or lower) value than 10 for  $|\Phi_v|_{min}$  could give better results. The optimisation of this value was not addressed in this work.

### 8.3 Conclusion

In this section, an algorithm was provided to give an upper boundary to  $|\Phi_v|_{min}$ , the minimal size for a set of valid rendez-vous to allow a prediction between the statuses faulty and non-faulty. We defined this upper boundary  $\max |\Phi_v|_{min}$  as a function of the number of instruments already diagnosed as faulty in the sensor network to adjust  $|\Phi_v|_{min}$  with concerns for the robustness of the diagnosis algorithm when instruments are diagnosed as faulty.

Afterwards, this algorithm was applied on the dataset used in the case studies presented in this chapter. It allowed first to identify that the value of  $|\Phi_v|_{min}$  initially chosen was higher than the overall upper boundary given by the algorithm. This could explain the predictions of instruments as ambiguous in previous results. Then, by choosing a lower value for  $|\Phi_v|_{min}$  to tolerate five instruments predicted as faulty in previous diagnosis procedures, the results of the case study were improved, showing the interest of the algorithm developed in this section.

## 9 Sensitivity of the algorithm to changes in the case study

In the previous sections, the diagnosis algorithm was introduced and a case study was presented to illustrate the application of the algorithm. The means to reduce false results were discussed just like the assumption regarding the instruments of class  $c_{max}$ . In Chapter 3, the influence on the results of various aspects of the case study was studied for the evaluation of *in situ* calibration strategies. In this section, the same type of studies are conducted regarding the case study that was presented in Section 5. Indeed, results are likely to be case specific according to the results of the two previous sections where we saw that the choice of the values of the parameters  $|\Phi_v|_{min}$ ,  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$ ,  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  and so on could change results obtained.

In this section, the influence of the drift values, of the true values and of the model used to build the true values is investigated. The role of the density of instruments is also questioned and a discussion regarding the robustness of the algorithm against other faults is provided.

In the following studies, **the parameters of the algorithm that allowed obtaining the best results in Section 8 are used**, e.g. with the predicted statuses kept as faulty from one diagnosis procedure to another, once the status of an instrument is predicted as faulty, and with  $|\Phi_v|_{min} = 10$ . It is explained when the value of a parameter is changed.

### 9.1 Influence of the values of drift

As the results of Section 8 and the following were obtained for one draw of drift values, the case study was repeated for 100 draws. Table 4.9.1 presents the mean and standard deviation of each metric.<sup>45</sup> This table shows that:

- the mean values are equal to the values obtained in Table 4.8.1. This indicates that on average, the observations and conclusions do not depend on the variability of the drift.
- the standard deviations of all the metrics are close to zero. This is small considering the mean values and the relative standard deviation varying from zero to at most 16% of the mean value, reached for  $FN$ .

Thus, in this case study, the results of the diagnosis algorithm are not dependent on the values of the drift<sup>46</sup>, with this drift model at least, and with this configuration of the algorithm.

<sup>45</sup>The values for the delay of positive detection  $\Delta\mathcal{D}$  are not provided here for the sake of clarity.

<sup>46</sup>Note that the range of magnitude of drift does not change here, only the variability is studied. A different range would be equivalent to operating with different measuring instruments in practice.

		True status				Prevalence		Accuracy	
		Non-faulty		Faulty					
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		36.00	0.00	117.00	0.00	0.76	0.00	0.91	0.00
Predicted status		<i>TN</i>		<i>FN</i>		<i>NPV</i>		<i>FOR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Non-faulty		36.00	0.00	13.88	0.69	0.72	0.01	0.28	0.01
Predicted status		<i>NDN</i>		<i>NDP</i>		<i>FDR</i>		<i>PPV</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$				
Ambiguous		0.00	0.00	0.06	0.34	0.00	0.00	1.00	0.00
Predicted status		<i>FP</i>		<i>TP</i>		<i>FNR</i>		<i>NDR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$				
Faulty		0.00	0.34	103.06	0.34	0.00	0.00	0.00	0.00
		<i>TNR</i>		<i>FNR</i>		<i>NDR</i>		<i>NDR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		1.00	0.00	0.12	0.01	0.00	0.00	0.00	0.00
		<i>NDNR</i>		<i>NDPR</i>		<i>NDR</i>		<i>NDR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		<i>FPR</i>		<i>TPR</i>		<i>NDR</i>		<i>NDR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00

**Table 4.9.1:** Confusion matrix for 100 simulations of the case study with drift values drawn again for each one (metrics are defined in Section 5.4).

## 9.2 Influence of the true values

3721

3722

3723

3724

3725

3726

3727

3728

In the same way, the results of Section 8 were obtained for one draw of true values. Therefore, the case study was repeated for 100 draws of true values. Table 4.9.2 presents the mean and standard deviation of each metric.<sup>47</sup> This table shows that on average the number of true positives and false negatives are respectively more and less important than with the initial draw. The standard deviations are also close to zero which is small compared to most of the mean values, except for the metrics *FP*, *NDP* and *FN*, plus some the associated metrics derived from them (*FNR*, *FPR*, *FOR*).

3729

3730

3731

3732

3733

Thus, again on average, the results obtained are globally better compared to the specific case study considered previously, in terms of *NPV*, *FOR*, accuracy and threat score notably.

In conclusion, the results obtained are dependent of the true values, at least when this model is considered. This is consistent with the studies carried out in Section 5.6 where we showed that the false results depends mostly on the measurement results associated to the rendez-vous.

<sup>47</sup>The values for the delay of positive detection  $\Delta\mathcal{D}$  are not provided here for the sake of clarity.

		True status				Prevalence		Accuracy	
		Non-faulty		Faulty					
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		36.47	1.38	116.53	1.38	0.76	0.01	0.95	0.02
Predicted status		<i>TN</i>		<i>FN</i>		<i>NPV</i>		<i>FOR</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Non-faulty		35.59	1.19	7.40	2.95	0.83	0.06	0.17	0.06
Ambiguous		<i>NDN</i>		<i>NDP</i>					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.00	0.00	0.10	0.33				
Faulty		<i>FP</i>		<i>TP</i>		<i>FDR</i>		<i>PPV</i>	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		0.88	2.75	109.03	2.75	0.01	0.01	0.99	0.01
		<i>TNR</i>		<i>FNR</i>					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.98	0.03	0.06	0.03				
		<i>NDNR</i>		<i>NDPR</i>		<i>NDR</i>			
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$		
		0.00	0.00	0.00	0.00	0.00	0.00		
		<i>FPR</i>		<i>TPR</i>					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.02	0.03	0.94	0.03				

**Table 4.9.2:** Confusion matrix for 100 simulations of the case study with true values following a 2D Gaussian model which parameters are randomly drawn for each simulation (metrics are defined in Section 5.4).

### 9.3 Influence of the model used for the true values

To reinforce the results on the influence of the values, the model used to derive the true values is changed in this section. Because the measurement results of the rendez-vous are samples of all the measurement results acquired during a time range, and because the predicted statuses are determined based on sets without taking into account the actual order of the measurements, we propose to randomly draw the true values at each time step and at each position following the uniform law  $\mathcal{U}(0, 400)$ .

The case study with this model for the true values was repeated for 100 draws of true values. Table 4.9.3 presents the mean and standard deviation of each metric.<sup>48</sup> The results of this table are similar to the ones of Table 4.9.2 in Section 9.2, except regarding false positives that are slightly higher. In general, the orders of magnitude are similar for the metrics. Thus, the model chosen for the true values does not seem to have a significant effect on the results. This statement does not take into account the fact that only random models were used here because of the

<sup>48</sup>The values for the delay of positive detection  $\Delta\mathcal{D}$  are not provided here for the sake of clarity.

		True status				Prevalence		Accuracy	
		Non-faulty		Faulty					
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		63.43	0.57	89.57	0.57	0.59	0.00	0.92	0.02
Predicted status		TN		FN		NPV		FOR	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Non-faulty		59.97	2.43	8.07	3.39	0.88	0.04	0.12	0.04
Ambiguous		NDN		NDP					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.00	0.00	0.15	0.36				
Faulty		FP		TP		FDR		PPV	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		3.46	3.42	81.35	3.42	0.04	0.02	0.96	0.02
		TNR		FNR					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.95	0.04	0.09	0.04				
		NDNR		NDPR		NDR			
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
		0.00	0.00	0.00	0.00	0.00	0.00		
		FPR		TPR					
		$\mu$	$\sigma$	$\mu$	$\sigma$				
		0.05	0.04	0.91	0.04				

**Table 4.9.3:** Confusion matrix for 100 simulations of the case study with true values randomly drawn following a uniform law (metrics are defined in Section 5.4).

3747 mathematical objects used by the diagnosis algorithm, e.g. sets. Indeed, the order between the  
3748 measured values is not exploited. We even observed in Section 5.6 that by considering only the  
3749 measurement results obtained when an instrument is in rendez-vous to predict its status, we were  
3750 actually performing a sampling of all its measurement results. This already breaks the continuity  
3751 of the measured signal. In Appendix C, an additional case study is provided to confirm that the  
3752 realism of the model for the true values does not have a major influence. Therefore, whether or  
3753 not the relationship between two consecutive measured values is realistic, this has no significance  
3754 for the diagnosis algorithm: it is aimed at determining a status over a time range and this time  
3755 range is considered as a whole. In this way, it is the values that have the most significant effect  
3756 as shown earlier in Section 5.6.

#### 3757 9.4 Influence of the density of instruments

3758 In previous studies, few ambiguous statuses were predicted. This is principally due to the  
3759 fact that most of the instruments of class 0 have often rendez-vous with the instruments of class  
3760 1 so that the size of the set of rendez-vous of an instrument  $s_i$  of class 0 with instruments of  
3761 class 1  $\Phi(s_i \rightarrow S^1, (t_d, \Delta t))$  always satisfies  $|\Phi(s_i \rightarrow S^1, (t_d, \Delta t))| \geq |\Phi_v|_{min}$ . In this way, with

3762 at most one or two executions of the main loop of the algorithm, all the instruments can be  
3763 diagnosed faulty or non-faulty. However, if instruments were meeting less often, we expect that  
3764 more ambiguous statuses would be produced.

3765 Consequently, to simulate the same sensor network of 10 instruments having less rendez-vous,  
3766 the number of accessible positions was increased from 100 to 2025<sup>49</sup>. From Figure 4.9.1, we  
3767 can observe that with 225 positions the values of the metrics deteriorate compared to the case  
3768 with 100 positions and are even poorer for 400 positions and more. This can be observed just  
3769 from the *ACC* going from more than 0.7 for 225 positions to less than 0.2 afterwards but the  
3770 other metrics give interesting insights. Indeed, from Figure 4.9.1a, we observe that for more than  
3771 400 positions, there are only *NDP* and *NDN*. This explains why the *TPR*, *FPR*, *TNR* and  
3772 *FNR* are equal to zero, and the *PPV*, *FDR*, *NPV* and *FOR* are not defined for more than  
3773 400 positions, in Figures 4.9.1b and 4.9.1c respectively. This does not only indicate that the  
3774 diagnosis algorithm provides poor results but also that there is something making the algorithm  
3775 unable to predict a different status than ambiguous for the measuring instruments.

3776 Therefore, the number of positions has an influence on the performances. This indicates  
3777 that the number of rendez-vous is at the heart of the diagnosis algorithm. If the number of  
3778 rendez-vous taking place cannot be increased by performing changes in the sensor network for  
3779 instance, the parameter that must be adjusted to have better performances is  $|\Phi_v|_{min}$ . Indeed, if  
3780 a lower value of  $|\Phi_v|_{min}$  can be acceptable, a sensor network with less rendez-vous occurring may  
3781 be diagnosable.

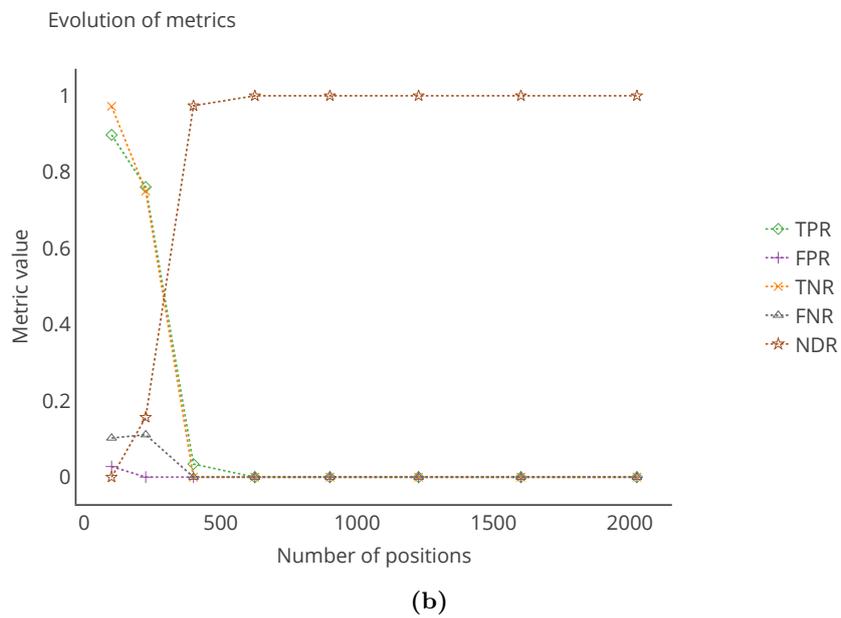
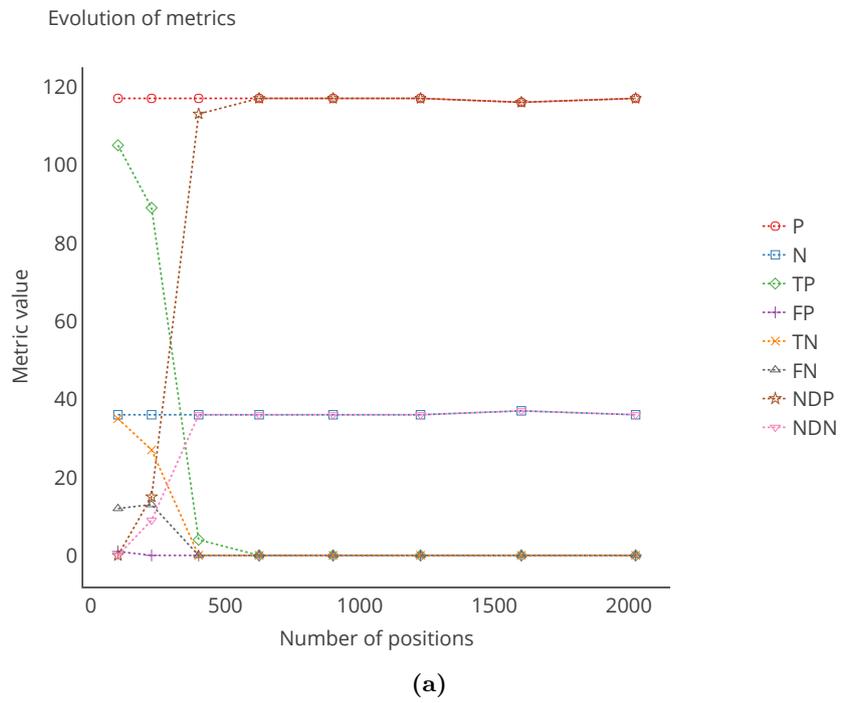
3782 In our case, while still varying the number of positions from 100 to 2025,  $|\Phi_v|_{min}$  is set to  
3783 6 instead of 10. Figure 4.9.2 shows that even better results of diagnosis are obtained with 225  
3784 positions than with 100 (more *TP*, less *FN*, thus a higher *TPR* and *PPV* and a lower *FNR*  
3785 and *FOR*, resulting in a higher *ACC* value notably).

3786 While the diagnosis algorithm still provides acceptable results with 400 positions, although  
3787 non-determined positives and negatives begin to occur, the performances remain still very  
3788 poor for a higher number of positions as the number of ambiguous predictions drastically  
3789 increases according to the evolution of *NDP* and *NDN* and the *NDR*. Thus, the value of *ACC*  
3790 deteriorates more slowly than in Figure 4.9.1 but still very quickly.

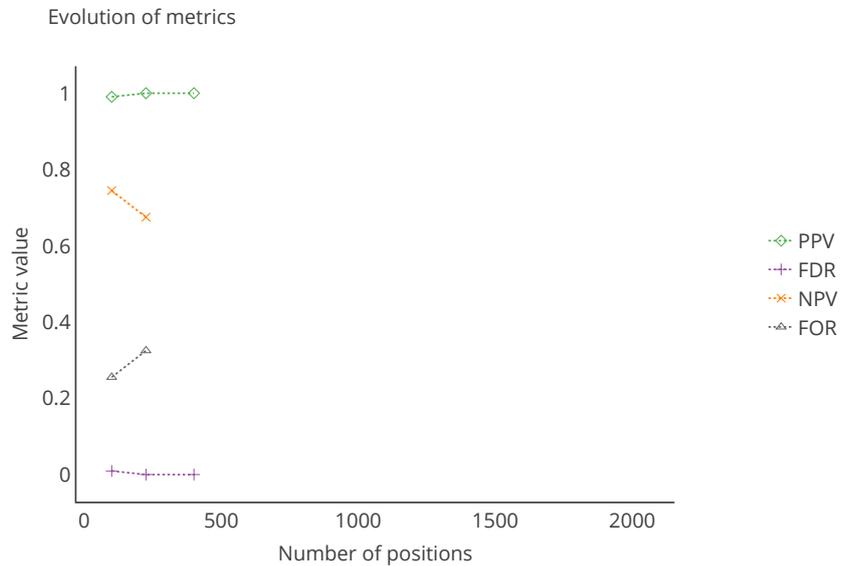
3791 Therefore, the diagnosis algorithm depends significantly of the density of instruments but  
3792 more in terms of rendez-vous than of a physical density of instrument. Indeed, in Section 5, we  
3793 defined the positions of the instruments in an area of  $1 \times 1$ km. However, we could have chosen  
3794 any other unit or subunit of length: this would have not changed the results we obtained. In  
3795 practice, it means the sensor network can cover any area as long as the nodes meet regularly for  
3796 diagnosis purposes.

---

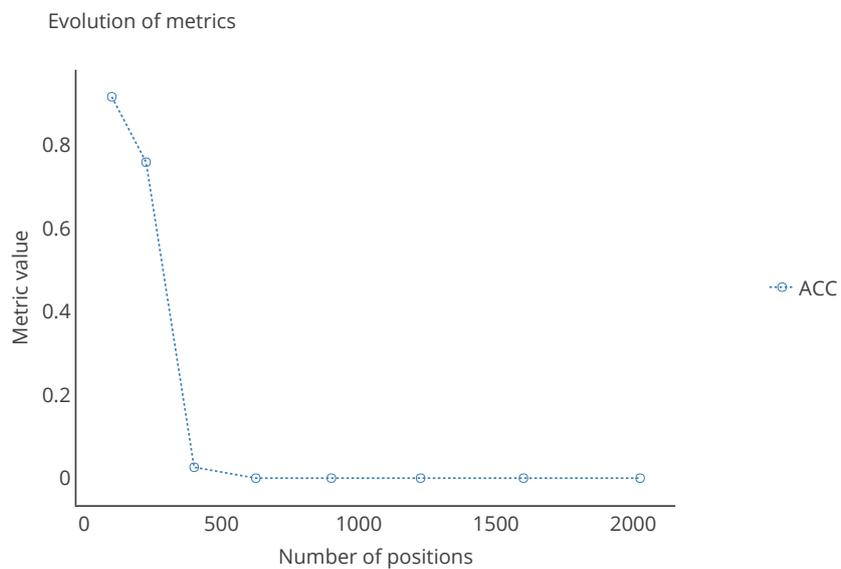
<sup>49</sup>The formula used to define these numbers of positions was the following:  $((1 + 0.5 \cdot i) \times 10)^2$  with  $i \in [0..7]$ .



