Rendez-vous Based Drift Diagnosis Algorithm For Sensor Networks Towards In Situ Calibration

Florentin Delaine, Bérengère Lebental, Hervé Rivano

Rendez-vous Based Drift Diagnosis Algorithm For Sensor Networks Towards In Situ Calibration

Florentin Delaine, Bérengère Lebental, and Hervé Rivano*†‡

June 9, 2022

Abstract

In recent years, low-cost sensors have raised strong interest for environmental monitoring applications. These instruments often suffer from degraded data quality. Notably, they are prone to drift. It can be mitigated with costly periodic calibrations. To reduce this cost, *in situ* calibration strategies have emerged, enabling the recalibration of instruments while leaving them in the field. However, they rarely identify which instruments actually need a calibration because of drift, so that *in situ* calibration may instead degrade performances. Therefore, a novel drift detection algorithm is presented in this work, exploiting the concept of rendez-vous between measuring instruments. Its originality lies mainly in the comparisons of values determining the state of the instruments, for which the quality of the measurement results is taken into account. It defines the concept of compatibility between measurement results. A case study is developed, showing an accuracy of 88% for correct detection of drifting instruments. The results of the diagnosis algorithm are then combined with calibration approaches. Results show a significant improvement of the measurement results. Notably, an increase of 15% of the coefficient of determination of the linear regression between their true values and the measured values is observed with the correction and the error on the slope and on the intercept respectively is reduced by 50% and 60% at least.

Abstract

[Note to practitioners] In this paper, we investigate the problem of drift detection in sensor networks. This work was motivated by the fact that faulty nodes are rarely detected in existing *in situ* calibration algorithm prior to the correction of the instruments. Moreover, existing

*Florentin Delaine and Bérengère Lebental are with Efficacity, F-77420 Champs-sur-Marne, France, COSYS-LISIS, Univ Gustave Eiffel, IFSTTAR, Marne-la-Vallée, F-77454, France and LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, F-91128, France.

†Hervé Rivano is with Université de Lyon, INSA Lyon, Inria, CITI, Villeurbanne, F-69621, France.

‡Corresponding author: Florentin Delaine (florentin.delaine@gmail.com)

fault diagnosis algorithms for sensor networks do not specifically target drift and are often applicable to either (dense) static or mobile sensor networks but not both. We propose an algorithm designed for the detection of drift faults regardless of the type of sensor network and of the measurand. Specific attention is paid to the metrological quality of the measurement results used to carry out the diagnosis. The output of the algorithm provides information that can be exploited for the recalibration of faulty instruments. In future work, we will aim at providing tools and recommendations for the adjusment of the parameters of the diagnosis algorithm but also more elaborated approaches based on the results of our diagnosis algorithm to calibrate faulty nodes.

1 Introduction

Low-cost sensor networks have raised a strong interest in the past years, notably for air pollution monitoring [1, 2, 3]. The deployment of large sensor networks at reasonable cost to provide information at high spatial resolution can now be envisioned, although several issues have yet to be tackled [4, 5]. Among the multiple challenges for these technologies, the improvement of their data quality is a major one [4, 6]. Reports on their performances [7, 8] show that drift is an important issue, although it is a well-known one. Measuring instruments and notably their sensors are subject to changes in their characteristics (sensitivity, linearity, repeatability, hysteresis...) over time due to ageing or to the conditions under which they operate for instance. Periodic calibration usually enables to mitigate its effect [9]. The traditional approach to calibration consists in exposing the devices to calibrate to standard values or to co-locate them next to a reference instrument. It is usually carried out in a controlled facility. Even when conducted in the field [10], this operation is often expensive compared to the cost of the instruments themselves, especially in the case of low-cost sensors. Various *in situ* calibration strategies have been developed in response to this issue [11, 12, 13]. They enable calibration to be carried out while leaving the sensors deployed in the field and without any physical intervention.

In previous work [13], we observed that most strategies perform the calibration of all instruments without evaluating first whether they really need it. It actually results in performance degradation for some sensors. Some calibration algorithms are providing this identification, notably in [14, 15]. 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 this paper, we investigate drift detection for measuring instruments in sensor networks with the help of measurement results from other devices. Our contribution is a drift diagnosis algorithm exploiting the concept of rendez-vous [16, 17]. They are used to determining if an instrument needs to be recalibrated. Their validity is studied beforehand, e.g. the measurements of one of the instru-

ments involved in a rendez-vous must be trustworthy for instance, so that only relevant rendez-vous are used to determine the state of an instrument. We apply this algorithm on a case study and show its efficiency for drift detection. In addition, the rendez-vous used to determine that an instrument is faulty can be exploited for its recalibration. Thus, two calibration approaches are combined with the diagnosis algorithm afterwards. The coefficient of determination of the linear regression between the measured values and their true values is notably increased by 15% on average with the correction and the error on the slope and on the intercept are lowered of 50% and 60% at least respectively. This gives promising results for the *in situ* calibration of sensor networks.

Roadmap. First, the related works are reviewed in Section 2. Then, concepts and notations used for the design of the diagnosis algorithm are introduced in Section 3, prior to its presentation in Section 4. A case study applying the diagnosis algorithm is developed in Section 5 before its combination with calibration approaches in Section 6. Finally, Section 7 gives a conclusion.

2 Related work

2.1 Overview on fault diagnosis algorithms for sensor networks

In the literature, the question of drift diagnosis in sensor network is generally encapsulated in the main theme of fault diagnosis in sensor networks. Various surveys [18, 19, 20, 21] reported multiple contributions on this subject, providing several insights.

First of all, the existing diagnosis algorithms are often targeting multiple faults. In [19], the 15 algorithms cited that could be able to address the drift fault (it was divided there between calibration, gain, and offset faults), detect stuck-at or out-of-range faults at the same time as drift without differentiating between faults.

Secondly, the diagnosis approaches reported target mainly static sensor networks. Mahapatro *et al.* [18] listed eight algorithms that can be applied both to static and mobile sensor networks. Zhang *et al.* [20] reported two references explicitly dealing with mobile sensor networks [22, 23]. They also stated that approaches designed for static networks behave poorly when applied to mobile ones. This indicates that algorithms exploiting specificities related to mobile networks have not been deeply investigated so far.

Finally, the methods used to carry out the diagnosis of any fault in sensor networks are diverse in terms of concepts and tools on which they are based on. Methods rely on comparisons between instruments, statistics and probabilities, as well as machine learning models (regressors, classifiers...). While dense sensor networks are envisioned nowadays, an approach performing direct comparisons of values measured by different instruments to track drift faults is straightforward. The concept of rendez-vous we want to exploit is a general frame to define the validity of such comparisons. Thus, we investigate in the following section

the algorithms compatible with this idea.

2.2 Diagnosis algorithms detecting faults through comparison of values

First, Chen *et al.* [24] introduced a faulty sensor detection algorithm consisting of four steps of evaluation. First, each instrument compares its measured value at the instant of diagnosis with the measured values of its neighbours. In a second step, each instrument determines if it is likely correct or faulty depending on the number of positive comparisons and its number of neighbours, and shares it with its neighbours. Then, based on its likely correct neighbours, an instrument determines if it is actually correct or not. The final step is dedicated to the management of the remaining undetermined instruments. Xu *et al.* [25] proposed an extension of this work dealing with the particular case of tree-like networks to reduce the number of communications. To avoid intermittent faults, they also proposed to compute the initial test result based on multiple comparisons of values between two instruments, instead of a unique comparison. Saha *et al.* [26] developed a similar algorithm but with comparisons for multiple quantities (measurand and remaining energy). A major drawback of these works is the need for a known sensor network structure, limiting their applicability to static sensor networks. Also, the decision on the state of the instrument is build on only one measured value per instrument, which is likely to generate false results considering drift fault. Finally, the question of the validity of the comparisons is not addressed, e.g. what has to be done when the value of a neighbour (or its own one) is out of the measurement range and may not be reliable. Note that this is also the case in the other works reported afterwards.

Ssu *et al.* [27] proposed an approach to diagnosis faults based on communications from source nodes to sink ones. The main idea is to send a request through two paths and to compare the results obtained at the sink node. If the results are different, then at least one faulty path exists. The algorithm tries to identify this path with the help of a third path and a majority voting procedure, but the faulty paths cannot always be identified. Such a method is particularly relevant to diagnose communication or hardware-related faults, but does not appear to be appropriate to track drift.

Lee *et al.* [28] presented a distributed algorithm to isolate faulty nodes. Initially, the nodes are all assumed as faulty. A comparison is made between the measured values of neighbouring nodes and if the result is higher than a threshold, then the test is positive. If less than a predefined number of positive tests has been obtained, or if the test with a non-faulty node is negative, then the diagnosed instrument is non-faulty. The algorithm works either based on single comparisons between neighbours or with comparisons repeated multiple times, which is more relevant to track drift. However, the structure of the sensor network has also to be known in this case.

Mahapatro *et al.* [29, 30] introduced a clustering-based diagnosis algorithm. The clustering part is used for the definition of the neighbours around cluster heads, the cluster heads being the instruments with the highest residual energy

levels. They also compare the measured values between instruments of a cluster and a majority voting strategy is used to determine the state of the nodes. The clustering part is overcoming the limitation of the work of Chen *et al.* [24] but in this algorithm, the decision on the state of the instrument is based again on one measured value only per instrument.

