A Novel Method for Sensor Data Validation based on the analysis of Wavelet Transform Scalograms

Francesco Cannarile[1,2], Piero Baraldi[1], Pierluigi Colombo[1] and Enrico Zio[1,3]

[1]*Energy Department, Politecnico di Milano, Via La Masa 34/3, 20156 Milano, Italy*
francesco.cannarile@polimi.it

[2]*Aramis Srl, Via Pergolesi 5, 20124, Milano, Italy*

[3]*Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supelec, France*

ABSTRACT

Sensor data validation has become an important issue in the operation and control of energy production plants. An undetected sensor malfunction may convey inaccurate or misleading information about the actual plant state, possibility leading to unnecessary downtimes and, consequently, large financial losses. The objective of this work is the development of a novel sensor data validation method to promptly detect sensor malfunctions. The proposed method is based on the analysis of data regularity properties, through the joint use of Continuous Wavelet Transform and image analysis techniques. Differently from the typical sensor data validation techniques which detect a sensor malfunction by observing variations in the relationships among measurements provided by different sensors, the proposed method validates the data collected by a given sensor only using historical data collected from the sensor itself. The proposed method is shown able to correctly detect different types and intensities of sensor malfunctions from energy production plants.

1. INTRODUCTION

Modern energy production plants are complex systems, equipped with hundreds of sensors to measure, at relative high frequency, physical parameters, such as pressures, temperatures and flows for operation control and diagnostic purposes. In practice, sensors may malfunction, i.e. they can provide inaccurate readings of the monitored physical parameters. The most common types of sensor malfunctions are: freezing (or constant), noise, spike (or short) and quantization (Sharma et al., 2010) (Tolle et al., 2005). They can lead to the incorrect intervention of plant operators and automatic control systems, causing undesirable consequences, such as unnecessary component downtimes, or even plant shutdowns with associated large financial losses. Thus, the task of promptly detecting the occurrence of a sensor malfunction, which is often referred to as sensor data validation, is of paramount importance. It has been addressed by a variety of methods including Auto Associative Neural Network (AANN) (Hines et al., 1998), Nonlinear Partial Least Squares Modeling (NLPLS) (Rasmussen et al., 2000), Principal Component Analysis (PCA) (Penha & Hines, 2001) (Baraldi et al., 2011), Auto Associative Kernel Regression (Baraldi et al., 2015) (Garvey et al., 2007), and Multivariate State Estimation Technique (MSET) (Gross et al., 1997) (Coble et al., 2012).

A limitation of these approaches is that they only detect the abnormal behavior of the measured signals, which, however, can be due to several causes, such as a sensor malfunction, a process anomaly, a failure of a plant component. The subsequent identification of the cause of the abnormal behavior is typically a time-consuming task, which requires an intervention of the plant personnel or the use of other dedicated diagnostic systems. Furthermore, data validation approaches typically detect the anomalous behavior of a sensor using information provided by other sensors. The basic idea is that a sensor malfunction causes a modification of the functional relationships among the measured signal values. The use of data collected from other sensors may cause difficulties from a practical point of view. For example, when hundreds of signals are monitored in a plant, it is necessary to group them into several subsets, since it has been shown in (Roverso et al., 2007), (Baraldi et al., 2011) that a single model based on all (hundreds) signals is not able to provide satisfactory performances. Although the problem of sensor grouping has been successfully addressed in (Baraldi

et al., 2011) and (Baraldi et al., 2014) by using ensembles of models dedicated to detecting of sensor malfunctions in a specific group of sensors, the proposed solutions still have some practical limitations:

1) the necessity of periodically updating the models and the corresponding signal grouping to take into account possible modifications of the signals relationships (Roverso et al., 2007);
2) the fact that these models are not easily scalable to a fleet of plants (Baraldi et al. 2011), since each plant has its own characteristics and, therefore, it requires a dedicated grouping of the signals.

To overtake these limitations, we aim at developing a completely different approach for detecting sensor malfunctions. The idea is to develop a dedicated data validation model for each sensor, based on historical data collected from the sensor itself in healthy conditions. Since the approach does not consider relationships among different signals, it can be systematically applied to a fleet of plants, without requiring sensor grouping.

The proposed sensor data validation method builds up from the idea that a sensor fault alters the regularity of a signal, i.e., its degree of smoothness. Continuous Wavelet Transforms (CWT) are able to characterize and quantify the local regularity of a signal (Mallat, 2008), and have been employed in many engineering applications. For example, the Lipshitz-exponent, which can be estimated from CWT by using the Wavelet Modulus Maxima (WMM) (Mallat & Hwang, 1992), has been used for bearing faults diagnostics (Li, 2010), machinery health monitoring (Miao et al., 2007) and signal denoising (Mallat & Hwang, 1992). A limitation of WMM is that it is sensible only to signal irregularities, whereas it does not allow detecting types of sensor malfunctions which add regularity to a signal, such as freezing. For this reason, in this work, we propose a novel method based on the use of CWT scalograms, which are two-dimensional images representing the time evolution of the squared magnitude (or power) of the CWT at different frequencies (Mallat, 2008).

The method combines the use of CWT with image analysis techniques for the identification of the similarity among the test data and an archive of historical data. It involves the following steps:i) performing the CWT of the test signal, ii) computing the corresponding scalogram image and iii) comparing this scalogram with those obtained from historical data of the signals collected by the sensor. With respect to the last step, the comparison between scalogram images is performed by defining a proper measure of similarity between images based on a pixel by pixel comparison.

The main contributions of this work are:

- the use of CWT scalogram images to detect sensor malfunctions;
- the development of a method which allows the detection of a sensor malfunction without using data measured by other sensors, is robust to different sensor malfunction types and intensities and able to graphically motivate the reasons of the detection through the use of scalograms.

The performance of the proposed method has been verified with respect to data taken from an energy production plant. Realistic examples of sensor malfunctions have been artificially injected in the data streams and the proposed method has been compared with a literature PCA-based approach from the point of view of the percentage of false and missed alarms.

The remainder of the paper is organized as follows. Section 2 highlights the main issues associated to sensor data validation and provides a description of the most common sensor malfunction types. In Section 3, the problem statement and notation are discussed. In Section 4, some mathematical features of CWT at the basis of the proposed method are discussed. Section 5 provides an in-depth discussion of the proposed method. The application of the proposed method to the case study is shown in Section 6. The methodology limitations and its possible developments are discussed in Section 7. Finally, in Section 8 conclusions are drawn.

2. SENSOR DATA VALIDATION

The objective of this work is the development of a sensor data validation method for online detecting sensor readings deviating from the ground truth values of the monitored physical parameters. Signal deviations can be triggered by a single sensor fault or by the failure of a node with attached several sensors, because of hardware failure or sensor internal malfunction (e.g., losing the connection with the sensor board). According to (Sharma et al., 2010), these types of malfunction are considered as non-functional faults since they only impact the fidelity of the reported data. The different types of sensor malfunctions are typically classified as (Ni et al., 2009) (Sharma et al., 2010):

- *Spike* (or *short*): a sharp change in the measured value between two successive measurements. It produces a single isolated sensor reading with a value that is significantly far from the signal ground truth (Figure 1).
- *Noise*: the variance of the sensor readings increases and the data becomes highly uncorrelated with the true signal values (Figure 2).
- *Freezing (or constant):* the sensor reports a constant value for a large number of successive samples. It may precede and/or follow an unexpected signal jump, with readings that may fall outside the range of the measured phenomenon. Figures 3 and 4 show some examples of freezing without and with jump, respectively.
- *Quantization*: a reduction of the analogue-to-digital resolution conversion. Quantization replaces signal ground truth values with their approximations into a

finite set of discrete levels. In practice, the sensor reading is characterized by intervals with constant values followed by sharp changes (Figure 5).

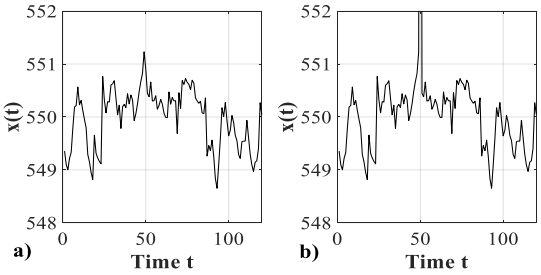

Figure 1. Example of sensor spike. Left: ground truth signal values; right: corresponding readings in case of sensor spike.