**Figure 4.9.1:** Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with  $|\Phi_v|_{min} = 10$  (metrics are defined in Section 5.4).  
(continued)

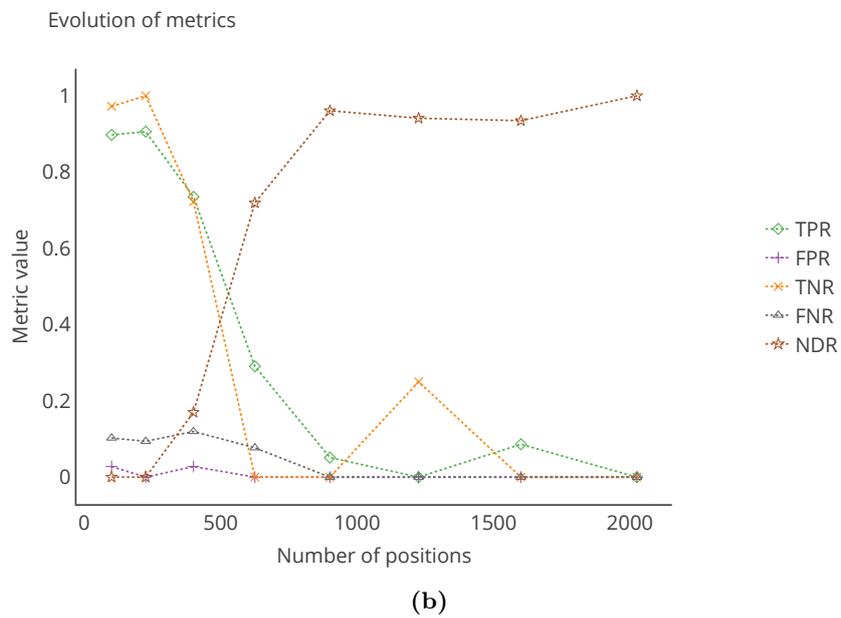
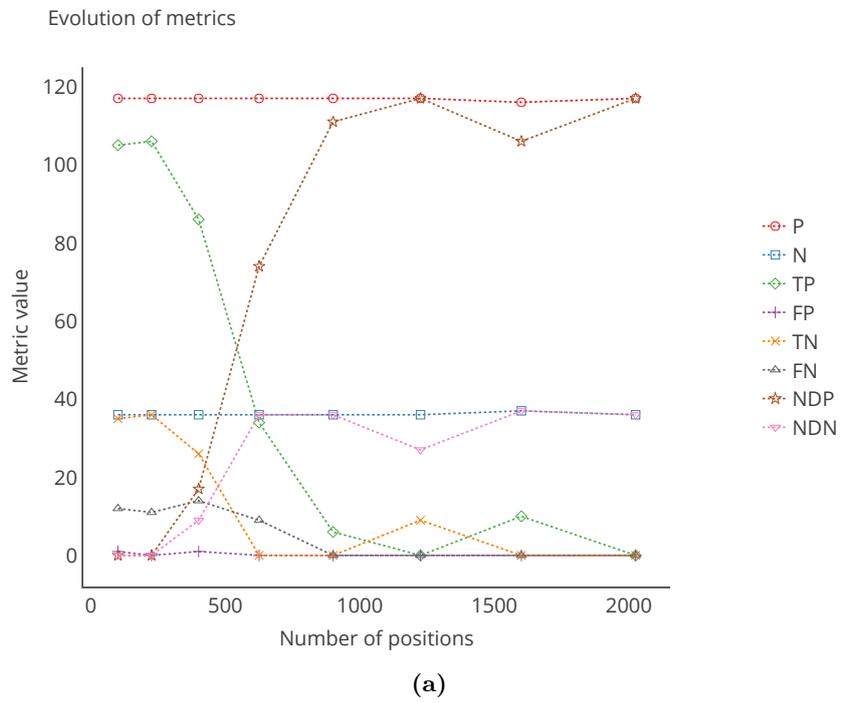


(c)

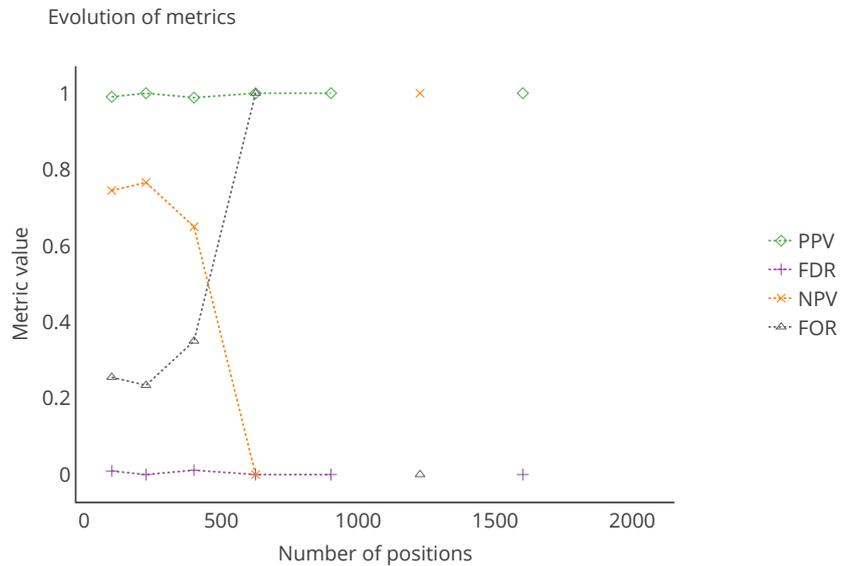


(d)

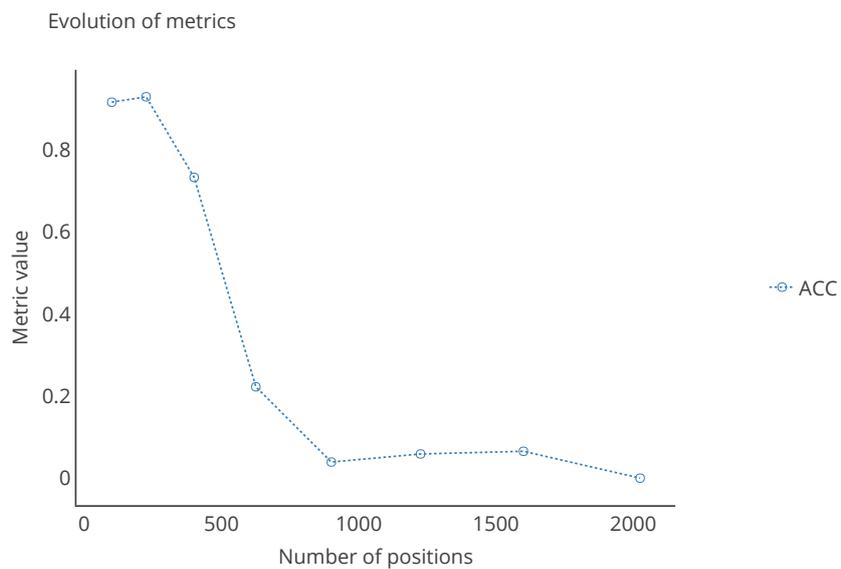
**Figure 4.9.1:** Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with  $|\Phi_v|_{min} = 10$  (metrics are defined in Section 5.4).



**Figure 4.9.2:** Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with  $|\Phi_v|_{min} = 6$  (metrics are defined in Section 5.4).  
(continued)



(c)



(d)

**Figure 4.9.2:** Evolution of the metrics computed over all the diagnosis procedures as a function of the number of accessible positions with  $|\Phi_v|_{min} = 6$  (metrics are defined in Section 5.4).

### 9.5 Influence of other faults

Concerning the influence of other faults on the results of diagnosis procedures, a reasoning similar to the one carried out in Section 5.6 for the study of false results can be made. Indeed, measurement results affected by other faults are likely to generate non-compatible measurement results with their true values. We can expect it is also the case regarding other instruments when such measurement results are associated to rendez-vous. Thus, it is very likely that, when they are valid, the rendez-vous of an instrument  $s_i$  associated to such measurement results affected by other faults are not belonging to the set of compatible valid rendez-vous of  $s_i$ ,  $\Phi_v^{\approx}(s_i, (t_d, \Delta t))$ . Therefore, regarding the rates  $r_{\Phi_v^{\approx}}(s_i \rightarrow S, (t_d, \Delta t))$ ,  $r_{\Phi_v^+}(s_i \rightarrow S, (t_d, \Delta t))$  and  $r_{\Phi_v^-}(s_i \rightarrow S, (t_d, \Delta t))$  and their maximal associated values, the presence of measurement results affected by other faults in valid rendez-vous for an instrument will tend to generate positive results if they are not treated in advance, e.g. the instruments should be predicted as faulty earlier.

To illustrate this, spikes and noise were added to the measured values following the same drift model and with the same parameters that were used in Chapter 3 Section 5.3.1. The results of the diagnosis are given in Tables 4.9.4 and 4.9.5. From them, we observe that  $N$  is lower and  $P$  is higher than in Table 4.8.1. Therefore, the instruments should be considered as faulty earlier. Regarding all the other metrics, their values are still good and similar to the results of Table 4.8.1. Thus, it appears that the supplementary faulty measured values are not affecting significantly the results of the diagnosis algorithm. More importantly, despite a slightly higher standard deviation than in Table 4.8.2, the average delay of positive detection is close to zero and even positive. Its maximal value is also positive. It indicates that the prediction of instruments as faulty happens earlier than in Section 8.

To conclude, the proposed diagnosis algorithm targeting initially drift faults is robust to other faults. Although instruments are predicted as faulty more times, regardless if such a prediction is true or false, they are less harmful than false negative results in the context of drift diagnosis. Indeed, it calls for an earlier maintenance of the devices in this case instead of a late one.

### 9.6 Conclusion

In this section, several studies were conducted to investigate the sensitivity of the diagnosis algorithm to different effects. The following insights were brought:

- the results are not dependent on the values of the drift
- the true values, and thus the measurement results involved in the rendez-vous, have an influence on the results. This is consistent with previous results of Section 5.6
- the model used to derive the true values have less influence on the results than the values themselves
- the density of instruments is critical in terms of frequency of rendez-vous with other instruments
- the algorithm is robust to other faults. Their only effect if they are not managed prior to a drift diagnosis procedure is to make the predictions of instruments as faulty happening earlier. This could generate false positive results, which are less prejudicial than false negatives

		True status		Prevalence	Accuracy
		Non-faulty $N = 23$	Faulty $P = 130$		
Predicted status	Non-faulty	$TN$ 20	$FN$ 11	$NPV$ 0.65	$FOR$ 0.35
	Ambiguous	$NDN$ 0	$NDP$ 2		
	Faulty	$FP$ 3	$TP$ 117	$FDR$ 0.03	$PPV$ 0.97
		$TNR$ 0.87	$FNR$ 0.08		
	$NDNR$ 0.00	$NDPR$ 0.02	$NDR$ 0.01		
	$FPR$ 0.13	$TPR$ 0.90			

**Table 4.9.4:** Confusion matrix of the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time,  $|\Phi_v|_{min} = 10$  and spikes and noise added to the measured values (metrics are defined in Section 5.4).

Delay of positive detection			
$\mu$	$\sigma$	min	max
0.2	1.7	-2	3

**Table 4.9.5:** Statistics of the delay of positive detection  $\Delta\mathcal{D}$  for the case study with the predicted statuses kept as faulty from one diagnosis procedure to another once the status of an instrument is predicted as faulty for the first time,  $|\Phi_v|_{min} = 10$  and spikes and noise added to the measured values

## 10 Combination with a simple calibration approach

The diagnosis algorithm is primarily oriented at detecting when instruments require calibration. We investigate now how this algorithm behaves when calibration is carried out on an instrument after it was predicted as faulty. The diagnosis algorithm is configured as in Section 5, except for  $|\Phi_v|_{min}$  which is set to 10 as a consequence of the study in Section 8. Therefore, **the status of an instrument that is predicted as faulty can change** as it is the trigger of a calibration procedure.

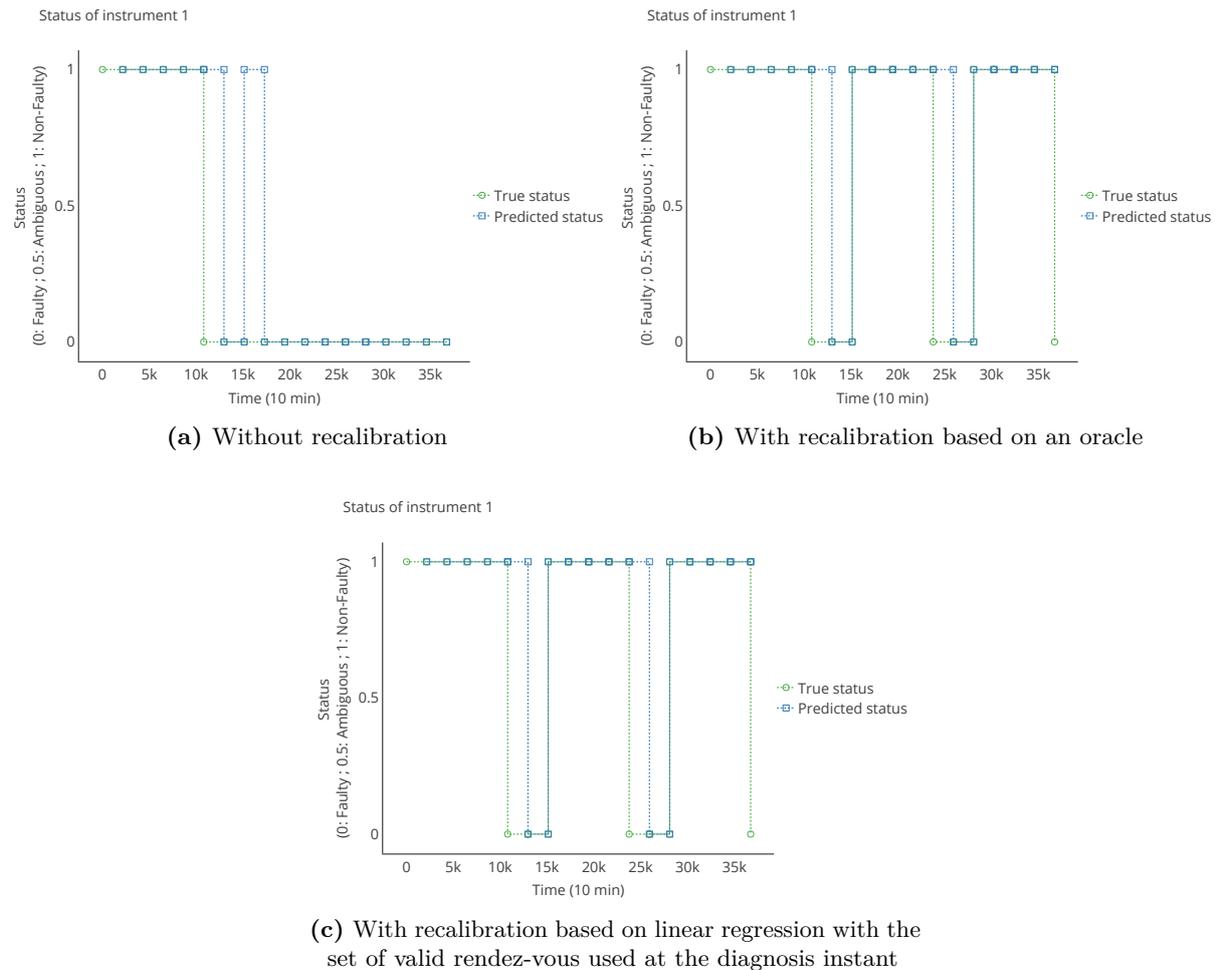
Three cases are considered when an instrument is detected as faulty:

- No calibration is carried out (equivalent to what was done in Section 5)
- Once an instrument is predicted as faulty, its gain and offset are respectively reinitialised to 1 and 0. It is equivalent to having an oracle that knows the gain and offset of each instrument at each instant. That is considered as a perfect calibration.
- A linear regression is carried out based on the couples of measured values of the faulty instruments and its diagnosers that are in the set of valid rendez-vous used for the diagnosis. This technique is usually one of the firsts to be tested for the calibration of measuring instruments [139].

Results for a single instrument are displayed in Figure 4.10.1. The instrument is recalibrated right after being diagnosed as faulty for both recalibration methods. Nevertheless, it is only visible at the next diagnosis procedure. Also from these curves we cannot assess the quality of the calibration, it only shows that it was efficient enough to be again above the threshold mentioned in the specification of the case study.

Table 4.10.1 represents the results for the case without calibration. Tables 4.10.2 and 4.10.3 display the values of the metrics computed on the case study with the oracle-based and linear regression-base calibration approaches respectively. Results are disturbing as the tables give slightly better results without calibration. However this is normal. In the cases where there is actually recalibration performed, instruments are changing of true status several times (this is shown by the values of  $N$ ,  $P$  and the prevalence in each table). In Section 5.5, we observed that most of the false results happen around the moment where an instrument actually becomes faulty. When instruments are recalibrated, they still continue to drift and they may change of status again over time. Therefore, more false results are likely to be obtained when a calibration strategy is applied. For this reason, the performances between the case study for which no calibration is carried out and the ones where it is, based on an oracle or on linear regression here, cannot be compared in terms of metrics without taking into account this fact.

That being said, the behaviour of the instruments is the same in terms of values and drift whether the calibration is performed with an oracle or with linear regression: it is only a different correction that is applied to the instruments. Thus, the results of the diagnosis algorithm can be compared between Tables 4.10.2 and 4.10.3. From these tables, it appears that the calibration based on linear regression provides better results in terms of improvement of the quality of the instruments' measurements:  $N$  is greater (consequently  $P$  is lower) and there are four less false negative for three more true positive and negative and one more false positive. Therefore, the accuracy with this method of calibration is slightly better. Although that the values of all the metrics are of the same order of magnitude, it is disturbing that the oracle-based method gives less accurate results, yet it corresponds to an ideal recalibration. In fact, like Figure 4.10.1, these tables do not assess the actual quality of the calibration but only the quality of the diagnosis. It only shows that indeed some instruments are corrected so that they are truly non-faulty again—and can be detected legitimately as so.



**Figure 4.10.1:** Evolution of true and predicted status for instrument  $s_1$  for different choices of calibration

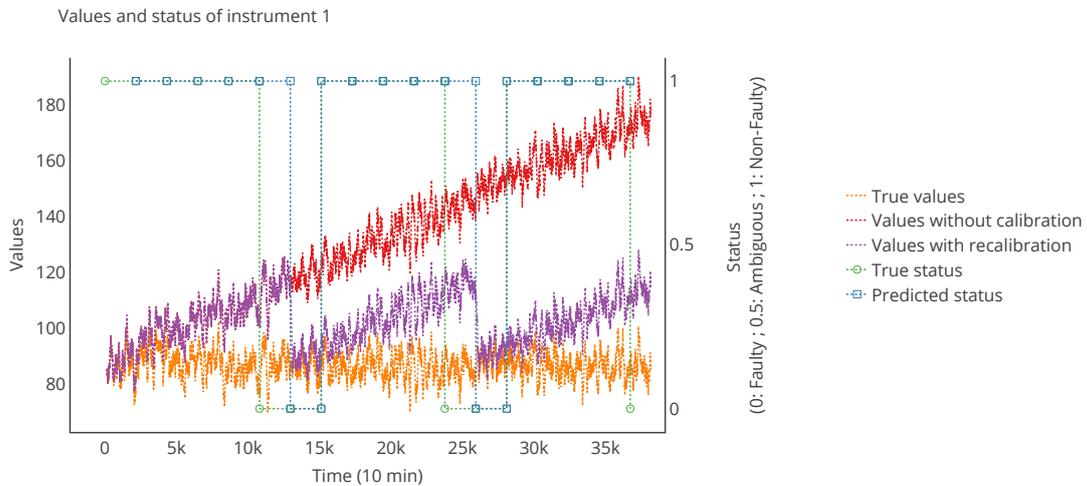
3884 To proceed to an evaluation of the impact of the calibration on the measurement results, the  
 3885 error model introduced in Chapter 3 is considered and its parameters are estimated for each  
 3886 instrument considering the entire time interval of study. Table 4.10.4 presents the statistics of the  
 3887 parameters and score of the model and Figure 4.10.3 represents a target plot of the instruments  
 3888 as a function of their slope and intercept. We observe that both the calibration approaches give  
 3889 equivalent results compared to the case without calibration. The error on the slope and on the  
 3890 intercept are reduced of at least 50% and 64% respectively.

3891 However, even with the oracle-based calibration, we observe that the remaining average offset  
 3892 can still be considered as significant. This residual is explained by the fact that the gain and  
 3893 offset of the instruments are corrected at specific moments and not at each time step: between  
 3894 two diagnosis procedures, the instruments continue to drift. Moreover, when an instrument is  
 3895 considered as faulty, the correction is applied on the values that are measured after the moment  
 3896 this prediction is made. This is illustrated in Figure 4.10.2. Thus, it explains why when we  
 3897 compute the parameters of the error model on the entire time range of study, the slope and  
 3898 intercept are not equal to their ideal values (one and zero respectively).

3899 More significantly, the average score of the linear regression on the error model increases  
 3900 from 0.84 to 0.97 at least indicating a good connection between the calibrated values and the

		True status			
		Non-faulty $N = 36$	Faulty $P = 117$	Prevalence 0.76	Accuracy 0.88
Predicted status	Non-faulty	$TN$ 36	$FN$ 17	$NPV$ 0.68	$FOR$ 0.32
	Ambiguous	$NDN$ 0	$NDP$ 2		
	Faulty	$FP$ 0	$TP$ 98	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.15		
		$NDNR$ 0.00	$NDPR$ 0.02	$NDR$ 0.01	
		$FPR$ 0.00	$TPR$ 0.84		

**Table 4.10.1:** Confusion matrix of the case study on calibration without recalibration (metrics are defined in Section 5.4).



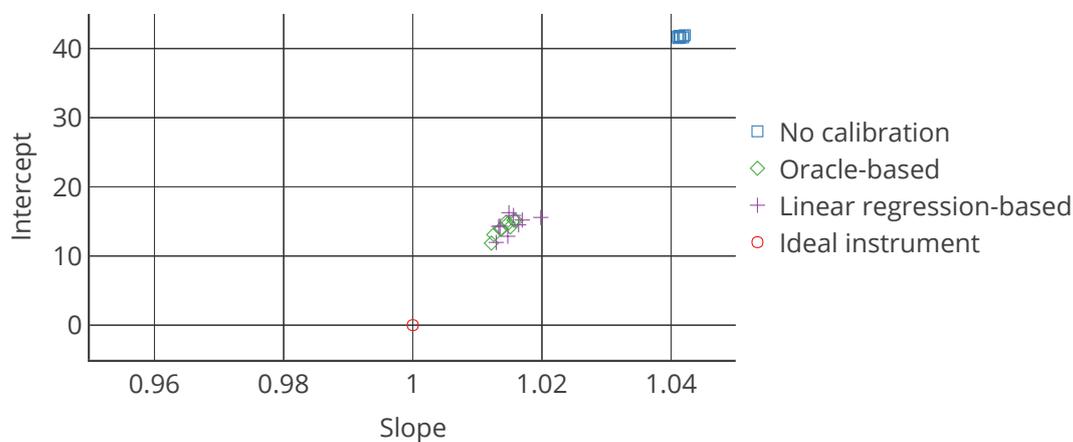
**Figure 4.10.2:** Evolution of the true values, of the measured values without and with recalibration by linear regression for  $s_1$ . The true and predicted statuses are also plotted to show when recalibrations are triggered.

3901 true values compared to the measured values and the true values if no calibration is performed.  
 3902 Thus, a simple calibration approach, based on the information contained in the rendez-vous used  
 3903 to predict the statuses of the instruments, allows obtaining an interesting improvement of their  
 3904 metrological performances.

3905 To conclude, this case study shows that, in addition to being able to detect quite accurately  
 3906 when an instrument does not meet a specification, the diagnosis algorithm can also provide useful

		True status			
		Non-faulty $N = 107$	Faulty $P = 46$	Prevalence 0.30	Accuracy 0.83
Predicted status	Non-faulty	$TN$ 107	$FN$ 26	$NPV$ 0.80	$FOR$ 0.20
	Ambiguous	$NDN$ 0	$NDP$ 0		
	Faulty	$FP$ 0	$TP$ 20	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.57		
	$NDNR$ 0.00	$NDPR$ 0.00	$NDR$ 0.00		
	$FPR$ 0.00	$TPR$ 0.43			

**Table 4.10.2:** Confusion matrix of the case study on calibration with an oracle-based calibration (metrics are defined in Section 5.4).



**Figure 4.10.3:** Target plot of the 9 nodes of class 0 of the network as a function of their slope and intercept following the error model of Chapter 3, computed on the entire time interval of study, for each calibration strategy.

3907 data to perform the recalibration of measuring instruments through the sets of valid rendez-vous  
3908 used during diagnosis procedures.

