Sustainability in Environment
ISSN 2470-637X (Print) ISSN 2470-6388 (Online)
Vol. 3, No. 4, 2018
www.scholink.org/ojs/index.php/se

Original Paper

Enhanced PEMS Performance and Regulatory Compliance through Machine Learning

Gregorio Ciarlo[1*], Daniele Angelosante[2], Marco Guerriero[2], Giorgio Saldarini[3] & Nunzio Bonavita[1]

[1] ABB S.p.A., Via Albareto, 35, 16153 Genoa, Italy

[2] ABB Switzerland, Corporate Research Center, Segelhofstrasse 1, 5405 Baden, AG, Switzerland

[3] ABB S.p.A., Via Luigi Vaccani, 4, 22010 Tremezzina CO, Italy

[*] Gregorio Ciarlo, ABB S.p.A., Via Albareto, 35, 16153 Genoa, Italy

Received: October 12, 2018 Accepted: October 28, 2018 Online Published: November 1, 2018

doi:10.22158/se.v3n4p329 URL: http://dx.doi.org/10.22158/se.v3n4p329

Abstract

Modeling technologies can provide strong support to existing emission management systems, by means of what is known as a Predictive Emission Monitoring System (PEMS). These systems do not measure emissions through any hardware device, but use computer models to predict emission concentrations on the ground of process data (e.g., fuel flow, load) and ambient parameters (e.g., air temperature, relative humidity). They actually represent a relevant application arena for the so-called Inferential Sensor technology which has quickly proved to be invaluable in modern process automation and optimization strategies (Qin et al., 1997; Kadlec et al., 2009). While lots of applications demonstrate that software systems provide accuracy comparable to that of hardware-based Continuous Emission Monitoring Systems (CEMS), virtual analyzers are able to offer additional features and capabilities which are often not properly considered by end-users. Depending on local regulations and constraints, PEMS can be exploited either as primary source of emission monitoring or as a back-up of hardware-based CEMS able to validate analyzers' readings and extend their service factor. PEMS consistency (and therefore its acceptance from environmental authorities) is directly linked to the accuracy and reliability of each parameter used as input of the models. While environmental authorities are steadily opening to PEMS, it is easy to foresee that major recognition and acceptance will be driven by extending PEMS robustness in front of possible sensor failures. Providing reliable instrument fail-over procedures is the main objective of Sensor Validation (SV) strategies. In this work, the capabilities of a class of machine learning algorithms will be presented, showing the results based on tests performed actual field data gathered at a fluid catalytic cracking unit.

Keywords

Inferential sensor, predictive emission monitoring systems, emission monitoring, sensor validation, machine learning

1. Introduction

The increasing attention devoted to air quality by legislative, scientific, industrial and public sectors has led to the development and enforcement of different emission monitoring strategies.

The traditional solution employed by the industry to comply with the legislation is to monitor the emissions via hardware based CEMS; such systems normally comprise analysers (to sample and identify the compositions of released flue gas) and an IT infrastructure to manage, record and store the emissions values (Arioni et al., 2013; Paira et al., 2017).

PEMS are an innovative technology able to exploit advanced mathematical models in order to estimate emission values with an accuracy comparable with that of CEMS at a fraction of the cost.

Several techniques are actually available for model building, however, a broad categorization into two groups can be made:

- Empirical or data-driven modelling, wherein available data are used to train "black-box" models;
- First principle modelling, wherein a theoretical model derived from the physics of the problem is adopted.

Indipendently from the approach, PEMS need to be fed with real-time process data, such as fuel flow, load, operating pressure and temperature.

On the ground of more than 15 years of successful field applications (Eisemann et al., 2014) and because of the inherent lower life-cycle costs and reduced footprint (Roth & Lawrence, 2010), PEMS have gained and keep gaining acceptance and popularity in a growing number of industries and countries. According to several environmental regulations (EPA, 2009; Netherlands Technical Agreement, 2014), plants are allowed to lease a portable CEMS to gather emissions data, which, combined with simultaneous process data, will be used to build and validate mathematical models.

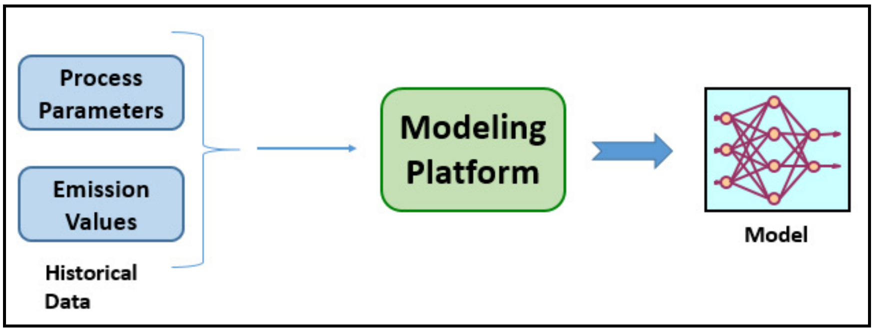

Figure 1. Inferential Modeling For Emission Monitoring

Once the models have been certified, the temporary CEMS is removed and replaced by the inferential system.

This approach is particularly suited when applied to combustion processes where fuel nature and composition are reasonably stable or measurable: for the same reason it is not advisable to use PEMS on urban incinerators where feed composition is "by design" highly variable and uncontrolled. Some countries have already released specific guidelines about where PEMS should be applied instead of CEMS and where the hardware-based solution will be maintained as the only allowed option (Ciarlo & Bonavita, 2016).

PEMS can also be used as a back-up if a CEMS is in place, and irrespective of which role it plays, it provides numerous benefits in different applications (Samdami, 1994).