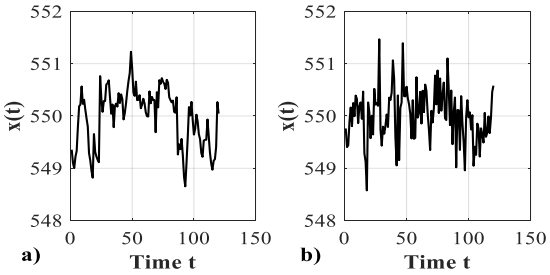

Figure 2. Example of sensor malfunction due to noise. Left: ground truth signal values; right: corresponding readings in case with noise.

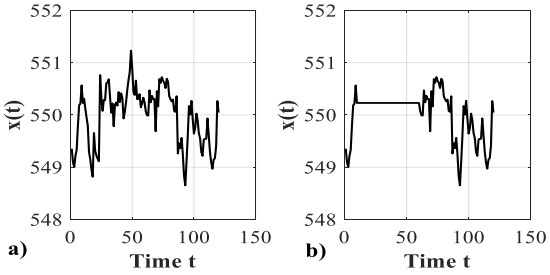

Figure 3. Example of sensor freezing. Left: ground truth signal values; right: corresponding readings in case of sensor freezing.

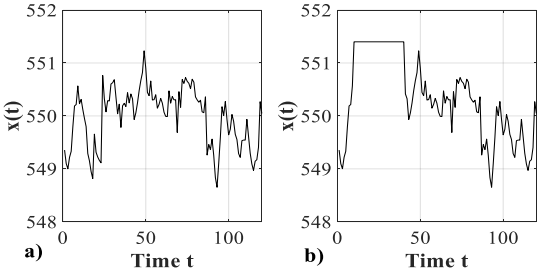

Figure 4. Example of sensor freezing with jump. Left: ground truth signal value; right: corresponding readings in case of sensor freezing with jump

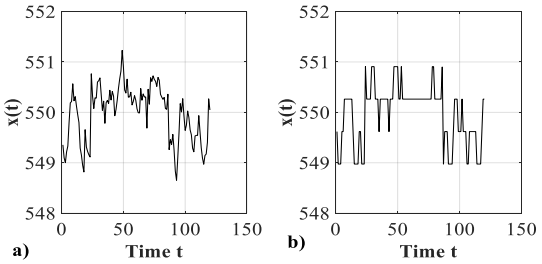

Figure 5. Examples of sensor quantization. Left: ground truth signal value; right: corresponding readings in case of sensor quantization.

3. PROBLEM STATEMENT AND NOTATION

Let $x(t)$ be the measurement of a generic plant sensor at time t. The objective of the present work is to develop a method for promptly detecting the occurrence of sensor malfunctions. We assume that:

i. historical measurements $x(\tau)$, $\tau < t$, taken by the sensor when it was in healthy conditions are available; this assumption requires that the training data are validated in advance by plant experts to guarantee that they have been collected by healthy sensors. This activity is typically performed by considering maintenance reports and by visual inspection of the acquired signals.

ii. the data in $x(\tau)$ are representative of the plant operating conditions, whether normal or anomalous, caused by the degradation and failure of components.

iii. Indeed, in real industrial applications, with sensor data collected for long periods of time (e.g. years), a large spectrum of plant operating conditions is registered, including plant anomalous ones.

The detection of the sensor malfunction is based on the analysis of the last L measurements collected in the time window $x^L(t) = \{x(t-L+1), \dots, x(t)\}$, which is hereafter referred to as test pattern. The historical measurements $x(\tau)$ are organized into S training vectors of length L, containing the measurements in the time window $x^{L,j} = \{x_j(1 + (j-1)\Delta), \dots, x_j((j-1)\Delta + L)\}$ with $j = 1, \dots, S$ and $0 \leq \Delta < L$. The difference $L - \Delta$ represents the overlapping between two consecutive time windows, i.e. the last $L - \Delta$ measurements of the $j-th$ vector $x^{L,j}$ coincide with the first $L - \Delta$ measurements of the vector $x^{L,j+1}$.

4. CONTINUOUS WAVELET TRANSFORMS FOR SENSOR MALFUNCTION DETECTION

Signal measurements in energy production plants may show transients and non-stationary behaviors. Therefore, time or frequency-domain methods, which have been developed for stationary signals, cannot be applied with success to the sensor data validation task. Due to the time-varying

frequency spectrum of the signals, suitable time–frequency decomposition tools are needed for real-time signal data validation. Time–frequency analysis can identify the signal frequency components and reveal their time-variant features. Various time–frequency analysis methods have been proposed and applied to fault detection, diagnostics and prognostics. Among these, Short-Time Fourier Transform (STFT), Wavelet Transform (WT), Hilbert–Huang Transform (HHT), and Wigner–Ville Distribution (WVD) are the most commonly used approaches.

Wavelet transform is a mathematical tool that converts a signal into a different form (Gao & Yan, 2011). The objective of the conversion is twofold: *i*) to reveal signal characteristics that are hidden in the time domain and *ii*) to provide a more succinct representation of the original signal. A base wavelet function $\psi(t)$ is needed in order to perform the wavelet transform. A wavelet is a small wave that has an oscillating wavelike characteristic and has its energy concentrated in time. A wavelet is used as template for analyzing time-varying or nonstationary signals by decomposing the signal into a 2D, time-frequency domain representation (Gao & Yan, 2011) (Mallat, 2008).

For any real signal $x(t) \in L^2(\mathbb{R})$, the Continuous Wavelet Transform (CWT) with scale parameter $s > 0$, translation parameter $u \in \mathbb{R}$ and wavelet function $\psi(t)$ is:

$$CWT_x^{\psi}(u,s) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt \qquad (1)$$

The reader interested in more mathematical details about wavelet transform can refer to Appendix A. The translation parameter u can be interpreted as the time instant around which the signal is analyzed. With respect to the scale parameter, at small s values $CWT_x^{\psi}(u,s)$ provides information on the details (i.e., the high frequency content) of the signal in the neighborhood of time instant u, whereas at large s values $CWT_x^{\psi}(u,s)$ provides information on the trend (i.e., the low frequency content) of the signal in the neighborhood of time instant u. A standard way of representing the CWT is to use a two-dimensional image, called scalogram, graphically representing the square of the CWT, $\left|CWT_x^{\psi}(u,s)\right|^2$, as function of the translation parameter u and scale parameter s. Since the signals considered are typically digital signals, a discrete approximation of Eq. (1) is typically computed (Torrence & Compo, 2010). The approximated scalogram is a matrix whose rows and columns correspond to different scales s and translation parameters u, respectively. Figure 6 shows a cosine signal with a sudden change of frequency at time $t = 25$ and its corresponding scalogram image, which clearly allows graphically identifying the time at which the change of frequency occurs.

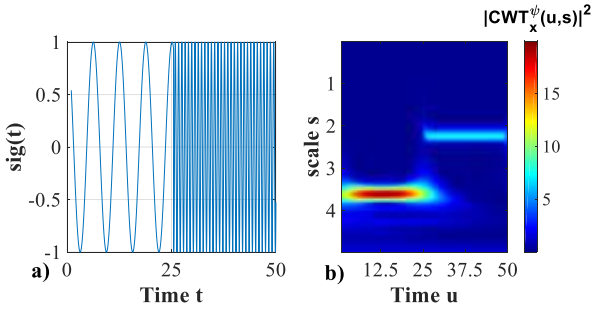

Figure 6. Left: Signal $x(t)$; right: and corresponding scalogram $\left|CWT_x^{\psi}(u,s)\right|^2$.

As mentioned earlier, a sensor malfunction alters the regularity of a signal, i.e., its degree of smoothness. For example: a sensor malfunction causing spikes adds irregularity to a signal, being a spike an approximation of a Dirac distribution, which is not differentiable (Mallat & Hwang, 1992); a sensor malfunction causing freezing of the sensor readings adds regularity to the signal, since a constant signal is differentiable infinite times. A measure of the local regularity of a signal is provided by the Lipshitz exponent α (Mallat & Hwang, 1992) which is introduced, from a mathematical point of view, in Appendix B. Considering a function $x(t)$, it is possible to show that:

- if $x(t)$ is uniformely Lipschitz $\alpha > n$ in the neighborhood of t_0, this implies that $x(t)$ is necessarily n times continuously differentiable in this neighborhood (Mallat, 2008);
- α equal to 1 implies that $x(t)$ is a continuously and differentiable function at t_0;
- $\alpha \in (0,1)$ implies that the function $x(t)$ is continuous at t_0 but the first derivative of the function at that point is not continuous;
- α equal to 0 implies that the function is discontinuous at t_0 but bounded in the neighborhood of t_0.