		<b>True status</b>			
		Non-faulty $N = 110$	Faulty $P = 43$	Prevalence 0.28	Accuracy 0.85
<b>Predicted status</b>	Non-faulty	$TN$ 109	$FN$ 22	$NPV$ 0.83	$FOR$ 0.17
	Ambiguous	$NDN$ 0	$NDP$ 0		
	Faulty	$FP$ 1	$TP$ 21	$FDR$ 0.05	$PPV$ 0.95
		$TNR$ 0.99	$FNR$ 0.51		
		$NDNR$ 0.00	$NDPR$ 0.00	$NDR$ 0.00	
		$FPR$ 0.01	$TPR$ 0.49		

**Table 4.10.3:** Confusion matrix of the case study on calibration with calibration based on linear regression (metrics are defined in Section 5.4).

<b>Calibration approach</b>	<b>Slope</b>		<b>Intercept</b>		<b>Score</b>	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
<b>No calibration</b>	1.04	0.00	42	0	0.84	0.00
<b>Oracle-based</b>	1.01	0.00	14	1	0.97	0.01
<b>Linear regression-based</b>	1.02	0.00	15	1	0.98	0.01

**Table 4.10.4:** Statistics of the slope, intercept and score of the error model of Chapter 3 computed for the nodes of class 0 of the network on the entire time interval of study and for each calibration strategy.

3909

## 11 Conclusion

3910

3911

3912

3913

3914

In this chapter, a novel algorithm for the diagnosis of drift faults in sensor networks was introduced. It exploits the concept of rendez-vous between measuring instruments. A key property of this algorithm is that it does not require any assumption of the type of sensor network. The only hypothesis needed concern the assumption as always non-faulty of at least a subset of instruments of the highest metrological class in the sensor network.

3915

3916

3917

3918

3919

3920

3921

3922

3923

Based on an initial case study that showed promising results, we discussed the adjustment of the different parameters of the algorithm. While we mostly provided guidelines to adjust them empirically, notably to reduce the number of false results, an algorithm was proposed to facilitate the adjustment of the minimal number of required valid rendez-vous to allow a prediction between the status faulty and non-faulty  $|\Phi_v|_{min}$ . It gives an upper boundary for the value of this parameter. In our case study, it allowed to identify that the value previously considered was too high to ensure a good diagnosis. Lowering the value of this parameter improved significantly the results. In summary, the recommendations provided across the chapter for all the parameters are summarised in Table 4.11.1.

3924

3925

3926

3927

3928

3929

3930

A study of the parameters of the case study influencing the results of the diagnosis algorithm was conducted. It showed the measurement results of the instruments, in particular those involved in rendez-vous, and the frequency of rendez-vous between instruments are key factors regarding respectively the results of the diagnosis algorithm and the adjustment of its parameters. The algorithm was also shown to be robust to other faults in the measurement results. Indeed, such faulty values are likely to generate false positives, which may be less prejudicial than false negatives.

3931

3932

3933

3934

3935

3936

3937

Finally, the diagnosis algorithm was associated to a simple calibration approach. When an instrument is diagnosed as faulty, the valid rendez-vous that were used to make the prediction of its status are exploited to make a linear regression between the values of the faulty instrument and the values from its diagnosers. This allowed to successfully correct the gain and offset of the faulty measuring instruments. Therefore, the diagnosis algorithm that was presented opens new perspectives on *in situ* calibration as this algorithm does more than indicating which instruments are faulty in a sensor network: it also provides information that can be exploited to correct them.

Parameter	Effect	Recommendation
Conditions of validity of measurement results	Determine if a measurement result can be compared to another.	It should be determined based on the properties of the measuring instruments deployed.
Conditions of spatiotemporal vicinity to be in rendez-vous	Determine if two measurement results can be compared to each other.	It can be determined based on expert specifications or on characteristics of the measurand and of the measuring instruments.
Instant of diagnosis $t_d$	Defines the reference time of the time range on which the diagnosis is performed.	Can be determined based on the characteristics of the instruments used and the context of deployment.
Periodicity of the diagnosis	Determines the frequency at which the algorithm is computed.	It can be determined based on the characteristics of the instruments used and the context of deployment.
Length of the time range $\Delta t$	Defines the length of the time range on which the diagnosis is performed.	It can be determined based on the characteristics of the instruments used.
Minimal relative difference of class $\Delta c_{min}^{\mathcal{D}}(k)$	Determines which classes of instruments can be the diagnosers of instruments belonging to a given class $k$ .	The higher the relative class is, the more accurate are the diagnosers. It should be set to +1 or 0 but not to a negative value.
Maximal tolerated values for the different rates $(r_{\Phi_v}^+)_{max}$ , $(r_{\Phi_v}^-)_{max}$ and $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$	Defines the triggers to determine if an instrument is faulty or non-faulty.	Too high values are likely to generate false negatives while too low values generate false positives. It should be adjusted to favour false positives instead of false negatives, while minimising their number.
Minimal required number of valid rendez-vous $ \Phi_v _{min}$	Determines if a given set of valid rendez-vous is large enough to determine if an instrument is faulty or non-faulty.	Considering a defined $\Delta t$ and using the algorithm provided, a value can be chosen based on the number of faulty instruments tolerated in the network. A sufficiently high $ \Phi_v _{min}$ should be chosen so that the content of the sets $\Phi_v$ are large enough to likely have representative sets.

**Table 4.11.1:** Recommendations for the different parameters of the diagnosis algorithm



# Conclusion and perspectives

---

## 1 Conclusion

In this thesis, we tackled the issue of calibration for sensor networks, a question that is particularly critical when low-cost instruments are deployed.

In the first part of this thesis, we reviewed the existing works concerning *in situ* calibration which are techniques aiming at calibrating measuring instruments while leaving them in the field, preferably without any physical intervention on them. Across this survey, we proposed a concise taxonomy enabling the classification of the algorithms. Four main groups of categories were defined, two of them following properties of the sensor networks that may be exploited, namely the presence of reference instruments in the network and the mobility of the nodes, and the other two concerning respectively the grouping strategy at stake in the algorithm and the kind of mathematical relationship that can be derived. This extensive survey showed that there are numerous contributions on the subject, covering a large spectrum of the cases that can be encountered when deploying sensor networks. However, a quantitative comparison of the performances of the algorithm was shown to be difficult because algorithms are often applied on specific case studies in their original publication. Indeed, there is no standard use cases or protocols to allow such comparisons.

Therefore, a framework for the simulation of sensor networks aimed at evaluating the performances of *in situ* calibration algorithms was proposed in the second part of this thesis. It was applied to the comparison of seven *in situ* calibration algorithms for blind and static sensor networks. We independently changed the model used to derive the true values of the measurand, the drift and fault model, the number of instruments and parameters of some calibration algorithms. We showed that even with a particular care to conduct a balanced comparison, different conclusions can be drawn by slightly changing the case study. Thus, in addition to being a protocol enabling the effective comparison of strategies, the proposed framework could be used to choose the most appropriate *in situ* calibration algorithm during the design of a sensor network.

However, we also observed that some calibration algorithms may correct measuring instruments regardless of their actual performances, particularly concerning those for blind sensor networks. Therefore, it is possible that instruments which measurement results are still correct regarding the metrological specifications of the instruments, have their values corrected. In the worst case, they may have their values degraded. Thus, a diagnosis algorithm for drift faults in sensor networks was proposed in the last part of this thesis. This algorithm is particularly interesting because it is based on rendez-vous between measuring instruments. This concept is more complex than a simple comparison of measured values as it encapsulates the idea of a necessary spatio-temporal vicinity so that measurement results can be relevantly compared. In addition, the algorithm demands no assumption on the type of sensor network used, requiring only that some of the best instruments in the network can be assumed as non-faulty. Across a case study, we showed that the algorithm allowed to successfully identify drifting instruments. In a second step, a better adjustment of the different parameters of the algorithm enabled to improve the results, although our guidelines are mainly empirical. Then, by investigating the influence of the design

3979 choices of our case study, we showed the measurement results involved in rendez-vous and the  
3980 frequency of rendez-vous between instruments, are key factors regarding respectively the results  
3981 of the diagnosis algorithm and the adjustment of its parameters. In addition, the algorithm  
3982 is robust to other faults: faulty values are likely to generate false positives, which may be less  
3983 prejudicial than false negatives in the context of environmental sensing. Finally, while combining  
3984 the diagnosis algorithm to a naive calibration approach, it was possible to successfully correct  
3985 the gain and offset of the faulty measuring instruments based on the measurement results of  
3986 rendez-vous between instruments, while leaving the measurement results of non-faulty ones  
3987 unchanged. Indeed, the diagnosis algorithm also provides information that can be exploited to  
3988 correct measuring instruments.

## 3989 **2 Perspectives**

3990 Several perspectives can be considered for the direct pursuit of this work but also for the  
3991 scientific community and the general public from a larger scope.

### 3992 **Direct perspectives**

#### 3993 **Extensions of the diagnosis algorithm**

3994 In the appendices, several extensions of the diagnosis algorithm are related. They aim  
3995 at being able to apply the diagnosis algorithm we proposed in any situation and with the  
3996 same performances. However, their development is not complete and demands supplementary  
3997 investigations. Moreover, the adjustment of some parameters of the algorithm is empirical and  
3998 may not be conducted through an automated optimisation procedure. Eventually, being able to  
3999 ensure that the specification stating when instruments should be considered as faulty or not, is  
4000 verified, or at least correctly implemented, would be valuable. Therefore, investigating if means  
4001 could be brought to improve the adjustment of the parameters of the algorithm and the formal  
4002 qualification of its performances would be of a major interest.

#### 4003 **Design of sensor networks**

4004 In this work, we proposed a framework that could be used to identify the relevant algorithms  
4005 for the *in situ* calibration of a sensor network in its environment of deployment. To facilitate the  
4006 development of sensor networks, there is a need for engineering tools to identify at the design  
4007 stage what the data processing algorithms are from which a sensor network would benefit, on top  
4008 of tools already optimising their deployment in the case of static sensor networks for instance [18].  
4009 This demands to aggregate the various results obtained by the scientific community from various  
4010 fields in a dedicated product. Such an integrated tool is not currently available in the literature.

#### 4011 **Perspectives for the scientific community and the general public**

##### 4012 **Standardisation of the vocabulary for the description of *in situ* calibration** 4013 **algorithms**

4014 In this work, we proposed a taxonomy for the classification of *in situ* calibration algorithms.  
4015 We observed that there were also other taxonomies. An intensive work of standardisation of this  
4016 vocabulary and on the practices to describe and evaluate these algorithms would be valuable  
4017 in order to structure the work of the scientific community on this subject. Calls were recently  
4018 published along the same lines [60, 133].

##### 4019 **Data processing standardisation**

4020 In scientific publications exploiting measurement results from low-cost sensor networks, data  
4021 processing that are carried out to have a clean dataset are not always fully described. For

4022 measuring instruments on the market, even for certified ones, the way the measurements are  
4023 processed is rarely mentioned or only as a black box. Concerning sensors as components of  
4024 measuring instruments, technical datasheets are often scarce and imprecise on how they behave  
4025 over time and depending on their operating conditions. In the scientific community, there are calls  
4026 for a standardisation of the data processing algorithms in order to determine how far processed  
4027 measurement results are from raw ones [60, 133]. We extend this call to the need for a clear  
4028 information on how measuring instruments and their components behave from a metrological  
4029 perspective, to facilitate the identification and use of adapted data processing algorithms.

### 4030 **Qualification of instruments and their processing**

4031 Although experiments under laboratory or field condition conducted during this thesis were  
4032 not reported in this manuscript, actual measuring instruments and systems were studied in the  
4033 early months of this thesis through the Sense-City platform [38] as it was done in the studies  
4034 reported in Chapter 1 Section 1.3.3. Regarding low-cost instruments and considering the cost  
4035 range we considered in this work (US \$100 to US \$10,000), there is a wide variety of quality  
4036 among these measuring instruments and systems. In the scientific community, there are reports  
4037 on the performances of measuring instruments but it is not mandatory for a manufacturer to  
4038 provide the results of such works with its devices. In fact, it is currently difficult for an end user  
4039 to identify clearly which instrument suits its needs and at which cost because the information  
4040 provided is quite poor. Therefore, an effort of qualification of the performances of measuring  
4041 instruments, with the data processing algorithms they use, and a transparent information of the  
4042 general public about it when instruments are put on the market would be valuable.

### 4043 **Environmental footprint of low-cost environmental sensor network**

4044 Currently, there is little information on the environmental footprint of low-cost measuring  
4045 instruments [78]. While their interest can be legitimately justified whatever the measurand  
4046 considered, it has to be counterbalanced by their impact on the environment. We are dealing  
4047 here most of the time with electronic devices that may be difficult to recycle and may require rare  
4048 materials. In addition, they also require a computer and telecommunication infrastructure, those  
4049 being known to have more and more impact on the greenhouse gases emissions around the world.  
4050 Being able to evaluate the global impact of the entire measuring chain, from the instruments  
4051 to the datacentres in which measurement results are stored and exploited, would be of a major  
4052 interest in order to determine the actual benefits of dense low-cost sensor networks compared  
4053 to what is currently achieved with the information brought by more traditional environmental  
4054 measuring systems.



## Diagnosis Algorithm for Drift Faults in Sensor Networks: Extensions

### 1 Formulation reducing the number of iterations of the algorithm

The formulation of Algorithm 1 given in Chapter 4 Section 4.2 is not optimised in terms of iterations. Indeed, considering the statuses of instruments are likely predicted based on the results obtained for instruments of a higher class than their own (based on the recommendation to set the minimal relative differences of class required  $\Delta c_{min}^{\mathcal{D}}(k)$  to +1 or 0 in Chapter 4 Section 5.7), it would be relevant to browse the set of instruments to diagnose  $S^{\mathcal{D}}$  by descending classes of instruments instead of browsing it regardless of the classes of the instruments. In addition, actualising the predicted statuses at the end of the evaluation of the instruments of each class would allow using the instruments predicted as non-faulty at the previous iteration.

Algorithm 3 provides this optimised formulation in terms of iterations.

---

**Algorithm 3:** Algorithm of the diagnosis procedure proposed for the detection of drift in sensor networks with its number of iterations optimised

---

```

Data:  $S, \Phi(S \rightarrow S, (t_d, \Delta t)), |\Phi_v|_{min}, \Delta c_{min}^{\mathcal{D}}$ 
Result:  $\widehat{\Omega}(S, t_d)$ 
/* Initiate the set of instruments to diagnose and the predicted statuses
   of the instruments */
 $S^{\mathcal{D}} \leftarrow S \setminus S^{c_{max}}$ 
 $\widehat{\Omega}(S^{c_{max}}, t_d) \leftarrow \text{NF}$ 
 $\widehat{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow \text{A}$ 
 $\widetilde{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow \widehat{\Omega}(S^{\mathcal{D}}, t_d)$  /* Initiate the actualised statuses */
/* Ignore the instruments that cannot have enough valid rendez-vous */
for  $s_i \in S^{\mathcal{D}}$  do
  | if  $|\Phi(s_i \rightarrow S, (t_d, \Delta t))| < |\Phi_v|_{min}$  then
  | |  $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$ 
end
/* Predict the status of each instrument to diagnose */
repeat
  |  $\widehat{\Omega}(S, t_d) \leftarrow \widetilde{\Omega}(S, t_d)$  /* The actualised statuses are now the predicted
  | statuses */

```

---

---

```

for  $k \leftarrow c_{max} - 1$  to 0 do /* For each class of instruments in descending
    order */
     $\widehat{\Omega}'(S, t_d) \leftarrow \widetilde{\Omega}(S, t_d)$  /* A local instance of the future predicted
        statuses is created, to take into account the results of the
        predictions for the previous values of  $k$  in this loop */
    for  $s_i \in S^{\mathcal{D}} \cap S^k$  do
        /* Build the current set of valid rendez-vous for  $s_i$ , using
             $\widehat{\Omega}'(S, t_d)$  instead of  $\widehat{\Omega}(S, t_d)$  */
         $c_{min}^{\mathcal{D}}(s_i) \leftarrow c(s_i) + \Delta c_{min}^{\mathcal{D}}(c(s_i))$ 
         $\Phi_v(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (t_d, \Delta t)), \text{ such as}$ 
             $s_j \notin S^{\mathcal{D}}, \widehat{\Omega}'(s_j, t') = NF, c(s_j) \geq c_{min}^{\mathcal{D}}(s_i),$ 
             $m(s_i, t) \in M^*(s_i, (t_d, \Delta t)) \text{ and } m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\}$ 
        /* If  $s_i$  have enough valid rendez-vous, then compute the
            different rates to actualise its status */
        if  $|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))| \geq |\Phi_v|_{min}$  then
             $r_{\Phi_v}^{\approx}(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^{\approx}(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 
             $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^+(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 
             $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^-(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 
            /* If a condition on the different rates is met, then the
                actualised status of  $s_i$  is set to faulty, otherwise, it is
                set to non-faulty */
            if  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^+)_{max}$  or  $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^-)_{max}$  or
                 $(1 - r_{\Phi_v}^{\approx}(s_i \rightarrow S, (t_d, \Delta t))) > (r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  then
                |  $\widetilde{\Omega}(s_i, t_d) \leftarrow F$ 
            else
                |  $\widetilde{\Omega}(s_i, t_d) \leftarrow NF$ 
            end
             $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$  /*  $s_i$  is diagnosed so it can be removed from  $S^{\mathcal{D}}$ 
                */
        end
    end
end
until  $\widehat{\Omega}(S, t_d) = \widetilde{\Omega}(S, t_d)$  /* Repeat until there is no difference between the
    predicted and actualised statuses */

```

---

## 2 Diagnosis with the prediction as non-faulty based on the highest sufficient class

4069  
4070

4071 In Chapter 4 Section 4.2, the prediction of an instrument  $s_i$  as non-faulty or faulty is based on  
 4072 the set of valid rendez-vous  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  if the condition regarding the minimal number  
 4073 of valid rendez-vous required  $|\Phi_v|_{min}$  is satisfied.  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  is notably defined with the  
 4074 help of the minimal class allowed to be a diagnoser of  $s_i$ ,  $c_{min}^{\mathcal{D}}(s_i)$ .

4075 However, suppose that  $\exists k > c_{min}^{\mathcal{D}}(s_i)$  such as  $|\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))| > |\Phi_v|_{min}$ . In this  
 4076 case, the status of  $s_i$  can be predicted with a smaller number of instruments. This is interesting

4077 because instruments of higher class usually have a lower measurement uncertainty. Thus, if the  
4078 predicted status of  $s_i$ ,  $\widehat{\Omega}(s_i, t_d)$ , is determined as non-faulty, a higher confidence in  $s_i$  can be  
4079 granted: it is predicted as non-faulty based on its compatibility with instruments of a higher  
4080 class than the minimal one required to diagnose it. Consequently, we can estimate that the  
4081 measurement results of  $s_i$  are more accurate than expected. By predicting the statuses of the  
4082 instruments in this manner, the results of the diagnosis algorithm would be enriched with a  
4083 useful information.

4084 Nevertheless, if  $s_i$  is predicted as faulty according to  $\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))$ , it is still possible  
4085 to obtain  $\widehat{\Omega}(s_i, t_d) = NF$  with the set of valid rendez-vous  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ . In this case, it is  
4086 still preferable to have  $s_i$  predicted as non-faulty despite a prediction as faulty with a subset of  
4087 the diagnoser instruments.

4088 It is also possible to obtain  $\widehat{\Omega}(s_i, t_d) = NF$  according to  $\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))$  but  $\widehat{\Omega}(s_i, t_d) =$   
4089  $F$  according to  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ . We estimate the prediction based on  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  is  
4090 more likely a false positive as less accurate instruments are considered than in  $S^{k+}$ . Such a case  
4091 may be very rare. Indeed, the less the diagnoser instruments are accurate, the more the predicted  
4092 statuses have a chance to be equal to non-faulty. This is due to mean used to determine the  
4093 compatibility between measurement results: it is based on the measurement uncertainty. This  
4094 time, it is preferable to have  $s_i$  predicted as non-faulty with a subset of the diagnoser instruments  
4095 despite a faulty prediction that can be made with all of them.

4096 Therefore, Algorithm 1 must be adapted to include this idea, expressed as a diagnosis  
4097 algorithm with negative prediction based on the highest sufficient class. It is described in  
4098 Algorithm 4<sup>50</sup>.

---

**Algorithm 4:** Algorithm of the diagnosis procedure proposed for the detection of drift  
in sensor networks with the prediction as non-faulty based on the highest sufficient class

---

```

Data:  $S, \Phi(S \rightarrow S, (t_d, \Delta t)), |\Phi_v|_{min}, \Delta c_{min}^{\mathcal{D}}$ 
Result:  $\widehat{\Omega}(S, t_d)$ 
/* Initiate the set of instruments to diagnose and the predicted statuses
   of the instruments */
 $S^{\mathcal{D}} \leftarrow S \setminus S^{c_{max}}$ 
 $\widehat{\Omega}(S^{c_{max}}, t_d) \leftarrow NF$ 
 $\widehat{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow A$ 
 $\widetilde{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow \widehat{\Omega}(S^{\mathcal{D}}, t_d)$  /* Initiate the actualised statuses */
/* Ignore the instruments that cannot have enough valid rendez-vous */
for  $s_i \in S^{\mathcal{D}}$  do
    if  $|\Phi(s_i \rightarrow S, (t_d, \Delta t))| < |\Phi_v|_{min}$  then
         $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$ 
    end
/* Predict the status of each instrument to diagnose */
repeat
     $\widehat{\Omega}(S, t_d) \leftarrow \widetilde{\Omega}(S, t_d)$  /* The actualised statuses are now the predicted
       statuses */
    for  $s_i \in S^{\mathcal{D}}$ 
         $c_{min}^{\mathcal{D}}(s_i) \leftarrow c(s_i) + \Delta c_{min}^{\mathcal{D}}(c(s_i))$ 
         $k \leftarrow c_{max}$ 

```

---

<sup>50</sup>The improvements brought with Algorithm 3 in Section 1 are not added for the sake of clarity.