Further than replicating what an HW-based CEMS can do, PEMS are able to offer additional features, like:

- Trace back causes of emissions, identifying process conditions resulting in limit exceedance;
- Reconstruct emission levels from historical data, in case of failure of the hardware device;
- Contribute to process optimization strategies.

Finally it should be underlined that PEMS, even if a well established technology since the 90's, are presently gaining momentum from the Digital Transformation wave and from the dramatic role that Machine Learning techniques are acquiring in process industry (Chui et al., 2018).

2. PEMS and Sensor Validation

While PEMS show a number of specific advantages over CEMS (reduced maintenance, no consumables, no energy consumption, etc.), because of their data-driven nature, they require distinctive attentions.

Great care has to be paid on the quality assurance of PEMS output; building an effective and reliable model is not enough as the ongoing accuracy of the model strictly relies on the quality of the input variables used by the model itself.

This means that each process parameter fed into the model shall be routinely verified to confirm that related instrumentation is performing according to expectations. In the case any anomalous behaviour is detected, a proper alarm shall be raised, warning about PEMS prediction quality.

Presently, some of the best commercial solutions exploit simple data range checks to verify, in real-time, if a sensor reading stays within predefined limits. Although reasonable, these checks are pretty relaxed and not very effective in identifying all possible failures (sensor freezing to a constant value is for example not discoverable through such a basic control). Additionally if sensor j is found out-of-range the typical action is to replace its value with the last useful sample at time t* (see Figure 2). Obviously this emergency gimmick becomes quickly less and less acceptable.

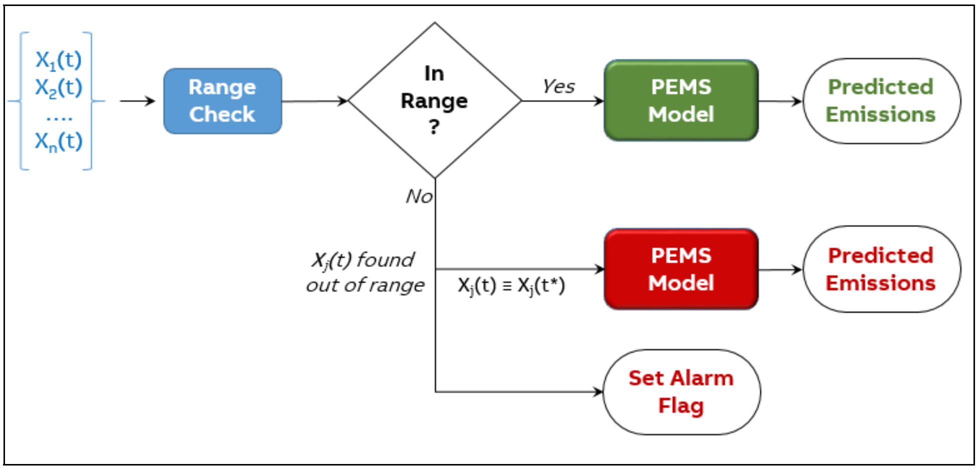

Figure 2. Basic Sensor Fault Detection in Present PEMS

Notwithstanding the above approach is actually accepted by most of the regulations, additional requirements will be increasingly present in legislation and in end-user specifications. The development and deployment of robust SV routines will become a relevant requirement for PEMS applications (Ciarlo et al., 2017): this additional software module will be responsible not only to identify possible sensors malfunctions, but also to provide reconstructed values to be temporarily used as input values. In fact, reconciled data-under proper conditions-could be automatically selected as a replacement of the affected sensor, maintaining PEMS accuracy and performances well within legislation limits.

Machine Learning techniques are now available and promise to greatly enahance data-driven application robustness and reliability. Authors have developed an approach that, starting from a very general sensor classification, is able to help end-users and practitioners in automatically:

a) Detect when a PEMS output should be labelled as "bad quality" because one of the related model input is faulty;

b) Identify which is the sensor to be blamed as faulty (in order to promptly alert maintenance department);

c) Whenever possible, substituting the faulty sensor reading with a reconciled value with the objective of preserving PEMS acceptable performances (Angelosante et al., 2018).

The final scope is to provide a resilient version of PEMS, able to properly manage model input failures. This Robust PEMS (or R-PEMS) could contribute to reassure Environmental and Health Authorities and public opinion at large about the reliability and effectiveness of software-based emission monitoring systems facilitating their adoption and operation.

3. R-PEMS: Concept and Procedure

At the beginning of each project, PEMS engineer have to select the most relevant process data showing good correlation with the emission variables.

The acquired parameters are normally chosen after a deep analysis and investigation on the P&IDs and other process documentation. As modern industrial plants are equipped with an extensive set of devices for control and monitoring purposes, the initial data-collection phase involves a very large number of signals (much more than those actually used as input variables for PEMS) that are then reduced during the engineering phases. During a PEMS project, it is quite common to acquire up to 80-100 different process variables and to progressively reduce the number up to use less than a dozen inputs for each model.

Under normal operations, the abundance of sensors naturally provides a frame rich of multiple correlations among instrument values. When a sensor encounters a fault (e.g., drifiting, freezing), the correlation of its values with other variables tends to change: this peculiar behaviour can be used to identify a faulty sensor.

This scenario paves the way for sensor validation techniques based on analytical redundancy: this approach consists in continuously comparing the actual readings from sensors with estimated values obtained through data-driven models exploiting the relationships with other plant parameters.

Unfortunately analytical redundancy is not always present among sensors. It is the case of variables that cannot be estimated by the readings of surrounding devices (e.g., the sulfur content in gas in combustion processes).

In thoses cases a different approach must be adopted: this is generally referred as hardware redundancy sensor validation, and requires the presence of multiple devices (usually at least three in order to allow "voting" mechanisms) so that it is possible to promptly identify failures and resort to reliable back-ups. Clearly, hardware redundancy, requiring additional instrumentation, is expensive and it is typical of critical processes (nuclear power plants, aerospace…) while it is limited to highly crucial variables in most industrial processes.

