Automatic detection of hardware failures in an air quality measuring station with low cost sensors

Sylvain Poupry, Cédrick Béler, Kamal Medjaher

Laboratoire Génie de Production, ENIT - Toulouse INP, 47 Avenue d'Azereix, Tarbes, 65000, France
sylvain.poupry@enit.fr
cedrick.beler@enit.fr
kamal.medjaher@enit.fr

Abstract

Monitoring air quality to protect the population is a challenge for cities with modest budgets. With this in mind, a measuring station has been developed using low-cost sensors (LCS) arranged in Triple Modular Redundancy (TMR). However LCS technology has limitations which lead to incomplete or inaccurate air quality measurements.

To improve the availability of the measuring station, and also to make the data gathered more reliable, a fault detection method is proposed in this paper. By comparing measurements collected by the LCS in TMR configuration, the proposed method synthesizes measurements for each monitored parameter and assesses the health state of the measuring station in real-time. This information can be used to promptly alert maintenance teams, facilitating timely interventions and ensuring the continuous monitoring of air quality.

1. Introduction

Monitoring air quality at a city level with conventional means is very costly and only cities with a large budget are equipped for that. Therefore, for small municipalities with a modest budget, it was proposed to develop air quality measuring stations with low cost sensors (LCS) as an alternative. Reducing costs by a factor of 10 to 100, LCS enable measurement points to be extended over vast areas, thanks to the integration of Internet of Things (IoT) technology. While they may be less precise than conventional measuring stations, LCS can be used in conjunction with them, as suggested by *Castell et al.* (Castell et al., 2017). They are highly accessible and easy to deploy without the need for specialized personnel. Taking advantage of these benefits, a proposed measurement station featuring triple redundancy using LCS sensors has been proposed in (Poupry, Béler, & Medjaher, 2022). The measuring station consists of 3 smart sensors (SmS) and an Aggregator component (Figure 1). All these elements are connected locally in WiFi to a Local Area Network (LAN) and connected to the Internet. The Aggregator's primary function is to receive, store, structure and process observation data transmitted by the SmS in order to be inform relevant authorities to make decisions. The role of each SmS is to transmit measurements structured in temporal matrix to the Aggregator at a period Te determined by monitoring objectives for specific parameters and the capacities of the LCS. Each SmS consists of a micro-controller and one or more LCS that measure the N physical parameters established during the city measurement definition.

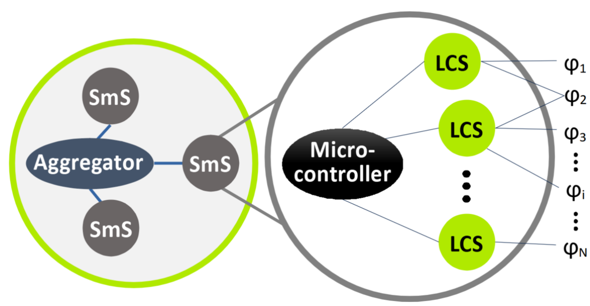

Figure 1. Measuring station composition.

The SmS have been designed in our research laboratory to augment LCS with enhanced computation and communication capabilities. At the heart of the station lies the Aggregator, which acts as a central hub that gather information and facilitates communication by means of an API, enabling smooth integration and data exchange between the Smart Sensors and the graphical user interface (GUI) used by with decision-makers. The initial deployment of one station has been successful, but currently, the initial deployment of the station has been beset by several hardware failures. Unfortunately, these faults can only be diagnosed by the individual who constructed the station, creating a need for a more streamlined and accessible maintenance process. Simplify-ing the detection of hardware failures is imperative, as maintenance personnel may lack extensive electronics knowledge and expertise in station reliability and fault diagnostics.

The main objective of the project is to enable long-term and large-scale air quality monitoring, supplying decision-makers with consistent data via an interface for well-informed decision-making regarding citizen protection. Nonetheless, initial deployment limitations show that LCS may encounter random and sudden failures during pollutant concentration measurement, requiring expert validation and network monitoring. To ensure consistent long-term pollutant level measurements, there is a need to simplify hardware failure detection, acknowledging that maintenance staff might lack expertise in electronics and station reliability and fault diagnostics. As such, this paper propose to extend the measurement station capabilities with self-confirmation detection and self-diagnostic features, facilitating more straightforward maintenance and system reconfiguration.

The rest of paper is organized into four sections. Section 2 discusses the current state of the field and highlights the novelty of the paper. Section 3 presents the proposed fault detection method, its assumptions, and its main components. Section 4 focuses on applying the method to a measuring station in Argelès-Gazost, France, and discussing the results. Finally, section 5 concludes the paper and suggests future directions.

