CHEMICAL ENGINEERING TRANSACTIONS
VOL. 57, 2017

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Sauro Pierucci, Jiří Jaromír Klemeš, Laura Piazza, Serafim Bakalis

ISBN 978-88-95608- 48-8; **ISSN** 2283-9216

DOI: 10.3303/CET1757056

Assessment and Testing of Sensor Validation Algorithms for Environmental Monitoring Applications

Gregorio Ciarlo[a], Elisa Bonica[b], Barbara Bosio[*b], Nunzio Bonavita[a]

[a]ABB S.p.A., Via Albareto 35, 16153 Genoa, Italy
[b]Dept. of Civil, Chemical and Environmental Engineering, University of Genova, Via Opera Pia 15, 16145 Genoa, Italy
barbara.bosio@unige.it

The increasing attention devoted to air quality by legislative, scientific, industrial and public sectors has led to the development of different control strategies for the emission level monitoring.
In this scenario, Predictive Emission Monitoring System (PEMS) is able to predict emission concentrations thanks to empirical or first principles models fed by real-time process data provided by measurement sensors. It follows that PEMS consistency (and, crucially, its acceptance from regulations-enforcing agencies) strictly depends on input accuracy and that reliable Sensor Validation (SV) strategies are fundamental.
In this work, the capability of two different SV techniques, Feed Forward Neural Networks and Locally Weighted Regression, is tested exploiting a commercial software package (ABB's IMP) on actual field data from a fluid catalytic cracking unit. The results showed that both techniques are suitable as complement to PEMS applications, but Locally Weighted Regression results are preferable for performance, economic and operating reasons.

1. Introduction

Acquiring proper, reliable and timely information about air emission levels from process and power plants is crucial in order to maintain an appropriate air quality and to guarantee local communities. New regulations have been issued or are currently under development in order to assure that adequate control actions are in place (Ribeiro and Kruglianskas, 2015, Saarinen, 2003, Di Natale et al., 2013).
In principle, there are a number of different control strategies (e.g. continuous, periodic, campaign monitoring) able to drive and to keep emissions inside the law-enforced limits, but, as a matter of fact, the most efficient, reliable and applied one is the Continuous Emission Monitoring System (CEMS). In CEMS a continuous stream of data is acquired by rapid-response analyzers, to be properly processed, displayed in real-time and stored for future evaluation (Arioni et al., 2013).
As an alternative, PEMS technology is able to estimate pollutant concentrations at the stack by means of advanced empirical or first principles models on the ground of other process data (e.g. flow rate, pressure, temperature). Recognized and accepted by a growing number of national regulations (EPA, 2009; Netherlands Technical Agreement, 2014), PEMS has gained interest in industry because many applications proved that software systems provide accuracy close to that of hardware-based CEMS (Eisenmann et al., 2014) and 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;
- the possibility to be used for process optimization purposes.

Additionally, PEMS lifecycle costs are a fraction of hardware-based CEMS, mainly because of extremely low operating costs (Roth and Lawrence, 2010). PEMS can be used autonomously for qualified emissions sources, as a replacement for CEMS, or as a backup to conventional CEMS, in order to extend its service factor and provide maintenance indications (Bonavita and Ciarlo, 2014). One of the main concerns unique to PEMS is related to process sensor (i.e model input) quality: once a reliable model has been built, validated and deployed, its performances still depend upon the quality of the input variables used by the model itself. This means that a lot of care has to be applied in order to routinely check if the all the process sensors used

as model inputs are performing as they should. If not, some level of alarms must be triggered to inform plant personnel and to warn about PEMS prediction quality.
In this scenario, even if present regulations allow also manual inspection and checking, there is an interest in developing and deploying robust automatic Sensor Validation (SV) applications, able to not only identify possible malfunctions, but also provide reconstructed values. These could be automatically selected as emergency and temporary replacement of the affected sensor, without compromising PEMS operation and, eventually, environmental monitoring.
Aim of this work is the analysis of two different SV techniques, Artificial Neural Networks and Locally Weighted Regression, in order to assess their performance and evaluate their viability as complement to PEMS applications.