The present paper focuses only on the analytical redundancy methods and how these solutions can help in maintaining PEMS accuracy within the regulatory-enforced limits: the normal target is to keep model relative accuracy within 10% limit even in case of input failures (different values can be applicable depending on actual pollutant concentrations).

From an operative point of view, the first step will be understanding if and where analytical redundancy among model inputs exists and how it can be used for SV purposes: in the following section an overview on how to classify the different sensors will be provided and eventually which SV strategy should be applied to each category.

3.1 Establishing a SV Procedure: The Sensor Taxonomy Classification

Building a PEMS always starts with identifying quantitative relationships among a number of sensors. From this perspective, sensors can be clustered in several types of sensor classes on the ground of their mutual interdependencies (Figure 3):

- Sensor Not Relevant for PEMS (NR)—must be discarded from the dataset
- Relevant for PEMS but Not Correlated to any other input (R-NC)—needed for PEMS model but unfit for any SV strategy -> Not recoverable

▪ Key Measurement (R-NC-KM)—without such a sensor the PEMS fails (requirement not met)

▪ Ancillary Measurement (R-NC-AM)—helps improving accuracy but PEMS can work even without this input

• Relevant for PEMS and Correlated to some other input (R-C)—fit both for PEMS and for SV

▪ Recoverable (REC) if it can be properly recovered from other sensors

▪ Not Recoverable otherwise (N-Rec) if the estimated surrogate is not good enough for PEMS purposes

− Key Measurement (N-Rec-KM) if without this sensor PEMS fails (backup needed)

− Ancillary Measurement (N-Rec-AM) if without this sensor PEMS performs within specs

In order to create robust PEMS, actions should be tailored on specific sensor category. While inputs coming from not recoverable sensors identified as key measurements (orange-dashed boxes in Figure 3) must be protected by HW redundancies, and inputs coming from not recoverable sensors identified as ancillary measurements (green-dotted boxes in Figure 3) can be discharged—and PEMS operating through reduced order models—very often the majority of sensors belongs to the Recoverable class, meaning that there is "enough" analytical redundancies among the input variables to grant the possibility to establish SW recovery strategies in case of failures.

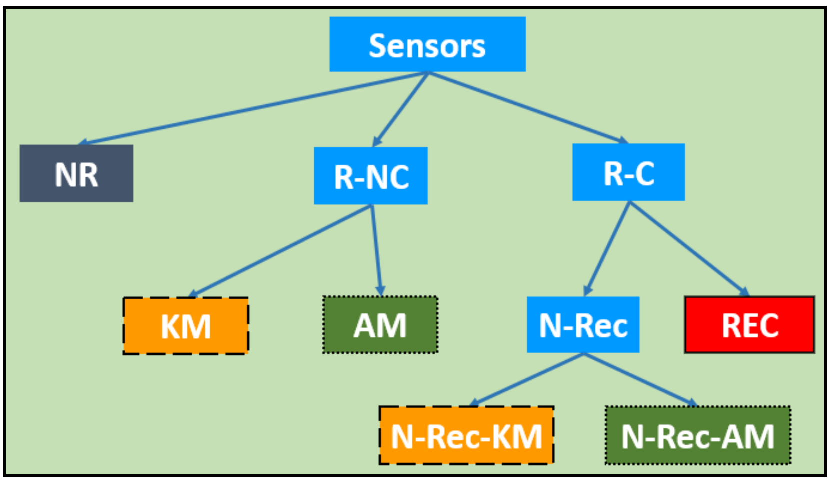

Figure 3. Sensor Taxonomy

3.2 Sensor Fault-Resilient PEMS

Sensor Validation and Reconciliation aims at detecting sensor failures based on multiple (redundant) sensor measurements and the constraint relations among them (Chow & Willsky, 1984; Lee & Park, 2005). Its principle is generally based on consistency checking between the observed behavior of the process provided by the sensors and the expected behavior given by a mathematical representation of the process.

The basic concept of R-PEMS is to add a number of computational functions to the standard PEMS and perform on-line checks in order to make the system as much robust as possible through advanced machine learning algorithms (see Table 1).

Table 1. The Three Steps of R-PEMS

	Function	Related Action
Fault Detection	Detect if any of the PEMS input is affected by any fault	Put PEMS output in Bad Quality status
Fault Isolation	Identify which is the sensor responsible for the problem	Alert Maintenance. Report sensor "i" as faulty
Sensor Reconciliation	Try to reconstruct a reliable value for sensor "i", based on the other (N-1) readings	Replace (temporarily) HW sensor reading with reconstructed value

Figure 4 depicts schematically the concept. Three software layers are added to the on-line implementation, each devoted to one specific function and producing a specific action.

In the following the three steps will be briefly discussed. All the following discussions are based on the assumption of a single sensor fault hypothesis. That is, we are assuming that persistent sensor failure are so sporadic that the probability that two sensors fails contemporaneously or in a short time difference is negligible. Therefore, our approach relies upon the fact that we can detect, isolate and reconcile a sensor before another one fails.