Chanak *et al.* [22] developed a comparison-based scheme using a reference mobile sink node moving between the static nodes of the network. When it is close to a node, several diagnoses are performed to detect hardware and software faults. This also enables a lower consumption of energy for the transmission of the measurement results, as it is no longer necessary to communicate with a distant gateway. The core of this contribution lies in the determination of an optimal path to meet with each node. While this approach of adding a mobile node is interesting for static sensor networks where the nodes cannot be considered as neighbours, it demands to have either the mobile node co-located long enough next to an instrument to detect drift or to add another layer of diagnosis combining the observations of multiple co-locations between the instruments. In general terms, this approach has to be generalized to more complex sensor networks with multiple mobile nodes and a specific protocol to track drift.

Luo *et al.* [31] proposed an approach using the concept of average consensus. Each instrument estimates first its state regarding the average consensus value built from the values of its neighbours. Each instrument also estimates the states of its neighbours. Then, the decisions of all the instruments are merged to make the final decision. While the concept at the core of this method if different from the previous ones, it remains in similar to the work of Chen *et al.* [24] and the derived ones. Thus, it has the same drawbacks.

In reaction to contributions computing the average value of the neighbours of an instrument by weighting their values with the inverse of the distance between them, Xiao *et al.* [32] argued that the distance does not control alone the relationship between the values of two instruments, notably if a closer one is actually faulty. Thus, they propose to take into account an indicator of trustworthiness computed for each node of the network. This confidence value is used in the voting procedures that are then used to determine the states of instruments. Ji *et al.* [33] also developed their algorithm around a weighted average of the values measured by an instrument and its neighbours. The weights represent a confidence level associated to each instrument. The difference between the measured value of an instrument and the average is compared to a threshold. If it is greater than the threshold, the confidence level of the instrument is decreased and once it reaches zero, the instrument is reported as faulty. These two works are among the firsts to add an idea of validity of the measured values used in diagnosis algorithms, but they have again similar drawbacks compared to the methods reported previously: the structure of the network has to be known and one value only per instrument is considered.

More generally, the idea of trust between instruments has been extensively studied, for instance in [34]. The trust model proposed is based on direct comparisons between instruments but also on recommendations from third parties,

with concerns on the security of the communications between the instruments. This trust framework is also adapted for mobile sensor networks. Wang *et al.* [35] later exploited Petri nets to introduce a trust-based formal model aimed at detecting faults in sensor networks. Although the subject is major in the context of the Internet of Things, such complex trust frameworks are quite high-level and should encapsulate algorithms like drift diagnosis.

Sharma *et al.* [36] also exploited the idea of confidence in a method similar to the one of Chen *et al.* [24]. The main difference with Chen *et al.* is that in a first step, each instrument analyses its own behaviour and that a confidence level is associated to the states determined for the instruments.

Feiyue *et al.* [37] proposed to combine the information resulting of a self-evaluation and of comparisons with the neighbours of each instrument. In a first step, the reliable nodes are determined through a majority voting procedure. Then, various metrics are computed and combined to make the decisions for the instruments that were not considered reliable in the first step.

Fu *et al.* [38] defined a lightweight fault detection strategy based on trend correlation and median values analysed against neighbouring nodes. The approach is distributed in terms of calculation to reduce the detection latency. However, all the nodes are assumed to be stationary.

2.3 Outcomes of the review

In general, fault diagnosis algorithms address multiple types of faults, concerning various aspects of measuring systems, but drift is not specifically targeted in existing works. Algorithms exploiting the same principle we want to rely our algorithm on, e.g. comparisons of measurement results between instruments, have been already presented in the literature. In addition, static sensor networks have been widely studied while mobile sensor networks have been much less covered. More generally, no algorithm can target both static and mobile sensor networks efficiently. A diagnosis algorithm not requiring any assumptions on the type of sensor network and its structure would be interesting.

Overall, the validity of the comparisons is rarely discussed from a metrological perspective in the literature. Moreover, the diagnosis is performed based on single values from a given set of instruments measured at the same time in most cases. Taking a decision on a short time range may not be correct if drift is targeted. Build a decision on the states of the instruments with long sequences of measured values over time, should be more appropriate to track drift faults. Finally, it would be valuable to take into account the quality of the measurement results and instead of exploiting the measured values only.

The algorithm we propose afterwards is built in response to these observations.

3 Concepts for a drift diagnosis algorithm based on rendez-vous

3.1 General definitions

A **sensor network** represented by a set of instruments S is considered. s_i or s_j denotes nodes of the network, e.g. **measuring instruments**. $m(s_i, t)$ is a **measurement result** obtained by s_i at t. $M(s_i, (t, \Delta t)$ is the set of measurement results for s_i, over the time range $[t - \Delta t; t]$. In this work, a measurement result $m(s_i, t)$ is composed of a **measured value** $v(s_i, t)$, and of a **measurement uncertainty** $\Delta v(s_i, t)$. $v_{true}(s_i, t)$ is the true value that should be measured by s_i at t if it were an ideal instrument.

In a sensor network, the instruments can be of different qualities, resulting in different levels of accuracy. Thus, $c(s_i)$ is the **accuracy class** (or class) of s_i and S^k is the set of measuring instruments where $c(s_i) = k$.

In addition, we introduce the notion of **validity for measurement results**. Indeed, 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*" [9]. 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.

$M^*(s_i, (t, \Delta t))$ refers to the set of measurement results for s_i, over the time range $[t - \Delta t; t]$ that are metrologically valid.

3.2 Compatibility of measurement results

Consider $s_i \in S$. A measurement result $m(s_i, t)$ is **compatible with true value** if $v_{true}(s_i, t) \in [v(s_i, t) \pm \Delta v(s_i, t)]$. $m(s_i, t)$ is **non-compatible with true value** if $v_{true}(s_i, t) \notin [v(s_i, t) \pm \Delta v(s_i, t)]$. The set of measurement results compatible with true values of s_i obtained during $[t - \Delta t; t]$ is noted $M^{\approx}(s_i, (t, \Delta t))$. These definitions can be extended for the comparison of measurement results between different instruments. Consider s_i and $s_j \in S$. $m(s_i, t)$ is **compatible with** $m(s_j, t')$ if $[v(s_i, t) \pm \Delta v(s_i, t)] \cap [v(s_j, t') \pm \Delta v(s_j, t')] \neq \varnothing$. It is noted $m(s_i, t) \approx m(s_j, t')$. $m(s_i, t)$ is **non-compatible with** $m(s_j, t')$ if $[v(s_i, t) \pm \Delta v(s_i, t)] \cap [v(s_j, t') \pm \Delta v(s_j, t')] = \varnothing$. The set of measurement results of s_i obtained during $[t - \Delta t; t]$ compatible with s_j is noted $M^{\approx}(s_i \rightarrow s_j, (t, \Delta t))$.

3.3 Rendez-vous

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

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

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

The concept of compatibility between measurement results can be extended to the concept of rendez-vous. Consider a rendez-vous $\varphi(s_i \to s_j, t)$. By joining together the definitions of compatible measurement results, a rendez-vous of s_i with s_j at t is stated as **compatible** if $m(s_i, t) \approx m(s_j, t')$. Otherwise it is stated as **non-compatible**. Consequently, $\Phi^{\approx}(s_i \to s_j, (t, \Delta t))$ is the set of compatible rendez-vous.

In practice, this general definition for rendez-vous can be applied in different ways. For instance, the condition of spatial vicinity can be a maximal distance between the instruments. Alternately, using the concept of representativity area [39], e.g. the area around the location of a measurement where different measurements provide identical (or similar) results, a rendez-vous may occur when there is a non-null intersection of the representativity areas for two measurements of two instruments. This latter definition may yield better results in practice than the one based on a distance as the area of representativity usually include the geometry of the area in which the instruments are, like buildings and crossings in an urban environment, but it requires an accurate derivation of the area of representativity, which may not be straightforward. More generally, the choice of how to define rendez-vous should depend on the measurand, the sensor network and its context of deployment. This is in fact a general requirement when designing sensor networks [40]. Thus, the diagnosis algorithm proposed in the following section considers only a conceptual definition of rendez-vous.

4 Algorithm for the diagnosis of drifts in a sensor network

4.1 General idea

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

To determine if instruments are correctly calibrated, the proposed approach

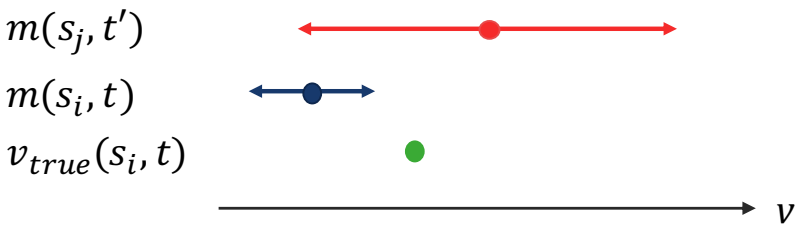


Figure 1: Example where the measurement results $m(s_i, t)$ and $m(s_j, t')$ are compatible with each other but where $m(s_i, t)$ is not compatible with its true value

consists in using the concept of rendez-vous introduced in Section 3.3. Consider two instruments s_i and s_j in rendez-vous. Suppose that one of the instruments, s_j for instance, is known as non-faulty and consider $m(s_i, t)$ and $m(s_j, t')$, the measurement results involved in a rendez-vous $\varphi(s_i \rightarrow s_j, t)$. If both results are metrologically valid, it means that $m(s_j, t')$ can be seen as a reference value for s_i: it is a known value with an associated uncertainty. Thus, if $m(s_i, t) \approx m(s_j, t')$, it means that s_i is correctly calibrated according to s_j. It invites to set its predicted state to non-faulty. Otherwise, it invites to predict it as faulty. As s_j helps to predict the state of s_i, it is called a **diagnoser** of s_i.