2. Sensor Validation for Predictive Emission Monitoring System

In a complex industrial process, typically the great number of different sensors causes the presence of relevant, although variable, correlations among their readings.
A faulty sensor is usually characterized by the absence of correlation with the remaining sensors; this feature is used in the identification of the faulty sensor when an abnormal operating condition is detected.
Two are the classical approaches for sensor validation: hardware redundancy and analytical redundancy.
The hardware redundancy approach increases the reliability of the system, but it is not always feasible for economical or availability reasons. The analytical redundancy approach consists in checking the consistency between the experimental measurements and the estimated values obtained through the relationships existing between the various process variables. In the latter case, the main model used for emission estimations can be coupled with a sensor validation model, according to the logic scheme in Figure 1. The estimation is compared with the physical sensor reading: if the difference exceeds a pre-defined threshold, the estimated value will replace the instrument reading as PEMS input, while an alarm will be generated and sent to the PEMS display, in order to alert maintenance technicians of a possible problem in the instrumentation.
With this feature, it is possible to maintain PEMS accuracy at the best level even when there is a fault in the field instrumentation.
Analytical redundancy methods use explicit input–output models usually built from system identification algorithms. However, in some cases this explicit formulation of the redundancy relationships may be difficult to obtain, because of process complexity and high dimensionality. As an alternative, implicit modelling approaches, which are data-driven techniques, are particularly suited for revealing relationships among the plant variables, avoiding the burden of deriving explicit mathematical equations.
During the whole development of the present work, the analytical approach based on data-driven techniques has been applied, exploiting the features and functionalities of a commercial software solution.