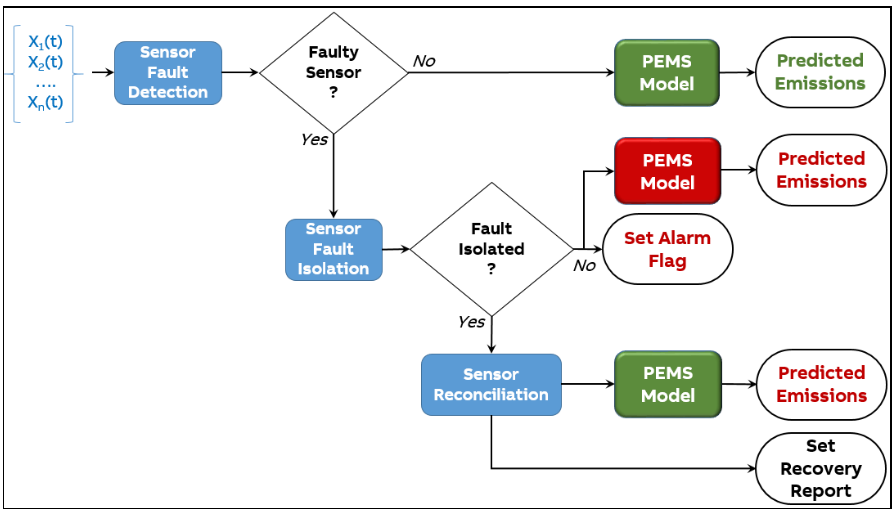

Figure 4. R-PEMS Concept

Step 1: Sensor Fault Detection (SFD)

The input of the SFD algorithm is the process vector $\boldsymbol{x}_t$ containing the values of the N process tags at time t, and the output of the SFD is a binary variable SFD_t which states whether the point $\boldsymbol{x}_t$ is faulty

($SFD_t = 1$) or not faulty ($SFD_t = 0$). In case the point $\boldsymbol{x}_t$ is not faulty, $\boldsymbol{x}_t$ is fed to the existing PEMS and the emission at time t corresponding to the process vector $\boldsymbol{x}_t$ is evaluated. If $SFD_t = 1$, we keep observing the evolution of SFD_τ for $t < \tau < t + t_{SFD}$, where t_{SFD} is an SFD activation length. If $SFD_\tau = 1$ for $t < \tau < t + t_{SFD}$, we declare that the process vector has one or more persistent faults, and we activate the SFI block.

Step 2: Sensor Fault Isolation (SFI)

Once the fault has been declared, the SFI aims at isolating the faulty sensor. The input of the SFI is a set of consecutive process vectors. Clearly, our rationale is to feed the SFI with points having a sensor fault. Therefore, we want to make sure that only the points after the sensor fault detection are fed to the SFI block are fed to the SFI block. The output of the SFI is an index i_{SFI} of the faulty sensor.

Step 3: Sensor Reconciliation (SR)

Once the faulty sensor index i_{SFI} has been identified, the last step of the R-PEMS pre-processing is the data reconciliation that is the substitution of the measurement of the faulty sensor with a reconciled value which is calculated from the non-faulty sensors. Several Machine Learning algorithms may be applied for such tasks, trying to exploit the wealth of information which is hidden in historical process data. Next section will provide some further details on the algorithms used for reaching the scope of building fault resilient PEMS.

4. Robust PEMS Algorithms

4.1 Operating Envelope Estimation

As mentioned in the previous sections, the idea of data validation and reconciliation is to detect sensor failures based on multiple (redundant) sensor measurements and the constraint relations among them. The mathematical representation of the process, that intrinsically contains the analytical redundancy of the system, can be derived from first principle equations, i.e., mass/energy conservation laws, Ohm/Kirchoff laws in electric networks, etc. However, in many situations, physics-based equations may be difficult to obtain due to the complexity of the process and high process dimensionality and therefore a data-driven approach has to be preferred.

In this section, we advocate a method for learning an approximate form of balance equations. It is worth pointing out here that we do not assume any process knowledge for performing this task. Any prior knowledge can be embedded in this formalism to improve the learning process.

To be specific, given the training data drawn from the process tags time series, we aim at discovering the operating envelope of the process, that is, the regions where the process tags must lie during normal operations. For the sake of simplicity, if we assume the process to be defined by $N = 3$ process tags, an example of the training data drawn from the process tags time series is depicted in Figure 5.

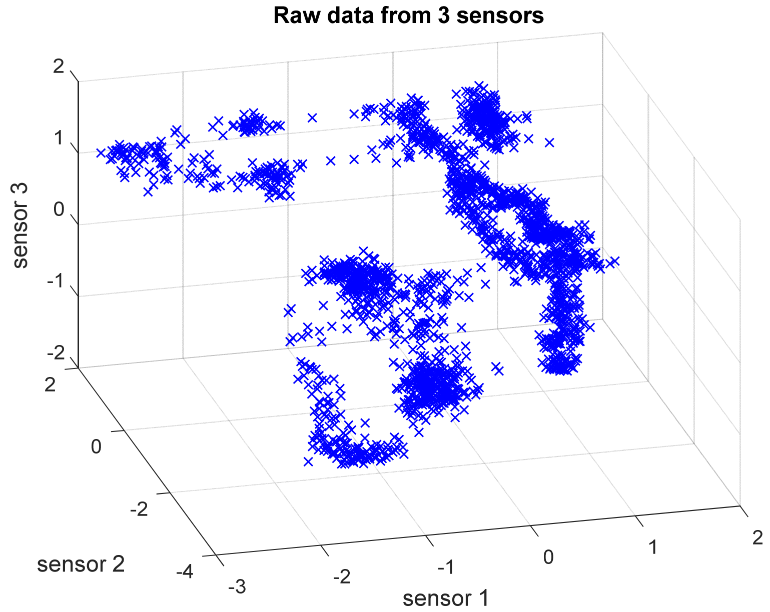

Figure 5. Training Set Example for N=3 Sensors

Given these training data, machine learning algorithms for density estimation can provide an estimate of the operating envelope. Figure 6 depicts the operating envelope of the training data an example of Figure 5. These algorithms include Nearest Neighbors, Neural Networks, Gaussian Mixture models, Locally Weighted Regressions (LWR), to name a few (Markou & Singh, 2003).