---

```

repeat
  /* Build the current set of valid rendez-vous for  $s_i$  with the
     instruments in  $S^{k+}$  */
   $\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t)) \leftarrow \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S^{k+}, (t_d, \Delta t)), \text{ such as }
     s_j \notin S^D, \widehat{\Omega}'(s_j, t') = NF, c(s_j) \geq c_{min}^D(s_i),
     m(s_i, t) \in M^*(s_i, (t_d, \Delta t)) \text{ and } m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\}$ 
  /* If  $s_i$  have enough valid rendez-vous, then compute the
     different rates to actualise its status */
  if  $|\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))| \geq |\Phi_v|_{min}$  then
     $r_{\Phi_v}^{\sim}(s_i \rightarrow S^{k+}, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^{\sim}(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}$ 
     $r_{\Phi_v}^+(s_i \rightarrow S^{k+}, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^+(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}$ 
     $r_{\Phi_v}^-(s_i \rightarrow S^{k+}, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^-(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S^{k+}, (t_d, \Delta t))|}$ 
    /* If a condition on the different rates is met, then the
       actualised status of  $s_i$  is set to faulty, otherwise, it is
       set to non-faulty */
    if  $r_{\Phi_v}^+(s_i \rightarrow S^{k+}, (t_d, \Delta t)) > (r_{\Phi_v}^+)_{max}$  or
        $r_{\Phi_v}^-(s_i \rightarrow S^{k+}, (t_d, \Delta t)) > (r_{\Phi_v}^-)_{max}$  or
        $(1 - r_{\Phi_v}^{\sim}(s_i \rightarrow S^{k+}, (t_d, \Delta t))) > (r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  then
      |  $\widetilde{\Omega}(s_i, t_d) \leftarrow F$ 
    else
      |  $\widetilde{\Omega}(s_i, t_d) \leftarrow NF$ 
    end
  end
   $k \leftarrow k - 1$ 
until  $k < c_{min}^D(s_i)$  or  $\widetilde{\Omega}(s_i, t) = NF$ 
if  $\widetilde{\Omega}(s_i, t_d) \neq A$  then
  |  $S^D \leftarrow S^D \setminus \{s_i\}$  /*  $s_i$  is diagnosed so it can be removed from  $S^D$  */
end
end
until  $\widehat{\Omega}(S, t_d) = \widetilde{\Omega}(S, t_d)$  /* Repeat until there is no difference between the
    predicted and actualised statuses */

```

---

### 3 From a centralised to a decentralised computation

From its description provided in Chapter 4 Section 4.2 and its pseudo-code in Algorithm 1, our diagnosis algorithm has to be computed in a centralised manner. Indeed, all the rendez-vous between the instruments must be known, with the associated measurement results and information to determine the validity of the rendez-vous, and along with the predicted and actualised status of the instruments during the different iterations of the main loop in the algorithm.

However, it is possible to consider a distributed calculation, notably if the measurement results are not collected in a centralised manner.

Consider an instrument  $s_i$ .

We assume  $s_i$  knows  $\Delta c_{min}^{\mathcal{D}}(s_i)$ ,  $|\Phi_v|_{min}$ ,  $(r_{\Phi_v}^+)_{max}$ ,  $(r_{\Phi_v}^-)_{max}$  and  $(r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$ .

It is also possible to assume it knows that  $\widehat{\Omega}(S^{c_{max}}, t_d) = NF$  due to the assumption of the instruments of class  $c_{max}$  as non-faulty required for the algorithm<sup>51</sup>.

Finally,  $s_i$  knows  $\Phi(s_i \rightarrow S, (t_d, \Delta T))$ . We assume when two instruments meet, they exchange information related on the rendez-vous they have (measurement results and so on).

Consider  $S^\varphi(s_i, (t_d, \Delta t))$  which is the set of instruments involved in a rendez-vous with  $s_i$  between  $[t_d - \Delta t; t_d]$ <sup>52</sup>:

$$S^\varphi(s_i, (t_d, \Delta t)) = \{s_j \text{ such as } \exists \varphi(s_i \rightarrow s_j, t') \in \Phi(s_i \rightarrow S, (t_d, \Delta T))\}$$

Based on it,  $s_i$  can define a local set of instruments to diagnose  $S^{\mathcal{D}} = (s_i \cup S^\varphi(s_i, (t_d, \Delta t))) \setminus S^{c_{max}}$ .

Depending on  $S^{\mathcal{D}}$ ,  $s_i$  can derive  $\Phi_v(s_i \rightarrow S, (t_d, \Delta T))$  if it knows  $\widehat{\Omega}(s_j, t_d)$  for  $s_j \in S^\varphi(s_i, (t_d, \Delta t)) \setminus S^{\mathcal{D}}$ . Then,  $s_i$  can determine its actualised status as indicated in Algorithm 1. If  $s_i \notin S^{\mathcal{D}}$  after this, its predicted status is known. Otherwise,  $s_i$  has to try to determine it again.

In fact, the most challenging step is to access to  $\widehat{\Omega}(s_j, t_d)$  for  $s_j \in S^{\mathcal{D}} \setminus \{s_i\}$  to actualise  $S^{\mathcal{D}}$  between each iteration of the algorithm.

Calls must be sent to these instruments asking if  $s_j$  determined if  $\widetilde{\Omega}(s_j, t_d)$  is equal to  $NF$  or  $F$ .

There is, however, a need for a synchronisation mechanism between the node. Indeed, because the prediction of each instrument's status depends on exchanges of information with others, some instruments may be able to perform iterations of the algorithm faster or slower than others. To avoid issues by mixing predicted statuses that were not determined after the same iteration of the diagnosis procedure, we define a loop counter  $n^{\mathcal{D}}$ . This counter is incremented at each repetition of the main loop of the algorithm. When  $s_i$  calls the instruments  $s_j \in S^{\mathcal{D}} \setminus \{s_i\}$  to ask for  $\widetilde{\Omega}(s_j, t_d)$ , it also indicates that this call is related to the loop  $n^{\mathcal{D}}$ .

After receiving the responses,  $s_i$  removes the  $s_j$  from its local  $S^{\mathcal{D}}$  if  $\widetilde{\Omega}(s_j, t_d) \neq A$  and it can again try to determine its status.

At some point,  $\widetilde{\Omega}(S^\varphi(s_i, (t_d, \Delta t), t_d))$  may not change anymore. Thus, like in Algorithm 1, if  $\widehat{\Omega}(S^\varphi(s_i, (t_d, \Delta t), t_d)) = \widetilde{\Omega}(S^\varphi(s_i, (t_d, \Delta t), t_d))$  after an iteration of the diagnosis procedure, the algorithm terminates and  $\widehat{\Omega}(s_i, t_d) = A$ .

The distributed formulation of the diagnosis algorithm is provided in Algorithm 5. By construction, due to its synchronous behaviour, it is certain that the results are identical to the ones obtained in a centralised manner.

The protocol of communication between the nodes, either when they are in rendez-vous or when they exchange actualised statuses, is not detailed in this pseudo-code. Moreover, we suppose that these communications are managed in a separate thread from the one in which is run the diagnosis algorithm to blockings. Finally, Algorithm 5 does not manage the case where:

<sup>51</sup>Or a subset of  $S^{c_{max}}$ . See Chapter 4 Section 6.

<sup>52</sup>Therefore,  $\Phi(s_i \rightarrow S^\varphi(s_i, (t_d, \Delta t)), (t_d, \Delta T)) = \Phi(s_i \rightarrow S, (t_d, \Delta T))$ .

- 4141 • the diagnosis algorithm run by  $s_j$  has terminated after the loop  $n^{\mathcal{D}'}$  and  $\widehat{\Omega}(s_j, t_d) = A$ .
- 4142 •  $\exists s_i$  such as it is still trying to determine its predicted status during a loop  $n^{\mathcal{D}} > n^{\mathcal{D}'}$ .

4143 Thus,  $s_j \in S^\varphi(s_i, (t_d, \Delta t))$  during the loop  $n^{\mathcal{D}}$ .  $s_i$  will ask  $s_j$  for its actualised status after the  
 4144 loop  $n^{\mathcal{D}}$  even if the status of  $s_j$  was definitively predicted as ambiguous. We can assume that in  
 4145 practice  $s_j$  can tell  $s_i$  after the loop  $n^{\mathcal{D}'}$  that  $\widehat{\Omega}(s_j, t_d)$  will not change anymore and thus avoid  
 4146 useless communications. This would require few changes in Algorithm 5 that were not written  
 4147 for the sake of clarity.

4148 In conclusion, it is possible to apply our diagnosis algorithm in a decentralised manner. The  
 4149 solution proposed is, however, not unique and other protocols could be considered.

---

**Algorithm 5:** Algorithm of the diagnosis procedure proposed for the detection of drift in sensor networks written in a distributed manner. Each node of the network runs this pseudo-code.

---

**Data:**  $S^\varphi(s_i, (t_d, \Delta t))$ ,  $\Phi(s_i \rightarrow S, (t_d, \Delta t))$ ,  $|\Phi_v|_{min}$ ,  $c_{min}^{\mathcal{D}}(s_i)$   
**Result:**  $\widehat{\Omega}(S, t_d)$

```

/* Initiate the set of instruments to diagnose and the predicted statuses
   of the instruments */
 $S^{\mathcal{D}} \leftarrow (s_i \cup S^\varphi(s_i, (t_d, \Delta t))) \setminus S^{c_{max}}$ 
 $\widehat{\Omega}(S^{c_{max}}, t_d) \leftarrow NF$ 
 $\widehat{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow A$ 
 $\widetilde{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow \widehat{\Omega}(S^{\mathcal{D}}, t_d)$  /* Initiate the actualised statuses */
/* Ignore the instruments that cannot have enough valid rendez-vous */
for  $s_i \in S^{\mathcal{D}}$  do
    if  $|\Phi(s_i \rightarrow S, (t_d, \Delta t))| < |\Phi_v|_{min}$  then
         $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$ 
    end
 $n^{\mathcal{D}} \leftarrow 0$ 
/* Predict the status of each instrument to diagnose */
repeat
     $n^{\mathcal{D}} \leftarrow n^{\mathcal{D}} + 1$ 
     $\widehat{\Omega}(S, t_d) \leftarrow \widetilde{\Omega}(S, t_d)$  /* The actualised statuses are now the predicted
       statuses */
    /* Build the current set of valid rendez-vous for  $s_i$  */
     $\Phi_v(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (t_d, \Delta t)), \text{ such as }
       s_j \notin S^{\mathcal{D}}, \widehat{\Omega}(s_j, t') = NF, c(s_j) \geq c_{min}^{\mathcal{D}}(s_i),
       m(s_i, t) \in M^*(s_i, (t_d, \Delta t)) \text{ and } m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\}$ 
    /* If  $s_i$  have enough valid rendez-vous, then compute the different
       rates to actualise its status */
    if  $|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))| \geq |\Phi_v|_{min}$ 
         $r_{\Phi_v}^{\approx}(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^{\approx}(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 
         $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^+(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 
         $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) \leftarrow \frac{|\Phi_v^-(s_i \rightarrow S, (t_d, \Delta t))|}{|\Phi_v(s_i \rightarrow S, (t_d, \Delta t))|}$ 

```

---

---

```

/* If a condition on the different rates is met, then the
   actualised status of  $s_i$  is set to faulty, otherwise, it is set to
   non-faulty */
if  $r_{\Phi_v}^+(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^+)_{max}$  or  $r_{\Phi_v}^-(s_i \rightarrow S, (t_d, \Delta t)) > (r_{\Phi_v}^-)_{max}$  or
 $(1 - r_{\Phi_v}^{\sim}(s_i \rightarrow S, (t_d, \Delta t))) > (r_{\Phi_v}^+ + r_{\Phi_v}^-)_{max}$  then
  |  $\tilde{\Omega}(s_i, t_d) \leftarrow F$ 
else
  |  $\tilde{\Omega}(s_i, t_d) \leftarrow NF$ 
end
 $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$  /*  $s_i$  is diagnosed so it can be removed from  $S^{\mathcal{D}}$  */
end
for  $s_j \in S^{\mathcal{D}} \setminus \{s_i\}$  do /* Actualise  $S^{\mathcal{D}}$  */
  | ask  $s_j$  if  $\tilde{\Omega}(s_j, t_d)$  has changed to  $NF$  or  $F$  during loop  $n^{\mathcal{D}}$ 
  | actualise  $\tilde{\Omega}(s_j, t_d)$  if necessary
  | if  $\tilde{\Omega}(s_j, t_d) \neq A$  then
  | |  $S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_j\}$ 
  | end
end
until  $\hat{\Omega}(S^\varphi(s_i, (t_d, \Delta t), t_d) = \tilde{\Omega}(S^\varphi(s_i, (t_d, \Delta t), t_d)$  or  $s_i \notin S^{\mathcal{D}}$  /* Repeat until there
   is no difference between the predicted and actualised statuses or if the
   status of  $s_i$  has been determined */

```

---

## 4 Multiple measurands

Low-cost sensors are prone to drift due to influence quantities affecting the indication provided by the instrument. Thus, it is frequent that the calibration relationship of such instruments has to take into account their influence quantities. This is easily applicable in a measurement system embedding a main instrument measuring the principal measurand targeted by the system and several other instruments that measure the values of the influence quantities of the main instrument.

Therefore, the sensor network addresses more than one measurand in this case. It can be seen as a "network of sensor networks", each one considering a specific measurand. In this way, some sensor networks are dependent on others to determine their measured values. The diagnosis of the instruments of such systems is discussed in this section.

### 4.1 General idea

We propose here a general concept of protocol to carry out the diagnosis of drift fault in sensor networks composed of measurements systems in which one or more instruments has influence quantities measured by other instruments of the same system.

The general idea is the following:

1. Carry out the diagnosis for each sub-network measuring a same quantity and for which the instruments have no influence quantities following the algorithm described in Chapter 4 Section 4.2.
2. Perform the diagnosis for each measuring a same quantity and for which the instruments have influence quantities measured by a sub-network previously diagnosed. The way to carry this diagnosis out is discussed in the following subsection.
3. Repeat step 2 until the diagnosis has been performed on each sub-network, e.g. for each measurand.

### 4.2 Diagnosis of drift faults in a sub-network with instruments having influence quantities

In fact, we aim at using the same algorithm already defined in Chapter 4 Section 4.2. Indeed, the presence of influence quantities questions only the validity of the measurement results.

We remind we assume that the sensor network is composed of measurement systems, in which the influence quantities of one or more of its instruments are measured by other instruments that also belong to the same system.

Consider a sub-network with one influence quantity, a diagnosis procedure  $d$  that happens at  $t_d$  and suppose we know the predicted statuses at  $t_d$  of the instruments measuring the influence quantities.

Consider a rendez-vous between two instruments  $s_i$  and  $s_j$ ,  $\varphi(s_i \rightarrow s_j, t)$  with  $t \in [t_d - \Delta t; t_d]$ <sup>53</sup>, with  $s_j$  supposed non-faulty and of class  $c(s_j) \geq c_{min}^D$ . The instruments measuring the influence quantity are denoted  $s'_i$  and  $s'_j$ <sup>54</sup>.

Let us discuss the validity of the rendez-vous. First regarding the measured values of  $s_i$ , three different cases appear:

---

<sup>53</sup>In this case, we could also speak of a rendez-vous between two systems because all the pairs of instruments measuring the same quantities in the systems  $i$  and  $j$  are in rendez-vous

<sup>54</sup>We also suppose when two systems are in rendez-vous, we have measurements for all the measurands. This assumption could be discussed in more details as it requires to extend the definition of rendez-vous and to discuss the case of instruments of a same system with different acquisition frequencies. Such particular cases are not treated here

- 4189 •  $s'_i$  is non-faulty: the measured value of  $s'_i$  can be used to determine the one of  $s_i$  with  
4190 its calibration relationship. In this case, the measurement result  $m(s_i, t)$  is valid and the  
4191 rendez-vous may be valid too<sup>55</sup>.
- 4192 •  $s'_i$  is faulty but  $s'_j$  is non-faulty: the measured value of  $s'_j$  can be used to determine the one  
4193 of  $s_i$  with its calibration relationship. In this case,  $m(s_i, t)$  is again valid the rendez-vous  
4194 may be valid too<sup>56</sup>.
- 4195 •  $s'_i$  and  $s'_j$  are faulty: the measured value of  $s_i$  cannot be determined as no value of its  
4196 influence quantity can be reliably used with the calibration relationship. In this case,  
4197  $m(s_i, t)$  is not valid. Therefore,  $\varphi(s_i \rightarrow s_j, t)$  is **not valid**.

4198 The same reasoning can be applied to  $s_j$ .<sup>5758</sup>

4199 In fact, the presence of influence quantities has only an influence on the validity of the  
4200 measurement results. Thus, the Equation 4.1 from Chapter 4 Section 4.1 used to determine  
4201  $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$  is still valid. We recall this equation:

$$\begin{aligned} \Phi_v(s_i \rightarrow S, (t_d, \Delta t)) = & \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (t_d, \Delta t)), \text{ such as} \\ & s_j \notin S^D, // s_j \text{ is not an instrument to diagnose} \\ & \widehat{\Omega}(s_j, t') = NF, // s_j \text{ is non-faulty} \\ & c(s_j) \geq c_{min}^D(s_i), // \text{The class of } s_j \text{ is higher or equal to the minimal} \\ & \text{class allowed to diagnose } s_i \\ & m(s_i, t) \in M^*(s_i, (t_d, \Delta t)), // \text{The measurement result of } s_i \text{ is valid} \\ & m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\} // \text{The measurement result of } s_j \text{ is valid} \end{aligned} \quad (\text{A.1})$$

4202 This is why, apart from the conditions on the validity of the measurement results that may  
4203 change, the diagnosis algorithm remains applicable in the context of measuring instruments with  
4204 influence quantities.

### 4205 4.3 Conclusion

4206 In this section, we presented a manner to extend our diagnosis algorithm for drift faults in  
4207 sensor networks to the case where

4208 However, a formulation enriching the results provided by the diagnosis algorithm could be  
4209 developed in future work. Indeed, the fact that values of measuring instruments from other  
4210 systems can be used to determine the status of an instrument in a given system could allow us  
4211 to determine a status at the level of the measuring system for instance or at least to qualify the  
4212 predicted statuses in light of how the validity of measurement results is determined.

## 4213 5 Diagnosis algorithm for drift faults in sensor networks with an event- 4214 based formulation

4215 In the first place, the diagnosis algorithm was presented as being applied at an instant  $t_d$  over  
4216  $[t_d - \Delta t; t_d]$ . It means that it is time which triggered the diagnosis and that the determination of

---

<sup>55</sup>Depending on the other conditions defined for the validity of the measurement results.

<sup>56</sup>See note 55

<sup>57</sup>This development can be extended to the case where there are multiple influences quantities but also to the case where more than two instruments meet at the same time, e.g. the values of the influence quantity measured by a third system could be used when  $s'_i$  and  $s'_j$  are faulty in our case.

<sup>58</sup>The definitions related the compatibility between two measurement results could be extended based on these three cases.

4217 the sets of rendez-vous is notably based on a time range.

4218 However, notably in the perspective of a real-time application, it could be valuable to trigger  
 4219 the diagnosis when rendez-vous occurs. A rendez-vous can be assimilated to an **event** able to  
 4220 trigger a diagnosis procedure. Such rendez-vous is noted  $\varphi(\dots)$ , " $\dots$ " standing for the instruments  
 4221 actually involved in the rendez-vous and the time instant at which it occurred.

4222 In the same way, the contents of the sets of rendez-vous could be determined based on a  
 4223 number of rendez-vous, noted  $|\Phi|^{\mathcal{D}}$ , that occurred before  $\varphi(\dots)$ . It can be useful if the number of  
 4224 rendez-vous on  $[t_d - \Delta t; t_d]$  varies a lot between the instruments. For instance, some of them  
 4225 may not be diagnosable with not enough (valid) rendez-vous on this time range. A solution is  
 4226 to increase  $\Delta t$ . However in this case, other instruments may have their statuses not correctly  
 4227 predicted with a greater  $\Delta t$  (see Chapter 4 Section 7). Thus, replacing  $\Delta t$  by  $|\Phi|^{\mathcal{D}}$  can help to  
 4228 overcome this situation: all the instruments will be diagnosed based on sets of rendez-vous that  
 4229 are equivalent in terms of size.

4230 In addition to the couple  $(t_d, \Delta t)$  defining the moment of the diagnosis and the time period  
 4231 on which it is made, there are other combinations possible to set these characteristics of the  
 4232 algorithm with the help of  $\varphi(\dots)$  and  $|\Phi|^{\mathcal{D}}$ :

- 4233 •  $(t_d, |\Phi|^{\mathcal{D}})$ : the diagnosis occurs at  $t_d$  and the last  $|\Phi|^{\mathcal{D}}$  rendez-vous are considered for each  
 4234 instrument to predict its status.
- 4235 •  $(t_d, \Delta t, |\Phi|^{\mathcal{D}})$ : the diagnosis occurs at  $t_d$  and the rendez-vous that happened in  $[t_d - \Delta t; t_d]$   
 4236 are considered, up to the  $|\Phi|^{\mathcal{D}}$  last ones.
- 4237 •  $(\varphi(\dots), \Delta t)$ : the diagnosis occurs on  $\varphi(\dots)$  and the rendez-vous that happened in  $[t_d - \Delta t; t_d]$   
 4238 are considered, with  $t_d$  being the instant at which  $\varphi(\dots)$  was generated.
- 4239 •  $(\varphi(\dots), |\Phi|^{\mathcal{D}})$ : the diagnosis occurs on  $\varphi(\dots)$  and the last  $|\Phi|^{\mathcal{D}}$  rendez-vous are considered  
 4240 for each instrument to predict its status.
- 4241 •  $(\varphi(\dots), \Delta t, |\Phi|^{\mathcal{D}})$ : the diagnosis occurs at  $\varphi(\dots)$  and the rendez-vous that happened in  
 4242  $[t_d - \Delta t; t_d]$  are considered, up to the  $|\Phi|^{\mathcal{D}}$  last ones<sup>59</sup>.

4243 The diagnosis algorithm can be easily adapted to consider these alternate definitions of the  
 4244 instant  $\alpha \in \{t_d, \varphi(\dots)\}$  at which is executed and of the duration  $\beta \in \{\Delta t, |\Phi|^{\mathcal{D}}, (\Delta t, |\Phi|^{\mathcal{D}})\}$  used  
 4245 to define the sets used by the algorithm. Indeed, the expression " $(t_d, \Delta t)$ " can be replaced by  
 4246 " $(\alpha, \beta)$ " straightforward in Algorithm 1<sup>60</sup>.

4247 To conclude, the proposed diagnosis algorithm can also be applied with an event-based  
 4248 definition of the instant at which it is executed and of the duration used to define the sets of  
 4249 rendez-vous it uses.

## 4250 6 Real-time diagnosis algorithm of drift faults in sensor networks

4251 Until then, the diagnosis was considered as a procedure carried out periodically or, with the  
 4252 definitions of Section 5, on events. If it is performed at each time step or on each event, the

---

<sup>59</sup>When  $\Delta t$  and  $|\Phi|^{\mathcal{D}}$  are combined it is possible that a set of rendez-vous:

- contains less than  $|\Phi|^{\mathcal{D}}$  rendez-vous if less than that occurred during  $[t_d - \Delta t; t_d]$ .
- does not contain all the rendez-vous that occurred during  $[t_d - \Delta t; t_d]$  if there were more rendez-vous than  $|\Phi|^{\mathcal{D}}$ .