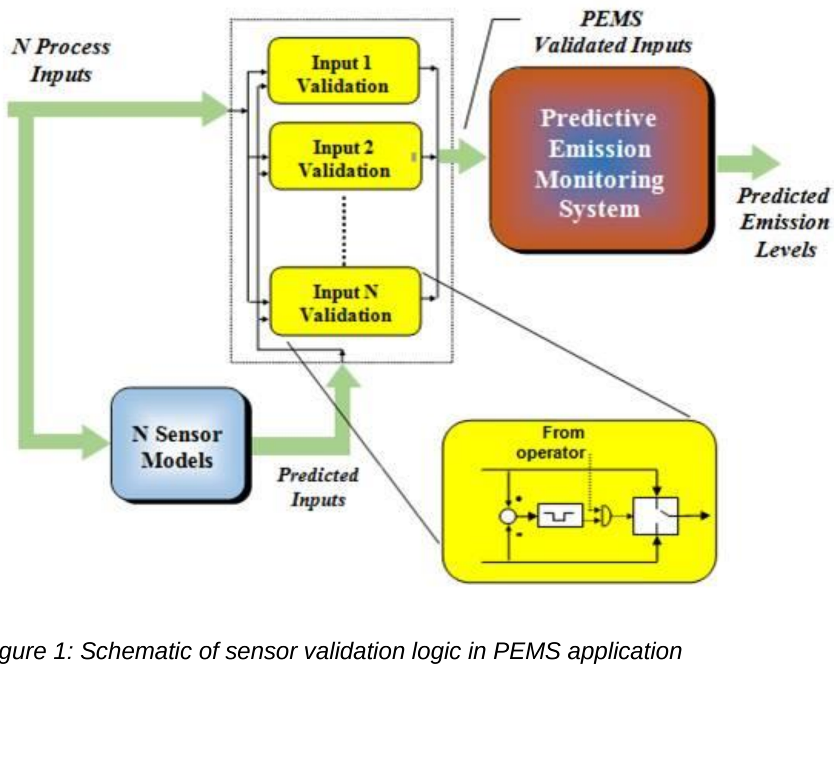

Figure 1: Schematic of sensor validation logic in PEMS application

Concerning sensor validation algorithms, in literature, several different techniques have been proposed: Principal Component Analysis, Bayesian Network, Partial Least Square (PLS), Gray models, Neural Networks (NN), Locally Weighted Regression (LWG). Aiming at identifying robust yet industrially applicable methods, authors' attention focused on the last two. In fact, NN have become a usual application in many areas of engineering (Santos et al., 2013) and in particular Feed Forward Neural Networks (FFNN) are widely used for model-building purposes in process control applications, resulting to be one of the main algorithms at the core of PEMS applications (Marlin, 2015). Using the same technology also for SV purposes would bring engineering benefits in terms of familiarity, easiness of implementation and maintenance. LWR model, on the contrary, is less popular but brings a specific advantage in terms of computational efficiency and speed of implementation: while NN are essentially Multi Input Single Output (MISO) technologies, LWR is a Multi Input Multi Output (MIMO) technique (Ting et al., 2010). This means that while a FFNN model allows predicting only one variable per model, LWR allows building a unique model able to check and validate each input variable.

3. The software tool

All the experimental work has been developed with the aid of the features and functionalities of the commercial package Inferential Modeling Platform (IMP) Model Builder, an ABB software solution specifically tailored to the development of data-driven applications (ABB S.p.A., 2006). Exploiting such an advanced and robust SW platform was crucial since it allowed automating and containing most of "error-prone" operations, which would have taken lot of time if manually performed.
During the initial data pre-processing and processing, the authors exploited the embedded functionalities of IMP in order to perform in a single, user-friendly environment all the tasks aimed at properly preparing the data-set for the model-building phase. In fact, IMP allowed easily importing and merging the initial process dataset and the emission data, and to perform a preliminary analysis of the behaviour of each variable in order to quickly eliminate not-properly working sensors and to easily detect the most relevant ones for prediction purposes. The model building as well as the SV sections of IMP provided all the elements to build and test models, using pre-compiled algorithms: the authors have been able to test several different techniques (from FFNN and NN to genetic algorithms, Multi Linear Regression and PLS) and identify the top performing one. In addition, IMP makes available the facilities to quickly compute model KPIs in order to compare different model performances and find out the most performing one.

4. The case study

The process unit selected to evaluate SV techniques applied in environmental applications was a Fluid Catalytic Cracking (FCC) unit of a major European oil refinery. FCC Unit is complemented with a SO_2 absorption unit (referred as Abs. unit herein) designed to abate the concentration of sulfur compounds released in the atmosphere (see Figure 2). Both the units have their own emission source, respectively identified as "FCC stack" and "Abs. Stack". In order to identify a proper subset of process variables to be analyzed with SV algorithms, the authors have developed predictive models for NO_x and O_2 emissions from the two stacks. After the identification of the most correlated (both on a statistical and process-related perspective) parameters, several modeling techniques have been applied: genetic algorithms, neural networks, multilinear regressions, etc. All the above operations have been made exploiting the features and the capabilities provided by the ABB proprietary software Inferential Modeling Platform (IMP).

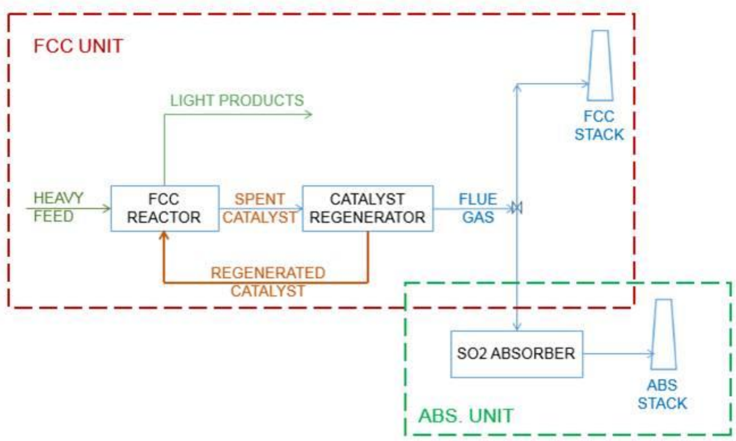

Figure 2: Layout of the units whose data were used for the test of the two SV algorithms.