2. State of the art and work positioning

LCS, being cost-effective, can be deployed in large quantities, which significantly enhances the spatial resolution of measurements across a designated area, refining the monitoring grid over the territory. On one hand, this improvement in measurement points can enrich prediction models, leading to better forecasts. On the other hand, it enables the detection of local pollution sources that may be difficult to predict using conventional spatial and topographic dispersion models. This approach is similar the work on waste treatment or industrial sites (Schneider et al., 2017; Morawska et al., 2018). However, the drawbacks of LCS include issues with their material quality, measurement drift, cross-interference with other pollutants, and their relatively short lifetime (Lewis et al., 2016). To overcome these shortcomings, the design of the measuring station is based on a triple modular redundancy (TMR) technique applied to SmS. The TMR method enhances the measurement horizon and and allows to yield coherent data. The principle of TMR involves three identical and independent modules operating in parallel and measuring the same physical parameter. The information output of these modules is subjected to an algorithm based on a voting process to generate a synthesis. The voting process can typically involve a majority vote, a median vote, or a weighted average vote (Lorczak, Caglayan, & Eckhardt, 1989). Although this configuration improves the reliability (Kucera, Hyncica, Cidl, & Vasatko, 2006), the fundamental algorithms have two negative aspects in the context of the proposed solution: on the one hand, these algorithms share a common feature of masking errors through aggregations and on the other hand the modules must be synchronized to enable comparison. However, the station's design ensures that the SmS operate independently, transmitting their measurements at their own frequency. As a result, the voting algorithm must be adjusted to account for this lack of synchronization and accept non-simultaneous measurements.

This paper focuses on fault detection in SmS by comparing them with each other. The closest domain to the application addressed in this paper is the Wireless Sensor Networks (WSN), where the Aggregator acts as a sink node and the SmS as sensor nodes. Among these various fault detection techniques presented in (Muhammed & Shaikh, 2017), the most appropriate for this study seems to be the centralized self-fault detection. Likewise, the median voting algorithm is chosen because of its measure of central tendency, and is less sensitive to extreme and missing values, which is a common problem with LCS. In data processing, the median is preferred to the arithmetic mean, as the mean is strongly influenced by extreme values (Leys, Ley, Klein, Bernard, & Licata, 2013). However, the median may have limitations with odd input values or when outliers are present (Bass, Latif-Shabgahi, & Bennett, 1997). These outliers, particularly those caused by hardware sensor failures, need to be excluded.

In conclusion, the measurement station uses TMR of SmS equipped with LCS that may have significant material defects. To enable self-diagnosis and auto-configuration, TMR techniques and fault detection in WSN are employed, with hardware maintenance performed by the maintenance team. The contribution of this paper consists in the formalization of a mathematical method executed in four steps: measurement collection, fault detection, diagnostic analysis with potential feedback from maintenance operators, and aggregation considering outliers exclusion and submission for aggregation.

3. Proposed method development

This section presents an exposition of the proposed method development, which is divided into three distinct parts. Initially, the measurement protocol and related definitions are provided, followed by an elaboration of the working assumptions pertaining to the method. Finally, a detailed account of the method is presented.

3.1. Measurement protocol

The measurement station described in this article measures $N = 4$ physical parameters: PM10 (φ_1), temperature (φ_2), humidity (φ_3), and pressure (φ_3). The station is composed of three SmS, each of which is a micro-controller connected to the LCS for computing and communication capabilities. Each SmS possesses its own sampling period Te_x of approximately one minute.

The station operates with an observation window of size τ, typically lasting around an hour. It down-samples the measurements from the SmS. The Aggregator plays a crucial role by collecting the measurements from the three SmS during the observation period and performing median aggregation. It facilitates communication among the SmS, detects failures, and synthesizes the observations for each τ period.

The station communicates with the external environment via the GUI that connects directly to the aggregator to obtain detection and diagnostic information, and to communicate feedback on any necessary maintenance interventions. The reliability of the station refers to the consistency and long-term availability of the data.

3.2. Working assumptions

This paper does not address sensor calibration policy due to resource limitations in the laboratory. The manufacturer's calibration is used, with adjustments made in a controlled environment to ensure consistency. This workaround does not significantly impact the study's focus on trend identification and failure detection. Calibration will be addressed in future studies.

For the measuring station's physical configuration, the time response is relatively slow, allowing for the exclusion of short-term pollution peaks. The sampling period is set at one hour ($\tau = 1$ hour).

As the distance between sensors in the station is small, the measured physical parameter is assumed to be the same. The SmS observations are merged at the same point for each parameter for all LCS.

The SmS has a sampling period of about a minute, requiring an observation window of at least three points for comparisons ($\tau > 2 * Te_x$). Its modular design facilitates easy replacement and reconfiguration.

The Aggregator is a single point of failure but is considered more reliable than the SmS. Local storage on each SmS prevents data loss in case of aggregator failure. Data recovery strategy will be considered in future work.