<sup>60</sup>In fact, behind " $\dots$ " in  $\varphi(\dots)$  lies a " $t_d$ " which is the instant at which the rendez-vous occurred. This  $t_d$  is used in practice to determine the sets involved in a diagnosis procedure. Also behind " $|\Phi|^{\mathcal{D}}$ " lies a " $t'_d$ " which is the instant when the first of the  $|\Phi|^{\mathcal{D}}$  rendez-vous occurred. This  $t'_d$  is different for each instrument. The definitions given in Chapter 4 Section 3 have to be adapted depending on the choice of  $\alpha$  and  $\beta$ .

4253 algorithm is not very efficient: it computes rates based on sets that are not often changing (and  
 4254 probably not a lot) between two time steps. This section aims at providing elements for the  
 4255 design of a more efficient way to perform in real-time the diagnosis of drift faults in a sensor  
 4256 network.

### 4257 6.1 Choice of an event-based approach

4258 To provide a real-time formulation of the diagnosis algorithm, we choose to rely on an  
 4259 event-based approach, at least regarding  $\alpha$  following the notations of Section 5.

4260 This choice is guided by the fact that the sets of rendez-vous  $\Phi(s_i \rightarrow S, (\alpha, \beta))$  for each  $s_i \in S$   
 4261 do not have to be actualised at each time step but only when a new rendez-vous happens.

4262 However, if  $\beta = \Delta t$ , it invites to determine the statuses of the instruments based on a sliding  
 4263 window of rendez-vous that changes at each time step. This case is discussed afterwards.

### 4264 6.2 General idea

4265 Consider two instruments  $s_i$  and  $s_j$ . When a rendez-vous happens between them, at  $t$  for  
 4266 instance,  $\Phi(s_i \rightarrow S, (\alpha, \beta))$  and  $\Phi(s_j \rightarrow S, (\alpha, \beta))$  are changed. Thus, the predicted statuses of  $s_i$   
 4267 and  $s_j$  have to be actualised.

4268 To determine the statuses of the instruments after this rendez-vous, the principles introduced  
 4269 in Chapter 4 Section 4.1 have to be extended.

4270 Indeed, we must determine which instrument between  $s_i$  and  $s_j$  is the diagnoser instrument  
 4271 for the other. The set of diagnosers of  $s_i$  at  $t$  is noted  $S_{diagnosers}(s_i, t)$ .

4272 At  $t$ , in a real-time approach, we have values for  $\widehat{\Omega}(s_i, t)$  and  $\widehat{\Omega}(s_j, t)$ . Depending on the case,  
 4273 the rendez-vous  $\varphi(s_i \rightarrow s_j, t)$  will be considered either in  $\Phi_v(s_i \rightarrow S, (\alpha, \beta))$  or  $\Phi_v(s_j \rightarrow S, (\alpha, \beta))$   
 4274 (sets of valid rendez-vous).

4275 The following cases are possible:

$\widehat{\Omega}(s_i, t)$	$\widehat{\Omega}(s_j, t)$	Decision
NF	NF	The diagnoser must be chosen between $s_i$ and $s_j$ .
NF	A	$s_i$ is a diagnoser for $s_j$ .
NF	F	$s_i$ is a diagnoser for $s_j$ .
A	NF	$s_j$ is a diagnoser for $s_i$ .
A	A	The rendez-vous cannot be considered as valid by any of the instruments.
A	F	The rendez-vous cannot be considered as valid by any of the instruments.
F	NF	$s_j$ is a diagnoser for $s_i$ .
F	A	The rendez-vous cannot be considered as valid by any of the instruments.
F	F	The rendez-vous cannot be considered as valid by any of the instruments.

**Table A.6.1:** Possible cases when a new rendez-vous occurs between  $s_i$  and  $s_j$  for a real-time diagnosis

4276 In fact, there is only one particular case: when  $\widehat{\Omega}(s_i, t) = \widehat{\Omega}(s_j, t) = NF$ . To choose the one  
 4277 to use as a diagnoser, we propose to rely on the class and on the duration since the instruments  
 4278 were predicted as non-faulty. The diagnoser is the one that has first the highest class and then  
 4279 the one that has been predicted as non-faulty first.

4280 Once this information is determined, e.g.  $s_j$  is a diagnoser for  $s_i$  for instance, the set of valid  
 4281 rendez-vous  $\Phi_v(s_i \rightarrow S, (\alpha, \beta))$  is actualised following a derived version of Equation 4.1 from

4282 Chapter 4 Section 4.1:

$$\begin{aligned}
 \Phi_v(s_i \rightarrow S, (\alpha, \beta)) = & \{\varphi(s_i \rightarrow s_j, t) \in \Phi(s_i \rightarrow S, (\alpha, \beta)), \text{ such as} \\
 & s_j \in S_{diagnosers}(s_i, t), // s_j \text{ is a diagnoser of } s_i \\
 & \widehat{\Omega}(s_j, t') = NF, // s_j \text{ is non-faulty} \\
 & c(s_j) \geq c_{min}^D(s_i), // \text{The class of } s_j \text{ is higher or equal to the minimal} \\
 & \text{class allowed to diagnose } s_i \\
 & m(s_i, t) \in M^*(s_i, (\alpha, \beta)), // \text{The measurement result of } s_i \text{ is valid} \\
 & m(s_j, t') \in M^*(s_j, (\alpha, \beta))\} // \text{The measurement result of } s_j \text{ is valid}
 \end{aligned}
 \tag{A.2}$$

4283 Then, the predicted status of  $s_i$  can be actualised following the same principles described in  
 4284 Chapter 4 Section 4.2 (rates, criteria...). Because the predicted status of only of the instruments  
 4285 involved in the rendez-vous is actualised, it would be valuable to use a decentralised approach.

### 4286 6.3 Initialisation of the algorithm

4287 Like in Chapter 4 Section 4.2, all the instruments in  $S^{c_{max}}$  are assumed as non-faulty, the  
 4288 set of instruments to diagnose  $S^D$  is equal to  $S \setminus S^{c_{max}}$  and the predicted statuses  $\widehat{\Omega}(s_i, t_d)$  are  
 4289 initialised to non-faulty for  $s_i \in S^{c_{max}}$  and to ambiguous for  $s_i \in S^D$ .

4290 At  $t = 0$ , for each  $s_i \in S$ ,  $\Phi(s_i \rightarrow S, (t, \beta)) = \emptyset$ . Overtime, these sets will be filled with  
 4291 rendez-vous. Only when  $|\Phi_v(s_i \rightarrow S, (\alpha, \beta))| \geq |\Phi_v|_{min}$  for  $s_i \in S^D$ , the predicted statuses can  
 4292 begin to change.

### 4293 6.4 Allowed changes of status

4294 With the principles presented, the predicted statuses of the instruments can change from  
 4295 one rendez-vous to another like in Chapter 4 Section 4.2. It may be preferable to consider an  
 4296 approach where the predicted status of an instrument is kept equal to faulty once it is first  
 4297 diagnosed as so as suggested in Chapter 4 Section 7 to avoid false results.

### 4298 6.5 Conclusion

4299 In this section, we provided avenues toward the definition of a real-time formulation of  
 4300 our diagnosis algorithm. We suggested using an event-based approach following the concepts  
 4301 introduced in Section 5. We introduced a general idea to actualise the predicted status of the  
 4302 instruments when a new rendez-vous occurs and discussed the particular cases. We presented the  
 4303 criteria of initialisation which are equivalent to those of Chapter 4 and we recommended using  
 4304 the approach of Chapter 4 Section 7 where the predicted status of an instrument is kept equal  
 4305 to faulty once it is first diagnosed. To validate the ideas presented, it would be necessary to  
 4306 fully develop the algorithm and compared its execution on a same case study where our original  
 4307 diagnosis algorithm<sup>61</sup> would be applied at each time step.

4308 Afterwards, the algorithm giving an upper boundary to  $|\Phi_v|_{min}$ , the minimal size for a set  
 4309 of valid rendez-vous to allow a prediction between the status faulty and non-faulty could be  
 4310 extended for the cases where the diagnosis is performed in real-time.

---

<sup>61</sup>With the predicted status of the instruments kept equal to faulty once first diagnosed as so.

# Appendix B

## On the Reason Why Assuming the Instruments of Class $c_{max}$ are not Drifting is Acceptable

---

In this appendix, more details are briefly given to understand why assuming the instruments of class  $c_{max}$  are not drifting is acceptable.

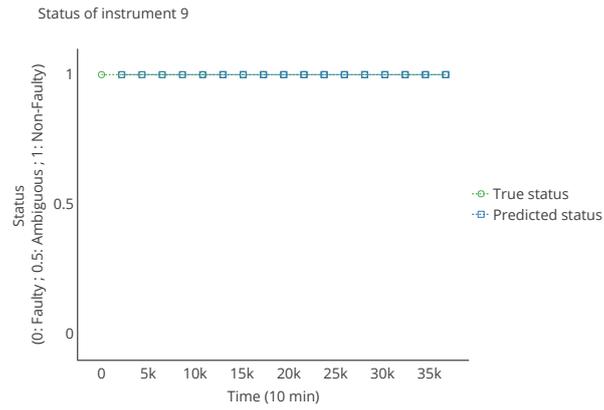
In Chapter 4 Section 6.2, it appeared that even with  $s_9$  drifting, the diagnosis algorithm still provides satisfying results when assuming this instrument as always non-faulty.

Suppose that there was a means to correctly predict the status of  $s_9$ . Following the reasoning that was used to design the algorithm in Chapter 4 Section 4, this instrument should not be considered at some point for the prediction of the statuses of other instruments. Nevertheless, based on the results obtained in the study we conducted, the information provided by this instrument with its measurement results is of a sufficient quality to help the prediction of the other instruments.

In fact, this can be explained by the definition used for  $r_{true}$  in Chapter 4 Section 5.3.  $r_{true}$  uses the set of measurements results compatible with the true values  $M^{\approx}(s_i, (t, \Delta))$ . The compatibility of a measurement with its true value is appreciated regarding the uncertainty on the measured value and the true value. For the instruments of class  $c_{max}$  here, the uncertainty, absolute or relative, is significantly smaller than the uncertainty of the instruments of lower class, particularly in this case study (see Table 4.5.1 in Chapter 4 Section 5 where  $\Delta_r v(s_9) \ll \Delta_r v(s_j)$  for all the other instruments  $s_j$ ). Thus, with this definition of  $r_{true}$ , when two instruments of different classes have their true statuses equal to faulty for the first time, it does not mean that their drift is equivalent. It is possible that an instrument of a given class is still performing well relatively to an instrument of a lower class, which is the case here with  $s_9$  drifting more slowly:  $\delta G_{max}(s_9) \ll \delta G_{max}(s_j)$  and  $\delta O_{max}(s_9) \ll \delta O_{max}(s_j)$ . This is the most important feature and such relationships would likely be valid in practice with instruments of different metrological class.

Therefore, to better appreciate the true status of a given instrument regarding one of a lower class (or not), the compatibility of its measurement results with its true values should be evaluated considering the highest measurement uncertainty between these two devices. In our case, the true status of  $s_9$  is always non-faulty if  $M^{\approx}(s_i, (t, \Delta))$  is determined with the relative uncertainty  $\Delta_r v$  of the instruments of class zero than with its own  $\Delta_r v$  as shown in Figure B.0.1.

In conclusion, assuming that  $s_9$  is not drifting and is always non-faulty is valid from the point of view of all the other instruments of lower class in the network because even when drifting, this instrument is still more accurate than the instruments of a lower class. The metric



**Figure B.0.1:** Evolution of the true status of  $s_9$ , computed with the relative uncertainty  $\Delta_r v$  on the results of the instruments of class 0 instead of its own  $\Delta_r v$ , and its predicted status when it is drifting

4347 used to determine if an instrument is truly faulty or not is not fairly appreciating the status  
4348 of an instrument when the goal is to compare its measurement results to the ones of another  
4349 instrument, potentially of a worse quality.

# Sensitivity of the Diagnosis Algorithm: Case Study with Values Related Over Time for the Parameters of the True Values' Model

---

## 1 Introduction

We extend here the results of Chapter 4 Section 9.3 on the influence of the model used for the true values in the case study developed in Section 5 of the same chapter. More precisely, an additional case study is provided to confirm that the realism of the model for the true values does not have a major influence.

## 2 Model used

We propose to use for the true values the same model as in Chapter 4 Section 3 but instead of using values drawn randomly for the amplitude and FWHM of the Gaussian curve at each time step, the dataset that was exploited in Chapter 3 Section 3 is considered here to derive these values. Thus, the values of the amplitude and FWHM of the Gaussian curve are related from one time step to another.

In this dataset, the values are given hourly. As we used a time step equivalent to 10min in Chapter 4 Section 3, the dataset has to be resampled. The missing values are interpolated linearly.

Apart from the model for the true values, the case study is identical to the one of Chapter 4 Section 8 (case study of Section 3 with improvements regarding false results and the adjustment of  $|\Phi_v|_{min}$ ).

## 3 Results

Table C.3.1 gives the values of each metric. Compared to the results of Table 4.8.1 and considering there is a slightly higher prevalence here (0.88 instead of 0.76), e.g. more faulty cases than non-faulty ones, we observe there are less false negative results (10 instead of 14). The number of true negatives  $TN$  is equal to the number of negative cases  $N$  (18). Thus, there are no false positives. There are also no predicted statuses equal to ambiguous.

Therefore, the  $TNR$  and  $TPR$  are high (1.00 and 0.93) and so is the  $PPV$  (1.00). Only the  $NPV$  and  $FOR$  are worse than with a random draw of the parameters of the model of the true values (0.64 and 0.36 respectively instead of 0.72 and 0.28). Although, considering the low

		True status		Prevalence	Accuracy
		Non-faulty $N = 18$	Faulty $P = 135$		
Predicted status	Non-faulty	$TN$ 18	$FN$ 10	$NPV$ 0.64	$FOR$ 0.36
	Ambiguous	$NDN$ 0	$NDP$ 0		
	Faulty	$FP$ 0	$TP$ 125	$FDR$ 0.00	$PPV$ 1.00
		$TNR$ 1.00	$FNR$ 0.07		
		$NDNR$ 0.00	$NDPR$ 0.00	$NDR$ 0.00	
		$FPR$ 0.00	$TPR$ 0.93		

**Table C.3.1:** Confusion matrix of the case study with true values based on the 2D Gauss model with values of its parameters derived from a real dataset

4382 number of  $TN$  due to the low value of  $N$ , this is not major.

4383 The overall performance from the point of view of the accuracy is slightly better than in  
4384 Table 4.8.1 (0.93 instead of 0.91)

#### 4385 4 Conclusion

4386 This case study confirms the conclusion from Chapter 4 Section 9.3, e.g. the model chosen for  
4387 the true values does not seem to have a significant effect on the results. Indeed, the mathematical  
4388 objects used by the diagnosis algorithm are sets: the order between the measured values is  
4389 not exploited. Combined to the observation of Chapter 4 Section 7 (the only measurement  
4390 results considered to the status of a measuring instrument are the ones obtained when it is in  
4391 rendez-vous. Thus, a sampling of all its measurement results is performed, breaking already the  
4392 continuity of the measured signal), the relationship between two consecutive measured values  
4393 has no significance for the diagnosis algorithm: it is aimed at determining a status over a time  
4394 range and this time range is considered as a whole. Therefore, it is the values that have the most  
4395 significant effect as shown earlier in Chapter 4 Section 7.

4396

- 4398 [1] Mohammed Abo-Zahhad, Sabah M. Ahmed, Nabil Sabor, and Shigenobu Sasaki. “Mobile  
4399 Sink-Based Adaptive Immune Energy-Efficient Clustering Protocol for Improving the  
4400 Lifetime and Stability Period of Wireless Sensor Networks”. In: *IEEE Sensors Journal*  
4401 15.8 (2015), pp. 4576–4586. DOI: 10.1109/JSEN.2015.2424296.
- 4402 [2] Airparif. *Airparif website*. 2020. URL: <https://www.airparif.asso.fr> (visited on  
4403 01/29/2019).
- 4404 [3] Airparif. *Laboratoire de métrologie*. 2020. URL: [https://www.airparif.asso.fr/  
4405 methodes-surveillance/laboratoire-metrologie](https://www.airparif.asso.fr/methodes-surveillance/laboratoire-metrologie) (visited on 01/29/2019).
- 4406 [4] Adrian Arfire, Ali Marjovi, and Alcherio Martinoli. “Model-based rendezvous calibration  
4407 of mobile sensor networks for monitoring air quality”. In: *2015 IEEE SENSORS*. 2015,  
4408 pp. 1–4. DOI: 10.1109/ICSENS.2015.7370258.
- 4409 [5] Algirdas Avizienis, Jean-Claude Laprie, Brian Randell, and Carl Landwehr. “Basic Con-  
4410 cepts and Taxonomy of Dependable and Secure Computing”. In: *IEEE Transactions on*  
4411 *Dependable and Secure Computing* 1.1 (2004), pp. 11–33. DOI: 10.1109/TDSC.2004.2.
- 4412 [6] Marek Badura, Piotr Batog, Anetta Drzeniecka-Osiadacz, and Piotr Modzel. “Regression  
4413 methods in the calibration of low-cost sensors for ambient particulate matter measure-  
4414 ments”. In: *SN Applied Sciences* 1.6 (2019). DOI: 10.1007/s42452-019-0630-1.
- 4415 [7] Laura Balzano and Robert Nowak. “Blind Calibration of Sensor Networks”. In: *2007 6th*  
4416 *International Symposium on Information Processing in Sensor Networks* (2007), pp. 79–88.  
4417 DOI: 10.1109/IPSN.2007.4379667.
- 4418 [8] Jose M. Barcelo-Ordinas, Messaoud Doudou, Jorge Garcia-Vidal, and Nadjib Badache.  
4419 “Self-calibration methods for uncontrolled environments in sensor networks: A reference  
4420 survey”. In: *Ad Hoc Networks* 88 (2019), pp. 142–159. DOI: 10.1016/j.adhoc.2019.01.  
4421 008.
- 4422 [9] Jose M. Barcelo-Ordinas, Pau Ferrer-Cid, Jorge Garcia-Vidal, Anna Ripoll, and Mar  
4423 Viana. “Distributed multi-scale calibration of low-cost ozone sensors in wireless sensor  
4424 networks”. In: *Sensors (Switzerland)* 19.11 (2019), pp. 1–25. DOI: 10.3390/s19112503.
- 4425 [10] Jose M. Barcelo-Ordinas, Jorge Garcia-Vidal, Messaoud Doudou, Santiago Rodrigo-Munoz,  
4426 and Albert Cerezo-Llavero. “Calibrating low-cost air quality sensors using multiple arrays  
4427 of sensors”. In: *IEEE Wireless Communications and Networking Conference, WCNC*.  
4428 2018, pp. 1–6. DOI: 10.1109/WCNC.2018.8377051.
- 4429 [11] Thomas Becnel, Tofigh Sayahi, Kerry Kelly, and Pierre Emmanuel Gaillardon. “A recursive  
4430 approach to partially blind calibration of a pollution sensor network”. In: *2019 IEEE*  
4431 *International Conference on Embedded Software and Systems, ICSS 2019*. 2019. DOI:  
4432 10.1109/ICSS.2019.8782523.

- 4433 [12] Antoine Berchet, Katrin Zink, Clive Muller, Dietmar Oettl, Juerg Brunner, Lukas  
4434 Emmenegger, and Dominik Brunner. “A cost-effective method for simulating city-wide air  
4435 flow and pollutant dispersion at building resolving scale”. In: *Atmospheric Environment*  
4436 158 (2017), pp. 181–196. DOI: 10.1016/j.atmosenv.2017.03.030.
- 4437 [13] Cagdas Bilen, Gilles Puy, Rémi Gribonval, and Laurent Daudet. “Convex Optimization  
4438 Approaches for Blind Sensor Calibration using Sparsity”. In: *IEEE Transactions on Signal*  
4439 *Processing* (2013), pp. 1–11. DOI: 10.1109/TSP.2014.2342651.
- 4440 [14] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML. *International vocabulary*  
4441 *of metrology - Basic and general concepts and associated terms (VIM)*. 3rd edn. JCGM  
4442 200:2012, 2012.
- 4443 [15] C. Borrego, A. M. Costa, J. Ginja, M. Amorim, M. Coutinho, K. Karatzas, Th Sioumis, N.  
4444 Katsifarakis, K. Konstantinidis, S. De Vito, et al. “Assessment of air quality microsensors  
4445 versus reference methods: The EuNetAir joint exercise”. In: *Atmospheric Environment*  
4446 147.2 (2016), pp. 246–263. DOI: 10.1016/j.atmosenv.2016.09.050.
- 4447 [16] C. Borrego, J. Ginja, M. Coutinho, C. Ribeiro, K. Karatzas, Th Sioumis, N. Katsifarakis,  
4448 K. Konstantinidis, Saverio De Vito, Elena Esposito, et al. “Assessment of air quality  
4449 microsensors versus reference methods: The EuNetAir Joint Exercise – Part II”. In:  
4450 *Atmospheric Environment* 193 (2018), pp. 127–142. DOI: 10.1016/j.atmosenv.2018.08.  
4451 028.
- 4452 [17] A. Botchkarev. “A New Typology Design of Performance Metrics to Measure Errors in  
4453 Machine Learning Regression Algorithms”. In: *Interdisciplinary Journal of Information,*  
4454 *Knowledge, and Management* 14 (2019), pp. 045–076. DOI: 10.28945/4184.
- 4455 [18] Ahmed Boubrima, Walid Bechkit, and Herve Rivano. “A new WSN deployment approach  
4456 for air pollution monitoring”. In: *2017 14th IEEE Annual Consumer Communications*  
4457 *and Networking Conference, CCNC 2017*. 2017, pp. 455–460. DOI: 10.1109/CCNC.2017.  
4458 7983151.
- 4459 [19] Ahmed Boubrima, Walid Bechkit, and Herve Rivano. “Optimal WSN Deployment Models  
4460 for Air Pollution Monitoring”. In: *IEEE Transactions on Wireless Communications* 16.5  
4461 (2017), pp. 2723–2735. DOI: 10.1109/TWC.2017.2658601.
- 4462 [20] Ahmed Boubrima, Walid Bechkit, and Hervé Rivano. “On the Deployment of Wireless  
4463 Sensor Networks for Air Quality Mapping: Optimization Models and Algorithms”. In:  
4464 *IEEE/ACM Transactions on Networking* 27.4 (Aug. 2019), pp. 1629–1642. DOI: 10.1109/  
4465 TNET.2019.2923737.
- 4466 [21] Ahmed Boubrima, Walid Bechkit, Hervé Rivano, and Lionel Soulhac. “Leveraging the  
4467 Potential of WSN for an Efficient Correction of Air Pollution Fine-Grained Simulations”.  
4468 In: *The 27th International Conference on Computer Communications and Networks*  
4469 *(ICCCN 2018)*. Hangzhou, China, 2018, pp. 1–9. DOI: 10.1109/ICCCN.2018.8487343.
- 4470 [22] Marcel Bruins, Jan Willem, Wendy W. J. Van De Sande, Alex Van Belkum, and Albert  
4471 Bos. “Enabling a transferable calibration model for metal-oxide type electronic noses”.  
4472 In: *Sensors and Actuators: B. Chemical* 188 (2013), pp. 1187–1195. DOI: 10.1016/j.snb.  
4473 2013.08.006.
- 4474 [23] Seamus Buadhachain and Gregory Provan. “A model-based control method for decen-  
4475 tralized calibration of wireless sensor networks”. In: *2013 American Control Conference*.  
4476 2013, pp. 6571–6576. DOI: 10.1109/ACC.2013.6580870.