In (Struzik, 2001), the estimation of the Lipschitz-exponent at a given point t_0 has been obtained through the use of the Wavelet Modulus Maxima (WMM). A WMM is defined as any point (u_0, s_0) such that $\left|CWT_t^{\psi}(u,s_0)\right|$ is a local maximum at $u = u_0$ and the maxima line consists of the points that are local maxima. The approximated estimation of α is provided by:

$$\alpha = 2^{\frac{1}{z-1}\sum_{s=1}^{s=z-1} log_2\left(\frac{CWT_x^{\psi}(u,s+1)}{CWT_x^{\psi}(u,s)}\right)} \qquad (2)$$

where z is the length of the maxima line that propagates from coarse scales to fine scales. This equation has been successfully applied in many engineering problems, like bearing faults diagnostics (Li, 2010), machinery health

monitoring problems (Miao et al., 2007) and signal denoising (Mallat & Hwang, 1992). These works typically rely on the fact that any irregularity can be detected by finding the translation parameter u at which WMM converge at fine scales (Mallat & Hwang, 1992). Notice, however, that methods for α estimation based on WMM are only able to provide a rough approximation, since they exploit only the information carried out by the first and last points of the maxima line (Miao et al., 2007). A common problem of WMM-based techniques for the estimation of α is that the limited resolution of a discrete signal implies that the scale s cannot be arbitrarily small, causing approximations which can lead to inaccurate Lipschitz exponent estimation (Tu et al., 2005). Therefore, the use of WMM for sensor data validation is applicable to detect only those types of sensor malfunctions adding irregularity to a signal, such as spike and noise, whereas those adding regularity to a signal, such as freezing, cannot be properly detected since none of the maxima lines converge to the u corresponding to the freezing (Mallat & Hwang, 1992).

To overcome these limitations of the use of WMM for sensor data validation, in this work we propose to directly work on scalogram images. This original approach is motivated also by the possibility of taking full advantage of the redundancy provided by the CWT, which allows avoiding loss of information (Kovačević & Chebira, 2007) and has been shown useful in many applications such as feature extraction (Sengüler, 2016).

With respect to the choice of the type of wavelet transform, notice that different sensor malfunctions influence the CWT coefficients in specific and different scale ranges, as it will be shown in Section 4.1 and Appendix B. For this reason, an efficient sensor validation tool should be based on a wavelet transform able to provide an accurate scale localization, such as Morlet wavelet:

$$\psi(t) = \frac{1}{\pi^{1/4}}\ e^{i\pi f_0 t}\, e^{-t^2/2} \qquad (3)$$

which has been shown to provide more accurate scale localization than other types of wavelet functions (Karacan & Olea, 2014).

4.1. Analysis of the scalogram characteristics in correspondence of different types of sensor malfunction

In this Section, we discuss the characteristics of the scalograms of the signals measured in case of different types of sensor malfunctions.

4.1.1. Spike

Figure 7 shows the scalograms obtained from a signal acquired by a healthy sensor (Figure 7a) and the same signal to which a spike has been artificially injected at time $t = 50$ (Figure 7b).

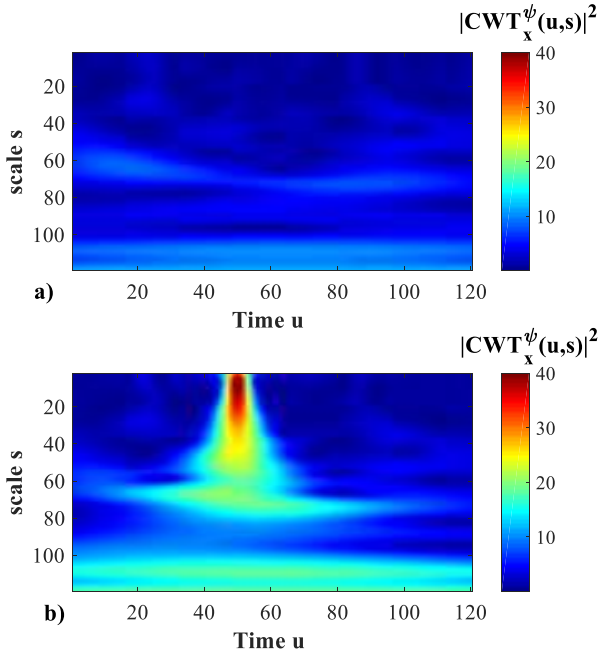

Figure 7. Top: scalogram of the signal of Figure 1(a) acquired by a healthy sensor; bottom: scalogram of the signal of Figure 1(b) corresponding to the same signal with a spike at $t = 50$.

As expected, the main difference between the two scalogram images is observed in the neighborhood of the time at which the spike has been injected and consists in the abrupt increasing of the wavelet coefficients at small scales. This result is coherent with the fact that, from a theoretical point of view, a spike can be seen as an approximation of a Dirac distribution which is characterized by a Lipschitz exponent equal to -1 (Mallat & Hwang, 1992). Thus, the wavelet transform modulus maxima increases proportionally to $\frac{1}{s}$ over a large range of scales in the corresponding neighborhood (Mallat & Hwang, 1992). In conclusion, a spike can be recognized for its large coefficients in the scalogram at small scales.

4.1.2. Noise

Figure 8 shows the scalograms obtained from a signal acquired by a healthy sensor (Figure 8a) and the same signal to which noise has been artificially injected (Figure 8b). The scalogram image shows larger CWT coefficients at all times in the case of presence of noise. According to (Qiu et al., 2006), this is due to the fact that noise adds irregularity to the signal in every sample, increasing its variance. In practice, a noisy signal shows sharper changes than the nominal one, which can be seen as a combination of many low intensity spikes. This implies CWT coefficients larger than in the case of a healthy sensor, but smaller than those observed in correspondence of the spike.

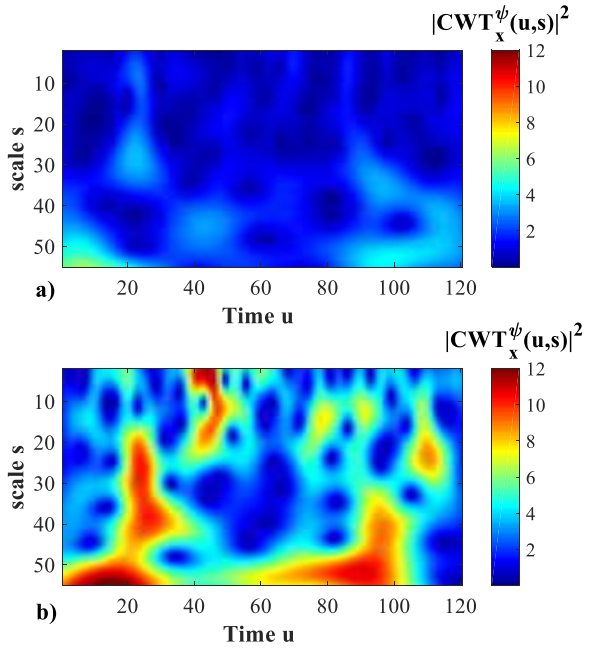

Figure 8. Top: scalogram of the signal of Figure 2(a) acquired by a healthy sensor; bottom: scalogram of the signal of Figure 2(b) corresponding to the same signal after artificially injecting a noise malfunction.

4.2. Freezing

Figure 9 shows the scalograms obtained from a signal acquired by a healthy sensor (Figure 3a) and the same signal to which a freezing has been artificially injected (Figure 3b). The scalogram obtained from the frozen signal is characterized by a large region with zero CWT coefficients at small scale. The zero CWT coefficients are due to the fact that when the wavelet atom $\psi_{u,s}(t)$ support includes that of a constant signal $x(t) = c_0$, Eq. (1) becomes:

$$CWT_x^{\psi}(u,s) = \int_{-\infty}^{+\infty} x(t)\,\psi^*{}_{u,s}(t)dt = \\ = \int_{support\ \psi_{u,s}} c_0\psi^*{}_{u,s}(t)dt = 0 \quad (4)$$

where the last equality holds for the vanishing moment property (Eq. 20 in Appendix B). Notice that, since the smaller is s, the smaller is the support of $\psi_{u,s}(t)$, we can conclude that for a fixed value of the translation parameter u, the support of the atom $\psi_{u,a}(t)$ is included in that of the atom $\psi_{u,b}(t)$ provided that $a < b$. Thus, if the support of $\psi_{u,b}(t)$ includes the frozen signal interval, then also the support of $\psi_{u,a}(t)$ includes the same interval and, consequently, has a zero CWT coefficient. For this reason, the region with zero CWT coefficient values becomes larger when s decreases to zero and tends to show a triangular shape (Figure 9).