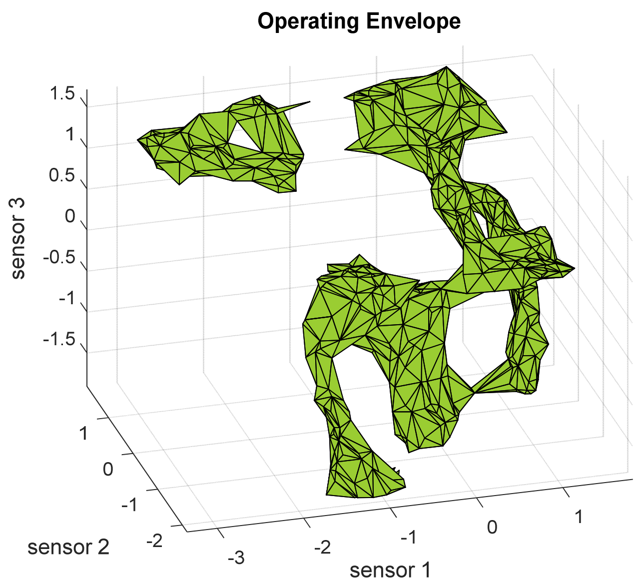

Figure 6. Operating Envelope of the Training Data

LWR have been adopted in the final implementation of the algorithm. Although a complete description of the algorithm is outside the scope of this work—the interested reader can refer to (Angelosante et al, 2018) for a more rigorous treatment of the sensor fault isolation, detection and reconciliation problem—in the next section a short introduction to LWR is given.

4.2 Locally Weighted Regression (LWR)

LWR is a nonparametric regression method wherein regression of a data input is performed by a linear (or quadratic) regression around a local subset of the training set around the point of interest (Cleveland, 1979).

An important parameter to be designed (typically via cross validation) is the bandwidth τ, that is, the amount of data around the point of interest used to perform the linear regression. A large bandwidth results in including a large portion of the whole training set in the linear regression, while a smaller bandwith results in a much more local weighted regression around the point of interest x^t indicated in Figure 7.

The major advantage of LWR is the insensitivity to training set imbalance, that is, the non-uniformity of the sample in the operating envelope. While classical parametric regression methods are biased by training set imbalance and require pre-processing, LWR are natively insensitive to this effect. Additionally LWR, as a memory-based method, entails favourable features which reduce the amount of engineering work and skills required to properly configure them. For these reasons, they have been chosen for the final implementation of the Robust PEMS.

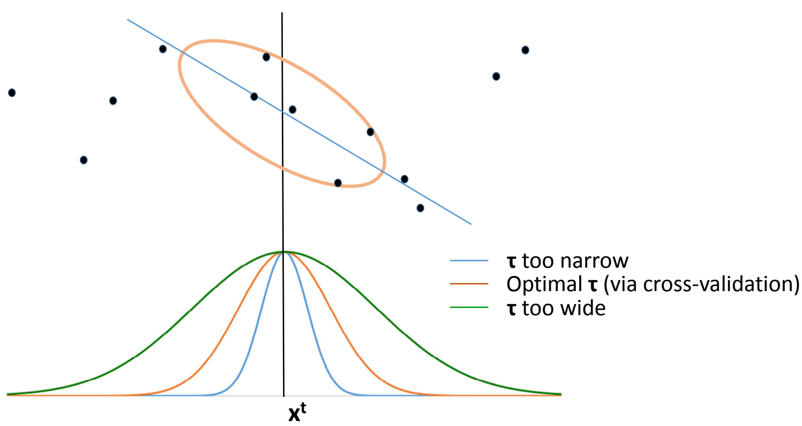

Figure 7. LWR Bandwidth Explaination

4.3 Practical Implementation of SFD

Once the operating envelope has been estimated, SFD can be performed. In fact, abnormal situations that occur due to sensor faults, induce changes in sensor measurements that can drive the process data outside the operating envelope.

The crux of the SFD is to monitor the distance between the process sensor data x and the operating envelope: a fault is declared in case the process data lies outside the operating envelope.

An example of a detected faulty case is depicted in Figure 8, which shows the operating envelope (green volume) and an anomalous process data point (red circle), which lies outside the estimated operating envelope. In such a case, the SFD will detect a sensor fault.

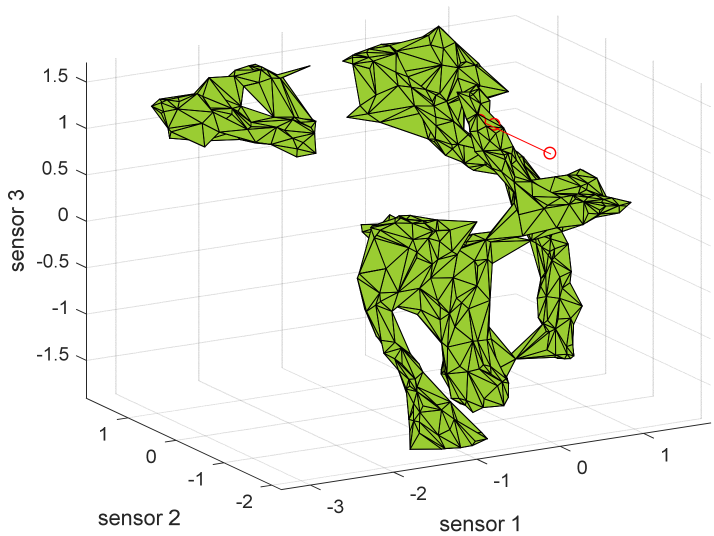

Figure 8. Operating Envelope and an Anomalous Process Data Point

4.4 Practical Implementation of SFI

The goal of SFI is to identify which sensor out of N has failed. In principle, the fault can occour in any of the N sensors. Isolating faulty sensors amounts to a combinatorial problem as all the possible combinations of sensors must be checked for consistency to find the anomalous sensors. Assuming that only one sensor at time can fail simplifies the search.

