Received November 6, 2017, accepted December 26, 2017, date of publication January 12, 2018, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2792838

Non-Stationary Process Monitoring for Change-Point Detection With Known Accuracy: Application to Wheels Coating Inspection

KARIM TOUT, FLORENT RETRAINT, AND RÉMI COGRANNE, (Member, IEEE)
Laboratory of System Modelling and Dependability, ICD, UMR 6281 CNRS, Troyes University of Technology, 10300 Troyes, France

Corresponding author: Karim Tout (karim.tout@utt.fr)

This work was supported by the Conseil Régional de Champagne-Ardenne through the project VISION under the industrial collaborative project framework under Grant E201314628.

ABSTRACT This paper addresses the problem of monitoring online a non-stationary process to detect abrupt changes in the process mean value. Two main challenges are addressed: First, the monitored process is non-stationary; i.e., naturally changes over time and it is necessary to distinguish those ''regular'' process changes from abrupt changes resulting from potential failures. Second, this paper aims at being applied for industrial processes where the performance of the detection method must be accurately controlled. A novel sequential method, based on two fixed-length windows, is proposed to detect abrupt changes with guaranteed accuracy while dealing with non-stationary process. The first window is used for estimating the non-stationary process parameters, whereas the second window is used to execute the detection. A study on the performances of the proposed method provides analytical expressions of the test statistical properties. This allows to bound the false alarm probability for a given number of observations while maximizing the detection power as a function of a given detection delay. The proposed method is then applied for wheels coating monitoring using an imaging system. Numerical results on a large set of wheel images show the efficiency of the proposed approach and the sharpness of the theoretical study.

INDEX TERMS Process control, statistical analysis, sequential analysis, parameter estimation, hypothesis testing theory, non-stationary process.

I. INTRODUCTION

In recent years, the change-point detection topic has been receiving increasing attention in various domains. It addresses the problem of detecting the point or multiple points at which a ''significant change'' occurs in a time series. These points are referred to as change points. The change-point detection process must be able to distinguish between a ''significant change'' indicating an abnormal event, and an ''insignificant change'' due to noise and that indicates a predicted or a normal behavior of data. Distinguishing change-points from spurious noise is very important in order to keep a false alarm rate. However surprisingly, sequential methods are hardly provided with established, or bounded, false-alarm probability and power functions.

A. STATE-OF-THE-ART

In general, change-point detection methods can be classified into ''offline'' and ''sequential'' methods. The choice of the appropriate class of methods depends heavily on the application.

''Offline'' methods, also referred to as retrospective methods, are considered in many applications, such as climate change study [1], biological applications [2], [3], econometric applications [4], and utility change in social media [5], to cite few topics. Such methods can only be applied after all the data, or observations, are received. Then, the objective is to detect all the change-points available in the data, along with estimating their locations. In applications for which these types of methods are used, the goal is usually to analyze time series and not to take immediate action after detecting the change points.

On the opposite, many other applications analyze data in real time with the goal to take an immediate response as soon as a change in the data is detected, as it can reveal a system failure which must be handled. In such cases, real-time data acquisition and analysis processes are required in order to raise an alarm as soon as a change-point is detected. Such problems fall within the scope of ''sequential'' methods, also referred to as online or real-time methods, in which it is assumed that the data are received sequentially, and that until a change-point is detected the process is allowed to continue. Contrariwise, when the data change, typically revealing a failure or a change in the underlying process, the aim is to detect the change-point with a minimal delay, in order to take the relevant actions, while also preserving a low false-alarm. Obviously, minimizing the detection delay and the false-alarm rate are contradictory goals. Sequential methods have been especially attracting attention from the industrial world, in which the term control chart is widely used, for quality control applications [6]–[8]. Industries have been pushing to produce higher quality and innovative products, which requires more and more manufacturing processes, while on the opposite, they are also required to reduce costs and production time. Hence, early fault detection for these industries is crucial to minimize downtime, reduce the product losses, and thus reduce manufacturing costs.

''Sequential'' change-point detection methods can be further categorized into ''parametric'' and ''non-parametric'' methods. On the one hand, ''non-parametric'' or data-driven methods have the advantage not to require any assumptions or any model on the data. They are based on statistical methods, especially supervised or non-supervised learning, to build detection rules based on large set of observations. Such decision rules are then applied to new data. While not requiring a model to describe the observations, those methods may, however, be limited, typically when the manufacturing process can largely change, and they are hardly provided with known statistical performances.

On the other hand, ''parametric'' methods are used when sufficient information on the monitoring process is available such that a statistical model of the observations can be designed. In other words, this approach requires that some distributional knowledge of the data is available and employed into the detection scheme. A common limitation of such methods is that they rely on pre-specified parametric models that are based on a priori information about the form of the data distribution, which is not always available.