- 4477 [24] Vladimir Bychkovskiy, Seapahn Megerian, Deborah Estrin, and Miodrag Potkonjak.  
4478 “A collaborative approach to in-place sensor calibration”. In: *Proceedings of the 2nd*  
4479 *International Conference on Information Processing in Sensor Networks*. 2003, pp. 301–  
4480 316. DOI: 10.1007/3-540-36978-3\_20.
- 4481 [25] Valerio Cambareri and Laurent Jacques. “A Non-Convex Blind Calibration Method for  
4482 Randomised Sensing Strategies”. In: *2016 4th International Workshop on Compressed*  
4483 *Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*. 2016,  
4484 pp. 16–20. DOI: 10.1109/CoSeRa.2016.7745690.
- 4485 [26] Joanna Gordon Casey and Michael P. Hannigan. “Testing the performance of field  
4486 calibration techniques for low-cost gas sensors in new deployment locations: across a  
4487 county line and across Colorado”. In: *Atmospheric Measurement Techniques* 11.11 (2018),  
4488 pp. 6351–6378. DOI: 10.5194/amt-11-6351-2018.
- 4489 [27] Nuria Castell, Franck R. Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak  
4490 Fishbain, David M. Broday, and Alena Bartonova. “Can commercial low-cost sensor  
4491 platforms contribute to air quality monitoring and exposure estimates?” In: *Environment*  
4492 *International* 99 (2017), pp. 293–302. DOI: 10.1016/j.envint.2016.12.007.
- 4493 [28] Prasenjit Chanak, Indrajit Banerjee, and R. Simon Sherratt. “Mobile sink based fault  
4494 diagnosis scheme for wireless sensor networks”. In: *Journal of Systems and Software* 119  
4495 (2016), pp. 45–57. DOI: 10.1016/j.jss.2016.05.041.
- 4496 [29] Jinran Chen, Shubha Kher, and Arun Somani. “Distributed fault detection of wireless  
4497 sensor networks”. In: *Proceedings of the 2006 workshop on Dependability issues in wireless*  
4498 *ad hoc networks and sensor networks - DIWANS '06*. 2006, p. 65. DOI: 10.1145/1160972.  
4499 1160985.
- 4500 [30] Tinghuan Chen, Bingqing Lin, Hao Geng, and Bei Yu. “Sensor drift calibration via spatial  
4501 correlation model in smart building”. In: *Proceedings - Design Automation Conference*.  
4502 2019. DOI: 10.1145/3316781.3317909.
- 4503 [31] Yun Cheng, Xiaoxi He, Zimu Zhou, and Lothar Thiele. “ICT: In-field Calibration Transfer  
4504 for Air Quality Sensor Deployments”. In: *Proceedings of the ACM on Interactive, Mobile,*  
4505 *Wearable and Ubiquitous Technologies* 3.1 (2019), pp. 1–19. DOI: 10.1145/3314393.
- 4506 [32] Yong-Seung Chung, Tae-Koon Kim, and Ki-Hyun Kim. “Temporal variation and cause of  
4507 acidic precipitation from a monitoring network in Korea”. In: *Atmospheric Environment*  
4508 30.13 (1996), pp. 2429–2435. DOI: 10.1016/1352-2310(95)00186-7.
- 4509 [33] Andrea L. Clements, William G. Griswold, Abhijit Rs, Jill E. Johnston, Megan M. Herting,  
4510 Jacob Thorson, Ashley Collier-Oxandale, and Michael Hannigan. “Low-Cost Air Quality  
4511 Monitoring Tools : From Research to Practice (A Workshop Summary)”. In: *Sensors* 17.11  
4512 (2017). DOI: 10.3390/s17112478.
- 4513 [34] José María Cordero, Rafael Borge, and Adolfo Narros. “Using statistical methods to carry  
4514 out in field calibrations of low cost air quality sensors”. In: *Sensors and Actuators, B:*  
4515 *Chemical* 267.2 (2018), pp. 245–254. DOI: 10.1016/j.snb.2018.04.021.
- 4516 [35] Eben S. Cross, Leah R. Williams, David K. Lewis, Gregory R. Magoon, Timothy B. Onasch,  
4517 Michael L. Kaminsky, Douglas R. Worsnop, and John T. Jayne. “Use of electrochemical  
4518 sensors for measurement of air pollution: Correcting interference response and validating  
4519 measurements”. In: *Atmospheric Measurement Techniques* 10.9 (2017), pp. 3575–3588.  
4520 DOI: 10.5194/amt-10-3575-2017.

- 4521 [36] Saverio De Vito, P Delli Veneri, Elena Esposito, Maria Salvato, W. Bright, R. L. Jones,  
4522 and Olalekan A.M. Popoola. “Dynamic Multivariate Regression For On-Field Calibration  
4523 Of High Speed Air Quality Chemical Multi-Sensor Systems”. In: *2015 XVIII AISEM*  
4524 *Annual Conference*. 2015, pp. 1–3. DOI: 10.1109/AISEM.2015.7066794.
- 4525 [37] Florentin Delaine, Bérengère Lebental, and Hervé Rivano. *Example case study applying a*  
4526 *“Framework for the Simulation of Sensor Networks Aimed at Evaluating In Situ Calibration*  
4527 *Algorithms”*. Version V1. 2020. DOI: 10.25578/CJCYMZ.
- 4528 [38] François Derkx, Bérengère Lebental, Tarik Bourouina, Frédéric Bourquin, Costel-Sorin  
4529 Cojocar, Enric Robine, and Henri Van Damme. “The Sense-City project”. In: *XVIIIth*  
4530 *Symposium on Vibrations, Shocks and Noise*. 2012.
- 4531 [39] Sharvari Deshmukh, Kalyani Kamde, Arun Jana, Sanjivani Korde, Rajib Bandyopadhyay,  
4532 Ravi Sankar, Nabarun Bhattacharyya, and R A Pandey. “Calibration transfer between  
4533 electronic nose systems for rapid In situ measurement of pulp and paper industry emissions”.  
4534 In: *Analytica Chimica Acta* 841 (2014), pp. 58–67. DOI: 10.1016/j.aca.2014.05.054.
- 4535 [40] Clément Dorffer. “Méthodes informées de factorisation matricielle pour l’étalonnage de  
4536 réseaux de capteurs mobiles et la cartographie de champs de pollution”. PhD thesis.  
4537 Université du Littoral Côte d’Opale, France, 2017.
- 4538 [41] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. “Blind Calibration of  
4539 Mobile Sensors Using Informed Nonnegative Matrix Factorization”. In: *12th International*  
4540 *Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2015)*. 2015,  
4541 pp. 497–505. DOI: 10.1109/ICASSP.2016.7472216.
- 4542 [42] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. “Blind mobile  
4543 sensor calibration using an informed nonnegative matrix factorization with a relaxed  
4544 rendezvous model”. In: *2016 IEEE International Conference on Acoustics, Speech and*  
4545 *Signal Processing (ICASSP)*. 2016, pp. 2941–2945. DOI: 10.1109/ICASSP.2016.7472216.
- 4546 [43] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. “Nonlinear mobile  
4547 sensor calibration using informed semi-nonnegative matrix factorization with a Vander-  
4548 monde factor”. In: *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop*  
4549 *(SAM)*. 2016. DOI: 10.1109/SAM.2016.7569735.
- 4550 [44] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. “Outlier-robust  
4551 calibration method for sensor networks”. In: *Proceedings of the 2017 IEEE Interna-*  
4552 *tional Workshop of Electronics, Control, Measurement, Signals and their Application to*  
4553 *Mechatronics, ECMSM 2017*. 2017. DOI: 10.1109/ECMSM.2017.7945907.
- 4554 [45] Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel. “Informed Nonneg-  
4555 ative Matrix Factorization Methods for Mobile Sensor Network Calibration”. In: *IEEE*  
4556 *Transactions on Signal and Information Processing over Networks* 4.4 (2018), pp. 667–682.  
4557 DOI: 10.1109/TSIPN.2018.2811962.
- 4558 [46] Bénédicte Dousset, Françoise Gourmelon, Karine Laaidi, Abdelkrim Zeghnoun, Emmanuel  
4559 Giraudet, Philippe Bretin, Elena Mauri, and Stéphanie Vandentorren. “Satellite monitoring  
4560 of summer heat waves in the Paris metropolitan area”. In: *International Journal of*  
4561 *Climatology* 31.2 (2011), pp. 313–323. DOI: 10.1002/joc.2222.
- 4562 [47] Elena Esposito, Saverio De Vito, Maria Salvato, V. Bright, R. L. Jones, and O. Popoola.  
4563 “Dynamic neural network architectures for on field stochastic calibration of indicative low  
4564 cost air quality sensing systems”. In: *Sensors and Actuators, B: Chemical* 231 (2016),  
4565 pp. 701–713. DOI: 10.1016/j.snb.2016.03.038.

- 4566 [48] Elena Esposito, Saverio De Vito, Maria Salvato, Grazia Fattoruso, V Bright, R L Jones,  
4567 Dte-fsn-dineneap E Fermi, and Portici Na. “Stochastic Comparison of Machine Learning  
4568 Approaches to Calibration of Mobile Air Quality Monitors”. In: *Sensors: Proceedings of*  
4569 *the Third National Conference on Sensors*. 2018. DOI: 10.1007/978-3-319-55077-0\_38.
- 4570 [49] Elena Esposito, Saverio De Vito, Maria Salvato, Grazia Fattoruso, and Girolamo Di  
4571 Francia. “Computational Intelligence for Smart Air Quality Monitors Calibration”. In:  
4572 *Computational Science and Its Applications - ICCSA 2017*. 2017, pp. 443–454. DOI:  
4573 10.1007/978-3-319-62398-6\_31.
- 4574 [50] Xinwei Fang and Iain Bate. “Using Multi-parameters for Calibration of Low-cost Sensors in  
4575 Urban Environment”. In: *Proceedings of the 2017 International Conference on Embedded*  
4576 *Wireless Systems and Networks*. 2017, pp. 1–11.
- 4577 [51] Xinwei Fang and Iain Bate. “An Improved Sensor Calibration with Anomaly Detection and  
4578 Removal”. In: *Sensors and Actuators, B: Chemical* 307.November 2019 (2020), p. 127428.  
4579 DOI: 10.1016/j.snb.2019.127428.
- 4580 [52] Yang Feiyue, Tao Yang, Zhang Siqing, Dai Jianjian, Xu Juan, and Hou Yao. “A Faulty node  
4581 detection algorithm based on spatial-temporal cooperation in wireless sensor networks”.  
4582 In: *Procedia Computer Science*. Vol. 131. Elsevier B.V., 2018, pp. 1089–1094. DOI: 10.  
4583 1016/j.procs.2018.04.266.
- 4584 [53] Pau Ferrer-Cid, Jose M. Barcelo-Ordinas, Jorge Garcia-Vidal, Anna Ripoll, and Mar  
4585 Viana. “A comparative Study of Calibration Methods for Low-Cost Ozone Sensors in IoT  
4586 Platforms”. In: *IEEE Internet of Things Journal* PP.X (2019), pp. 1–1. DOI: 10.1109/  
4587 jiot.2019.2929594.
- 4588 [54] Barak Fishbain, Uri Lerner, Nuria Castell, Tom Cole-Hunter, Olalekan A.M. Popoola,  
4589 David M. Broday, Tania Martinez-Iñiguez, Mark Nieuwenhuijsen, Milena Jovašević-  
4590 Stojanović, Dušan Topalović, et al. “An evaluation tool kit of air quality micro-sensing  
4591 units”. In: *Science of the Total Environment* 575 (2017), pp. 639–648. DOI: 10.1016/j.  
4592 scitotenv.2016.09.061.
- 4593 [55] Barak Fishbain and Erick Moreno-Centeno. “Self Calibrated Wireless Distributed Envi-  
4594 ronmental Sensory Networks”. In: *Scientific Reports* 6 (2016), pp. 1–10. DOI: 10.1038/  
4595 srep24382.
- 4596 [56] J. Fonollosa, L. Fernandez, A. Gutierrez-Galvez, R. Huerta, and S. Marco. “Calibration  
4597 transfer and drift counteraction in chemical sensor arrays using Direct Standardization”.  
4598 In: *Sensors and Actuators, B: Chemical* 236 (2016), pp. 1044–1053. DOI: 10.1016/j.snb.  
4599 2016.05.089.
- 4600 [57] Kaibo Fu, Wei Ren, and Wei Dong. “Multihop Calibration for Mobile Sensing : k-hop  
4601 Calibratability and Reference Sensor Deployment”. In: *IEEE INFOCOM 2017 - IEEE*  
4602 *Conference on Computer Communications*. 2017, pp. 1–9. DOI: 10.1109/INFOCOM.2017.  
4603 8056962.
- 4604 [58] Meiling Gao, Junji Cao, and Edmund Seto. “A distributed network of low-cost continuous  
4605 reading sensors to measure spatiotemporal variations of PM2.5 in Xi’an, China”. In:  
4606 *Environmental Pollution* 199 (2015), pp. 56–65. DOI: 10.1016/j.envpol.2015.01.013.
- 4607 [59] Hugh S. Gorman and Erik M. Conway. “Monitoring the environment: Taking a historical  
4608 perspective”. In: *Environmental Monitoring and Assessment* 106.1-3 (2005), pp. 1–10.  
4609 DOI: 10.1007/s10661-005-0755-0.

- 4610 [60] Gayle S.W. Hagler, Ronald Williams, Vasileios Papapostolou, and Andrea Polidori. “Air  
4611 Quality Sensors and Data Adjustment Algorithms: When Is It No Longer a Measurement?”  
4612 In: *Environmental Science and Technology* 52.10 (2018), pp. 5530–5531. DOI: 10.1021/  
4613 acs.est.8b01826.
- 4614 [61] S.R. Hanna, G.A. Briggs, and R.P. Jr. Hosker. *Handbook on atmospheric diffusion*. Tech.  
4615 rep. National Oceanic, Atmospheric Administration, Oak Ridge, TN (USA). Atmospheric  
4616 Turbulence, and Diffusion Lab., Jan. 1982. DOI: 10.2172/5591108.
- 4617 [62] Jane K. Hart and Kirk Martinez. “Environmental Sensor Networks: A revolution in  
4618 the earth system science?” In: *Earth-Science Reviews* 78.3-4 (2006), pp. 177–191. DOI:  
4619 10.1016/j.earscirev.2006.05.001.
- 4620 [63] David Hasenfrazt, Olga Saukh, and Lothar Thiele. “On-the-fly calibration of low-cost gas  
4621 sensors”. In: *Proceedings of the 9th European Conference on Wireless Sensor Networks*.  
4622 Ed. by Gian Pietro Picco and Wendi Heinzelman. Springer Berlin Heidelberg, 2012,  
4623 pp. 228–244. DOI: 10.1007/978-3-642-28169-3\_15.
- 4624 [64] D. M. Holstius, A. Pillarisetti, K. R. Smith, and E. Seto. “Field calibrations of a low-cost  
4625 aerosol sensor at a regulatory monitoring site in California”. In: *Atmospheric Measurement*  
4626 *Techniques* 7.4 (2014), pp. 1121–1131. DOI: 10.5194/amt-7-1121-2014.
- 4627 [65] Alexander T. Ihler, J W Fisher III, Randolph L. Moses, and Alan S. Willsky. “Nonpara-  
4628 metric belief propagation for self-calibration in sensor networks”. In: *Third International*  
4629 *Symposium on Information Processing in Sensor Networks, 2004. IPSN 2004*. 2004,  
4630 pp. 225–233. DOI: 10.1145/984622.984656.
- 4631 [66] Marian Emanuel Ionascu, Iasmina Gruicin, and Marius Marcu. “Towards Wearable Air  
4632 Quality Monitoring Systems - Initial Assessments on Newly Developed Sensors”. In: *2018*  
4633 *26th Telecommunications Forum, TELFOR 2018 - Proceedings*. IEEE, 2018, pp. 1–4. DOI:  
4634 10.1109/TELFOR.2018.8611832.
- 4635 [67] International Organization for Standardization. *ISO 9001:2015(E), Quality management*  
4636 *systems - Requirements*. Tech. rep. Geneva, CH: International Organization for Standard-  
4637 ization, Oct. 2015.
- 4638 [68] Rohan Jayaratne, Xiaoting Liu, Phong Thai, Matthew Dunbabin, and Lidia Morawska.  
4639 “The influence of humidity on the performance of a low-cost air particle mass sensor and  
4640 the effect of atmospheric fog”. In: *Atmospheric Measurement Techniques* 11.8 (2018),  
4641 pp. 4883–4890. DOI: 10.5194/amt-11-4883-2018.
- 4642 [69] Sai Ji, Shen Fang Yuan, Ting Huai Ma, and Chang Tan. “Distributed fault detection for  
4643 wireless sensor based on weighted average”. In: *NSWCTC 2010 - The 2nd International*  
4644 *Conference on Networks Security, Wireless Communications and Trusted Computing*.  
4645 Vol. 1. 2010, pp. 57–60. DOI: 10.1109/NSWCTC.2010.21.
- 4646 [70] Jinfang Jiang, Guangjie Han, Feng Wang, Lei Shu, and Mohsen Guizani. “An Efficient  
4647 Distributed Trust Model for Wireless Sensor Networks”. In: *IEEE Transactions on Parallel*  
4648 *and Distributed Systems* 26.5 (2015), pp. 1228–1237. DOI: 10.1109/TPDS.2014.2320505.
- 4649 [71] Wan Jiao, Gayle Hagler, Ronald Williams, Robert Sharpe, Ryan Brown, Daniel Garver,  
4650 Robert Judge, Motria Caudill, Joshua Rickard, Michael Davis, et al. “Community Air  
4651 Sensor Network (CAIRSENSE) project : evaluation of low-cost sensor performance in a  
4652 suburban environment in the southeastern United States”. In: *Atmospheric Measurement*  
4653 *Techniques* 9 (2016), pp. 5281–5292. DOI: 10.5194/amt-9-5281-2016.
- 4654 [72] Joint Committee For Guides In Metrology. *Evaluation of measurement data — Guide to*  
4655 *the expression of uncertainty in measurement*. Vol. 50. September. 2008, p. 134.

- 4656 [73] Federico Karagulian, Maurizio Barbieri, Alexander Kotsev, Laurent Spinelle, Michel  
4657 Gerboles, Friedrich Lagler, Nathalie Redon, Sabine Crunaire, and Annette Borowiak.  
4658 “Review of the performance of low-cost sensors for air quality monitoring”. In: *Atmosphere*  
4659 10.9 (2019). DOI: 10.3390/atmos10090506.
- 4660 [74] Jinsol Kim, Alexis A. Shusterman, Kaitlyn J. Lieschke, Catherine Newman, and Ronald C.  
4661 Cohen. “The Berkeley Atmospheric CO<sub>2</sub> Observation Network: Field calibration and  
4662 evaluation of low-cost air quality sensors”. In: *Atmospheric Measurement Techniques* 11.4  
4663 (2018), pp. 1937–1946. DOI: 10.5194/amt-11-1937-2018.
- 4664 [75] Fadi Kizel, Yael Etzion, Rakefet Shafran-Nathan, Ilan Levy, Barak Fishbain, Alena  
4665 Bartonova, and David M. Broday. “Node-to-node field calibration of wireless distributed  
4666 air pollution sensor network”. In: *Environmental Pollution* 233 (2018), pp. 900–909. DOI:  
4667 10.1016/j.envpol.2017.09.042.
- 4668 [76] Dheeraj Kumar, Sutharshan Rajasegarar, and Marimuthu Palaniswami. “Automatic sensor  
4669 drift detection and correction using spatial kriging and kalman filtering”. In: *2013 IEEE*  
4670 *International Conference on Distributed Computing in Sensor Systems*. 2013, pp. 183–190.  
4671 DOI: 10.1109/DCOSS.2013.52.
- 4672 [77] Dheeraj Kumar, Sutharshan Rajasegarar, and Marimuthu Palaniswami. “Geospatial  
4673 Estimation-Based Auto Drift Correction in Wireless Sensor Networks”. In: *ACM Transac-*  
4674 *tions on Sensor Networks* 11.3 (2015), 50:1–50:39. DOI: 10.1145/2736697.
- 4675 [78] Prashant Kumar, Lidia Morawska, Claudio Martani, George Biskos, Marina Neophytou,  
4676 Silvana Di Sabatino, Margaret Bell, Leslie Norford, and Rex Britter. “The rise of low-cost  
4677 sensing for managing air pollution in cities”. In: *Environment International* 75 (2015),  
4678 pp. 199–205. DOI: 10.1016/j.envint.2014.11.019.
- 4679 [79] Intel Berkeley Research lab. *Intel Lab Data*. 2004. URL: [http://db.lcs.mit.edu/  
4680 labdata/labdata.html](http://db.lcs.mit.edu/labdata/labdata.html) (visited on 10/22/2019).
- 4681 [80] Rachid Laref, Etienne Losson, Alexandre Sava, and Maryam Siadat. “Support vector  
4682 machine regression for calibration transfer between electronic noses dedicated to air  
4683 pollution monitoring”. In: *Sensors* 18.11 (2018). DOI: 10.3390/s18113716.
- 4684 [81] Alain Le Tertre, Agnès Lefranc, Daniel Eilstein, Christophe Declercq, Sylvia Medina,  
4685 Myriam Blanchard, Benoît Chardon, Pascal Fabre, Laurent Filleul, Jean-françois Jusot,  
4686 et al. “Impact of the 2003 Heatwave on All-Cause Mortality in 9 French Cities”. In:  
4687 *Epidemiology* 17.1 (2014), pp. 75–79. DOI: 10.1097/01.ede.0000187650.36636.1f.
- 4688 [82] Byung Tak Lee, Seung Chul Son, and Kyungran Kang. “A blind calibration scheme  
4689 exploiting mutual calibration relationships for a dense mobile sensor network”. In: *IEEE*  
4690 *Sensors Journal* 14.5 (2014), pp. 1518–1526. DOI: 10.1109/JSEN.2013.2297714.
- 4691 [83] Myeong-Hyeon Lee and Yoon-Hwa Choi. “Distributed diagnosis of wireless sensor net-  
4692 works”. In: *TENCON 2007 - 2007 IEEE Region 10 Conference*. 2007, pp. 2159–2165. DOI:  
4693 10.1109/TENCON.2007.4429026.
- 4694 [84] Alastair C. Lewis, Erika Von Schneidemesser, and Richard E. Peltier. *Low-cost sensors for*  
4695 *the measurement of atmospheric composition: overview of topic and future applications*.  
4696 Tech. rep. May. World Meteorological Organization (WMO), 2018.
- 4697 [85] Hongyong Li, Yujiao Zhu, Yong Zhao, Tianshu Chen, Ying Jiang, Ye Shan, Yuhong Liu,  
4698 Jiangshan Mu, Xiangkun Yin, Di Wu, et al. “Evaluation of the performance of low-cost  
4699 air quality sensors at a high mountain station with complex meteorological conditions”.  
4700 In: *Atmosphere* 11.2 (2020), pp. 1–16. DOI: 10.3390/atmos11020212.