Predicting the state of an instrument based on only one value often leads to false predictions because drift is a fault having usually a longer characteristic time than others, such as 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 state 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 state of s_i can reduce the impact of these particular cases. Moreover, if the class of s_i is such as $c(s_i) \gg c(s_j)$, predicting the state of s_i based on s_j, even if $\widehat{\Omega}(s_j, t') = NF$, 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 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 state. This minimal class is noted $c^{\mathcal{D}}_{min}(s_i)$ and is defined by:

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

where $\Delta c^{\mathcal{D}}_{min}(k)$ is the relative difference of class required between instruments of class k and the instruments allowed to be their diagnosers.

Thus, to determine the state 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:

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

It is also possible that the input information of a diagnosis algorithm is not sufficient to choose between F and NF states. Thus, a third option for the predicted state of a system is **ambiguous** (A).[1] In the present case, it is when $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ does not contain enough rendez-vous to allow a prediction with enough confidence. This minimal size for a set of valid rendez-vous for any instrument is noted $|\Phi_v|_{min}$.

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

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 Δt 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 used to predict the states of the instruments of S. This forms a scene as defined by [17].

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

Then, the predicted states $\widehat{\Omega}(s_i, t_d)$ for each $s_i \in S$ are initialised. The predicted states 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 states of all the instruments, noted $\widehat{\Omega}(S, t_d)$, are actualised. These actualised states are noted $\widetilde{\Omega}(S, t_d)$.

They are determined by iteration until the predicted state of each instrument $s_i \in S^{\mathcal{D}}$ remains unchanged. 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 state, whatever the predicted states of the other instruments.

[1] The true state of an instrument cannot be ambiguous. It is either faulty or non-faulty.

Therefore, it is not possible to actualise its predicted state with another value than ambiguous, which is already the value of $\widetilde{\Omega}(s_i, t_d)$, and s_i is removed from $S^{\mathcal{D}}$.

Otherwise, the set of valid rendez-vous $\Phi_v(s_i \rightarrow S, (t_d, \Delta t))$ is determined. If its size is lower than $|\Phi_v|_{min}$, the actualised predicted state 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 state can be determined between non-faulty and faulty.

To do so, we compute $r^{\approx}_{\Phi_v}(s_i, (t_d, \Delta t))$, 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

$$r^{\approx}_{\Phi_v}(s_i, (t_d, \Delta t)) = \frac{|\Phi^{\approx}_v(s_i, (t_d, \Delta t))|}{|\Phi_v(s_i, (t_d, \Delta t))|}$$

Based on $(r^{\approx}_{\Phi_v})_{min}$, which is the minimal tolerated value associated to the rate $r^{\approx}_{\Phi_v}(s_i, (t_d, \Delta t))$, if the condition $r^{\approx}_{\Phi_v}(s_i \rightarrow S, (t_d, \Delta t))) < (r^{\approx}_{\Phi_v})_{min}$ is true, then $\widetilde{\Omega}(S, t_d)$ is equal to faulty. Otherwise $\widetilde{\Omega}(s_i, t_d)$ is set to non-faulty. In the end, s_i is removed from $S^{\mathcal{D}}$.

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

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

4.3 Relevance of the algorithm depending on the type of sensor network and the characteristics of the instruments

In this section, no assumption is made on the type of the sensor network (with or without reference instruments, with or without mobile nodes [13]). Indeed, according to the general definition of a rendez-vous, there is no reason that any of the instruments should be mobile or static. In the following sections, this concept is applied to sensor networks with mobile nodes, but as long as the conditions regarding the spatiotemporal vicinity are respected as exposed in Section 3.3, rendez-vous can happen, even for sensor networks with static nodes only.

In addition, the algorithm does not determine the best way to identify the state of the instruments, e.g. a list of instruments that can be compared consecutively to determine their state, that could exploit the structure of the sensor network. Instead, it takes into account all the rendez-vous the instruments had with others on a given time range, regardless of any possible underlying relationship between the nodes. Then, it investigates iteratively per instrument which rendez-vous are valid, e.g. notably those with instruments which are known to be non-faulty (due to the initial assumption on the instruments of the

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

procedure DIAGNOSIS(S, $\Phi(S \to S, (t_d, \Delta t))$, $|\Phi_v|_{min}$, $\Delta c^{\mathcal{D}}_{min}$)

Initiate the set of instruments to diagnose

$S^{\mathcal{D}} \leftarrow S \setminus S^{c_{max}}$

Initiate the predicted states

$\widehat{\Omega}(S^{c_{max}}, t_d), \widehat{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow$ NF, A

Initiate the actualised states

$\widetilde{\Omega}(S^{\mathcal{D}}, t_d) \leftarrow \widehat{\Omega}(S^{\mathcal{D}}, t_d)$

Ignore the instruments that cannot have enough valid rendez-vous

for $s_i \in S^{\mathcal{D}}$ **do**

if $|\Phi(s_i \to S, (t_d, \Delta t))| < |\Phi_v|_{min}$ **then**

$S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$

end if

end for

In this main loop, the state of each instrument to diagnose is predicted

repeat

The actualised states are now the predicted states

$\widehat{\Omega}(S, t_d) \leftarrow \widetilde{\Omega}(S, t_d)$

for $s_i \in S^{\mathcal{D}}$ **do**

$c^{\mathcal{D}}_{min}(s_i) \leftarrow c(s_i) + \Delta c^{\mathcal{D}}_{min}(c(s_i))$

Build the current set of valid rendez-vous

$$\begin{aligned}
&\Phi_v(s_i \to S, (t_d, \Delta t)) \leftarrow \\
&\{\varphi(s_i \to s_j, t) \in \Phi(s_i \to S, (t_d, \Delta t)), \\
&\text{such as } s_j \notin S^{\mathcal{D}}, \widehat{\Omega}(s_j, t') = NF, \\
&c(s_j) \geq c^{\mathcal{D}}_{min}(s_i), \\
&m(s_i, t) \in M^*(s_i, (t_d, \Delta t)), \\
&m(s_j, t') \in M^*(s_j, (t_d, \Delta t))\}
\end{aligned}$$

If s_i have enough valid rendez-vous, then compute the different rates to actualize its state

if $|\Phi_v(s_i \to S, (t_d, \Delta t))| \geq |\Phi_v|_{min}$ **then**

$r^{\approx}_{\Phi_v}(s_i \to S, (t_d, \Delta t)) \leftarrow \frac{|\Phi^{\approx}_v(s_i \to S, (t_d, \Delta t))|}{|\Phi_v(s_i \to S, (t_d, \Delta t))|}$

If the condition on the rate is met, then the actualised state of s_i is set to faulty, otherwise, it is set to non-faulty

if $r^{\approx}_{\Phi_v}(s_i \to S, (t_d, \Delta t))) < (r^{\approx}_{\Phi_v})_{min}$ **then**

$\widetilde{\Omega}(s_i, t_d) \leftarrow$ F

else

$\widetilde{\Omega}(s_i, t_d) \leftarrow$ NF

end if

$S^{\mathcal{D}} \leftarrow S^{\mathcal{D}} \setminus \{s_i\}$ # s_i is diagnosed so it can be removed from $S^{\mathcal{D}}$

end if

end for

The main loop is repeated until there is no difference between the predicted and actualised states

until $\widehat{\Omega}(S, t_d) = \widetilde{\Omega}(S, t_d)$

return $\widehat{\Omega}(S, t_d)$

end procedure

highest class in the network and the preceding iterations of the main loop of the algorithm) and builds its decision if enough valid rendez-vous were encountered.

Finally, no assumption is also made on the characteristics of the instruments such as the way they drift or their uncertainty for instance. Their characteristics are used to determine if a measurement result is valid and if two measurement results are compatible but for instance, it is not necessary to have instruments drifting linearly. Such a feature is not exploited by the algorithm.

For these reasons, the proposed algorithm can be applied to any sensor network with instruments without specific characteristics as long as:

- a practical definition for rendez-vous allowing their observation is chosen, followed by the choice of an acceptable minimal size for sets of valid rendez-vous.
- characteristics of the instruments allowing to determine the validity and compatibility of measurement results are known.
- a requirement can be exploited to determine notably the criterion to use to derive the states of instruments based on their sets of valid rendez-vous, in our case the minimal tolerated value $(r^{\approx}_{\Phi_v})_{min}$ associated to the rate of compatible rendez-vous in the set of valid rendez-vous of s_i over the time range $[t_d - \Delta t; t_d]$.[2]
- instruments of class c_{max} can be assumed as non-faulty. This is a significant assumption, but it is a completely realistic one, which in the future could be added as design principle for sensor networks if drift of performances is expected. For instance, in the context of air pollution monitoring in cities, there are already multiple monitoring stations used for regulatory purposes (see Airparif's website for the city of Paris for instance). Indeed, due to the applicable European Directives in Europe [41], these stations have to be maintained (and thus calibrated) regularly to keep their certification. Beyond the field of air pollution monitoring, considering the usually high investment, deployment and maintenance costs of large sensor networks, adding to the network the cost of acquisition and maintenance of a single high-quality measuring instrument can be easily envisioned. By contrast, the maintenance of hundreds of instruments is not for both technical and economic reasons.