B. CONTRIBUTION AND ORGANIZATION OF THIS PAPER

The present paper falls within the scope of ''parametric sequential'' with the goal to monitor a non-stationary process in real time in order to detect an abrupt change in its mean. In an industrial situation, it is required to detect the change within a given maximal detection delay (number of observations after the change) and it is wished to control the false-alarm probability over a fixed run length. In this operational context, a two fixed-length windows sequential method (2FLW-SEQ) based on the well-known CUSUM procedure is proposed for which the statistical performances are bounded. This sequential method is then applied for wheels coating monitoring. In fact, when a spray gun nozzle partially clogs, or gets blocked, this will be translated into a sudden change in the paint intensity caused by the lack of paint on the wheel.

The main contributions of the present paper are the following:

1) A two fixed-length windows sequential method (2FLW-SEQ) is proposed for monitoring a non-stationary process in real time. The first window is considered to deal with the non-stationarity of the process, while the second window is the one used for the sequential detection procedure.
2) The proposed sequential procedure operates under the non-classical criteria of minimizing the worst-case probability of missed detection under the constraint of a maximal detection delay, while controlling the false alarm probability for a given number of observations.
3) A statistical study of the proposed method is established that allows to lower bound the detection power as a function of the maximal allowed detection delay, and enables to upper bound the false alarm probability for a given number of observations.

One can note that the present submission is an extended version of the conference paper [9]. In comparison to this prior publication, this paper includes a statistical study that allows the calculation of an upper bound of false detection probability and a lower bound of power function. The sharpness of those theoretical findings is verified on a large dataset. This paper also includes a practical study on the impact of the parameters of the proposed methodology in order to select the most relevant ones.

The present paper is organized as follows. Section II briefly recalls the well-known cumulative sum (CUSUM) procedure [10] and states the problem of change-point detection for a non-stationary process emphasizing on the main difficulties and limitations of the CUSUM in this context. Next, section III presents the proposed method; first, the model used to deal with observations' non-stationarity is presented. Second, the ensuing statistical test is detailed. Then, to comply with requirements on low false alarm probability and highest change-point detection performance under a maximal delay constraint, the performance of the proposed method are studied in Section IV. Afterwards, section V presents the problem of paint coating intensity variation on produced wheels. Finally, Section VI presents numerical results obtained on a wide range of real data and studies the sharpness of the theoretical performance for the proposed method. Finally, Section VII concludes the paper.

II. CHANGE-POINT DETECTION PROBLEM STATEMENT

This section formally states the usual problem of abrupt change-point detection and recalls the well-known CUSUM method before highlighting the main particularity of the problem addressed in this paper.

The sequential change-point detection problem can be formulated as follows.[1] Let us consider $\{x_n\}_{n\geq 1}$ a sequence of

[1]The reader is referred to [6]–[8] for detailed introduction on sequential and change-point detection.

independent and identically distributed (*i.i.d*) observations that are acquired sequentially. At the beginning, the sequence is considered in a normal state, and the observations follow a probability distribution $\mathcal{P}_{\theta_0}$. Then, at an unknown point $v \geq 0$ (the change-point), the sequence reaches an abnormal state, in which the observations follow a different probability distribution $\mathcal{P}_{\theta_1}$. The problem formulation can be rewritten as follows:

$$x_n \sim \begin{cases} \mathcal{P}_{\theta_0} & \text{if } 1 \leq n \leq v, \\ \mathcal{P}_{\theta_1} & \text{if } n \geq v+1, \end{cases} \tag{1}$$

The sequential change-point detection consists of detecting the change-point v as soon as it occurs, while at the same time preserving a low false alarm rate.

For the online continuous inspection, for each new observation received, a decision rule is computed to test between the two following hypotheses:

$$\begin{cases} \mathcal{H}_0 : \{\theta = \theta_0\}, \\ \mathcal{H}_1 : \{\theta = \theta_1\}, \end{cases} \tag{2}$$

As long as the test (also called the stopping rule) fails to reject $\mathcal{H}_0$, the data acquisition continues. When the observations x_i are statistically independent, a usual approach to decide between the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ is to use the cumulative sum (CUSUM) procedure which can be defined, for observations up to N as follows [10]:

$$\delta_N = \begin{cases} 0 & \text{if } S_1^N = \max(S_1^{N-1} + s_N - \lambda; 0) < \tau, \\ 1 & \text{if } S_1^N = \max(S_1^{N-1} + s_N - \lambda; 0) \geq \tau, \end{cases} \tag{3}$$

where δ_N is the CUSUM statistical test, λ is a constant that avoids spurious false-alarms, τ is a conveniently pre-defined threshold and, for initialization, $S_1^0 = 0$. Though the decision statistics s_N and the constant λ were not defined in [10], the logarithm of the well-known likelihood ratio is commonly used:

$$s_n = \log\left(\frac{p_{\theta_1}(x_n)}{p_{\theta_0}(x_n)}\right), \tag{4}$$

where p_{θ_0} and p_{θ_1} are the probability density functions (PDF) under hypotheses associated with distributions $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ respectively, which are assumed to be known, and the constant λ is usually the average of the expected values $\lambda = 1/2\left(\mathbb{E}_{\mathcal{H}_0}[s] + \mathbb{E}_{\mathcal{H}_1}[s]\right)$.