Referring to the example in Figure 8, it is clear that the faulty sensor is sensor 2. In fact, if we eliminate sensor 2 from the problem, so creating operating envelope without sensor 2, the process tag data will lie inside this reduced operating envelope (Figure 9). This procedure also shows a pratical manner for identifying sensor anomalies, that is, if the fault is caused by the i-th sensor, it is expected that all but the i-th projection are still outside the projected operating envelope.

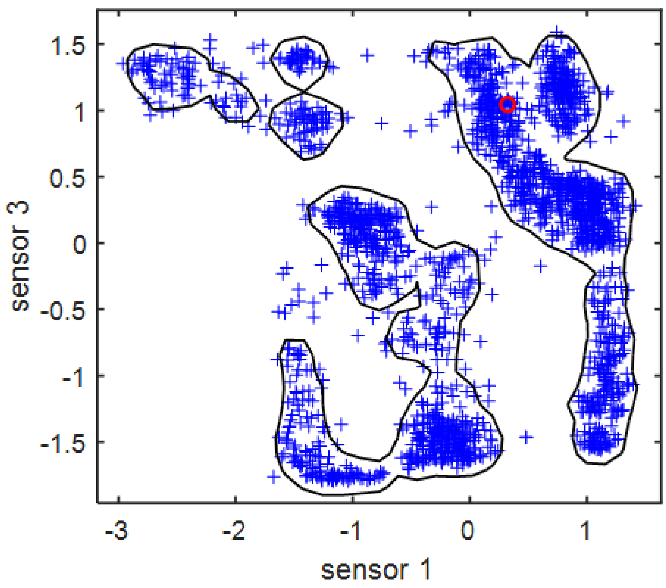

Figure 9. Projected Process Tag Vector (Red Circle) Vs Projected Operating Envelope (Black Line)

4.5 Practical Implementation of the Reconciliation Algorithm

The output of the SFI will return the index i_{SFI} of the most likely faulty sensor. The goal of the recovery algorithm is to reconstruct this sensor reading, using the rest of the healthy data. For this task, we advocate the usage of non-linear regression methods (e.g., LWR) that, using training data, create a soft-sensor of the faulty tag, using the rest of the healthy tags, as sketched in Figure 10.

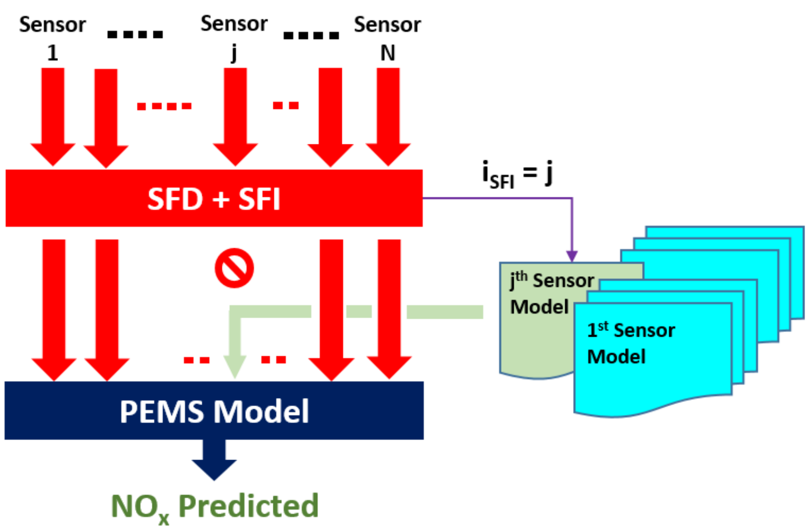

Figure 10. R-PEMS Schematics

5. Tests and Results

In this section, we empirically verify the performances of the SFD/SFI algorithms that have been introduced in the previous sections. We are considering a data set collected during a real PEMS project in Europe. The raw dataset consists in more than 100 sensor data. Out of these tags, only 29 are used for the development of PEMS application. In this document, we are focusing only in one particular gas for the sake of brevity, the O_2, and we will be dealing only with the 8 tags in input to the O_2 PEMS, that is, N = 8.

The dataset consists of 4160 time ordered observations from 8 different sensors tags. The dataset is split into a training set, validation set and testing set. The training, validation and testing datasets will have the size of 2774, 693 and 693 observations, respectively. The training set is used to develop both the PEMS and the SFD/SFI and Recovery algorithm. The Validation set is used for fine tuning the models. The test set is used for performance evaluation purpose.

5.1 Sensor Fault Simulation Setting

Different types of faults pose different requirements on the design of the SFD/SFI algorithms that should be flexible enough to handle various types of faults and handle them accordingly. Calibration errors are probably the key source of many faults manifesting themselves as a bias or a drift in the sensor readings. Moreover, faults can be classified based upon the temporal persistence as permanent faults that are continuous and stable in time; intermittent faults and transient faults.

In this study we focus our attention on the permanent faults characterized by continuity of the fault signature occurrence leading to a mismatch that we can observe over time. Within the class of permanent faults we restrict our study to the following fault types:

- bias,
- drift,
- freeze.

While large biases and drifts may be discovered from simple range check, the sensor freeze will be never discovered by a range check. Therefore, for the sake of brevity, we highlight the performance of the proposed SFD/SFI and Recovery algorithm for sensor freeze. In this paper, we want to analyze how "general purpose" SFD/SFI methods would perform on sensor freeze. In this respect, the proposed algorithm has shown also good performance against sensor biases and drifts.

In our sensor fault simulation setting, we retain the unchanged training set used for the evaluation of the PEMS and the SFD, SFI, and Recovery algorithm and we simulate freeze in the test set. This assumption is not limited since any fault or anomalous data in the training set can be manually or automatically removed by the PEMS designer.