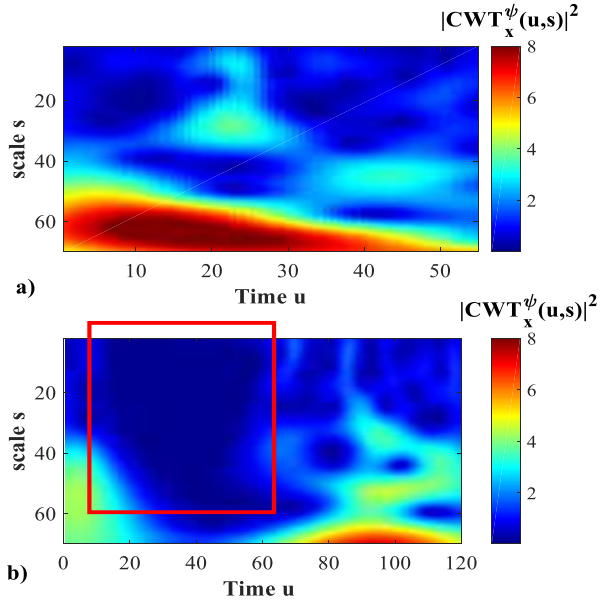

Figure 9. Top: scalogram of the signal of the Figure 3(a) acquired by a healthy sensor; bottom: scalogram of the signal of Figure 3(b) corresponding to the same signal after artificially injecting a freeze without jump malfunction.

4.3. Quantization

Figure 10 shows the scalograms obtained from a signal acquired by a healthy sensor (Figure 5a) and the same signal to which a quantization has been artificially injected (Figure 5b). The comparison of these two Figures shows that the CWT coefficients at large scales are very similar whereas there are differences at small scales. In detail, the effect of the quantization is twofold:

- when the quantized signal is constant for several successive samples, the CWT coefficients become smaller with respect to the same case without quantization (dashed region in Figure 10b). This is due to the fact that the quantized signal behaves like a frozen signal in this time interval;
- when quantization induces sudden jumps, the CWT coefficients become larger than those of the same case without quantization. This is due to the fact that a quantized signal behaves like a low intensity spike in these time intervals.

Thus, a quantized signal can be viewed as a signal in which short periods of freezing are alternated to low intensity spikes.

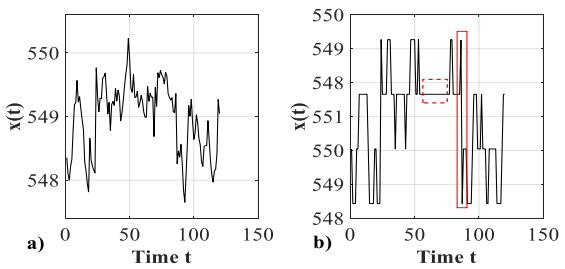

Figure 10. Signal in nominal condition (left) corresponding to the same signal after artificially injecting a quantization malfunction (right).

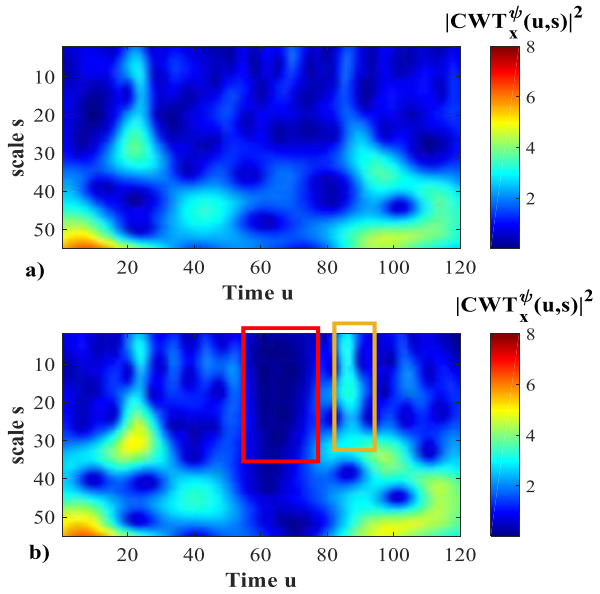

Figure 11. Top: scalogram of the signal of the Figure 5(a) acquired by a healthy sensor; bottom: scalogram of the signal of Figure 5(b) corresponding to the same signal after artificially injecting a quantization malfunction.

Notice that freezing and quantization malfunctions can also be detected by computing a first-order forward finite difference approximation of the test signal first derivative, and then, setting a detection threshold on the number of consecutive chunks of 0's. A drawback of this approach is that it would require the development of dedicated models and the setting of the corresponding detection thresholds for each type of sensor malfunctions, whereas the proposed method allows dealing with all the considered sensor malfunctions using just one model and one detection threshold.

5. SENSOR MALFUNCTION DETECTION METHOD

The method proposed in this work is based on the idea of comparing the scalogram obtained from the test vector $x^L(t) = \{x(t-L+1), \dots, x(t)\}$ to the scalograms obtained from the training vectors $x^{L,j} = \{x_j(1+(j-1)\Delta), \dots, x_j((j-1)\Delta + L)\}, j = 1, \dots, S, 0 \le \Delta < L$.

First, each training vector $x^{L,j}, = 1, \dots, S$, is transformed in the corresponding scalogram by applying the following procedure:

Step 1: Compute the CWT, $CWT^{\psi}_{x^{L,j}}(u,s)$, and the corresponding scalogram image $I^j(t)$. The scalogram $I^j(t)$ is a matrix of size NxM, where N and M depend on the discretization of the scale parameter s and translation parameter u (typically $M = L$), respectively. According to the results of the analysis of Section 3, large scale values do not provide useful information for sensor malfunction detection and, consequently, the analysis focuses on scale values lower than a prefixed threshold, i.e., only scale values $s < \tilde{s}$ are retained. Notice that this results in a reduction of the original scalogram dimensions from NxM to $\tilde{N}x\,M$, being $\tilde{N} < N$.

Step 2: Process the scalogram image to:

a) enhance the differences at low scales, which have been shown to be relevant for the identification of a sensor malfunction caused by freezing or quantization (Sections 3.3 and 3.4);

b) normalize the intensities $CWT^{\psi}_{x^{L,j}}(u,s)$ in the range $[0,1]$.

Step a) transforms the scalogram image $I(t)$ into a new scalogram image:

$$\tilde{I}^j(t)_{p,q} = \begin{cases} I^j(t)_{p,q} & if\ I^j(t)_{p,q} \le a_{max} \\ a_{max} & if\ I^j(t)_{p,q} > a_{max} \end{cases} \quad (5)$$

where a_{max} is a predefined threshold.

Step b) converts the scalogram $\tilde{I}^j(t)$ into a greyscale image $G^j(t)$ by scaling its entries in the interval [0,1] as follows:

$$G^j(t)_{p,q} = \frac{\tilde{I}^j(t)_{p,q} - \tilde{I}^j(t)_{min}}{\tilde{I}^j(t)_{max} - \tilde{I}^j(t)_{min}} \quad (6)$$

where $\tilde{I}^j(t)_{min}$ and $\tilde{I}^j(t)_{max}$ are the minimum and the maximum values of the matrix $\tilde{I}^j(t)$ in all the training scalograms.

Since two consecutive training vectors, $x^{L,j}$ and $x^{L,j+1}$, overlap of $\Delta - L$ components (Section 3), i.e., the last $L - \Delta$ measurements of the j^{th} vector $x^{L,j}$ coincide with the first $L - \Delta$ measurements of the vector $x^{L,j+1}$, the effect on the scalogram of the occurrence of an event, such as a plant transient, will be visible at different times in different consecutive scalograms. This allows obtaining in the training scalograms an overall representation of the signal measured by healthy sensors that is invariant from the shift of the events.