These are the only requirements and assumption necessary to put in practice the presented algorithm. Although, a performance level is not guaranteed by design.

[2] Alternate manners to determine the predicted state of an instrument according to compatible and non-compatible valid rendez-vous could be chosen. For instance, the value of the predicted state could be chosen according to the number of compatible and non-compatible valid rendez-vous. Thus, $(r^{\approx}_{\Phi_v})_{min}$ would be replaced by a minimal number of compatible rendez-vous regarding $|\Phi^{\approx}_v(s_i \to S, (t_d, \Delta t))|$.

5 Application of the diagnosis algorithm

To showcase how to apply the diagnosis algorithm, this section provides a case study based on simulation with an analysis of its performances.

5.1 Definition of the case study

5.1.1 Sensor network

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

The instruments of class $k = 0$ move randomly in a discrete space of 100 positions at discrete time. The positions of the instruments are geometrically defined according to a grid of 10×10 with a 100m step, and centered on $(0, 0)$ as shown in Figure 2. At each time step a new position is chosen randomly for each instrument following a uniform law. Instruments may remain in place and multiple instruments may share a single position. Two instruments are in rendez-vous when they are at the same position at the same time. In addition, the instrument of class c_{max} is static. Its position is randomly drawn among the 100 positions.

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

5.1.2 Instruments

Instruments are assumed to be initially calibrated. In this case study, the instrument of class c_{max} is assumed as perfect, e.g. it does not drift. The instruments of class zero all follow a random gain and offset increase (RGOI) drift model [42]. 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 the uniform law $\mathcal{U}(0, \delta G_{max})$, and:

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

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

δG_{max} and δO_{max} are respectively the maximal gain and offset possible increase per time step.

Measured values are expressed following:

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

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

Table 1: Values of the parameters of the case study. The values of δG_{max} and δO_{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.

Parameter		**Value**	**Unit**
Class 1			
	$\Delta_r v$	1	%
	v_{min}	0.752	µg m^{-3}
Class 0			
	δG_{max}	2/(24*6*30)	%/10min
	δO_{max}	18.8/(24*6*30)	µg/m^3/10min
	$\Delta_r v$	30	%
	v_{min}	37.6	µg m^{-3}
True values model			
	A_{max}	200	µg m^{-3}
	σ_{max}	3826	m

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

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

A detection limit $v_{min}(c_i)$ is also defined to balance the effect of high uncertainties for low measured values. All the measurement results that have a measured value below the detection limit are not considered valid.

All the values of the parameters were defined according to information available in actual instrument datasheets. The following devices were considered to provide the values: AC32M Nitrogen oxides analyser (analyser, high quality), Cairsens NO_2 (low-cost measuring instrument), NO2-B43F Nitrogen Dioxide Sensor (low-cost sensor). However, the actual drift model is in general not provided in such datasheets; we chose an arbitrary, realistic drift model compatible with the outcomes of the considered documents. Nevertheless, though it was not validated, the drift model is similar to other ones that can be found out in the literature [43, 44].

Finally, relative uncertainty, limit of detection, gain drift and offset drift were used to model the behaviour of measuring instruments. Other characteristics could have been considered to assess the validity of the measurement results or to better model them, but this question is out of scope of this paper and shall be managed along with the adjustment of the parameters of the diagnosis algorithm to provide fine guidelines for sensor network designers.

The values used for the parameters considered in this case study are listed in Table 1.

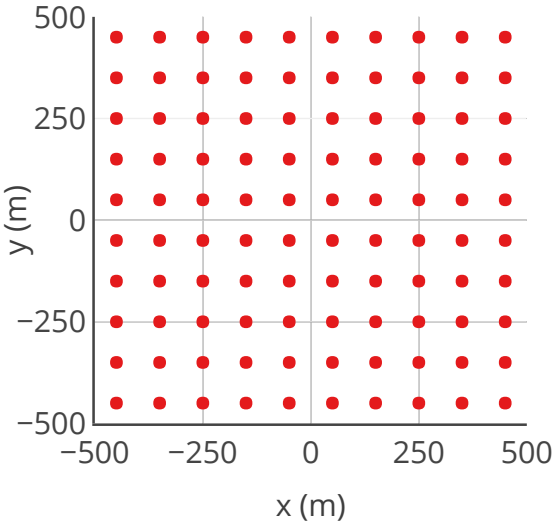


Figure 2: Map of the 100 positions available for the instruments in the case study. It is a 10×10 grid with a 100m step.

5.1.3 True values

To model the true values of the instruments, a pollutant source is considered at the center of the area of study. The concentration of pollutant, 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 σ for x and y. This model is not very realistic but has been used in other papers [45] for its ease of implementation. An example of a pollution map is displayed in Figure 3.

$A(t)$ and $\sigma(t)$ are drawn randomly at each time step following the uniform laws $\mathcal{U}(0, A_{max})$ and $\mathcal{U}(0, \sigma_{max})$ respectively. The values of A_{max} and σ_{max} are reported in Table 1.

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**

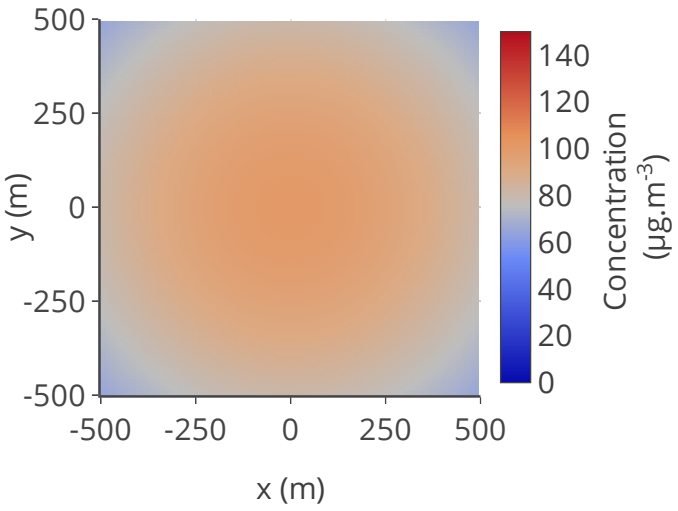


Figure 3: Example of concentration map used following $C(x, y, t) = A(t) \exp\left(-\frac{x^2 + y^2}{\sigma(t)^2}\right)$ for $A = 100\mu g\,m^{-3}$ and $\sigma = 637m$

values over the past 15 days. According to this specification, we choose $\Delta t = 15$ days and the rate threshold $(r^{\approx}_{\Phi_v})_{min}$ is set to 25%.

The minimal number of valid rendez-vous to conclude with a predicted state different from ambiguous $|\Phi_v|_{min}$ is set to 10.

5.3 Definition of the true state of an instrument

As it is a simulation, it is possible to know the true state 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, $\Omega(s_i, t) = NF$.

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 [46, 47] are appropriate but must be adapted as the predicted states can take three values instead of two.

Regarding the possible true states and predicted states, there are six cases:

- A non-faulty state predicted for an instrument which true state is non-faulty is a **true negative** (**TN**)
- A faulty state predicted for an instrument which true state is non-faulty is a **false negative** (**FN**)
- An ambiguous state predicted for an instrument which true state is non-faulty is a **non-determined negative** (**NDN**)
- A non-faulty state predicted for an instrument which true state is faulty is a **false positive** (**FP**)
- A faulty state predicted for an instrument which true state is faulty is a **true positive** (**TP**)
- An ambiguous state predicted for an instrument which true state is faulty is a **non-determined positive** (**NDP**)

To these primary metrics, P and N are added, which are respectively **the number of positives** (faulty true state) and **the number of negatives** (non-faulty true state). Metrics derived from the primary metrics are defined in Table 2 to facilitate the analysis of the results.

For the computation of the metrics, only the instruments in $S \setminus S^{c_{max}}$ are considered. The instruments of class c_{max} being always assumed as non-faulty and being not drifting in the following sections, taking them into account would bias the metrics.

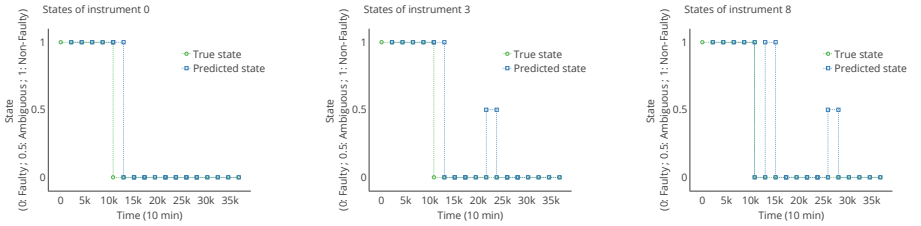
Another metric added is the **delay of first positive detection for an instrument** s_i, noted $\Delta\mathcal{D}(s_i)$. It is the difference between the index of the diagnosis procedure d where the true state $\Omega(s_i, t_d)$ changes from NF to F and the index of the diagnosis procedure d' where the predicted state $\widehat{\Omega}(s_i, t_{d'}) = F$ for the first time. Thus:

$$\Delta\mathcal{D}(s_i) = d - d'$$

This value can be positive or negative as there can be early or late positive detection. In the following results, the average and the standard deviation of this metric over $S \setminus S^{c_{max}}$ are considered.