We have simulated a deterministic sensor freeze in each of the sensor. From the original test set, we have created 8 test sets wherein in the i-th test set, the i-th sensor is frozen at its value at time 300 (we have performed test for various freezing time but they will not be reported in this paper).

Figure 11 shows the case Sensor 1 freezes at time 300 vs the original sensor value. The other input sensor to the PEMS, i.e., Sensor 2 - Sensor 8 are unchanged.

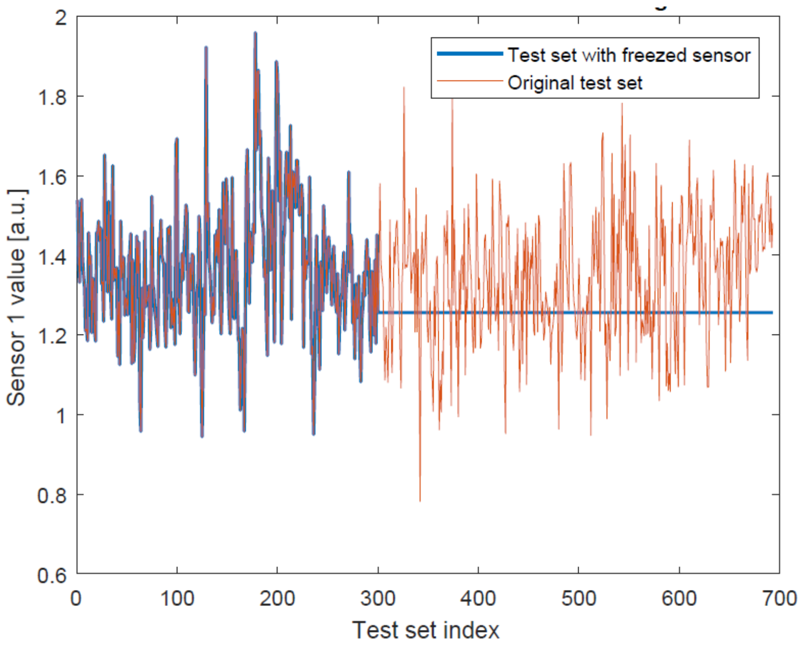

Figure 11. Sensor Tag 1 Freeze vs Original

5.2 Sensor Fault Simulation Results

Figures 12 and 13 aim at showing the performance improvement of the R-PEMS with reconcilied data vs the original PEMS in presence of sensor freeze. More specifically, Figure 11 shows the true O_2 concentration in the test set (magenta line), the predicted PEMS O_2 concentration in case of absence of freeze (blue line), the predicted PEMS O_2 concentration in case of freeze (red line), and the predicted Robust PEMS O_2 concentration in case of freeze (yellow line).

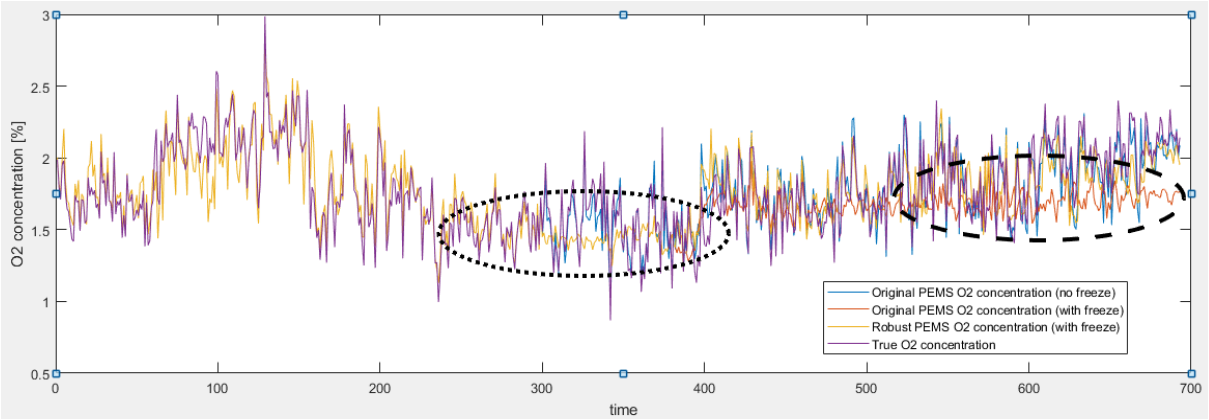

Figure 12. PEMS vs R-PEMS Performance Comparison

The dotted ellipses denotes a point around freeze time 300 wherein there is a significant discrepancy between the true O_2 concentration and the performance of the original PEMS and R-PEMS with freeze. At this point, the SFD/SFI has not decided about the fault. The dash ellipses show a significant discrepancy between the true O_2 concentration and the performance of the original PEMS, but the R-PEMS output is in line with the target performance. To show this, Figure 12 depicts the absolute relative error of the original PEMS with no freeze, the original PEMS with freeze, and the R-PEMS with freeze. Out of the 693 test points, we have created 9 block of 70 consucutive point and the last of 63 points, and we have calculated the relative absolute error (i.e., the relative accuracy).