Then, for the test vector x^L, we repeat Steps 1 and 2 to obtain its corresponding scalogram $I(t)$. Notice that entries of matrix $\tilde{I}(t)$ (Eq. (5)) close to a_{max} at low scales indicate sensor malfunctions which add irregularity (i.e., noise and spike) and entries lower than a_{max} indicate sensor malfunctions which add regularity (i.e., quantization and freezing). With respect to matrix $G(t)$ in Eq. (6), entries close to 1 at low scales are typical of noise and spike malfunctions, whereas entries close to 0 are typical of quantization and freezing malfunctions. Once the training and test grayscale images, G^j and G, have been obtained, they are compared by applying the following procedure:

A1 Compute the dissimilarities d_j between the greyscale image $G(t)$ and all the greyscale images G_j obtained from the historical signals $x^{L,j}$, pre-processed according to Steps 1-2:

$$d_j = \left\| G(t) - G_j \right\| \quad j = 1, \dots, S \tag{7}$$

where the matrix norm of the scalogram is:

$$\|G(t)\| = \sum_{p=1}^{\tilde{N}} \sum_{j=q}^{M} \left| G(t)_{p,q} \right| \tag{8}$$

A2 Identify the scalogram of the training set most similar to the one currently tested, i.e. the one characterized by the minimum dissimilarity d^*:

$$d^* = \min_{j=1,\dots,S} d^j \tag{9}$$

A3 Compare d^* with a fixed detection threshold T: if d^* is greater than T, then a sensor malfunction is detected.

Parameters T, a_{max}, and $\tilde{s}$ are set by minimizing a weighted sum of the number of false and missed alarms, f_a and m_a, on a validation set:

$$\tilde{o}(f_a, m_a) = w_1 f_a + w_2 m_a \tag{10}$$

where the weights w_1 and w_2 are set by considering a proper trade off between missed and false alarms. The validation set is formed by:

i. historical data collected when the sensor was healthy, different from those used for the model training.

ii. data representative of sensor malfunctions. If these latter data are not available, they can be simulated using the procedure described in Appendix C.

5.1 TIME COMPLEXITY

Resorting to the big O notation typically employed for evaluating algorithm complexity (Wegener, 2005), the computational complexity of the different steps of the proposed method for testing a signal segment of L samples and setting $M = L$ is:

- *Step 1*: $O(\tilde{N} L \log L)$ for computing the wavelet transform of the test signal $x^L(t)$, with $O(L \log L)$ representing the time complexity required per scale (Torrence & Compo, 2010);
- *Step 2*: $O(\tilde{N} L)$ for scalogram preprocessing;
- A1: $O(\tilde{N} L S)$ for computing all distances d_j;
- A2: $O(S)$;
- A3: $O(1)$.

Notice that the computational complexity is $O(L^3)$ in the worst case, i.e., when $L = \tilde{N} = S$.

6. CASE STUDY

We consider a dataset containing real temperature measurements recorded at a sampling frequency $f_s = 1\ Hz$ from a component of an electricity production plant (Baraldi et al., 2015). The data have been validated by plant experts to guarantee that they have been collected by healthy sensors when the plant components were in healthy conditions. The temperature signal has been segmented using a fixed time window of length $L = 120$ samples (corresponding to 120 seconds), with overlapping of 20 samples. The overlapping of the training pattern has been introduced to deal with the fact that a malfunction can occur at any time of the test window. Therefore, in order to detect it, various shifted training vectors with an overlap of $L - \Delta = 20$ samples are considered in the training set. Since the available data have been collected by a healthy sensor, we have artificially simulated sensor malfunctions of different types and intensities, according to the procedure proposed in (Sharma et al., 2010) and reported in Appendix C. Figure 12 shows an example of signal behavior and Figure 13 examples of simulated low-intensity sensor malfunctions. Since the available historical signal vectors $x^{L,j}$ have been collected from a plant in normal condition, the case study does not fully meet the second assumption of the problem statement, i.e., that the available historical data are representative of all the plant operating conditions, including anomalies caused by degradation and failure of the component. Therefore, component malfunctions can be erroneously detected as sensor malfunction by the data validation method. Notice, however, that component failures and malfunctions typically involve several signals at the same time and therefore can be distinguished from sensor malfunctions.

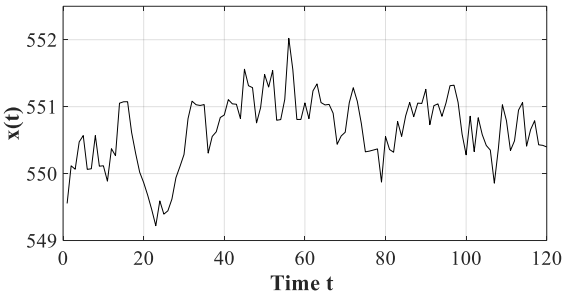

Figure 12. Signal measurements obtained from a healthy sensor.

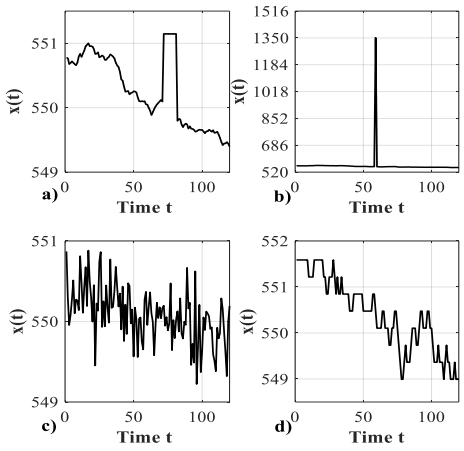

Figure 13. Simulated sensor malfunctions: freezing (top-left), spike (top-right), noise (down-left), quantization (down-right).

6.1. Dataset partitioning

We have partitioned the available data into three subsets: i) a training set, ii) a validation set, and iii) a test set. The training set is formed by 67 signal segments measured from a healthy sensor and constitutes the set of vectors $x^{L,j}$ from which the dissimilarity of the test segment is computed in Step 3 (Section 4). The validation and test sets are formed by 400 and 460 signal segments, respectively, and contain measurements from the healthy sensor and artificially injected sensor malfunctions of different types and intensities, according to the proportions of Tables 1 and 2. The validation set has been used to determine the values of the parameters of the method: wavelet coefficient threshold a_{max} (Step 2a), maximum scale $\tilde{s}$ (Step 1) and detection threshold T (Step 5), whereas the test set has been used to evaluate the performance of the proposed methodology. To better mimic a real application, the signal segments of the training set temporally preceed those of the validation set, which preceed those of the test set.

Type of sensor malfunction	Number of signals in the validation set
Freezing	100
Spike	100
Noise	100
Quantization	50
Healthy	50

Table 1. Validation set partition.

Sensor malfunctions	Number of signals in the test set
Freeze	100
Spike	100
Noise	100
Quantization	80
Healthy	80

Table 2. Test set partition.

6.2. Results

Wavelet coefficient threshold a_{max}, scale $\tilde{s}$ and detection threshold T have been set by minimizing the function $\tilde{o}$ Eq. (9) assuming $w_1 = w_2 = 1$, i.e., by giving same importance to the contributes. By setting $a_{max} = 0.06$, $\tilde{s} = 2.8$ and $T = 884$, we have obtained the optimal trade off 1% of missed alarms and 4% of false alarms in the validation set. This choice of the scale parameter $\tilde{s}$ results in a reduction of the original scalogram dimensions from 591x120 to 50x120, with evident benefits on the computational burden. Figure 14 shows the variations of the false alarm rates and of different types of missed alarm rate with respect to variation of the detection threshold T. It is interesting to observe that if the threshold T is progressively increased, the first types of missed alarms that occur are those caused by quantization and freezing malfunctions, whereas spike and noise malfunctions are correctly recognized. This is due to the fact that the scalograms corresponding to quantization and freezing malfunctions are more similar to those obtained from a healthy sensor than those corresponding to spike and noise malfunctions, as shown in Figures 7, 8, 9 and 11. Thus, the identification of quantization and freezing malfunctions is more sensible to the threshold value than that of the spike and noise malfunctions.

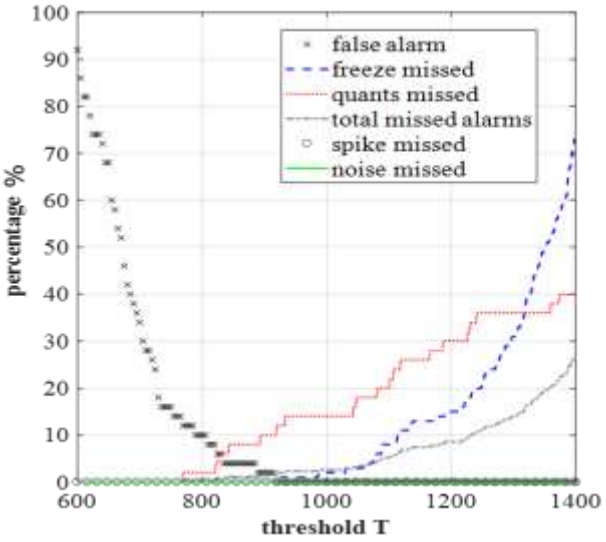

Figure 14. Variations of the false alarm rate (cross-dotted black line) and of the missed alarm rates due to freezing (dashed blue curve), quantization malfunctions (dotted red curve), spike malfunctions (circle-dotted purple curve), noise malfunctions (continuous green curve). The total variation of the missed alarm rate is referred using the (dash-dot grey curve).