3.3. Detailed method and related formulas

The proposed method has been incorporated into the Aggregator to ensure the reliability of the observations made by the SmS. This method involves a sequence of eight steps, which are outlined in Figure 2. The processing steps culminate in the production of synthesis measurements of the physical parameters, along with the detection of hardware faults.

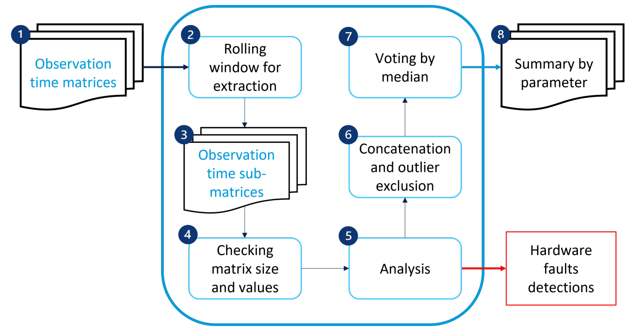

Figure 2. The functionalities of the aggregator.

Step 1: Initialization.
The initial step involves the creation of temporal observation matrices, which are also called as raw data. The observations are sent from all the SmS of the station and are slightly asynchronous due to the varying observation periods of each SmS. Indeed, each one measures the four physical parameter with its own sampling period, denoted by Te_x. All these observations (M_x from SmS x) are stored in a temporal matrix Y_x of size ($M_x * N$). The physical measurements are not synchronized since each SmS is autonomous and independent. The physical parameter measured by the SmS number x at time k is denoted by $y_x^{\varphi_i}(k)$. Thus, an observation of the four physical parameter ($N = 4$) at time k is noted $y_{x,k}$ as shown in Eq. (1). The observation matrix Y_x is described in matrix (2).

$$y_{x,k} = \{y_x^{\varphi_1}(k), y_x^{\varphi_2}(k), y_x^{\varphi_3}(k), y_x^{\varphi_4}(k)\} \tag{1}$$

$$Y_x = \begin{bmatrix} y_x^{\varphi_1}(1) & y_x^{\varphi_2}(1) & y_x^{\varphi_3}(1) & y_x^{\varphi_4}(1) \\ y_x^{\varphi_1}(2) & y_x^{\varphi_2}(2) & y_x^{\varphi_3}(2) & y_x^{\varphi_4}(2) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi_1}(k) & y_x^{\varphi_2}(k) & y_x^{\varphi_3}(k) & y_x^{\varphi_4}(k) \\ y_x^{\varphi_1}(k+1) & y_x^{\varphi_2}(k+1) & y_x^{\varphi_3}(k+1) & y_x^{\varphi_4}(k+1) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi_1}(M_x) & y_x^{\varphi_2}(M_x) & y_x^{\varphi_3}(M_x) & y_x^{\varphi_4}(M_x) \end{bmatrix} \tag{2}$$

Step 2: Rolling window for extraction.
A rolling observation window of size τ starts at a certain time t and ends at $t + \tau$. It remains fixed in size and moves along the collected data by the SmS. This window serves both real-time and offline measurements, and is also called as an observation window. It has two advantages: grouping measurements of the same parameter to enhance data reliability and reducing computation time by processing only the data within the time period τ without repeating calculations for each new data point. The rolling window technique confirms fault detection through comparison and validates hardware updates.

Step 3: Observation time sub-matrix.

The observation window includes q_x observations from SmS x due to its own sampling period Te_x. As depicted in the matrix (3), the purple lines are the sub-matrix and are extracted from the Y_x observation matrix. At time $t = k * Te_x$, the sub-observation matrix $y^{\tau}_{x,k}$ of size ($q_x * 4$) is extracted from the temporal matrix Y_x of each SmS at a specific moment k. The Sub-matrices are extracted according to the condition given in Eq. (4) to ensure that a minimum of two points per SmS sub-matrix are extracted.

$$Y_x = \begin{bmatrix} y_x^{\varphi 1}(1) & y_x^{\varphi 2}(1) & y_x^{\varphi 3}(1) & y_x^{\varphi 4}(1) \\ y_x^{\varphi 1}(2) & y_x^{\varphi 2}(2) & y_x^{\varphi 3}(2) & y_x^{\varphi 4}(2) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(k{+}1) & y_x^{\varphi 2}(k{+}1) & y_x^{\varphi 3}(k{+}1) & y_x^{\varphi 4}(k{+}1) \\ y_x^{\varphi 1}(k{+}2) & y_x^{\varphi 2}(k{+}2) & y_x^{\varphi 3}(k{+}2) & y_x^{\varphi 4}(k{+}2) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(k{+}h) & y_x^{\varphi 2}(k{+}h) & y_x^{\varphi 3}(k{+}h) & y_x^{\varphi 4}(k{+}h) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(k{+}q_x) & y_x^{\varphi 2}(k{+}q_x) & y_x^{\varphi 3}(k{+}q_x) & y_x^{\varphi 4}(k{+}q_x) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(M_x) & y_x^{\varphi 2}(M_x) & y_x^{\varphi 3}(M_x) & y_x^{\varphi 4}(M_x) \end{bmatrix} \quad (3)$$