Table 1: PEMS models results - KPI comparison

KPI	FCC STACK	FCC STACK	ABS STACK	ABS STACK
	O_2	NO_x	O_2	NO_x
R^2	86.20%	84.85%	86.30%	84.70%
|E|	0.053 % Vol.	6.97 mg/m3	0.085 % Vol	6.77 mg/m3
$E_{\%}$	3.27%	2.06%	5.22%	2.40%

The PEMS models performances are briefly summarized in Table 1, where:
- R^2 is the coefficient of determination, i.e. a statistical index which provides a measure of how good is the model in replicating observed values (i.e. how much of the data variance is explained by the model);
- |E| is the average discrepancy between the predicted and the measured value;
- $E_{\%}$ is the average percentage error, obtained dividing the absolute discrepancy at each record by the related measured value.

The table above shows that FFNN technique provides very accurate predictions with a relative error around 2-5% where for such applications the requirements is for an error below 10%.

5. Results and discussion

The tests dedicated to SV algorithms were performed reconstructing the main meaningful input variables of the reference plant. In this section, a summary of the results obtained will be presented referring, for assuring generalities, to the analysis on two very different variables: the content of oxygen at the outlet of the CO boiler and the flow rate of the sour water in the absorber unit.
Figure 3 shows the results obtained on the O_2 analyzer values with the FFNN and LWR approaches, providing a visual indication about the good prediction capabilities. The above results have been obtained on a “fresh” set of data (i.e. data that were not used for model building purpose, a shrewdness needed to avoid producing overfitting and underperforming models, so as in Bonavita et al., 2003).
An objective measure of the prediction capabilities of the two algorithms is given by the three usual KPIs reported in Table 2, while Figure 4 shows an analysis of error distribution.
The relative error proves an additional indication of the prediction quality: as mentioned above this is largely below the 10% value.
The two algorithms provide similar results and performances, allowing an accurate and effective prediction of the target variable. In particular, the FFNN is able to estimate the oxygen concentration with a relative error within the 10% threshold for around 85% of the samples; LWR grants same performances for over 93% of the samples.

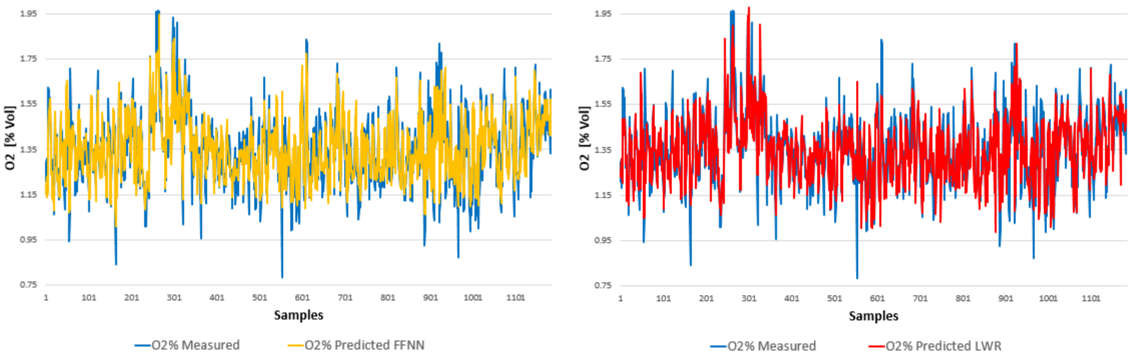

Figure 3 – Reconstruction of the O_2 analyzer values with FFNN (left) and LWR (right).

Table 2: O_2 analyzer reconstruction - KPI comparison

KPI	FFNN	FWR
R^2	80.21%	77.58%
|E|	0.056 % Vol.	0.052 % Vol.
$E_{\%}$	4.18%	3.87%