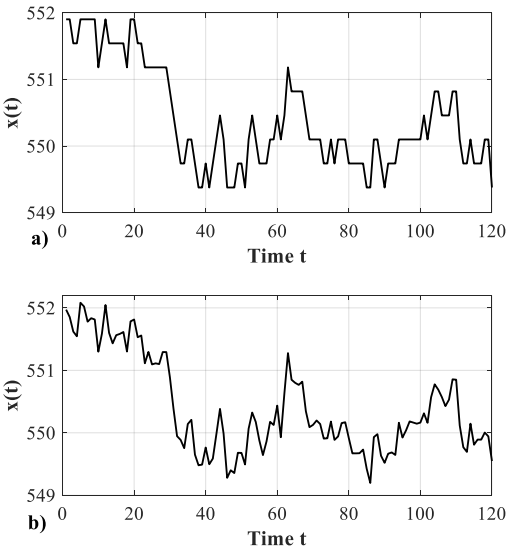

Figure 15. Example of missed alarm: the quantized signal segment (Top) and the corresponding signal segment before the malfunction injection (Down).

The application of the proposed method to the signal segments of the test set gives a 0% rate of false alarms and a 1.5% rate of missed alarms, caused by quantization, whereas freezing, spikes and noise are always correctly detected. Figure 15 shows an example of a missed alarm caused by a quantized signal segment incorrectly considered as healthy. Notice that the degree of quantization of this signal segment (intensity of the malfunction) is very small and the quantized signal segment appears very similar to the corresponding segment before the injection of the malfunction (Figure 15, Top).

We have compared the results of the proposed methodology with those obtained by applying a) a sensor data validation approach based on the use of Principal Component Analysis (PCA) (Penha & Hines, 2001) and b) a two classes SVM classifier with Gaussian kernel directly applied to the raw measurements.

The PCA approach relies on the following steps:

- the extraction of 87 lumped features, such as statistical metrics (e.g., means, standard deviations, etc.) and analytics (e.g., derivatives, elongation, etc.), signal transforms in the frequency domain (e.g., Fourier Transform, Laplace Transform) and/or in the time-frequency domain (e.g., Short Time Fourier Transform (STFT). The considered set of features have been shown able to catch the dynamic behavior of the signals in prognostics and health management applications in (Baraldi et al., 2016) (Cannarile et al., 2017);
- the application of PCA to the training data, which correspond to measurements obtained from a healthy sensor;
- the identification of the number of principal components to be used for the signal reconstruction (Penha & Hines, 2001). This is performed by looking for the most satisfactory trade-off between false and missed alarm rates in the validation set;
- the reconstruction of the test set data and the comparison of the Square Prediction Error (SPE) (Lee et al, 2004) (also referred to as Q-statistic or residual (Lee et al., 2004)) with a fixed threshold (Lee et al., 2004).

With respect to the SVM classifier we have considered two classes: "normal condition" (class 1) and "sensor malfunctioning" (class -1). The SVM classifier has been built using a dataset which will be referred to as T_r^{New}, formed by the union of the training and validation sets introduced in Section 6.1. T_r^{New} is, therefore, made by $N_1 = 117$ and $N_{-1} = 350$ patterns of class 1 and -1, respectively. The two parameters of the SVM with Gaussian kernel, i.e., the scale parameter σ^2 of the Gaussian kernel and the cost parameter C, controlling the tradeoff between error penalization and the complexity of the classification function, have been set by trial-and-error with the objective of minimizing the cost function in Eq. (10) with weights $w_1 = w_2 = 0.5$. Table 3 reports the considered values of the parameters σ^2 and C. The set Tr^{new} has been randomly partitioned in a training-set (75% of total patterns) and a validation set (25% of total patterns). Being the training-set unbalanced ($N_{-1}^{tr} = 0.75 * N_{-1} > N_1^{tr} = 0.75 * N_1$), we have used the Different Error Cost (DEC) method (Batuwita & Palade, 2013) where the

parameter C has been scaled by $N^{tr}/(2N_1^{tr})$ for class 1 patterns and by $N^{tr}/(2N_{N_{-1}}^{tr})$ for class -1 patterns, with $N^{tr} = N_1^{tr} + N_{-1}^{tr}$. We have found that the solution minimizing the considered cost function (Eq. (10)) is: $(\sigma^2, C) = (16,64)$. Finally, we have trained a new SVM using T_r^{New} and tested it on the test set (Table 2).

Hyperparameter	Considered values
σ^2	$\{2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}$ $2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8\}$
C	$\{2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}$ $2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8\}$

Table 3. Considered values of hyperparameters σ^2 and C.

Table 4 compares the results obtained by the proposed method, the PCA-based method and the SVM classifier.

METHOD	Percentage of Missed Alarm	Percentage of False Alarm
Proposed Method	0%	1.25%
PCA-based Method ($P = 90\%$)	10.8%	1.25%
SVM using raw data	22.5%	45.5%

Table 4. Comparison of the performance of the proposed method with the PCA based approach.

From Table 4 we can conclude that: 1) the PCA proposed approach is less accurate than the proposed method: the percentage of missed alarms increases from 0% to 10.8% (Table 4), with the same percentage of false alarms; 2) the SVM method performs poorly compared to the proposed method with larger rates of missed alarms and false alarms.

We have evaluated the robustness of the proposed method with respect to different intensities of the malfunctions, simulated according to (Sharma et al., 2010) (see Appendix C). Table 5 reports the results in term of missed alarms for the different types of sensor malfunctions.

Sensor malfunctions	Low Intensity	Medium intensity	High intensity
Freezing	0%	1%	0%
Spike	0%	0%	0%
Noise	0%	0%	0%
Quantization	6%	2%	0%

Table 5. Percentage of missed alarms considering sensor malfunctions of low, medium and high intensities.

We can conclude that the method provides satisfactory performances and, as expected, the overall percentage of missed alarms decreases as the malfunction intensity increases.

Furthermore, we have tested the proposed method on 100 signal segments characterized by the simultaneous presence of two sensor malfunctions, obtained by randomly sampling their times of occurrence and their intensities from the same probability distributions used for sampling low intensity single sensor malfunctions. Table 6 reports the results in terms of missed alarms, for the different combinations of two sensor malfunctions. It is interesting to observe that the percentage of missed alarms, in case of quantization malfunction decreases to 0% (it was 6% in case of single low intensity malfunction). This is due to the fact that scalogram modifications caused by spike or noise malfunctions (Figures 7 and 8) are easier to detect than those caused by to the quantization anomaly (Figure 11), and, therefore, the detection of the quantization malfunction is facilitated by the simultaneous presence of spike and noise malfunctions.

Sensor malfunctions	Number of segments	Missed alarms
Quantization+Spike	20	0%
Quantization+Noise	20	0%
Freezing+Spike	20	0%
Noise+Spike	20	0%
Noise+Freeze	20	0%

Table 6. Percentage of missed alarms considering pairs of sensor malfunctions.

With regards to the computational time, testing signal segments of $L = 120$ samples has required on average 0.052 seconds using an Intel Core i5-M430 @ 2.26 GHz processor with 4 Gb RAM in a MATLAB 2017b environment. Therefore, the proposed approach is suitable for being used in field operation.

7. DISCUSSION AND OUTLOOKS

In this work, we have not considered the possible influence of one sensor malfunction on other sensor readings. In complex systems characterized by many interconnected components in which the readings of some sensors are used for system control, one sensor malfunction can cause non-optimal decisions of the control system, which results in anomalous behaviors of other signals. In this case, although the proposed method correctly identifies the sensor affected by the malfunction, it will also erroneously detect malfunctions of other sensors. Furthermore**,** multiple sensor malfunctions can be detected at the same time in cases of plant abnormal conditions caused by degradation or failures of plant components which are not included in the training set. The discrimination between a sensor malfunction and a

plant abnormal condition requires the development of a supervisor system integrating the data validation tool with a module for the detection of abnormal plant conditions.