$$(k + q_x) * Te_x - (k + 1) * Te_x \leq \tau \quad (4)$$

$$Y^{\tau}_{x,k} = \begin{bmatrix} y_x^{\varphi 1}(k{+}1) & y_x^{\varphi 2}(k{+}1) & y_x^{\varphi 3}(k{+}1) & y_x^{\varphi 4}(k{+}1) \\ y_x^{\varphi 1}(k{+}2) & y_x^{\varphi 2}(k{+}2) & y_x^{\varphi 3}(k{+}2) & y_x^{\varphi 4}(k{+}2) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(k{+}h) & y_x^{\varphi 2}(k{+}h) & y_x^{\varphi 3}(k{+}h) & y_x^{\varphi 4}(k{+}h) \\ \vdots & \vdots & \vdots & \vdots \\ y_x^{\varphi 1}(k{+}q_x) & y_x^{\varphi 2}(k{+}q_x) & y_x^{\varphi 3}(k{+}q_x) & y_x^{\varphi 4}(k{+}q_x) \end{bmatrix} \quad (5)$$

Step 4: Matrix checking.
Once extracted, the sub-matrices are then checked for their size and values using the algorithm presented in Figure 3. The check of the matrix size allows to determine whether the SmS concerned was able to perform observations or not. During this step, it should be noted that software failures are also a possibility. However, it is necessary to emphasize that our focus is solely on hardware failures. The absence of observations indicates that a hardware failure affects the SmS x. This check is then carried out on all matrices in order to count the number of SmS in service.

For the matrices containing observations, each column (corresponding to the observations of a specific parameter with from the sensors of the identified SmS) is extracted. The obtained data are valid if they contain at least two points, they

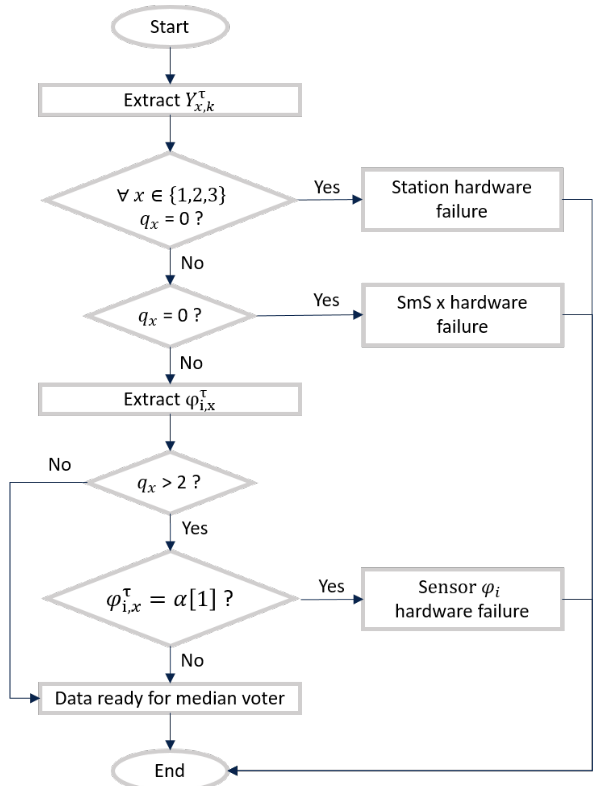

Figure 3. Checking sub-matrix for detections

are used later. If they contain more than three points, their values are compared to detect if they are close enough. When all values are identical, the constant is identified and named α. This means that the sensor is blocked and only gives the α value, and that a hardware problem has occurred. These data are excluded from the concatenation. If the values are different, they are valid and concatenated. From these checks, a detection matrix is built. The detections are represented by a matrix denoted $D^{\tau}_{s,k}$, of the form of matrix given in Eq. (6). In this matrix, the following notations are used: n_i is the number of active SmS for parameter φ_i, m_i is the list of identifiers of SmS without observations (missed observation) during period τ and b_i is also a list of identifiers of SmS with a failure of type blocked LCS (blocked observation) during period τ.

$$D^{\tau}_{s,k} = \begin{bmatrix} n_1 & n_2 & n_3 & n_4 \\ m_1 & m_2 & m_3 & m_4 \\ b_1 & b_2 & b_3 & b_4 \end{bmatrix} \quad (6)$$

Step 5: Analysis of detection matrix.
The analysis of the detection matrix enables the identification and transmission of detection and diagnostic processes via the GUI, based on the interpretation of the expert. The detection matrix also facilitates the detection and early diagnostic by the maintenance team. It is compared to the previous one, if available, and only changes are reported to the GUI. Depend-

ing on its value of α, a diagnostic can be established. Differences with the previous detection matrix prevents redundant alerts (every τ), which is particularly useful for maintenance. Therefore, only changes are identified, which enable failures and reconfigurations pertinent and verified.