Table 2: Metrics derived from P, N, TP, TN, FP, FN, NDP and NDN

Name	**Expression**	**Proportion of:**
Prevalence	$Prev = \frac{P}{P+N}$	- positive cases.
True positive rate	$TPR = \frac{TP}{P}$	- positive cases correctly detected.
True negative rate	$TNR = \frac{TN}{N}$	- negative cases correctly detected.
False positive rate	$FPR = \frac{FP}{N}$	- positive cases incorrectly detected.
False negative rate	$FNR = \frac{FN}{P}$	- negative cases incorrectly detected.
Non-determined positive rate	$NDPR = \frac{NDP}{P}$	- positive cases detected as ambiguous.
Non-determined negative rate	$NDNR = \frac{NDN}{N}$	- negative cases detected as ambiguous.
Non-determined rate	$NDR = \frac{NDP+NDN}{P+N}$	- all cases detected as ambiguous.
Positive predictive value	$PPV = \frac{TP}{TP+FP}$	- correctly detected positive cases per positive call.
False discovery rate	$FDR = \frac{FP}{TP+FP}$	- incorrectly detected positive cases per positive call.
Negative predictive value	$NPV = \frac{TN}{TN+FN}$	- correctly detected negative cases per negative call.
False omission rate	$FOR = \frac{FN}{TN+FN}$	- incorrectly detected negative cases per negative call.
Accuracy	$ACC = \frac{TP+TN}{P+N}$	- correct detection.



(a) Instrument 0, class 0 (b) Instrument 3, class 0 (c) Instrument 8, class 0

Figure 4: Evolution of the true state and of the predicted state for several instruments

5.5 Results

5.5.1 Observations from the results of different instruments

In Figure 4, the evolution of the true and predicted states over time are represented for different instruments. It shows that the diagnosis algorithm can produce different results over the instruments:

- The state of the instrument at each diagnosis procedure is often correctly predicted, except for a delay right after it becomes faulty (Figures 4a and 4b)
- There are predicted states equal to ambiguous (Figures 4b and 4c)
- Changing decisions happen, e.g. an instrument predicted as faulty may be predicted as non-faulty later (Figure 4c)

In all these particular cases, false results are observed, e.g. for instance an instrument is predicted as non-faulty when it is faulty. This is what happens at $t \approx 14k$ for s_8 in Figure 5: the value of $r^{\approx}_{\Phi_v}$ is not coherent with r_{true} because of valid rendez-vous issues. It is expected because the true state of an instrument is based on **all its measurement results against the true values** on the considered time range according to its definition in Section 5.3, whereas its predicted state is based on **the measurement results of its valid rendez-vous with other devices**. Thus, instruments which are actually non-faulty can be predicted as faulty and *vice versa* because the sets of measurement results used to compute the true and predicted state of an instrument are different.

5.5.2 Overall performances

Globally, the values of the metrics computed over all the diagnosis procedures are listed in Table 3. In this study, the prevalence of positive cases, e.g. the proportion of cases where the instruments are actually faulty over all the diagnosis procedures, is equal to 76%, for 153 cases (9 instruments diagnosed 17 times each).

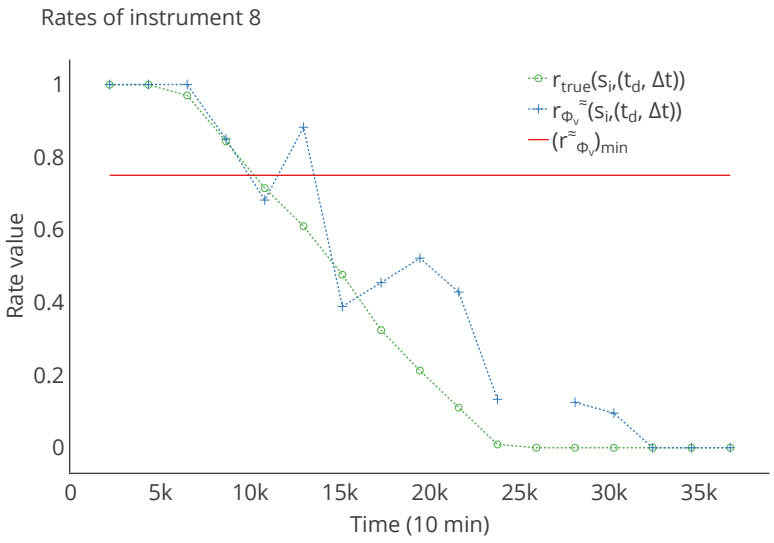


Figure 5: Evolution of the rates r_{true} and $r^{\approx}_{\Phi_v}$ for s_8 in the case study. The missing points for $r^{\approx}_{\Phi_v}$ are due to the ambiguous cases where $r^{\approx}_{\Phi_v}$ could not be computed.

Table 3 shows that:

- Most of the predicted states are true negatives (36 cases) and true positives (98 cases)
- Few predicted states are false negatives (17 cases) and non-determined positives (2 cases)
- No non-determined negative and false positive are predicted

Consequently, the TNR is equal to 1 and the FPR is equal to zero whereas the TPR is important (0.84) and the FNR is moderate (0.15). Moreover, the $NDNR$ is equal to zero and the $NDPR$ is equal to 0.02, thus $NDR = 1\%$ of ambiguous predictions were made which is low.

However, it appears that the algorithm is mostly correct when it predicts that instruments are faulty ($PPV = 1$, $FDR = 0$). It is often wrong when it predicts that instruments are non-faulty, about one third of the time ($NPV = 0.68$, $FOR = 0.32$). According to Figure 4, there is a delay when an instrument actually becomes faulty and the first prediction as faulty. It confirmed by the statistics of the delay of positive detection $\Delta\mathcal{D}$ in Table 3: the average value is negative and in fact it is always lower or equal to zero according to the maximal value of this metric. Nevertheless, the accuracy is equal to 88% which shows that there are mostly correct predictions overall.

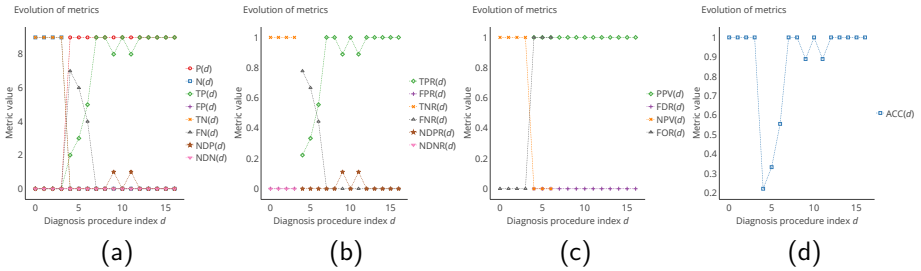


Figure 6: Evolution of the metrics computed for each diagnosis procedure as a function of the current diagnosis procedure id

5.5.3 Evolution over time of the metrics

The values of the metrics given in the previous subsection are representative of the results for all the diagnosis 17 procedures. They do not allow understanding how this overall result is achieved over the case study. Therefore, we study here their evolution.

Figure 6 shows the evolution of the metrics as a function of the index of the diagnosis procedure. Each curve gives the values obtained for the corresponding metric at the diagnosis d with the results of the nine instruments, where d is the index of the diagnosis procedure.

Figure 6a shows that first there are only non-faulty instruments (curve of N), and faulty ones appear at the 5^{th} diagnosis ($d = 4$). After the 6^{th} diagnosis, all the instruments of class zero are faulty (curve of P). The transition is quite abrupt, but this is expected because all the instruments of class zero follow the same drift model with the same parameters. The curves of the true positives TP and true negatives TN follow mostly the ones of P and N respectively. The false positives or negatives are in fact occurring around the transition where instruments of class zero become faulty. This behaviour is more clearly represented in Figures 6b and 6c with the curves of the TPR and FNR, and the curves of the NPV and FOR respectively. Finally, with Figure 6d, we observe that the global accuracy, initially equals to one, decreases until the transition phase and then increase again, converging to 1.

To conclude, this particular study of the evolution of the metrics allowed to understand the global results over the 17 diagnosis procedures: the false results occur mainly when the instruments actually become faulty and the more there are instruments actually predicted as faulty, the more the number of predicted states equal to ambiguous increases. Thus, on the first hand, worse overall results could have been displayed by considering the results of few diagnosis procedures around the transition phase. On the other hand, better results would be obtained by increasing the number of diagnosis procedures in the case study, without changing any parameter of the algorithm, e.g. by increasing the duration of the case study. However, this would not reduce the number of false

Table 3: Confusion matrix of states-related metrics and statistics of the delay of positive detection $\Delta\mathcal{D}$ for the case study

		True state			
		Non-faulty $N = 36$	Faulty $P = 117$	Prevalence 0.76	Accuracy 0.88
Predicted state	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		

Delay of positive detection			
μ	σ	min	max
-1.6	1.2	-3	0

that may happen in absolute terms.