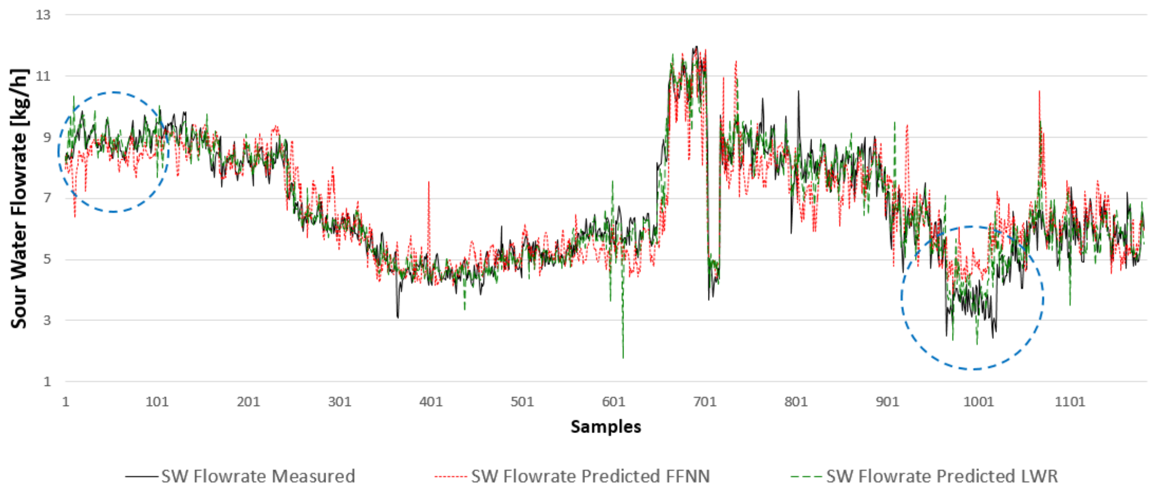

Figure 4: Reconstruction of the sour water flowrate through FFNN and LWR

Table 3: Sour water flowmeter reconstruction: KPI comparison

KPI	FFNN	FWR
R^2	77.17%	91.04%
$\|E\|$	0.67 kg/h	0.38 kg/h
$E_{\%}$	11.30%	6.48%

The same test and analysis have been performed also on the sour water flow rate, Figure 4 provides a comparison between the models obtained with FFNN and LWR.
Also in this case, models built with both technologies are quite well aligned with the measurement from the physical device; however in some specific regions (marked by the dotted circle) the LWR model is more accurate and closer to measured variable. This consideration is confirmed by the analysis of the relevant indicators (Table 3). Although LWR algorithm provides significantly better results, if a 20% level is considered (which can be considered as the target threshold for such kind of applications), NN estimations are within these limits for more than 85% of the samples, with more than 61% of the records within the 10% threshold (against 83% for LWR).