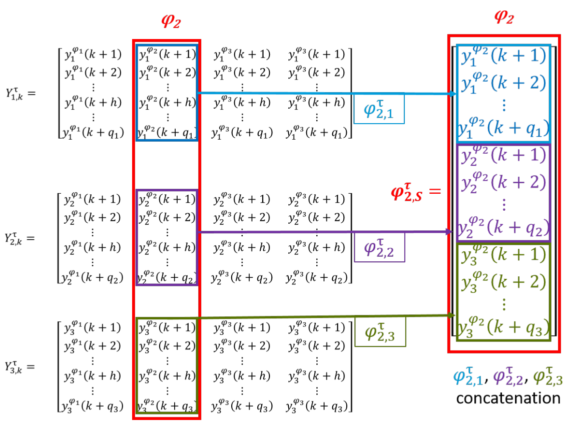

Figure 4. $\varphi^{\tau}_{i,x}$ extraction and concatenation.

Step 6: Concatenation and outlier exclusion.
After the analysis step, all valuable data are concatenated and the invalid data are excluded. As shown in figure 4, these data are stored in a column. $\varphi^{\tau}_{i,x}$ represent observations of a physical parameter φ_i from the SmS x. All data from all SmS are concatenated in one column noted $\varphi^{\tau}_{i,s}$. This column represents observation data of a physical parameter φ_i from the station s. This concatenation step is applied to all measured parameters of the station. Once the concatenations are done, the values contained in the $\varphi^{\tau}_{i,s}$ columns are submitted to the median vote.

Step 7: Median vote.
The voting algorithm employed in this study utilizes the median to synthesize all parameter values and generate a summary matrix for the station. The median-based voting method involves arranging the values of $\varphi^{\tau}_{i,s}$ in ascending order and selecting the value at position $q_x/2$ in the case of an odd number of values, while for an even number of values, the average of the values at position $q_x/2$ and $(q_x + 1)/2$ is computed.

Step 8: Synthesis by parameter.
The resulting set of values obtained by the median vote represents a temporal sub-matrix observation of the station, encompassing all physical parameters, denoted as $S^{\tau}_{s,k}$, collected at the station level for a specific sampling interval τ. This step serves as a data resampling process, whereby the sub-matrix is concatenated with the previous one on a general temporal matrix denoted Y_s which is consisting of $S^{\tau}_{s,k}$. Afterwards, the resulting matrix and detection matrix are transformed into a suitable set of alerts and curves, which are made available and transmitted to the relevant authorities via the GUI. This allows for real-time monitoring and prompt interventions in the event of significant pollution incidents.

4. Application

The study being presented was conducted as part of the BOLDAIR project, which is funded by the Occitanie region and involves the participation of ADEME. The project is being carried out within the CCPVG (Communauté de communes Pyrenees vallées des gaves) in the central Pyrenees mountain area in the South-West of France. The initial deployments of the project were implemented in the city of Argelès-Gazost, and the research described in this paper was conducted in this specific context.

4.1. Measuring station technical setup

The measurement station is installed on an external wall of an administrative building, positioned one meter away from the wall and protected from environmental factors such as wind and sunlight that may cause measurement inaccuracies. The sampling duration is set to approximately 12 minutes (Te_x), and the observation window duration is one hour (τ) to have at least three observations per each SmS. The monitored physical parameters are Temperature (**T**), Humidity (**H**), Pressure (**P**), and Particulate Matter less than 10 µm in diameter (**PM10**).

From a technical standpoint, the measurement station comprises three SmS, a power supply unit, and an Aggregator. The aggregator is based on a Raspberry Pi connected in WiFi to an Internet access point. The power supply unit provides electricity to the SMS and Aggregator. The SMS module consists of two LCS (SdS011 and BME280) selected for their established use in similar projects. These sensors are connected to an ESP32 microcontroller via an Inter Integrated Circuit (I²C) bus which is a three-wire connection that facilitates sensor replacement. Furthermore, this type of connection allows the ESP32 to detect sensor presence and to provide a specific value in the case of a malfunction. For instance, a value of zero indicates the absence of a sensor.

In terms of communication, the desktop application for decision makers, which is a web client, connects directly to the aggregator node. This allows for convenient access to physical parameter readings and alerts, all through a single interface. The interface itself displays various curves and widgets, which are represented in the Figure 5.