5.6 Complementary studies

Based on these first results, multiple studies can be conducted to better understand the behaviour of our algorithm and identify its parameters of influence. A detailed in-depth study is the object of future work, but we present here two complementary results on the influence over the diagnosis performances of the model of the measurand and of the fault model used to derive the measured values.

5.6.1 Influence of the model of the measurand

The 2D Gaussian function used for the case study produces concentration fields that are quite homogeneous. To determine the behaviour of our algorithm with

a more variable measurand, the same case study was conducted by replacing only the model of the measurand with a random draw of the true values at each time step and at each position following the uniform law $\mathcal{U}(0, 400)$. A draw is represented in Figure 7. The results are reported in Table 4. We observe the results are similar to those of Table 3: mainly correct detections (TN, TP), a few false results but mainly false negatives. PPV, NPV, TNR and TPR have the same order of magnitude as previously and consequently so does their complementary metrics. The differences are a bit more significant for the delay of positive detection and its statistics. On average, it is better regarding its mean (-0.5 instead of -1.6 for the previous single study) and worse for its maximal value (1.9 instead of 0). The other metrics (min and σ) are equivalent on average. However, the most interesting result lies in the standard deviations of each metric. They are all close to zero, demonstrating that whatever the values drawn, it does not influence significantly the performance of the diagnosis algorithm.

Thus, through the comparison of the average results obtained with a uniform law as model for the true values, against the results obtained with a 2D Gaussian function as model for the true values, we observed first that the model chosen for the true values does not seem to have a significant effect on the results. In a second step, as the case study was repeated 100 times, we observed by studying the standard deviations of the metrics that similar results were obtained whatever the draw. Thus, the true values themselves also do not have a major influence on the results.

We explain this behaviour by the use of the concept of rendez-vous: only the values measured when instruments meet count, regardless of the general spatial and temporal variability of the measurand as long as the spatio-temporal conditions are respected for instruments to be in rendez-vous. Although these conditions are very simple in the present case study (instruments are in rendez-vous when they are at the same position at the same time) and easy to meet through simulation, it highlights once again that the choice of the practical definition for rendez-vous allowing their observation is major for the well-behaviour of our algorithm. Nevertheless, further studies are required to identify if there are conditions under which the true values or the behaviour of the measurand have an influence on the results.

5.6.2 Robustness of the algorithm

In the first place, only drift was added to true values to simulate the measured values. In practice, other faults like noise or spikes would be present in the measurement results, if not detected and corrected before the application of the drift diagnosis algorithm. It may have a major influence on the results as measurement results affected by other faults are likely to generate non-compatible measurement results with their true values. We can expect that such measurement results, if involved in valid rendez-vous, would generate non-compatible valid rendez-vous, influencing the value of the rate $r^{\approx}_{\Phi_v}$. This should likely generate positive results, e.g. the instruments are predicted as faulty earlier than

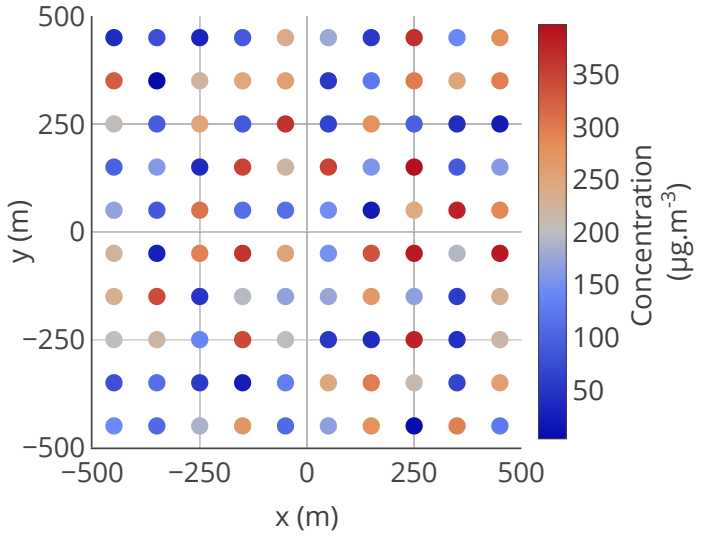


Figure 7: Example map of drawn values following the uniform law $\mathcal{U}(0, 400)$ at each measuring instrument's possible position

expected.

To illustrate this, spikes and noise were added to the measured values. The noise is modelled 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 $\mathcal{N}(0, \varepsilon_{max})$ at each time step, with $\varepsilon_{max} = 20$ here. For spikes, they occur 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:

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

The results are given in Table 5. We observe that N is lower and P is higher than in Table 3. 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 3. 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 3, the average delay of positive detection is close to zero and even

Table 4: Confusion matrix of states-related metrics and statistics of the delay of positive detection $\Delta\mathcal{D}$ with true values drawn from the uniform law $\mathcal{U}(0, 400)$, repeated 100 times

		True state							
		Non-faulty		Faulty		Prevalence		Accuracy	
		μ	σ	μ	σ	μ	σ	μ	σ
		63.34	0.59	89.66	0.59	0.59	0.00	0.83	0.03
Predicted state	N.-Faulty	TN		FN		NPV		FOR	
		μ	σ	μ	σ	μ	σ	μ	σ
		60.87	1.50	22.63	3.46	0.73	0.03	0.27	0.03
	Ambig.	NDN		NDP					
		μ	σ	μ	σ				
		0.00	0.00	0.17	0.38				
	Faulty	FP		TP		FDR		PPV	
		μ	σ	μ	σ	μ	σ	μ	σ
		2.47	3.39	66.86	3.39	0.04	0.02	0.97	0.02
		TNR		FNR					
		μ	σ	μ	σ				
		0.96	0.02	0.18	0.03				
		$NDNR$		$NDPR$		NDR			
		μ	σ	μ	σ	μ	σ		
		0.00	0.00	0.00	0.00	0.00	0.00		
		FPR		TPR					
		μ	σ	μ	σ				
		0.04	0.02	0.75	0.04				

Statistics on the delays of positive detection over the 100 draws				
	μ	σ	min	max
$\Delta\mathcal{D}$ mean	-0.5	0.5	-1.4	1.2
$\Delta\mathcal{D}$ std dev	1.6	0.4	0.7	2.6
$\Delta\mathcal{D}$ min	-3.1	0.9	-5.0	-1.0
$\Delta\mathcal{D}$ max	1.9	1.0	0.0	4.0

positive. Its maximal value is also positive. It indicates that the prediction of instruments as faulty happens earlier as expected.

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, it is 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.

5.7 Summary and outcomes of the case studies

In this section, a case study was conducted to put in practice the algorithm of diagnosis presented in Section IV. It provides satisfying results regarding the initial specification, with a few false or ambiguous results (only false negatives and non-determined positives). The overall accuracy shows a large majority of correct predictions (88%). It is worth mentioning that results cannot be extrap-

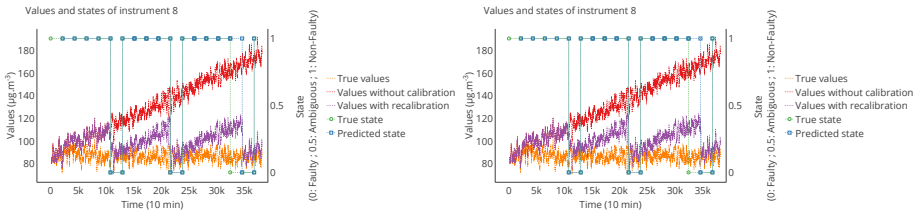
Table 5: Confusion matrix of states-related metrics and statistics of the delay of positive detection $\Delta\mathcal{D}$ with spikes and noise added to the measured values

		True state			
		Non-faulty	Faulty	Prevalence	Accuracy
		$N = 23$	$P = 130$	0.85	0.90
Predicted state	Non-faulty	TN	FN	NPV	FOR
		20	11	0.65	0.35
	Ambiguous	NDN	NDP		
		0	2		
	Faulty	FP	TP	FDR	PPV
		3	117	0.03	0.97
		TNR	FNR		
		0.87	0.08		
		$NDNR$	$NDPR$	NDR	
		0.00	0.02	0.01	
		FPR	TPR		
		0.13	0.90		

Delay of positive detection			
μ	σ	min	max
0.2	1.7	-2	3

olated directly to another case study as this case study is rather elementary in terms of sensor network, measurand and rendez-vous conditions. A prospect of this work is to provide automated means to adjust the parameters of the algorithm depending on the specific features of the use case. Two additional case studies are provided to give insights on the sensitivity of the method to case study parameters: we showed first that the spatio-temporal variability of the measurand has no influence on the diagnosis results, as long as the conditions for instruments to be in rendez-vous are well-defined and respected. In a second step, we showed that the algorithm is robust to other faults like spikes or noise, the only side effect being an earlier prediction of the instruments as faulty, generating false positives that are less harmful than false negatives in this case.

The diagnosis algorithm is primarily oriented toward detecting when instruments require calibration. Considering that it would be combined to a cali-



(a) With recalibration based on an oracle

(b) With recalibration based on linear regression with the set of valid rendez-vous used at the diagnosis instant

Figure 8: Evolution of the true values, of the measured values without and with recalibration for different choices of calibration approach for s_8. The true and predicted states are plotted to show when recalibrations are triggered. The true and predicted states for the case without recalibration corresponds to those of Figure 4c. A rolling averaging of the values over 24 h was performed to make the graph easier to read.