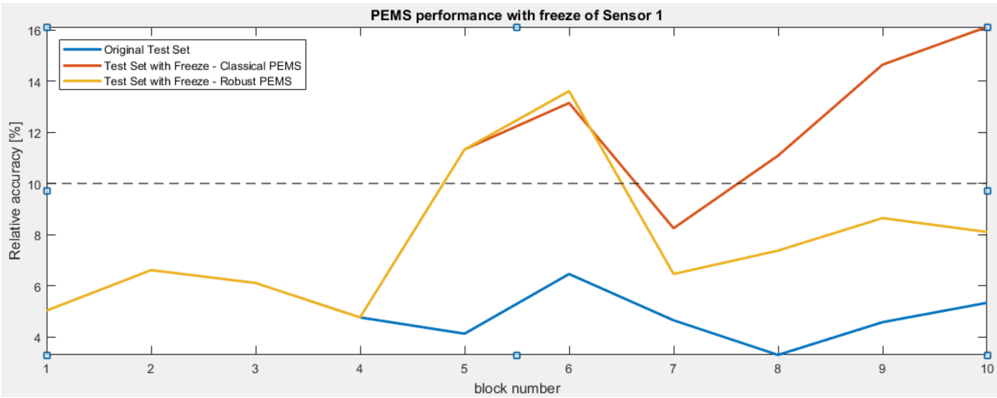

Figure 13. PEMS vs R-PEMS Relative Accuracy Comparison

As already pointed out, the target is to keep this relative accuracy within 10% limit. While the original PEMS with no fault is well below this limit, when Sensor 1 freezes, the relative accuracy of the original PEMS exceed the 10% limit. On the other hand, the R-PEMS, after a transient phase required for detecting the fault, recovers the performance to an acceptable limit.

6. Conclusions

Predictive Emission Monitoring Systems have shown the capability to provide a valuable alternative to traditional HW-based analysers, offering comparable level of performances and reliability at reduced life-cycle cost and with additional benefits.

Such paybacks are making plant owners more and more aware of PEMS potential and are largely expanding the interest around this technology.

However some additional features are expected to be required to remove any residual concern from Authorities which would jeopardize their complete acceptance. Other than a problem in the integrity of the model—which can be assessed on a regular basis—PEMS performances may degrade only for input values reliability issues.

Focus of this paper has been to introduce a possible systematic approach to sensor validation for environmental monitoring and to outline what a modern software tool should be able to provide. Starting from a detailed classification of sensor types and roles it is possible to classify sensors whose readings can be reconstructed through analytical redundancy from the ones which have little or no correlation with other measured variables. While the latter can be validated only by hardware redundancy, sensor belonging to the first category are usually the largest majority and are the typical target variables used for advanced data-driven algorithms.

The paper shows how Machine Learning techniques can effectively contribute to solve the tasks of Fault Detetction, Fault Isolation and Sensor Reconciliation. Extensive tests on real plant data showed how a careful application of several algorithms is able to properly manage the problem of single-sensor fault, driving back the PEMS prediction inside the acceptance limits and largely increasing the reliability of the system even in presence of a sensor malfunction. Among several alternative algorithms, Local Weighted Regression was selected as the most appropriate because of the smaller amount of engineering requested and the easiness of implementation.

References

Angelosante, D., Guerriero, M., Ciarlo, G., & Bonavita, N. (2018). A sensor fault-resilient framework for Predictive Emission Monitoring Systems. *Proc. Of Fusion 2018*. https://doi.org/10.23919/ICIF.2018.8455445

Ciarlo, G., & Bonavita, N. (2016). Fulfilling Evolving End-Users Expectations for Site-Wide Emission Monitoring: the Role of PEMS. In *Proc. of "CEM 2016 - 12th International Conference and Exhibition on Emission Monitoring* (pp. 18-20). Lisbon, Portugal.

Ciarlo, G., Bonica, E., Bosio, B., & Bonavita, N. (2017). Assessment and Testing of Sensor Validation Algorithms for Environmental Monitoring Applications. *Chemical Engineering Transactions* (Vol. 57).

Chow, E., & Willsky, A. (1984). Analytical redundancy and the design of robust failure detection systems. In *IEEE Transactions on Automatic Control* (Vol. 29, No. 7, pp. 603-614). https://doi.org/10.1109/TAC.1984.1103593

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insight from hundreds of use cases. In *McKinsey Global Institute Discussion Paper*.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(829). https://doi.org/10.1080/01621459.1979.10481038

Einsenmann, T., Bianchin, R., & Triebel, D. (2014). Predictive Emission Monitoring (PEM): Suitability and Application in View of U.S. EPA and European Regulatory Frameworks. *Proc. of "CEM 2014—International Conference and Exhibition on Emission Monitoring*. Istanbul, Turkey.

EPA. (2009). *Performance Specification 16—Specifications and Test Procedures for Predictive Emission Monitoring Systems in Stationary Sources*.

Kadlec, P., Gabrys, B., & Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers and Chemical Engineering*, *33*, 795-814. https://doi.org/10.1016/j.compchemeng.2008.12.012

Lee, S. C., & Park, C. E. (2005). Sensor value validation based on implicit sensor redundancy for reliable operation of power plants. In *IEEE Transactions on Energy Conversion* (Vol. 20, No. 2, pp. 373-380).

Markou, M., & Singh, S. (2003). Novelty detection: A review—Part 1: statistical approaches. *Signal Processing*, *83*(2481).

Netherlands Technical Agreement. (2014). *Guidelines for Predictive Emission Monitoring Systems (PEMS)—Execution and quality assurance*. NTA7379, Delft.

Paira, R., Ganapathiraman, G., & Avery, A. (2017). *Emission monitoring systems global market research study*. Retrieved August, 2017, from https://www.arcweb.com/market-studies/emission-monitoring-systems

Qin, J., Hongyu, Yue., & Ricardo, D. (1997). Self-validating Inferential Sensors with Application to Air Emission Monitoring. *Ind. Eng. Chem. Res.*, *36*(5), 1675-1685. https://doi.org/10.1021/ie960615y

Roth, M., & Lawrence, P. (2010). A cost-effective alternative to continuous emission monitoring systems. *Environmental Science & Engineering Magazine*, *56-57*.

Samdani, G. S. (1994). Software takes on air monitoring. *Chemical Engineering*, *December*, 30-33.

Arioni, E., Bonavita, N., & Paco, M. (2013, October). Keeping an eye on emissions. In *Hydrocarbon Engineering* (pp. 43-47).