4.2. Results

The data presented in this section are observations from the measurement station in Argelès-Gazost municipality from February to May 2022. The graphs displayed were generated from the GUI connected to the aggregator. Figure 5 compiles the raw data for Humidity and PM10, excluding Temperature and Pressure due to similar behavior in terms of detection. In-

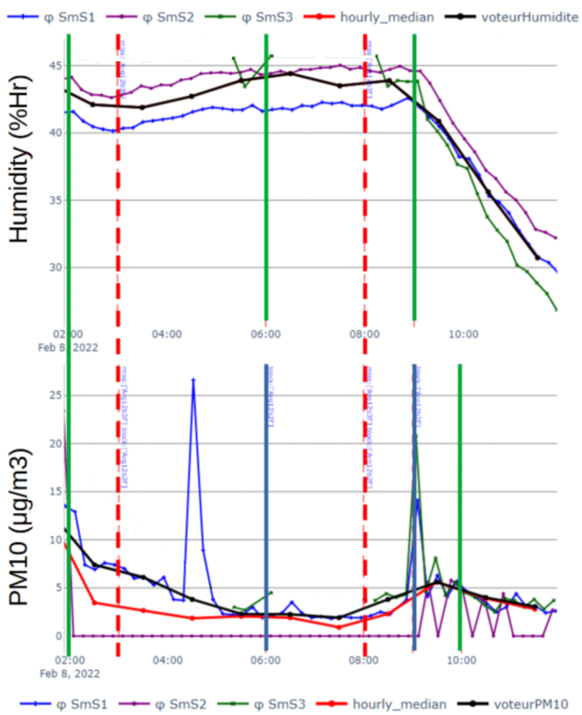

Figure 5. Aggregator output with detections.

deed, only complete SMS hardware failures were detectable since no other problems occurred. However, PM10 LCS are more sensitive to external conditions and have shown different types of failures. Therefore, the following figures will be presented for the physical parameter of PM10 due to its relevance to the proposed method employed in this paper. SmS 1, 2, and 3 are respectively named Arg12b1f, Arg12b2f, and Arg12b3f. Hourly medians (in red) are derived from resampling all observations of PM10 using the proposed method with a rolling window of size τ. The aggregator output (in black) and detections (red vertical bars) are shown. In Figure 5, red bars indicate missing sensor observations, green bars indicate restoration to normal, and blue bars indicate sensors blocked at a constant value identified as α. This Figure highlights missing SMS detections for all parameters and shows different detections depending on the measured parameter, enabling targeted intervention at the station level. Notably, elusive failures are detected through alternating failure and return-to-normal patterns, as shown for SmS 2 in the figure.

The voter and hourly median curves are indistinguishable when everything is functioning properly, as is the case for the humidity parameter in Figure 5. It is interesting to examine the difference observed between the voter and the hourly median during the first detection of a missing sensor and a blocked sensor. Figure 6 shows a reproduction of the previously observed configuration over a longer period of time. The detection matrix confirms the blocked values: LCS from SmS 1 at $\alpha = 4$ and LCS from SmS 2 at $\alpha = 0$ for non-response. So, there is a divergence between the voter and the hourly median. The method presented in this paper excludes blocked values, providing central values closer to the contextualization proposed by the detection matrix.

When blocked values are excluded, the voter is concentrated on the central value corresponding to SmS number 3, whereas the median is influenced by the blocking of SmS 1 at $\alpha = 4$. The situation is worse when the PM10 sensor on SmS 2 also fails to communicate with $\alpha = 0$: the median corresponds to the blocked value in the middle $\alpha = 4$, while the last SmS continues to operate correctly. With this method, the voter gives the result of the last SmS in service.

The methodology employed involves utilizing a measurement station equipped with LCS sensors, which not only enhances the reliability of observations but also provides contextual information through fault detection. This approach ultimately contributes to the overall robustness of the observations made.

The methodology described allows maintenance teams to effectively locate faulty sensors and, through feedback, to identify observation patterns related to specific failures. This approach also facilitates the detection of elusive failures and their complete failure in the future, as some values may evolve to be locked to a specific value over time during observations. It should be noted that a long-term goal of this research is to integrate the detection method with a knowledge base, in order to improve the automatic detection and explanation of failures by exploiting the accumulated expertise.

Finally, connection issues associated with missing SmS are detected based on the number of sensors in service for each parameter. A connection problem of an SmS is identified, in particular, by the elongation of its sampling period. Thus, hardware failures and needed reconfigurations are directly identified by the aggregator, along with annotations done in the GUI on the graphs of each physical parameter.

4.3. Discussion

This study seeks to show that even with potentially inconsistent low-cost sensors (LCS), long-term measurements can still be successfully carried out. Despite their random failure and the proposed method to identify failing sensors while maintaining measurement continuity even in the occurrence of a failure event.