Table 6: Mean (μ) and standard deviation (σ) of the slope, intercept and coefficient of determination r^2 computed for the nodes of class 0 of the network on the entire time interval of study and for each calibration strategy.

Calibration approach	**Slope**		**Intercept**		**r²**	
	μ	σ	μ	σ	μ	σ
No calibration	1.04	0.00	42	0	0.84	0.00
Oracle-based	1.01	0.00	14	1	0.97	0.01
Linear regression-based	1.02	0.00	15	1	0.98	0.01

bration approach in practice, we investigate in the following section how this algorithm behaves when calibration is actually carried out on an instrument after it is predicted as faulty. Moreover, the valid rendez-vous used to predict the state of an instrument provide associations of its measured values with the values from other instruments considered as references. We use them in a second calibration approach.

6 Combination with calibration approaches

In the study presented in this section, 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 respec-

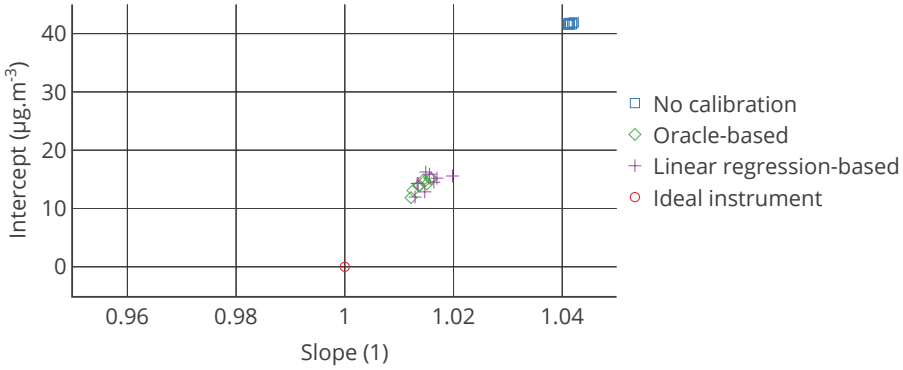


Figure 9: Target plot of the 9 nodes of class 0 of the network as a function of their slope and intercept, computed on the entire time interval of study, for each calibration strategy.

tively reset to their initial values, 1 and 0 respectively. 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 often tested in the literature for the calibration of measuring instruments [10].

The case study is identical to the one described in Section 5 regarding the sensor network, the true and measured values and the configuration of the diagnosis algorithm.

Results for a single instrument are displayed in Figure 8. The instrument is recalibrated right after being diagnosed as faulty for both recalibration methods. Nevertheless, it is only visible at the next diagnosis procedure (instrument is predicted non-faulty again). From the curves of the predicted states we cannot assess the quality of the calibration, it only shows that it was efficient enough to be classified again as non-faulty.

To proceed to an evaluation of the impact of the calibration on the measurement results, a linear regression between the true values and the corrected values is carried out for each instrument. It is computed on the entire time interval of study. Table 6 presents the statistics of the slope, intercept and coefficient of determination of the regression and Figure 9 represents a target plot of the instruments as a function of their slope and intercept. We observe that both the calibration approaches give equivalent results compared to the case without calibration. The error on the slope and on the intercept are respectively reduced of at least 50%, e.g. from 1.04 to 1.02, and 60%, e.g. from 41 to 16.

However, even with the oracle-based calibration, we observe that the remaining average offset can still be considered as significant. This residual is explained by the fact that the gain and offset of the instruments are corrected at specific moments and not at each time step: between two diagnosis procedures, the instruments continue to drift. Moreover, when an instrument is considered as faulty, the correction is applied only on the values that are measured after the moment this prediction is made. This is illustrated in Figure 8.

More significantly, the average coefficient of determination of the linear regression increases of 15 % (from 0.84 to 0.97) indicating at least a better connection between the calibrated values and the true values than between the measured values and the true values if no calibration is performed. Thus, a simple calibration approach, based on the information contained in the rendez-vous used to predict the states of the instruments, allows obtaining an interesting improvement of their metrological performances.

To conclude, this case study shows that, in addition to being able to detect when an instrument does not meet a specification and even without a fine adjustment of its parameters, the diagnosis algorithm proposed provides also useful data to perform the recalibration of measuring instruments, namely through the sets of valid rendez-vous used during diagnosis procedures. Nevertheless, the calibration approach we used in this section and more generally the case studies we conducted in this paper are quite simple. Extensive studies, through experiments or simulation, are needed to figure out how the best *in situ* calibration results can be obtained with our diagnosis algorithm and a calibration approach, depending on the measurand, the sensor network and its context of deployment.

7 Conclusion

A novel algorithm for the diagnosis of drift faults in sensor networks is introduced. It exploits the concept of rendez-vous between measuring instruments. The algorithm studies the validity of the rendez-vous before using them to determine the states of the instruments. It can be applied to any type of sensor network in terms of composition of the network and mobility of the nodes. The major assumption needed is to be able to assume the instruments of the highest metrological class in the sensor network as always non-faulty. For practical use of the algorithm, an end-user will have to define the proper criteria for validity and compatibility between measurement results based on the monitored phenomenon and on instruments' datasheets; then to establish, based on the specificity of the use case (mobility of the instruments, spatial variability of the phenomenon) and of the instruments (for instance the time they require to perform a measurement), the conditions under which instruments are in rendez-vous, the minimal number of valid rendez vous required to derive the state of an instrument, and the requirement indicating in which state an instrument is depending on its number of valid rendez-vous over a given time range.

A case study based on a phenomenon following a 2D Gaussian law is con-

ducted to showcase the algorithm and to provide insights on its performances. It shows satisfying results with mostly correct prediction, although there are a few false negatives. The overall accuracy of correct prediction is equal to 88% for this use case. Complementary studies provide a preliminary result on the sensitivity of the method to the specifics of each case study. One shows that the spatio-temporal variability of the measurand has no influence on the diagnosis results and that the algorithm is robust to other faults such as spikes or noise.

Afterwards, the diagnosis algorithm is associated to two calibration approaches. When an instrument is diagnosed as faulty, the valid rendez-vous that were used to make the prediction of its state 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, with a reduction of the error up to 50% on average for the gain. Therefore, the diagnosis algorithm that is 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.

Future work aims at providing means to adjust the different parameters of the algorithm toward optimized performances of diagnosis. Following on from the first complementary case studies we exposed, the identification of the main performance factors of the sensor network and its environment of deployment is a major perspective for the use of this algorithm in real-case applications.

Resources

Codes and data files used to produce the results of Sections 5 and 6 are available online under licence AGPL 3.0 and ODbL 1.0 respectively at https://doi.org/10.25578/IAZYBN.

References

[1] N. Castell, F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. M. Broday, and A. Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?" *Environment International*, vol. 99, pp. 293–302, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.envint.2016.12.007

[2] P. Kumar, L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. Di Sabatino, M. Bell, L. Norford, and R. Britter, "The rise of low-cost sensing for managing air pollution in cities," *Environment International*, vol. 75, pp. 199–205, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.envint.2014.11.019

[3] F. Derkx, B. Lebental, T. Bourouina, F. Bourquin, C.-S. Cojocaru, E. Robine, and H. Van Damme, "The Sense-City project," in *XVIIIth Symposium on Vibrations, Shocks and Noise*, 2012. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00860842/file/doc00015219.pdf

[4] J. K. Hart and K. Martinez, "Environmental Sensor Networks: A revolution in the earth system science?" *Earth-Science Reviews*, vol. 78, no. 3-4, pp. 177–191, 2006.

[5] P. W. Rundel, E. A. Graham, M. F. Allen, J. C. Fisher, and T. C. Harmon, "Environmental sensor networks in ecological research," *New Phytologist*, vol. 182, no. 3, pp. 589–607, 2009.

[6] X. Fang and I. Bate, "Issues of using wireless sensor network to monitor urban air quality," in *International Workshop on the Engineering of Reliable, Robust, and Secure Embedded Wireless Sensing Systems (FAILSAFE). ACM.*, 2017.

[7] C. Borrego, A. M. Costa, J. Ginja, M. Amorim, M. Coutinho, K. Karatzas, T. Sioumis, N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, P. Smith, N. André, P. Gérard, L. A. Francis, N. Castell, P. Schneider, M. Viana, M. C. Minguillón, W. Reimringer, R. P. Otjes, O. von Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, and M. Penza, "Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise," *Atmospheric Environment*, vol. 147, no. 2, pp. 246–263, 2016.

[8] F. Karagulian, M. Barbiere, A. Kotsev, L. Spinelle, M. Gerboles, F. Lagler, N. Redon, S. Crunaire, and A. Borowiak, "Review of the performance of low-cost sensors for air quality monitoring," *Atmosphere*, vol. 10, no. 9, 2019.

[9] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, and OIML, *International vocabulary of metrology - Basic and general concepts and associated*

terms (VIM), 3rd ed. JCGM 200:2012, 2012. [Online]. Available: http://www.bipm.org/vim

[10] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, "Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide," *Sensors and Actuators, B: Chemical*, vol. 215, pp. 249–257, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.snb.2015.03.031

[11] B. Maag, Z. Zhou, and L. Thiele, "A Survey on Sensor Calibration in Air Pollution Monitoring Deployments," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4857–4870, 2018.