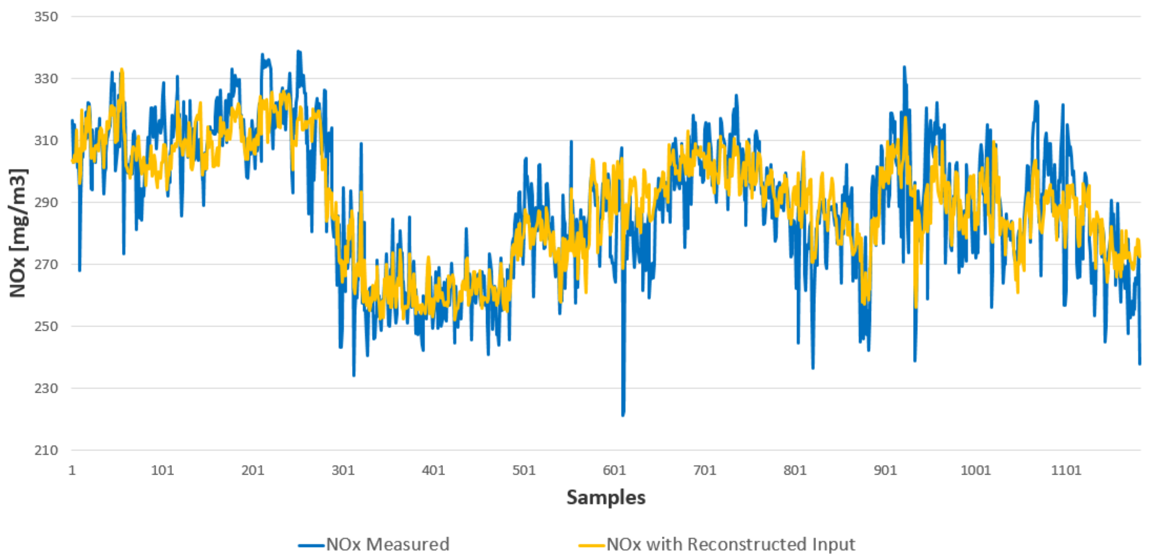

Figure 5 – NO_x predictions with reconstructed input values

As a final check on the LWR performance, the reconstructed values (for the two above discussed variables as well as for additional ones obtained with similar accuracy results) have been used as input variables for the original emissions models: this allowed simulating a complete PEMS package (i.e. PEMS models and SV models), under the worst possible conditions: all the sensor readings substituted by reconstructed values.
As an example, the following picture reports the results obtained for the NO_x emissions at the Absorption unit.
Employing reconstructed model inputs determines just a limited reduction in model accuracy compared with the same model fed with physical sensor readings, with the related error increasing from 2.4% to 2.85% of the NO_x measurement. Similar results have been obtained also for the oxygen emissions where the percentage error moved from the original 5.2% to 6.9%, still well within the tolerance limits for PEMS models.

6. Conclusions

Feed Forward Neural Networks and Locally Weighted Regression techniques have been tested in the framework of Sensor Validation aimed to reconstruct process data which, in case of sensor failure, can be used by PEMS as an emergency substitution of the affected input data.
While both FFNN and LWR provide acceptable performances for SV implementation, the latter, together with the higher accuracy provides tangible advantages in terms of ease of use and practical deployment.
The most relevant advantage of LWR technique relies in the algorithms itself: LWR, requiring a single MIMO model, results much more efficient both in development effort and in real-time execution requirements (notwithstanding a slightly higher calculation complexity).
Concluding the assessment on LWR performances, a final simulation has been performed: values reconstructed by means of these techniques have been fed to the models developed for the estimation of the emissions in order to verify model behaviour.
The final simulation in the “worst-case” scenario further proved LWR viability also in extremely challenging conditions. Even with all inputs reconstructed, average discrepancies in PEMS predictions were well below 10%, the most stringent requirement in actual legislations.

References

ABB S.p.A., 2006. Inferential Modeling Platform 3.1 Model Builder for Sensor Validation, User Manual.

Arioni E., Bonavita N., Paco M., 2013, Keeping an eye on emissions, Hydrocarbon Engineering, 43-47.

Bonavita N., Ciarlo G., 2014, Inferential Sensors for Emission Monitoring: an Industrial Perspective, Frontiers in Environmental Engineering, 3, 21-28.

Bonavita N., Martini R., Matsko T., 2003, Improvement in the Performance of Online Control Applications via Enhanced Modeling Techniques, Proc. of ERTC Computing 2003, Milan, Italy.

Di Natale F., Carotenuto C., D’Addio L., Lancia A., Antesb T., Szudygab M., Jaworekc A., Gregoryd D., Jacksone M., Volpef P., Belecag R., Manivannang N., Abbodg M., Balachandrang W., 2013, New Technologies for Marine Diesel Engine Emission Control, Chemical Engineering Transactions, 32, 361-366, DOI: 10.3303/CET1332061.

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 – Internation 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, Rule 74 FR 12575, Washington.

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

Marlin T., 2015, Process Control: Designing Processes and Control Systems for Dynamic Performance, 2nd ed. McGraw-Hill.

Ribeiro F.M., Kruglianskas I., 2015, Principles of environmental regulatory quality: a synthesis from literature review. Journal of Cleaner Production, 96, 58-76.

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

Saarinen K., 2003, Monitoring total emissions from industrial installations, Environmental Science & Policy, 6, 4, 367–376.

Santos R. R. C., Santos B.F., Fileti A. M. F., Silva F.V., Zemp R. J., 2013, Application of Artificial Neural Networks in an Experimental Batch Reactor of Styrene Polymerization for Predictive Model Development, Chemical Engineering Transactions, 32, 1399-1404, DOI: 10.3303/CET1332234.

Ting J. A., Vijayakumar S., Schaal S., 2010, Locally weighted regression for control, in: Sammut, C., Webb, G. (Eds.), Encyclopedia of Machine Learning, Springer.