The objective is to highlight that in the context of LCS deployment, these sensors typically have an average lifespan of no more than four months. However, the goal is to perform measurements throughout the year. It appears that with a knowledgeable maintenance team ready to intervene, the station's lifespan can be extended through these failure detections and preliminary diagnoses. The practical applications of

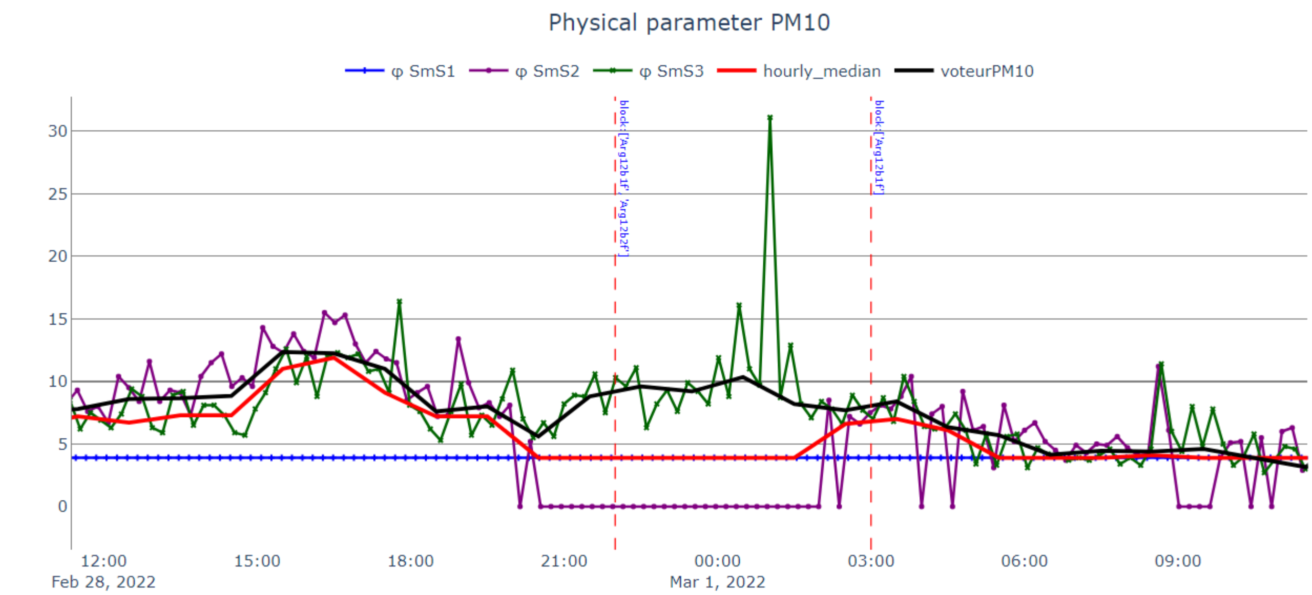

Figure 6. Typical influence on median.

these findings aim to enhance the longevity of an LCS-based measurement station. The implications suggest that the success of a self-configuring station will rely on the quality of its alerts and the maintenance team's ability to address issues before all SmS units fail and cease to function.

These findings are in line with existing literature on how to use LCS to monitor air quality, and the implications suggest that it is necessary to continue to find ways to circumvent LCS failures until the technology can be made more reliable.

4.4. Results synthesis

The results demonstrate the relevance of using the proposed method to enhance the observations of the measuring station while taking into account the failures of the LCS. The method enables the Aggregator to adapt the synthesis according to the events affecting the LCS, and also provides detection for contextualizing the data. Furthermore, it facilitates comprehension of the LCS malfunctions to ease diagnostics and interventions by the maintenance team.

5. Conclusion

This paper presented a method to improve data reliability for an air quality measurement station using LCS. The method detects LCS failures by comparing observations from SmS in TMR. This approach identifies sudden hardware failures, informing the maintenance team. Combining this method with successful interventions ensures consistent data availability.

However, the method has limitations. It may not effectively detect sensor drift and degradation-related failures in LCS. Variability among LCS measurements can indicate the beginning of failure. The spatial configuration and design of the measurement station boxes may affect LCS observations. The assumption that the aggregator never fails is unrealistic, and its failure can lead to data loss.

In the future and to overcome the limitations, it would be beneficial to investigate potential sensor degradation or construct failure indices for other types of sensors to predict the quality of station components. Outlier detection methods will be implemented to introduce mobile thresholds for detecting outliers and to construct failure indices based on SmS measurement errors. Comparing such observations with those of conventional stations could be a topic of discussion for the validation data and even for some better calibration feature. Additionally, developing confidence or quality indexes for measurements would be valuable to detect sensors that are failing due to continuous degradation.