Another issue not investigated in this work is the classification of the type of sensor malfunction. We expect that a multi-class classifier (e.g., K-Nearest Neighbors (KNN), Decision Trees (DT), etc.), performs poorly in practice, due to the difficulty of discriminating between freezing and quantization sensor malfunctions which produce very similar variations in the signal values and in the associated scalograms, as discussed in Section 4. A possible approach to overtake this difficulty is the development of a One-Vs-All (OVA) classification system, where different binary classifiers are developed, each one trained to distinguish patterns of a single class from those of all the remaining classes.

A limitation of the proposed method is that it cannot identify sensor malfunctions which cause drifts of the sensor readings. This is due to the fact that drifts do not alter the regularity of a signal (i.e., its smoothness), but introduce a (typically monotone) trend in the signal, which has no effect at the high frequency (low scale values). Therefore, the detection of the drift malfunction would require the development of method able to distinguish malfunctions which, differently from the malfunction types considered in this work, have effects at low frequencies. Finally, notice that if the current operational condition of the plant remarkably differs from those in which the training set data have been recorded, the performance of the method for detecting the occurrence of a sensor malfunctioning is expected to deteriorate. This problem can be overtaken by periodically retraining the model.

8. CONCLUSION

In this work, we have developed a novel method for sensor data validation, which combines the use of CWT with an image analysis technique. Sensor validation is performed by comparing the CWT scalogram obtained from the test signal with those obtained from historical data of the same signal. The performance of the method, measured in terms of false and missed alarm rates, is shown superior to that of a PCA-based approach and binary SVM classifier for data validation. From a practical point of view, the method, differently from the traditional sensor data validation approaches which consider the correlations among plant signals, is easily applicable to all the sensors of a fleet of plants being the validation of the data measured from a sensor independent to that of other sensors. Furthermore, it has been shown that the analysis of the obtained scalograms allows distinguishing among the different types of sensor malfunction.

APPENDIX A: CONTINUOUS WAVELET TRANSFORMS

In mathematical terms, a wavelet is a function $\psi(t) \in L^2(\mathbb{R})$ satisfying the admissibility condition (Mallat & Hwang, 1992):

$$\int_0^{+\infty} \frac{|\xi(\omega)|^2}{\omega} d\omega = \int_{-\infty}^{0} \frac{|\xi(\omega)|^2}{|\omega|} d\omega = C < \infty \qquad (11)$$

where $L^2(\mathbb{R})$ denotes the space of square-integrable functions and $\xi(\omega)$ the Fourier transform of the wavelet function $\psi(t)$. The admissibility condition implies that the Fourier transform of the function $\psi(t)$ vanishes at zero frequency:

$$|\xi(\omega)|_{\omega=0}{}^2 = 0 \qquad (12)$$

and that the average value of the wavelet $\psi(t)$ is zero (Mallat & Hwang, 1992):

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \qquad (13)$$

A dictionary of time-frequency atoms is defined from the wavelet function $\psi(t)$ by scaling $\psi(t)$ by s (referred to as the scale parameter) and translating it by u (referred to as translation parameter):

$$\psi_{u,s}(t) := \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \quad u \in \mathbb{R}, \quad s > 0 \qquad (14)$$

For any real signal $x(t) \in L^2(\mathbb{R})$, the Continuous Wavelet Transform (CWT) with scale parameter s and translation parameter u is:

$$CWT_x^{\psi}(u,s) = \int_{-\infty}^{+\infty} x(t) \psi_{u,s}(t) dt \qquad (15)$$

The factor $\frac{1}{\sqrt{s}}$ in Eq. (14) guarantees that the wavelet transform in Eq. (1) is directly comparable at different scales.

APPENDIX B: LIPSCHITZ EXPONENT

A function $x(t)$ is pointwise Lipschitz $\alpha \geq 0$ at t_0, if there exist $A > 0$ and a polynomial P_{t_0} of degree $n_\alpha = \lfloor \alpha \rfloor$, the greatest integer less than or equal to α, such that (Mallat, 2008):

$$\left| x(t) - P_{t_0}(t) \right| \leq A |t - t_0|^{\alpha} \quad \forall t \epsilon \mathbb{R} \qquad (16)$$

- The function $x(t)$ is uniformly Lipschitz α over the interval $[a, b]$ if it satisfies Eq. (16) for all $t_0 \epsilon [a, b]$, with a constant A that is independent of t_0 (Mallat, 2008).
- The Lipschitz regularity of $x(t)$ at t_0 or over $[a, b]$ is the greatest value of α such that $x(t)$ is Lipschitz-α, i.e. the least real number that is greater than or equal to all α (Mallat 2008).

The Lipshitz coefficient can be interpreted by considering the Taylor formula. Suppose that $x(t)$ is m times differentiable in the interval $[t_0 - \delta\,, t_0 + \delta\,]$. Let P_{t_0} be the Taylor polynomial in the neighborhood of t_0:

$$P_{t_0}(t) = \sum_{k=0}^{m-1} \frac{x^{(k)}(t_0)}{k!}(t - t_0)^k \quad (17)$$

The approximation error:

$$\varepsilon_{t_0} = x(t) - P_{t_0}(t) \quad (18)$$

satisfies:

$$\forall t \epsilon [t_0 - \delta, t_0 + \delta],$$
$$\left|\varepsilon_{t_0}\right| \leq \frac{|t - t_0|^m}{m!} \sup_{w \epsilon [t_0 - \delta, t_0 + \delta]} \left|x^{(m)}(w)\right| \quad (19)$$

Since the Taylor formula relates the differentiability of a signal to local polynomial approximations (Mallat, 2008), the m^{th}order differentiability of $x(t)$ in the neighborhood of t_0 yields an upper bound of the error ε_{t_o} when t tends to t_0. The Lipschitz regularity refines this upper bound with non-integer exponents and, thus, it provides uniform regularity measurements over time intervals and at specific points t_0. If $x(t)$ has a singularity at t_0 then, the Lipschitz exponent at t_0 characterizes the singularity behavior (Mallat, 2008). CWT have been used to estimate the Lipschitz exponent, and, thus, to characterize the local regularity of functions (Mallat & Hwang, 1992). According to (Holschneider & Tchamitchian, 1989), the asymptotic decay of the wavelet transform at small scales is related to the local Lipschitz regularity through the following theorem:

Theorem 1:

Let $x(t) \in L^2(\mathbb{R})$ and $[a, b]$ an interval of $\mathbb{R}$. Let $0 < \alpha < 1$. For any $\epsilon > 0$, a function $x(t)$ is uniformly Lipschitz α over intervals $(a + \epsilon, b - \epsilon)$, if and only if for any $\epsilon > 0$ there exists a constant A_ε such that for any $u \in (a + \epsilon, b - \epsilon)$ and scale s:

$$\left|CWT_x^{\psi}(u, s)\right| \leq A_\varepsilon s^\alpha \quad (20)$$

In order to extend Theorem 1 to Lipschitz exponents α larger than 1, it is necessary to impose that the wavelet $\psi(t)$ has enough vanishing moments (Mallat & Hwang, 1992). A wavelet $\psi(t)$ is said to have n vanishing moments if and only if for all positive integers $\beta < n$ it satisfies (Mallat, 2008):

$$\int_{-\infty}^{+\infty} t^\beta \ \psi(t)\, dt = 0 \quad (21)$$

If the wavelet $\psi(t)$ has n vanishing moments, then, Theorem 1 remains valid for any non-integer value α such that $0 < \alpha < n$ (Mallat & Hwang, 1992).

APPENDIX C: SENSOR MALFUNCTIONS SIMULATION

Different sensor malfunction intensities have been simulated according to (Sharma et al., 2010), using fixed time window $x^L = \{x(1), \dots, x(L)\}$ of L samples. According to (Sharma et al., 2010), we distinguish among low, medium and high intensity malfunctions, where low intensity malfunctions are harder to detect since faulty samples do not differ significantly from normal sensor readings. Low intensity sensor malfunctions have been simulated by setting the parameters $f, g, h, \tilde{L}$ and Q in Eq. (22), Eq. (23), Eq. (24) and Eq. (25) to the values used in (Sharma et al., 2010) and reported in Table 6. To simulate medium and high intensity malfunctions the parameters f, g, h and Q in Eq. (22), Eq. (23), Eq. (24) and Eq. (25), have been set as in (Sharma et al., 2010) and are reported in Table 7.