[12] J. M. Barcelo-Ordinas, M. Doudou, J. Garcia-Vidal, and N. Badache, "Self-Calibration Methods for Uncontrolled Environments in Sensor Networks: A Reference Survey," *Ad Hoc Networks*, 2019. [Online]. Available: https://doi.org/10.1016/j.adhoc.2019.01.008

[13] F. Delaine, B. Lebental, and H. Rivano, "In Situ Calibration Algorithms for Environmental Sensor Networks: A Review," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 5968–5978, aug 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8686160/

[14] Y. Wang, A. Yang, Z. Li, P. Wang, and H. Yang, "Blind drift calibration of sensor networks using signal space projection and Kalman filter," in *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2015, pp. 1–6.

[15] Z. Li, Y. Wang, A. Yang, and H. Yang, "Drift detection and calibration of sensor networks," in *2015 International Conference on Wireless Communications and Signal Processing, WCSP 2015*, 2015, pp. 1–6.

[16] O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele, "On rendezvous in mobile sensing networks," in *In Springer RealWSN*, 2013, pp. 29–42.

[17] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, "Blind mobile sensor calibration using an informed nonnegative matrix factorization with a relaxed rendezvous model," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2941–2945.

[18] A. Mahapatro and P. Khilar, "Fault Diagnosis in Wireless Sensor Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. Early Acce, pp. 1–27, 2013.

[19] T. Muhammed and R. A. Shaikh, "An analysis of fault detection strategies in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 78, pp. 267–287, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.jnca.2016.10.019

[20] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11 349–11 364, 2018.

[21] A. Mehmood, N. Alrajeh, M. Mukherjee, S. Abdullah, and H. Song, "A survey on proactive, active and passive fault diagnosis protocols for WSNs: Network operation perspective," *Sensors (Switzerland)*, vol. 18, no. 6, 2018.

[22] P. Chanak, I. Banerjee, and R. S. Sherratt, "Mobile sink based fault diagnosis scheme for wireless sensor networks," *Journal of Systems and Software*, vol. 119, pp. 45–57, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jss.2016.05.041

[23] M. Abo-Zahhad, S. M. Ahmed, N. Sabor, and S. Sasaki, "Mobile Sink-Based Adaptive Immune Energy-Efficient Clustering Protocol for Improving the Lifetime and Stability Period of Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4576–4586, 2015.

[24] J. Chen, S. Kher, and A. Somani, "Distributed fault detection of wireless sensor networks," in *Proceedings of the 2006 workshop on Dependability issues in wireless ad hoc networks and sensor networks - DIWANS '06*, 2006, p. 65. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1160972.1160985

[25] X. Xu, W. Chen, J. Wan, and R. Yu, "Distributed fault diagnosis of wireless sensor networks," in *International Conference on Communication Technology Proceedings, ICCT*, 2008, pp. 148–151.

[26] T. Saha and S. Mahapatra, "Distributed Fault diagnosis in wireless sensor networks," in *2011 International Conference on Process Automation, Control and Computing*, 2011.

[27] K. F. Ssu, C. H. Chou, H. C. Jiau, and W. T. Hu, "Detection and diagnosis of data inconsistency failures in wireless sensor networks," *Computer Networks*, vol. 50, no. 9, pp. 1247–1260, 2006.

[28] M.-H. Lee and Y.-H. Choi, "Distributed diagnosis of wireless sensor networks," in *TENCON 2007 - 2007 IEEE Region 10 Conference*, 2007, pp. 2159–3450.

[29] A. Mahapatro and P. M. Khilar, "Online distributed fault diagnosis in wireless sensor networks," *Wireless Personal Communications*, vol. 71, no. 3, pp. 1931–1960, 2013.

[30] A. Mahapatro and A. K. Panda, "Choice of detection parameters on fault detection in wireless sensor networks: A multiobjective optimization approach," *Wireless Personal Communications*, vol. 78, no. 1, pp. 649–669, 2014.

[31] X. Luo and M. Dong, "Distributed Faulty Sensor Detection in Sensor Networks," in *Artificial Neural Networks – ICANN 2009*, 2009, pp. 964–975.

[32] X. Y. Xiao, W. C. Peng, C. C. Hung, and W. C. Lee, "Using sensorranks for in-network detection of faulty readings in wireless sensor networks," in *International Workshop on Data Engineering for Wireless and Mobile Access*, 2007, pp. 1–8.

[33] S. Ji, S. F. Yuan, T. H. Ma, and C. Tan, "Distributed fault detection for wireless sensor based on weighted average," in *NSWCTC 2010 - The 2nd International Conference on Networks Security, Wireless Communications and Trusted Computing*, vol. 1, 2010, pp. 57–60.

[34] J. Jiang, G. Han, F. Wang, L. Shu, and M. Guizani, "An Efficient Distributed Trust Model for Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1228–1237, 2015.

[35] N. Wang, J. Wang, and X. Chen, "A trust-based formal model for fault detection in wireless sensor networks," *Sensors (Switzerland)*, vol. 19, no. 8, pp. 1–20, 2019.

[36] K. P. Sharma and T. P. Sharma, "rDFD: reactive distributed fault detection in wireless sensor networks," *Wireless Networks*, vol. 23, no. 4, pp. 1145–1160, 2017.

[37] Y. Feiyue, T. Yang, Z. Siqing, D. Jianjian, X. Juan, and H. Yao, "A Faulty node detection algorithm based on spatial-temporal cooperation in wireless sensor networks," in *Procedia Computer Science*, vol. 131. Elsevier B.V., 2018, pp. 1089–1094. [Online]. Available: https://doi.org/10.1016/j.procs.2018.04.266

[38] X. Fu, Y. Wang, W. Li, Y. Yang, and O. Postolache, "Lightweight fault detection strategy for wireless sensor networks based on trend correlation," *IEEE Access*, vol. 9, pp. 9073–9083, 2021.

[39] D. Rodriguez, M. Valari, S. Payan, and L. Eymard, "On the spatial representativeness of NOX and PM10 monitoring-sites in Paris, France," *Atmospheric Environment: X*, vol. 1, p. 100010, 2019. [Online]. Available: https://doi.org/10.1016/j.aeaoa.2019.100010

[40] Y. Li, Z. Yuan, L.-W. A. Chen, A. Pillarisetti, V. Yadav, M. Wu, H. Cui, and C. Zhao, "From air quality sensors to sensor networks: Things we need to learn," *Sensors and Actuators B: Chemical*, vol. 351, p. 130958, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925400521015264

[41] EU, "Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe," *Official Journal of the European Communities*, vol. 152, pp. 1–43, 2008.

[42] F. Delaine, B. Lebental, and H. Rivano, "Framework for the Simulation of Sensor Networks Aimed at Evaluating In Situ Calibration Algorithms," *Sensors*, vol. 20, no. 16, p. 4577, aug 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/16/4577

[43] S. Moltchanov, I. Levy, Y. Etzion, U. Lerner, D. M. Broday, and B. Fishbain, "On the feasibility of measuring urban air pollution by wireless distributed sensor networks," *Science of the Total Environment*, vol. 502, pp. 537–547, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.scitotenv.2014.09.059

[44] A. Arfire, A. Marjovi, and A. Martinoli, "Model-based rendezvous calibration of mobile sensor networks for monitoring air quality," in *2015 IEEE SENSORS*, 2015, pp. 1–4.

[45] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, "Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 4, pp. 667–682, 2018.

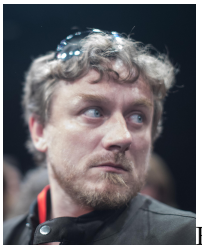
[46] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210832718301546

[47] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning*, 1st ed. Springer Publishing Company, Incorporated, 2011.

Florentin Delaine obtained his Ph.D. in Computer Science from Institut Polytechnique de Paris, France, in 2020. He graduated from École Normale Supérieure Paris-Saclay, France, in 2017 from which he obtained a first M.S. in Applied Physics and a second M.S. in Design and Control of Critical Systems. His research interests include data science for sensor networks with a focus on metrology and dependability for these systems.

Bérengère Lebental graduated from Ecole Polytechnique's (Institut Polytechnique de Paris, France) Engineering Program in 2006 and

both its Physics and its Nanotechnology Master of Science Programs in 2007. She obtained her Ph.D. in Civil Engineering from Université Paris-Est, France, in 2010. Since 2010, she is research scientist at Laboratory of Physics of Interfaces and Thin Films (École Polytechnique, CNRS) and French Institute of Science and Technology for Transport, Development and Network (IFSTTAR). Her research focuses on nano-enabled sensors for smart city applications, including design, fabrication, reliability analysis, IoT integration, field deployment and data exploitation. She has coordinated several large scale national and EU research projects on sensors for smart cities, such as Sense-City (`http://sense-city.ifsttar.fr/`) or Proteus (`http://www.proteus-sensor.eu/`). She is cofounder of the startup Altaroad for connected roads.



Hervé Rivano is University Professor at INSA Lyon. He is the head of the Inria/INSA Lyon common team Agora of the CITI lab. The team focuses on wireless networks for digital cities. He obtained his PhD in november 2003 from the University of Nice-Sophia Antipolis and graduated from the École Normale Supérieure de Lyon. His research interests include combinatorial optimization applied to network design and provisioning. He focuses on capacity/energy tradeoff for urban cellular and mesh networks design and low cost and dense wireless sensor networks for environmental sensing.