Acknowledgment

The authors would like to thank Rose-Marie GRENOUILLET in charge of the Environment and Serge DUTHU in charge of operations for their advice and for the setting up of the measuring station on the field. This work is co-funded by Région Occitanie and Communauté de Communes Pyrénées Vallées des Gaves.

NOMENCLATURE

N	number of physical parameters monitored, $N \in \mathbb{N}$
x	SmS number with $x \in \{1, ..., 3\}, x \in \mathbb{N}$
Te_x	sensor n° x sampling period
τ	station sampling period x such that $\tau > 2Te_x$
τ	observation window size
k	k^{th} sampling period with $t = k * Te_x$
φ_i	physical parameter number i, $i \in \{1, ..., N\}, i \in \mathbb{N}$
M_x	number of SmS x observations
Y_x	SmS x observation time matrix $(M_x * N)$
$y_x^{\varphi_i}(k)$	parameter φ_i measurement by SmS n° x at time k
$y_{x,k}$	SmS x observations at time k
q_x	number of SmS x observations during τ
$y_{x,k}^{\tau}$	Y_x sub-matrix $(q_x * N)$ at time k during τ
$\varphi_{i,x}^{\tau}$	$y_{x,k}^{\tau}$ column containing the observations of φ_i
$\varphi_{i,s}^{\tau}$	$\varphi_{i,x}^{\tau}$ concatenation from all SmS
$S_{s,k}^{\tau}$	summary of SmS observation between t and $t+\tau$
$D_{s,k}^{\tau}$	detection matrix between t and $t+\tau$
α	constant observation of a blocked LCS during τ

REFERENCES

Bass, J., Latif-Shabgahi, G. R., & Bennett, S. (1997). *Experimental comparison of voting algorithms in cases of disagreement*. (Pages: 523) doi: 10.1109/EURMIC.1997.617368

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., … Bartonova, A. (2017, February). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, *99*, 293–302. doi: 10.1016/j.envint.2016.12.007

Kucera, P., Hyncica, O., Cidl, J., & Vasatko, J. (2006, February). Realibility model of TMR system with fault detection. *IFAC Proceedings Volumes*, *39*(21), 468–472. doi: 10.1016/S1474-6670(17)30233-1

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., … White, A. (2016, July). Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discussions*, *189*(0), 85–103. (Publisher: The Royal Society of Chemistry) doi: 10.1039/C5FD00201J

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013, July). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. doi: 10.1016/j.jesp.2013.03.013

Lorczak, P., Caglayan, A., & Eckhardt, D. (1989, June). A theoretical investigation of generalized voters for redundant systems. In *[1989] The Nineteenth International Symposium on Fault-Tolerant Computing. Digest of Papers* (pp. 444–451). doi: 10.1109/FTCS.1989.105617

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., … Williams, R. (2018, July). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International*, *116*, 286–299. doi: 10.1016/j.envint.2018.04.018

Muhammed, T., & Shaikh, R. A. (2017, January). An analysis of fault detection strategies in wireless sensor networks. *Journal of Network and Computer Applications*, *78*, 267–287. doi: 10.1016/j.jnca.2016.10.019

Poupry, S., Béler, C., & Medjaher, K. (2022). Development of a reliable measurement station for air quality monitoring based on low-cost sensors and active redundancy. *IFAC-PapersOnLine*, *55*(5), 7–12.

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017, September). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment International*, *106*, 234–247. doi: 10.1016/j.envint.2017.05.005

BIOGRAPHIES

Sylvain Poupry was born in France in 1986. He received a Generalist Engineering Degree with Integrated Systems Design Option at Tarbes National School of Engineering (ENIT), in July 2019. He is a third-year doctoral student within the Production Engineering Laboratory (LGP). The topic of his Ph.D. research is "Contribution to the design and implementation of a reflexive Cyber-Physical System: application to air quality prediction in the "vallées des gaves"". His current research interests are low-cost sensors, air pollution sensing, the internet of things, data science, and Prognostics and Health Management.

Cédrick Béler is assistant professor at ENIT (Ecole Nationale d'ingénieurs de Tarbes). His researches are conducted in the field of Social-Cyber-Physical System and Digital Twin and related to data science, knowledge management. He is especially interested in the way information is organized in distributed networks of information systems with human in the loop. Application are developed in the context of industry 4.0 but also in the public space.

Kamal Medjaher received the Engineering degree in electronics from Mouloud Mammeri University, Tizi Ouzou, Algeria, the M.S. degree in control and industrial computing from Ecole Centrale de Lille, Villeneuve-d'Ascq, France, in 2002, and the Ph.D. degree in control and industrial computing from University of Lille 1, Villeneuve-d'Ascq, France, in 2005. He is currently Full Professor at Tarbes National School of Engineering (ENIT), France. He conducts his research activities within the Production Engineering Laboratory. His current research interests include prognostics and health management of industrial systems and predictive maintenance.