Parameter	Coefficient
h	1
$\tilde{L}$	19
f	1.5
g	0.5
Q	8

Table 6. Parameters values used to simulate low intensity sensor malfunctions

Parameters	Coefficient for medium intensity	Coefficient for high intensity
h	1	1
$\tilde{L}$	40	80
f	5	10
g	1.5	3
Q	6	3

Table 7. Parameters values used to simulate low intensity sensor malfunctions

- **Spike**

Spike malfunctions have been simulated by randomly drawing a sample r and replacing the reported value $x(k)$ with

$$\tilde{x}(r) = x(r) + fx(r) \quad (22)$$

where the multiplicative factor f determines the intensity of the spike faults.

- **Noise**

Noise malfunctions have been simulated selecting a set of successive samples W and added a random draw from a normal distribution, $N(0, g^2\sigma^2)$, to each sample $x(r)$ in W, i.e.,

$$\tilde{x}(r) = x(r) + \sqrt{g^2\sigma^2}N(0,1) \quad (23)$$

where σ^2 is the variance of the signal in nominal condition and g is a multiplicative factor, which allows controlling the intensity of noise malfunction.

- **Freezing**

Freezing malfunctions have been simulated selecting the time length $\tilde{L} < L$ for which the signal measurement is affected by freezing, randomly sampling the time of occurrence of the malfunction $\tilde{k} = 1, \dots, L - \tilde{L}$, and replacing the sensor reading with

$$\tilde{x}(r) = x(\tilde{k}) + h \qquad r = \tilde{k}, \dots, \tilde{k} + \tilde{L} - 1 \qquad (24)$$

where h indicates the size of the sudden jump at the beginning of the freezing.

- **Quantization**

To inject quantization faults, we have firstly computed the minimum min and the maximum max values within the time window x^L; then, we have selected the number Q of discrete levels, so that the possible values that the quantized signal can assume are

$$y_l = (l-1)\left(\frac{max - min}{Q}\right) + min \quad l = 1, \dots, Q \qquad (25)$$

Finally, the reported value $x(r)$ is replaced with $y_{q^*(r)}$, where the index $q^*(r)$ satisfies

$$q^*(r) = argmin_{q=1,\dots,Q} |y_l - x(r)| \qquad (26)$$

REFERENCES

Baraldi, P., Di Maio, F., Turati, P., Zio, E. (2015). Robust signal reconstruction for condition monitoring of industrial components via a modified Auto Associative Kernel Regression method. *Mechanical Systems and Signal Processing*, 60, pp. 29-44.

Baraldi, P., Cannarile, F., Di Maio, F., Zio, E. (2016). Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. *Engineering Applications of Artificial Intelligence*, 56, pp. 1-13.

Baraldi, P., Gola, G., Zio, E., Roverso, D., Hoffmann, M. (2011). A randomized model ensemble approach for reconstructing signals from faulty sensors. *Expert Systems with Applications*, 38 (8), pp. 9211-9224.

Baraldi, P., Mangili, F., Gola, G., Nystad, B.H., Zio, E. (2014). A hybrid ensemble-based approach for process parameter estimation and degradation assessment in offshore oil platforms. *International Journal of Performability Engineering*, 10 (5), pp. 497-509.

Batuwita, R. and Palade, V. (2013). Class Imbalance Learning Methods for Support Vector Machines. In *Imbalanced Learning: Foundations. Algorithms and Applications.* John Wiley & Sons.

Cannarile, F., Baraldi, P., Compare, M., Zio, E. (2017). An unsupervised clustering method for assessing the degradation state of cutting tools in the packaging industry. *Safety and Reliability: Theory and Application-Proceedings of the European Safety and Reliability* Conference, ESREL 2017.

Coble, J.B., Meyer, R.M., Ramuhalli, P., Bond, L.J., Hashemian, H., Shumaker, B., et al. (2012). A Review of Sensor Calibration Monitoring for Calibration Interval Extension in Nuclear Power Plants. Pacific Northwest National Laboratory.

Gao, R. X., Yan, R., (2011). *Wavelets: Theory and Applications for Manufacturing*. New York: Springer.

Garvey, J., Garvey, D., Seibert, R., Hines, J.W. (2007). Validation of on-line monitoring techniques to nuclear plant data. *Nuclear Engineering Technologies*, 39 (2), pp. 149-158.

Gross, K.C., Singer, R.M., Wegerich, S.W., Herzog, J.P. (1997). Application of a model-based fault detection system to nuclear plant signals. *Proceedings of the intelligent system applications to power systems, ISAP, Seoul, Korea*, pp 66–70.

Hines, J.W., Uhrig, R.E., Wrest, D.J. (1998). Use of autoassociative neural networks for signal validation. *Journal of Intelligent and Robotic System*, 21, pp. 143-154.

Holschneider, M., Tchamitchian, P. (1989). Regularite locale de la fonction non-differentiable de Riemann. Les ondelettes. *Lecture notes in Mathematics***.** NewYork: Springer-Verlag.

Karacan, C.O., Olea, R.A. (2014). Inference of strata separation and gas emission paths in longwall overburden using continuous wavelet transform of well logs and geostatistical simulation. *Journal of Applied Geophysics*, 105, pp. 147-158.

Kovačević, J., Chebira, A. (2007). Life beyond bases: The advent of frames (Part I). *IEEE Signal Processing Magazine*, 24 (4), pp. 86-104.

Lee, J.-M., Yoo, C.K., Choi, S.W., Vanrolleghem, P.A., Lee, I.-B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59 (1), pp. 223-234.

Li, H. (2010). Gear Fault Diagnosis Based on Continuous Morlet Wavelet Amplitude and PhaseMap. *Seventh International Conference on Fuzzy Systems and Knowledge Discovery 2010*, 6, pp. 2619-2622.

Mallat, S., (2008). *A Wavelet Tour of Signal Processing.* Academic Press.

Mallat, S., Hwang, W.L. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38 (2), pp. 617-643.

Miao, Q., Huang, H.-Z., Fan, X. (2007). Singularity detection in machinery health monitoring using Lipschitz exponent function. *Journal of Mechanical Science and Technology*, 21 (5), pp. 737-744.

Ni, K., Ramanathan, N., Chehade, M.N.H., Balzano, L., Nair, S., Zahedi, S., Kohler, E., Pottie, G., Hansen, M., Srivastava, M. (2009). Sensor network data fault types. *ACM Transactions on Sensor Networks*, 5 (3), pp. 1-29.

Penha, R., Hines, J. (2001). Using principal component analysis modeling to monitor temperature sensors in a nuclear research reactor. *Maintenance and reliability conference (MARCON 2001).*

Qiu, H., Lee, J., Lin, J., Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, Volume 289, (4-5), pp. 1066-1090.

Rasmussen, B., Hines, J.W., Uhrig, R.E. (2000). Nonlinear partial least squares modeling for instrument surveillance and calibration verification. *Proceedings of the maintenance and reliability conference (MARCON 2000).*

Roverso, D., Hoffmann, M., Zio, E., Baraldi, P., Gola, G. (2007). Solutions for plant-wide on-line calibration monitoring. *Proceedings of the European Safety and Reliability Conference 2007, ESREL 2007 - Risk, Reliability and Societal Safety*.

Sengüler, T., Seker, S. (2016). Continuous wavelet transform for ferroresonance detection in power systems. *Electrical Engineering*, 99 (2), pp. 595-600.

Sharma, A.B., Golubchik, L., Govindan, R. (2010). Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Transactions on Sensor Networks*, 6 (3), art. no. 23.

Struzik, Z.R. (2001). Wavelet methods in (financial) time-series processing. *Physica A: Statistical Mechanics and its Applications*, 296 (1-2), pp. 307-319.

Tolle, G., Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., Burgess, S., Dawson, T., Buonadonna, P., Gay, D., and Hong, W. (2005). A Macroscope in the Redwoods. *Proceedings of the 2nd international conference on Embedded networked sensor systems (SenSys)*, pp. 51–63.

Torrence, C., Compo, G.P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79 (1), pp. 61-78.

Tu, C.-L., Hwang, W.-L., Ho, J. (2005). Analysis of singularities from modulus maxima of complex wavelets. *IEEE Transactions on Information Theory*, 51 (3), pp. 1049-1062.

Wegener, I., (2005). Complexity theory: Exploring the limits of efficient algorithms. *Complexity Theory: Exploring the Limits of Efficient Algorithms*, pp. 1-308.