- 4701 [86] Jason Jingshi Li, Boi Faltings, Olga Saukh, David Hasenfratz, and Jan Beutel. “Sensing  
4702 the Air We Breathe-The OpenSense Zurich Dataset.” In: *Proceedings of the Twenty-Sixth*  
4703 *AAAI Conference on Artificial Intelligence*. 2012, pp. 323–325.
- 4704 [87] Zhan Li, Yuzhi Wang, Anqi Yang, and Huazhong Yang. “Drift detection and calibration  
4705 of sensor networks”. In: *2015 International Conference on Wireless Communications and*  
4706 *Signal Processing, WCSP 2015*. 2015, pp. 1–6. DOI: 10.1109/WCSP.2015.7341138.
- 4707 [88] Chun Lin, Nicola Masey, Hao Wu, Mark Jackson, David J Carruthers, Stefan Reis, Ruth M  
4708 Doherty, Iain J Beverland, and Mathew R Heal. “Practical Field Calibration of Portable  
4709 Monitors for Mobile Measurements of Multiple Air Pollutants”. In: *Atmosphere* 8.12  
4710 (2017). DOI: 10.3390/atmos8120231.
- 4711 [89] John Lipor and Laura Balzano. “Robust blind calibration via total least squares”. In: *2014*  
4712 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.  
4713 2014, pp. 4244–4248. DOI: 10.1109/ICASSP.2014.6854402.
- 4714 [90] Xiao Liu, Sitian Cheng, Hong Liu, Sha Hu, Daqiang Zhang, and Huansheng Ning. “A  
4715 survey on gas sensing technology”. In: *Sensors* 12.7 (2012), pp. 9635–9665. DOI: 10.3390/  
4716 s120709635.
- 4717 [91] Byoung Gook Loh and Gi Heung Choi. “Calibration of Portable Particulate Matter–Monitoring  
4718 Device using Web Query and Machine Learning”. In: *Safety and Health at Work* 10.4  
4719 (2019), pp. 452–460. DOI: 10.1016/j.shaw.2019.08.002.
- 4720 [92] Xuanwen Luo and Ming Dong. “Distributed Faulty Sensor Detection in Sensor Networks”.  
4721 In: *Artificial Neural Networks – ICANN 2009*. 2009, pp. 964–975. DOI: 10.1007/978-3-  
4722 642-04277-5\_97.
- 4723 [93] Balz Maag, Olga Saukh, David Hasenfratz, and Lothar Thiele. “Pre-Deployment Testing,  
4724 Augmentation and Calibration of Cross-Sensitive Sensors”. In: *Proceedings of the 2016*  
4725 *International Conference on Embedded Wireless Systems and Networks*. 2016, pp. 169–180.
- 4726 [94] Balz Maag, Zimu Zhou, Olga Saukh, and Lothar Thiele. “SCAN: Multi-Hop Calibration  
4727 for Mobile Sensor Arrays”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*  
4728 1.2 (2017), 19:1–19:21. DOI: 10.1145/3090084.
- 4729 [95] Balz Maag, Zimu Zhou, and Lothar Thiele. “A Survey on Sensor Calibration in Air  
4730 Pollution Monitoring Deployments”. In: *IEEE Internet of Things Journal* 5.6 (2018),  
4731 pp. 4857–4870. DOI: 10.1109/JIOT.2018.2853660.
- 4732 [96] Sachit Mahajan and Prashant Kumar. “Evaluation of low-cost sensors for quantitative  
4733 personal exposure monitoring”. In: *Sustainable Cities and Society* 57. January (2020),  
4734 p. 102076. DOI: 10.1016/j.scs.2020.102076.
- 4735 [97] Arunanshu Mahapatro and Pabitra Mohan Khilar. “Online distributed fault diagnosis in  
4736 wireless sensor networks”. In: *Wireless Personal Communications* 71.3 (2013), pp. 1931–  
4737 1960. DOI: 10.1007/s11277-012-0916-8.
- 4738 [98] Arunanshu Mahapatro and Ajit Kumar Panda. “Fault Diagnosis in Wireless Sensor  
4739 Networks: A Survey”. In: *IEEE Communications Surveys Tutorials Early Acce* (2013),  
4740 pp. 1–27. DOI: 10.1109/SURV.2013.030713.00062.
- 4741 [99] Arunanshu Mahapatro and Ajit Kumar Panda. “Choice of detection parameters on  
4742 fault detection in wireless sensor networks: A multiobjective optimization approach”. In:  
4743 *Wireless Personal Communications* 78.1 (2014), pp. 649–669. DOI: 10.1007/s11277-014-  
4744 1776-1.

- 4745 [100] Carl Malings, Rebecca Tanzer, Aliaksei Hauryliuk, Srinivasa P. N. Kumar, Naomi Zim-  
4746 merman, Levent B. Kara, Albert A. Presto, and R. Subramanian. “Development of a  
4747 General Calibration Model and Long-Term Performance Evaluation of Low-Cost Sensors  
4748 for Air Pollutant Gas Monitoring”. In: *Atmospheric Measurement Techniques Discussions*  
4749 (2018), pp. 1–30. DOI: 10.5194/amt-2018-216.
- 4750 [101] Ali Marjovi, Adrian Arfire, and Alcherio Martinoli. “High Resolution Air Pollution Maps  
4751 in Urban Environments Using Mobile Sensor Networks”. In: *2015 International Conference  
4752 on Distributed Computing in Sensor Systems* (2015), pp. 11–20. DOI: 10.1109/DCOSS.  
4753 2015.32.
- 4754 [102] Jan-Frederic Markert, Matthias Budde, Gregor Schindler, Markus Klug, and Michael Beigl.  
4755 “Privacy-Preserving Collaborative Blind Macro-Calibration of Environmental Sensors  
4756 in Participatory Sensing”. In: *EAI Endorsed Transactions on Internet of Things* 18.10  
4757 (2018), pp. 1–8. DOI: 10.4108/eai.15-1-2018.153564.
- 4758 [103] Cory R. Martin, Ning Zeng, Anna Karion, Russell R. Dickerson, Xinrong Ren, Bari N.  
4759 Turpie, and Kristy J. Weber. “Evaluation and environmental correction of ambient CO2  
4760 measurements from a low-cost NDIR sensor”. In: *Atmospheric Measurement Techniques*  
4761 10.7 (2017), pp. 2383–2395. DOI: 10.5194/amt-10-2383-2017.
- 4762 [104] Measurement specialities. *HS1101LF – Relative Humidity Sensor Datasheet*. 2012. URL:  
4763 <https://docs-emea.rs-online.com/webdocs/142c/0900766b8142cdcd.pdf> (visited  
4764 on 09/26/2018).
- 4765 [105] Measurement specialities. *NO2-B43F – Nitrogen Dioxide Sensor Datasheet*. 2019. URL:  
4766 <http://www.alphasense.com/WEB1213/wp-content/uploads/2019/09/NO2-B43F.pdf>  
4767 (visited on 10/22/2019).
- 4768 [106] Amjad Mehmood, Nabil Alrajeh, Mithun Mukherjee, Salwani Abdullah, and Houbing Song.  
4769 “A survey on proactive, active and passive fault diagnosis protocols for WSNs: Network  
4770 operation perspective”. In: *Sensors (Switzerland)* 18.6 (2018). DOI: 10.3390/s18061787.
- 4771 [107] Emiliano Miluzzo, Nicholas D. Lane, Andrew T. Campbell, and Reza Olfati-Saber. “Cali-  
4772 Bree: A self-calibration system for mobile sensor networks”. In: *Distributed Computing in  
4773 Sensor Systems*. 2008, pp. 314–331. DOI: 10.1007/978-3-540-69170-9\_21.
- 4774 [108] Georgia Miskell, Jennifer A. Salmond, and David E. Williams. “Solution to the Problem of  
4775 Calibration of Low-Cost Air Quality Measurement Sensors in Networks”. In: *ACS Sensors*  
4776 3.4 (2018), pp. 832–843. DOI: 10.1021/acssensors.8b00074.
- 4777 [109] Sharon Moltchanov, Ilan Levy, Yael Etzion, Uri Lerner, David M. Broday, and Barak  
4778 Fishbain. “On the feasibility of measuring urban air pollution by wireless distributed  
4779 sensor networks”. In: *Science of the Total Environment* 502 (2015), pp. 537–547. DOI:  
4780 10.1016/j.scitotenv.2014.09.059.
- 4781 [110] Lidia Morawska, Phong K Thai, Xiaoting Liu, Akwasi Asumadu-sakyi, Godwin Ayoko,  
4782 Alena Bartonova, Andrea Bedini, Fahe Chai, Bryce Christensen, Matthew Dunbabin, et al.  
4783 “Applications of low-cost sensing technologies for air quality monitoring and exposure  
4784 assessment : How far have they gone ?” In: *Environment International* 116 (2018), pp. 286–  
4785 299. DOI: 10.1016/j.envint.2018.04.018.
- 4786 [111] Naser Hossein Motlagh, Eemil Lagerspetz, Petteri Nurmi, Xin Li, Samu Varjonen, Julien  
4787 Mineraud, Matti Siekkinen, Andrew Rebeiro-Hargrave, Tareq Hussein, Tuukka Petäjä, et  
4788 al. “Toward Massive Scale Air Quality Monitoring”. In: *IEEE Communications Magazine*  
4789 58.2 (2020), pp. 54–59. DOI: 10.1109/MCOM.001.1900515.

- 4790 [112] M. Mourad and J.-L. Bertrand-Krajewski. “A method for automatic validation of long  
4791 time series of data in urban hydrology”. In: *Water Science and Technology* 45.4-5 (2002),  
4792 pp. 263–270. DOI: 10.2166/wst.2002.0601.
- 4793 [113] Michael Mueller, Jonas Meyer, and Christoph Hueglin. “Design of an ozone and nitrogen  
4794 dioxide sensor unit and its long-term operation within a sensor network in the city of  
4795 Zurich”. In: *Atmospheric Measurement Techniques* 10.10 (2017), pp. 3783–3799. DOI:  
4796 10.5194/amt-10-3783-2017.
- 4797 [114] Thaha Muhammed and Riaz Ahmed Shaikh. “An analysis of fault detection strategies in  
4798 wireless sensor networks”. In: *Journal of Network and Computer Applications* 78 (2017),  
4799 pp. 267–287. DOI: 10.1016/j.jnca.2016.10.019.
- 4800 [115] Kevin Ni, Mani Srivastava, Nithya Ramanathan, Mohamed Nabil Hajj Chehade, Laura  
4801 Balzano, Sheela Nair, Sadaf Zahedi, Eddie Kohler, Greg Pottie, and Mark Hansen. “Sensor  
4802 network data fault types”. In: *ACM Transactions on Sensor Networks* 5.3 (2009), pp. 1–29.  
4803 DOI: 10.1145/1525856.1525863.
- 4804 [116] Thor Bjørn Ottosen and Prashant Kumar. “Outlier detection and gap filling methodologies  
4805 for low-cost air quality measurements”. In: *Environmental Science: Processes and Impacts*  
4806 21.4 (2019), pp. 701–713. DOI: 10.1039/c8em00593a.
- 4807 [117] F. J. Pierce and T. V. Elliott. “Regional and on-farm wireless sensor networks for  
4808 agricultural systems in Eastern Washington”. In: *Computers and Electronics in Agriculture*  
4809 61.1 (2008), pp. 32–43. DOI: 10.1016/j.compag.2007.05.007.
- 4810 [118] Olalekan A.M. Popoola, Gregor B. Stewart, Mohammed I. Mead, and Roderic L. Jones.  
4811 “Development of a baseline-temperature correction methodology for electrochemical sensors  
4812 and its implications for long-term stability”. In: *Atmospheric Environment* 147 (2016),  
4813 pp. 330–343. DOI: 10.1016/j.atmosenv.2016.10.024.
- 4814 [119] Matthieu Puigt, Gilles Delmaire, Gilles Roussel, Matthieu Puigt, Gilles Delmaire, and  
4815 Gilles Roussel. “Environmental signal processing : new trends and applications”. In:  
4816 *European Symposium on Artificial Neural Networks, Computational Intelligence and*  
4817 *Machine Learning (ESANN)*. 2017.
- 4818 [120] Xiaoliang Qin, Lujian Hou, Jian Gao, and Shuchun Si. “The evaluation and optimization  
4819 of calibration methods for low-cost particulate matter sensors: Inter-comparison between  
4820 fixed and mobile methods”. In: *Science of the Total Environment* 715 (2020), p. 136791.  
4821 DOI: 10.1016/j.scitotenv.2020.136791.
- 4822 [121] Aakash C. Rai, Prashant Kumar, Francesco Pilla, Andreas N. Skouloudis, Silvana Di,  
4823 Carlo Ratti, Ansar Yasar, and David Rickerby. “End-user perspective of low-cost sensors  
4824 for outdoor air pollution monitoring”. In: *Science of the Total Environment* 607-608  
4825 (2017), pp. 691–705. DOI: 10.1016/j.scitotenv.2017.06.266.
- 4826 [122] Naveen Ramakrishnan, Emre Ertin, and Randolph L. Moses. “Distributed signature  
4827 learning and calibration for large-scale sensor networks”. In: *2010 Conference Record of*  
4828 *the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. 2010, pp. 1545–  
4829 1549. DOI: 10.1109/ACSSC.2010.5757796.
- 4830 [123] Nithya Ramanathan, Laura Balzano, Marci Burt, Deborah Estrin, Thomas C. Harmon,  
4831 Charlie Harvey, Jenny Jay, Eddie Kohler, Sarah Rothenberg, and Mani Srivastava. *Rapid*  
4832 *Deployment with Confidence: Calibration and Fault Detection in Environmental Sensor*  
4833 *Networks*. Tech. rep. UCLA, 2006.

- 4834 [124] Nithya Ramanathan, Laura Balzano, Deborah Estrin, Mark Hansen, Thomas C. Harmon,  
4835 Jenny Jay, William Kaiser, and Gaurav Sukhatme. “Designing wireless sensor networks  
4836 as a shared resource for sustainable development”. In: *2006 International Conference on*  
4837 *Information and Communication Technology and Development, ICTD2006*. 2006, pp. 256–  
4838 265. DOI: 10.1109/ICTD.2006.301863.
- 4839 [125] Duarte Raposo, André Rodrigues, Jorge Sá Silva, and Fernando Boavida. “A Taxonomy  
4840 of Faults for Wireless Sensor Networks”. In: *Journal of Network and Systems Management*  
4841 25.3 (2017), pp. 591–611. DOI: 10.1007/s10922-017-9403-6.
- 4842 [126] Delphy Rodriguez, Myrto Valari, Sébastien Payan, and Laurence Eymard. “On the spatial  
4843 representativeness of NOX and PM10 monitoring-sites in Paris, France”. In: *Atmospheric*  
4844 *Environment: X* 1 (2019), p. 100010. DOI: 10.1016/j.aeaoa.2019.100010.
- 4845 [127] Philip W. Rundel, Eric A. Graham, Michael F. Allen, Jason C. Fisher, and Thomas C.  
4846 Harmon. “Environmental sensor networks in ecological research”. In: *New Phytologist*  
4847 182.3 (2009), pp. 589–607. DOI: 10.1111/j.1469-8137.2009.02811.x.
- 4848 [128] Tamal Saha and Sudipta Mahapatra. “Distributed Fault diagnosis in wireless sensor net-  
4849 works”. In: *2011 International Conference on Process Automation, Control and Computing*.  
4850 2011. DOI: 10.1109/PACC.2011.5978857.
- 4851 [129] Françoise Sailhan, Valérie Issarny, and Otto Tavares-Nascimento. “Opportunistic Mul-  
4852 tiparty Calibration for Robust Participatory Sensing”. In: *Proceedings - 14th IEEE*  
4853 *International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2017*. 2017,  
4854 pp. 435–443. DOI: 10.1109/MASS.2017.56.
- 4855 [130] Claude Sammut and Geoffrey I. Webb. *Encyclopedia of Machine Learning*. 1st. Springer  
4856 Publishing Company, Incorporated, 2011. DOI: 10.1007/978-0-387-30164-8.
- 4857 [131] Olga Saukh, David Hasenfratz, and Lothar Thiele. “Reducing multi-hop calibration errors  
4858 in large-scale mobile sensor networks”. In: *Proceedings of the 14th International Conference*  
4859 *on Information Processing in Sensor Networks*. 2015, pp. 274–285. DOI: 10.1145/2737095.  
4860 2737113.
- 4861 [132] Olga Saukh, David Hasenfratz, Christoph Walser, and Lothar Thiele. “On rendezvous in  
4862 mobile sensing networks”. In: *In Springer RealWSN*. 2013, pp. 29–42. DOI: 10.1007/978-  
4863 3-319-03071-5\_3.
- 4864 [133] Philipp Schneider, Alena Bartonova, Nuria Castell, Franck R. Dauge, Michel Gerboles,  
4865 Gayle S.W. Hagler, Christoph Hüglin, Roderic L. Jones, Sean Khan, Alastair C. Lewis,  
4866 et al. “Toward a unified terminology of processing levels for low-cost air-quality sensors”.  
4867 In: *Environmental Science and Technology* 53.15 (2019), pp. 8485–8487. DOI: 10.1021/  
4868 acs.est.9b03950.
- 4869 [134] Krishna P. Sharma and T. P. Sharma. “rDFD: reactive distributed fault detection in  
4870 wireless sensor networks”. In: *Wireless Networks* 23.4 (2017), pp. 1145–1160. DOI: 10.  
4871 1007/s11276-016-1207-1.
- 4872 [135] Emily G. Snyder, Timothy H. Watkins, Paul A. Solomon, Eben D. Thoma, Ronald W.  
4873 Williams, Gayle S.W. Hagler, David Shelow, David A. Hindin, Vasu J. Kilaru, and Peter W.  
4874 Preuss. “The changing paradigm of air pollution monitoring”. In: *Environmental Science*  
4875 *and Technology* 47.20 (2013), pp. 11369–11377. DOI: 10.1021/es4022602.
- 4876 [136] Jae Ho Sohn, Michael Atzeni, Les Zeller, and Giovanni Pioggia. “Characterisation of  
4877 humidity dependence of a metal oxide semiconductor sensor array using partial least  
4878 squares”. In: *Sensors and Actuators, B: Chemical* 131.1 (2008), pp. 230–235. DOI: 10.  
4879 1016/j.snb.2007.11.009.

- 4880 [137] Seung-chul Son and Byung-tak Lee. “A Blind Calibration Scheme using a Graph Model  
4881 for Optical Mobile Sensor Network”. In: *2014 12th International Conference on Optical  
4882 Internet 2014 (COIN)*. 2014, pp. 1–2. DOI: 10.1109/COIN.2014.6950632.
- 4883 [138] Laurent Spinelle, Manuel Aleixandre, and Michel Gerboles. *Protocol of evaluation and  
4884 calibration of low-cost gas sensors for the monitoring of air pollution*. Tech. rep. European  
4885 Commission, 2013.
- 4886 [139] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto  
4887 Bonavitacola. “Field calibration of a cluster of low-cost available sensors for air quality  
4888 monitoring. Part A: Ozone and nitrogen dioxide”. In: *Sensors and Actuators, B: Chemical*  
4889 215 (2015), pp. 249–257. DOI: 10.1016/j.snb.2015.03.031.
- 4890 [140] Laurent Spinelle, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto  
4891 Bonavitacola. “Field calibration of a cluster of low-cost commercially available sensors  
4892 for air quality monitoring. Part B: NO, CO and CO<sub>2</sub>”. In: *Sensors and Actuators, B:  
4893 Chemical* 238 (2017), pp. 706–715. DOI: 10.1016/j.snb.2016.07.036.
- 4894 [141] Laurent Spinelle, Clothilde Mantelle, Caroline Marchand, Benoît Herbin, Sabine Crunaire,  
4895 and Nathalie Redon. *Résultats du premier Essai national d’Aptitude des micro-Capteurs  
4896 pour la surveillance de la qualité de l’air*. 2020. URL: [https://www.lcsqa.org/fr/  
4897 rapport/resultats-du-premier-essai-national-daptitude-des-micro-capteurs-  
4898 eamc-pour-la-surveillance](https://www.lcsqa.org/fr/rapport/resultats-du-premier-essai-national-daptitude-des-micro-capteurs-eamc-pour-la-surveillance).
- 4899 [142] Kuo Feng Ssu, Chih Hsun Chou, Hewijin Christine Jiau, and Wei Te Hu. “Detection  
4900 and diagnosis of data inconsistency failures in wireless sensor networks”. In: *Computer  
4901 Networks* 50.9 (2006), pp. 1247–1260. DOI: 10.1016/j.comnet.2005.05.034.
- 4902 [143] Miloš S. Stanković, Srdjan S. Stanković, and Karl H. Johansson. “Asynchronous Distributed  
4903 Blind Calibration of Sensor Networks under Noisy Measurements”. In: *IEEE Transactions  
4904 on Control of Network Systems* 5.1 (2018), pp. 571–582. DOI: 10.1109/TCNS.2016.  
4905 2633788.
- 4906 [144] Miloš S. Stanković, Srdjan S. Stanković, and Karl Henrik Johansson. “Distributed Blind  
4907 Calibration in Lossy Sensor Networks via Output Synchronization”. In: *IEEE Transactions  
4908 on Automatic Control* 60.12 (2015), pp. 3257–3262. DOI: 10.1109/TAC.2015.2426272.
- 4909 [145] John M. Stockie. “The mathematics of atmospheric dispersion modeling”. In: *SIAM  
4910 Review* 53.2 (2011), pp. 349–372. DOI: 10.1137/10080991X.
- 4911 [146] Li-Juan Sun, Yang-Qing Su, Shu Shen, Ru-Chuan Wang, and Wen-Juan Li. “Cooperative  
4912 Calibration Scheme for Mobile Wireless Sensor Network”. In: *2019 15th International  
4913 Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*. IEEE, 2020, pp. 143–150.  
4914 DOI: 10.1109/msn48538.2019.00038.
- 4915 [147] Li Sun, Dane Westerdahl, and Zhi Ning. “Development and Evaluation of A Novel and  
4916 Cost-Effective Approach for Low-Cost NO<sub>2</sub> Sensor Drift Correction”. In: *Sensors* 17.8  
4917 (2017). DOI: 10.3390/s17081916.
- 4918 [148] Maen Takruri, Khalid Aboura, and Subhash Challa. “Distributed recursive algorithm for  
4919 auto calibration in drift aware wireless sensor networks”. In: *Innovations and Advanced  
4920 Techniques in Systems, Computing Sciences and Software Engineering* (2008), pp. 21–25.  
4921 DOI: 10.1007/978-1-4020-8735-6\_5.
- 4922 [149] Maen Takruri and Subhash Challa. “Drift aware wireless sensor networks”. In: *2007 10th  
4923 International Conference on Information Fusion*. 2007, pp. 1–7. DOI: 10.1109/ICIF.2007.  
4924 4408091.

- 4925 [150] Maen Takruri, Sutharshan Rajasegarar, Subhash Challa, Christopher Leckie, and Marimuthu  
4926 Palaniswami. “Online Drift Correction in Wireless Sensor Networks Using Spatio-Temporal  
4927 Modeling”. In: *2008 11th International Conference on Information Fusion*. 2008, pp. 1–8.
- 4928 [151] Rui Tan, Guoliang Xing, Zhaohui Yuan, Xue Liu, and Jianguo Yao. “System-level Cali-  
4929 bration for Data Fusion in Wireless Sensor Networks”. In: *ACM Transactions on Sensor  
4930 Networks* 9.3 (2013), 28:1–28:27. DOI: 10.1145/2480730.2480731.
- 4931 [152] Christopher Taylor, Ali Rahimi, Jonathan Bachrach, Howard Shrobe, and Anthony Grue.  
4932 “Simultaneous localization, calibration, and tracking in an ad hoc sensor network”. In:  
4933 *2006 5th International Conference on Information Processing in Sensor Networks*. 2006,  
4934 pp. 27–33. DOI: 10.1145/1127777.1127785.
- 4935 [153] Alaa Tharwat. “Classification assessment methods”. In: *Applied Computing and Informat-  
4936 ics* (2018). DOI: <https://doi.org/10.1016/j.aci.2018.08.003>.
- 4937 [154] Yudong Tian, Grey S. Nearing, Christa D. Peters-Lidard, Kenneth W. Harrison, and  
4938 Ling Tang. “Performance Metrics, Error Modeling, and Uncertainty Quantification”. In:  
4939 *Monthly Weather Review* 144.2 (2016), pp. 607–613. DOI: 10.1175/MWR-D-15-0087.1.
- 4940 [155] Dušan B. Topalović, Miloš D. Davidović, Maja Jovanović, A. Bartonova, Z. Ristovski,  
4941 and Milena Jovašević-Stojanović. “In search of an optimal in-field calibration method of  
4942 low-cost gas sensors for ambient air pollutants: Comparison of linear, multilinear and  
4943 artificial neural network approaches”. In: *Atmospheric Environment* 213.May (2019),  
4944 pp. 640–658. DOI: 10.1016/j.atmosenv.2019.06.028.
- 4945 [156] Wataru Tsujita, H. Ishida, and Toyosaka Moriizumi. “Dynamic gas sensor network for air  
4946 pollution monitoring and its auto-calibration”. In: *SENSORS, 2004 IEEE*. Vienna, 2004,  
4947 pp. 56–59. DOI: 10.1109/ICSENS.2004.1426098.
- 4948 [157] Wataru Tsujita, Akihito Yoshino, Hiroshi Ishida, and Toyosaka Moriizumi. “Gas sensor  
4949 network for air-pollution monitoring”. In: *Sensors and Actuators, B: Chemical* 110.2  
4950 (2005), pp. 304–311. DOI: 10.1016/j.snb.2005.02.008.
- 4951 [158] European Union. *Directive 2008/50/EC of the European Parliament and of the Council  
4952 of 21 May 2008 on ambient air quality and cleaner air for Europe*. 2008.
- 4953 [159] Marijn van der Velde, Gunter Wriedt, and Fayçal Bouraoui. “Estimating irrigation use and  
4954 effects on maize yield during the 2003 heatwave in France”. In: *Agriculture, Ecosystems  
4955 and Environment* 135.1-2 (2010), pp. 90–97. DOI: 10.1016/j.agee.2009.08.017.
- 4956 [160] Alexander Vergara, Shankar Vembu, Tuba Ayhan, Margaret A. Ryan, Margie L. Homer,  
4957 and Ramón Huerta. “Chemical gas sensor drift compensation using classifier ensembles”.  
4958 In: *Sensors and Actuators, B: Chemical* 166-167 (2012), pp. 320–329. DOI: 10.1016/j.  
4959 snb.2012.01.074.
- 4960 [161] Chao Wang, Parameswaran Ramanathan, and Kewal K. Saluja. “Calibrating nonlinear  
4961 mobile sensors”. In: *2008 5th Annual IEEE Communications Society Conference on Sensor,  
4962 Mesh and Ad Hoc Communications and Networks*. 2008, pp. 533–541. DOI: 10.1109/  
4963 SAHCN.2008.70.
- 4964 [162] Chao Wang, Parameswaran Ramanathan, and Kewal K. Saluja. “Moments based blind  
4965 calibration in mobile sensor networks”. In: *2008 IEEE International Conference on  
4966 Communications*. 2008, pp. 896–900. DOI: 10.1109/ICC.2008.176.
- 4967 [163] Chao Wang, Parameswaran Ramanathan, and Kewal K. Saluja. “Blindly Calibrating  
4968 Mobile Sensors Using Piecewise Linear Functions”. In: *2009 6th Annual IEEE Communi-  
4969 cations Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*.  
4970 2009, pp. 1–9. DOI: 10.1109/SAHCN.2009.5168912.

- 4971 [164] Na Wang, Jiacun Wang, and Xuemin Chen. “A trust-based formal model for fault  
4972 detection in wireless sensor networks”. In: *Sensors (Switzerland)* 19.8 (2019), pp. 1–20.  
4973 DOI: 10.3390/s19081916.
- 4974 [165] Rao Wang, Qingyong Li, Haomin Yu, Zechuan Chen, Yingjun Zhang, Ling Zhang, Houxin  
4975 Cui, and Ke Zhang. “A Category-Based Calibration Approach with Fault Tolerance  
4976 for Air Monitoring Sensors”. In: *IEEE Sensors Journal* 1748.c (2020), pp. 1–1. DOI:  
4977 10.1109/jsen.2020.2994645.
- 4978 [166] Yuzhi Wang, Anqi Yang, Xiaoming Chen, Pengjun Wang, Yu Wang, and Huazhong Yang.  
4979 “A Deep Learning Approach for Blind Drift Calibration of Sensor Networks”. In: *IEEE*  
4980 *Sensors Journal* 17.13 (2017), pp. 4158–4171. DOI: 10.1109/JSEN.2017.2703885.
- 4981 [167] Yuzhi Wang, Anqi Yang, Zhan Li, Xiaoming Chen, Pengjun Wang, and Huazhong Yang.  
4982 “Blind drift calibration of sensor networks using sparse Bayesian learning”. In: *IEEE*  
4983 *Sensors Journal* 16.16 (2016), pp. 6249–6260. DOI: 10.1109/JSEN.2016.2582539.
- 4984 [168] Yuzhi Wang, Anqi Yang, Zhan Li, Pengjun Wang, and Huazhong Yang. “Blind drift  
4985 calibration of sensor networks using signal space projection and Kalman filter”. In:  
4986 *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and*  
4987 *Information Processing (ISSNIP)*. 2015, pp. 1–6. DOI: 10.1109/ISSNIP.2015.7106904.
- 4988 [169] Peng Wei, Zhi Ning, Sheng Ye, Li Sun, Fenhuan Yang, Ka Chun Wong, Dane Westerdahl,  
4989 and Peter K.K. Louie. “Impact analysis of temperature and humidity conditions on  
4990 electrochemical sensor response in ambient air quality monitoring”. In: *Sensors* 18.2  
4991 (2018). DOI: 10.3390/s18020059.
- 4992 [170] Peng Wei, Li Sun, Anand Abhishek, Qing Zhang, Zong Huixin, Zhiqiang Deng, Ying Wang,  
4993 and Zhi Ning. “Development and evaluation of a robust temperature sensitive algorithm  
4994 for long term NO<sub>2</sub> gas sensor network data correction”. In: *Atmospheric Environment*  
4995 230.April (2020), p. 117509. DOI: 10.1016/j.atmosenv.2020.117509.
- 4996 [171] Lena Weissert, Elaine Miles, Georgia Miskell, Kyle Alberti, Brandon Feenstra, Geoff S.  
4997 Henshaw, Vasileios Papapostolou, Hamesh Patel, Andrea Polidori, Jennifer A. Salmond, et  
4998 al. “Hierarchical network design for nitrogen dioxide measurement in urban environments”.  
4999 In: *Atmospheric Environment* 228.2 (2020). DOI: 10.1016/j.atmosenv.2020.117428.
- 5000 [172] Kamin Whitehouse and David Culler. “Calibration as parameter estimation in sensor  
5001 networks”. In: *Proceedings of the 1st ACM International Workshop on Wireless Sensor*  
5002 *Networks and Applications*. ACM, 2002, pp. 59–67. DOI: 10.1145/570743.570747.
- 5003 [173] Jerome J Workman Jr. “A Review of Calibration Transfer Practices and Instrument  
5004 Differences in Spectroscopy”. In: *Applied Spectroscopy* 72.3 (2018), pp. 340–365. DOI:  
5005 10.1177/0003702817736064.
- 5006 [174] World Meteorological Organization. *Technical regulations. Volume II, Meteorological*  
5007 *service for international air navigation*. 2008.
- 5008 [175] Jiawen Wu and Guanghui Li. “Drift Calibration Using Constrained Extreme Learning  
5009 Machine and Kalman Filter in Clustered Wireless Sensor Networks”. In: *IEEE Access* 8  
5010 (2020), pp. 13078–13085. DOI: 10.1109/ACCESS.2019.2949878.
- 5011 [176] Chaocan Xiang and Panlong Yang. “An Iterative Method of Sensor Calibration in Par-  
5012 ticipatory Sensing Network”. In: *2013 IEEE 10th International Conference on Mobile*  
5013 *Ad-Hoc and Sensor Systems*. 2013, pp. 431–432. DOI: 10.1109/MASS.2013.19.

- 
- 5014 [177] Yun Xiang, Lan Bai, Ricardo Piedrahita, Robert P. Dick, Qin Lv, Michael Hannigan, and  
5015 Li Shang. “Collaborative calibration and sensor placement for mobile sensor networks”.  
5016 In: *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor*  
5017 *Networks (IPSN)*. 2012, pp. 73–83. DOI: 10.1145/2185677.2185687.
- 5018 [178] Xiang Yan Xiao, Wen Chih Peng, Chih Chieh Hung, and Wang Chien Lee. “Using  
5019 sensor ranks for in-network detection of faulty readings in wireless sensor networks”. In:  
5020 *International Workshop on Data Engineering for Wireless and Mobile Access*. 2007, pp. 1–  
5021 8. DOI: 10.1145/1254850.1254852.
- 5022 [179] Xianghua Xu, Wanyong Chen, Jian Wan, and Ritai Yu. “Distributed fault diagnosis of  
5023 wireless sensor networks”. In: *International Conference on Communication Technology*  
5024 *Proceedings, ICCT*. 2008, pp. 148–151. DOI: 10.1109/ICCT.2008.4716155.
- 5025 [180] Ke Yan and David Zhang. “Calibration transfer and drift compensation of e-noses via  
5026 coupled task learning”. In: *Sensors and Actuators: B. Chemical* 225 (2016), pp. 288–297.  
5027 DOI: 10.1016/j.snb.2015.11.058.
- 5028 [181] Jielong Yang, Wee Peng Tay, and Xionghu Zhong. “A dynamic Bayesian nonparametric  
5029 model for blind calibration of sensor networks”. In: *2017 IEEE International Conference*  
5030 *on Acoustics, Speech and Signal Processing (ICASSP)*. 2017, pp. 4207–4211. DOI: 10.  
5031 1109/ICASSP.2017.7952949.
- 5032 [182] Haibo Ye, Tao Gu, Xianping Tao, and Jian Lu. “SBC : Scalable Smartphone Barometer  
5033 Calibration through Crowdsourcing”. In: *11th International Conference on Mobile and*  
5034 *Ubiquitous Systems: Computing, Networking and Services*. 2014. DOI: 10.1002/wcm.2706.
- 5035 [183] Wei Ying Yi, Kin Ming Lo, Terrence Mak, Kwong Sak Leung, Yee Leung, and Mei Ling  
5036 Meng. “A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems”.  
5037 In: *Sensors* 15.12 (2015), pp. 31392–31427. DOI: 10.3390/s151229859.
- 5038 [184] Haomin Yu, Qingyong Li, Rao Wang, Zechuan Chen, Yingjun Zhang, Yangli-ao Geng,  
5039 Ling Zhang, Houxin Cui, and Ke Zhang. “A Deep Calibration Method for Low-Cost  
5040 Air Monitoring Sensors with Multi-Level Sequence Modeling”. In: *IEEE Transactions*  
5041 *on Instrumentation and Measurement* 9456.c (2020), pp. 1–1. DOI: 10.1109/tim.2020.  
5042 2978596.
- 5043 [185] Zeyu Zhang, Amjad Mehmood, Lei Shu, Zhiqiang Huo, Yu Zhang, and Mithun Mukherjee.  
5044 “A survey on fault diagnosis in wireless sensor networks”. In: *IEEE Access* 6 (2018),  
5045 pp. 11349–11364. DOI: 10.1109/ACCESS.2018.2794519.
- 5046 [186] Naomi Zimmerman, Albert A. Presto, Srinivasa P.N. Kumar, Jason Gu, Aliaksei Hauryliuk,  
5047 Ellis S. Robinson, Allen L. Robinson, and R. Subramanian. “A machine learning calibration  
5048 model using random forests to improve sensor performance for lower-cost air quality  
5049 monitoring”. In: *Atmospheric Measurement Techniques* 11.1 (2018), pp. 291–313. DOI:  
5050 10.5194/amt-11-291-2018.
-

**Titre :** Étalonnage in situ de l'instrumentation bas coût pour la mesure de grandeurs ambiantes : méthode d'évaluation des algorithmes et diagnostic des dérives

**Mots clés :** Réseaux de capteurs, Qualité de mesure, Étalonnage, Evaluation, Diagnostic, Mesures ambiantes

**Résumé :** Dans de nombreux domaines allant de l'agriculture à la santé publique, des grandeurs ambiantes doivent être suivies dans des espaces intérieurs ou extérieurs. On peut s'intéresser par exemple à la température, aux polluants dans l'air ou dans l'eau, au bruit, etc. Afin de mieux comprendre ces divers phénomènes, il est notamment nécessaire d'augmenter la densité spatiale d'instruments de mesure. Cela pourrait aider par exemple à l'analyse de l'exposition réelle des populations aux nuisances comme les polluants atmosphériques.

Le déploiement massif de capteurs dans l'environnement est rendu possible par la baisse des coûts des systèmes de mesure, qui utilisent notamment des éléments sensibles à base de micro ou nano technologies. L'inconvénient de ce type de dispositifs est une qualité de mesure insuffisante. Il en résulte un manque de confiance dans les données produites et/ou une hausse drastique des coûts de l'instrumentation causée par les opérations nécessaires d'étalonnage des instruments ou de remplacement périodique des capteurs.

Il existe dans la littérature de nombreux algorithmes qui offrent la possibilité de réaliser l'étalonnage des instruments en les laissant déployés sur le terrain, que l'on nomme techniques d'étalonnage *in situ*.

L'objectif de cette thèse est de contribuer à l'effort de recherche visant à améliorer la qualité des données des instruments de mesure bas coût à travers leur étalonnage *in situ*.

En particulier, on vise à 1) faciliter l'identification des techniques existantes d'étalonnage *in situ* applicables à un réseau de capteurs selon ses propriétés et les caractéristiques des instruments qui le composent ; 2) aider au choix de l'algorithme le plus adapté selon le réseau de capteurs et son contexte de déploiement ; 3) améliorer l'efficacité des stratégies d'étalonnage *in situ* grâce au diagnostic des instruments qui ont dérivé dans un réseau de capteurs.

Trois contributions principales sont faites dans ces

travaux. Tout d'abord, une terminologie globale est proposée pour classer les travaux existants sur l'étalonnage *in situ*. L'état de l'art effectué selon cette taxonomie a montré qu'il y a de nombreuses contributions sur le sujet, couvrant un large spectre de cas. Néanmoins, le classement des travaux existants selon leurs performances a été difficile puisqu'il n'y a pas d'étude de cas de référence pour l'évaluation de ces algorithmes.

C'est pourquoi dans un second temps, un cadre pour la simulation de réseaux de capteurs est introduit. Il vise à guider l'évaluation d'algorithmes d'étalonnage *in situ*. Une étude de cas détaillée est fournie à travers l'évaluation d'algorithmes pour l'étalonnage *in situ* de réseaux de capteurs statiques et aveugles. Une analyse de l'influence des paramètres et des métriques utilisées pour extraire les résultats est également menée. Les résultats dépendant de l'étude de cas, et la plupart des algorithmes réétalonnant les instruments sans évaluer au préalable si cela est nécessaire, un outil d'identification permettant de déterminer les instruments qui sont effectivement fautifs en termes de dérive serait précieux.

Dès lors, la troisième contribution de cette thèse est un algorithme de diagnostic ciblant les fautes de dérive dans les réseaux de capteurs sans faire d'hypothèse sur la nature du réseau de capteurs considéré. Basé sur le concept de rendez-vous, l'algorithme permet d'identifier les instruments fautifs tant qu'il est possible de supposer qu'un instrument n'est pas fautif dans le réseau de capteurs. À travers l'analyse des résultats d'une étude de cas, nous proposons différents moyens pour diminuer les faux résultats et des recommandations pour régler les paramètres de l'algorithme. Enfin, nous montrons que l'algorithme de diagnostic proposé, combiné à une technique simple d'étalonnage, permet d'améliorer la qualité des résultats de mesure. Ainsi, cet algorithme de diagnostic ouvre de nouvelles perspectives quant à l'étalonnage *in situ*.

**Title :** In situ calibration of low-cost instrumentation for the measurement of ambient quantities: evaluation methodology of the algorithms and diagnosis of drifts

**Keywords :** Sensor networks, Measurement quality, Calibration, Evaluation, Diagnosis, Environmental sensing

**Abstract :** In various fields going from agriculture to public health, ambient quantities have to be monitored in indoors or outdoors areas. For example, temperature, air pollutants, water pollutants, noise and so on have to be tracked. To better understand these various phenomena, an increase of the density of measuring instruments is currently necessary. For instance, this would help to analyse the effective exposure of people to nuisances such as air pollutants.

The massive deployment of sensors in the environment is made possible by the decreasing costs of measuring systems, mainly using sensitive elements based on micro or nano technologies. The drawback of this type of instrumentation is a low quality of measurement, consequently lowering the confidence in produced data and/or a drastic increase of the instrumentation costs due to necessary recalibration procedures or periodical replacement of sensors.

There are multiple algorithms in the literature offering the possibility to perform the calibration of measuring instruments while leaving them deployed in the field, called *in situ* calibration techniques.

The objective of this thesis is to contribute to the research effort on the improvement of data quality for low-cost measuring instruments through their *in situ* calibration.

In particular, we aim at 1) facilitating the identification of existing *in situ* calibration strategies applicable to a sensor network depending on its properties and the characteristics of its instruments; 2) helping to choose the most suitable algorithm depending on the sensor network and its context of deployment; 3) improving the efficiency of *in situ* calibration strategies through the diagnosis of instruments that have drifted in a sensor network.

Three main contributions are made in this work. First,

a unified terminology is proposed to classify the existing works on *in situ* calibration. The review carried out based on this taxonomy showed there are numerous contributions on the subject, covering a wide variety of cases. Nevertheless, the classification of the existing works in terms of performances was difficult as there is no reference case study for the evaluation of these algorithms.

Therefore in a second step, a framework for the simulation of sensors networks is introduced. It is aimed at evaluating *in situ* calibration algorithms. A detailed case study is provided across the evaluation of *in situ* calibration algorithms for blind static sensor networks. An analysis of the influence of the parameters and of the metrics used to derive the results is also carried out. As the results are case specific, and as most of the algorithms recalibrate instruments without evaluating first if they actually need it, an identification tool enabling to determine the instruments that are actually faulty in terms of drift would be valuable.

Consequently, the third contribution of this thesis is a diagnosis algorithm targeting drift faults in sensor networks without making any assumption on the kind of sensor network at stake. Based on the concept of rendez-vous, the algorithm allows to identify faulty instruments as long as one instrument at least can be assumed as non-faulty in the sensor network. Across the investigation of the results of a case study, we propose several means to reduce false results and guidelines to adjust the parameters of the algorithm. Finally, we show that the proposed diagnosis approach, combined with a simple calibration technique, enables to improve the quality of the measurement results. Thus, the diagnosis algorithm opens new perspectives on *in situ* calibration.