A. DIFFICULTIES OF NON-STATIONARITY AND CRITERION OF OPTIMALITY

In the present paper, the studied process is non-stationary in the mean. As a consequence, the problem of detecting an abrupt change in an i.i.d random sequence is not relevant anymore because (1) the hypotheses are composite, that is they are characterized by a set of possible parameters Θ_0 and Θ_1 and (2) for observation x_n the PDFs p_{θ_0} and p_{θ_1} are unknown. In fact, when monitoring a non-stationarity process whose distribution parameters may "naturally" change over time, the change-point detection problem as stated in (1)–(2) is no longer relevant. Indeed, since under the hypothesis $\mathcal{H}_0$ the distribution parameter θ_0 may change within the set Θ_0, the hypotheses are defined by:

$$\begin{cases} \mathcal{H}_0 : \{\theta \in \Theta_0\}, \\ \mathcal{H}_1 : \{\theta \in \Theta_1\}, \end{cases} \tag{5}$$

and one should instead consider the following sequential test problem:

$$x_n \sim \begin{cases} \mathcal{P}_{\theta_{0,n}}, & \theta_{0,n} \in \Theta_0 \text{ if } 1 \leq n \leq v, \\ \mathcal{P}_{\theta_{1,n}}, & \theta_{1,n} \in \Theta_1 \text{ if } n \geq v+1, \end{cases} \tag{6}$$

The main issue to tackle those scientific difficulties is to have an accurate model of Θ_0 and Θ_1; in other words, to be able to model with enough accuracy the set of "regular" changes in the process of the abrupt changes that reveals a malfunctioning.

Regarding the scientific difficulties, when the distribution parameters $\theta_{0,n}$ and $\theta_{1,n}$ are unknown, in such a context the likelihood ratio (4) cannot be calculated for a given observation x_n. A usual solution that is adopted in the present paper is to use a generalized likelihood ratio that consists of substituting unknown parameters $\theta_{0,n}$ and $\theta_{1,n}$ by their estimations using the maximum likelihood estimation.

The second main challenge addressed in the present paper is the introduction of an unusual criterion of optimality. Indeed the CUSUM has been shown to be asymptotically optimal with respect to the criterion that consists in minimizing the average worst case detection, see [23]–[25] for details on the so-called Lorden's criterion and CUSUM optimality.

However, a minimal average delay is not equivalent to a maximal detection accuracy for a given detection delay. Focusing on a practical industrial context, the present paper aims at maximizing the probability of change-point detection for a fixed maximal delay; this is justified for cost-reduction purposes as the change point corresponds in practice to a malfunction in a production process.

III. PROPOSED CHANGE-POINT DETECTION METHOD

As discussed in section II-A, the purpose of this article is to design a change point detection method in the case of a non-stationary process with a constraint on the maximal detection delay. This section first presents how to deal with the process non-stationarity that represents a nuisance parameter; then the novel two fixed-length windows sequential method (2FLW-SEQ) is presented, that fits with the constraint on the detection delay, and rejects online the nuisance parameter generated by the process non-stationarity.

To facilitate the theory elaboration of the proposed method, Table 1 provides a nomenclature that lists all the important symbols that are to be used in what follows.

A. PROCESS MODELING

Let us consider a sliding window of size L. After the first L observations, for each new received data x_N, the window slides by one point to contain the observations from x_{N-L+1}

TABLE 1. Nomenclature.

x_i	Observation of index i
N	Index of the last acquired observation
L	Length of the first window
Y_N	Vector containing the last L observations
μ_N	Expectation of vector Y_N
$\mathbf{I}_L$	Identity matrix of size L
σ^2	Constant variance of all vectors Y_N, $Y_N, \forall N \geq L$
$\mathbf{H}$	Expectation model
q	Number of basis vectors of the model $\mathbf{H}$
$\mathbf{d}_N$	Contributions of the q basis vectors
M	Length of the second window (maximal detection delay)
K_M	Vector of size L that describes the change in mean value
a	Amplitude of the change
$\mathbf{W}$	Orthogonal projection onto the null space of $\mathbf{H}$
$\mathbf{r}_N$	Vector of residuals after the projection of Y_N with $\mathbf{W}$
$\boldsymbol{\theta}_M$	Projection of K_M with $\mathbf{W}$
δ_N	Decision rule of the proposed statistical test
τ	Detection threshold
$\widetilde{S}_{N-L+1}^{N}$	Results of the proposed method
T	Stopping time
v	Change point index
$\mathbb{P}_{md}(M)$	Worst-case probability of missed detection
$\beta(M)$	Detection power
$\widetilde{\beta}(M)$	Lower bound of the detection power
Φ	Standard normal cumulative distribution function
R	Run length
$\mathbb{P}_{fa}(R)$	Worst-case probability of false alarm
$\alpha(R)$	False alarm probability
$\widetilde{\alpha}(R)$	Upper bound of the false alarm probability
H_D	Hellinger distance

to x_N. Let $Y_N = (x_{N-L+1}, \ldots, x_{N-1}, x_N)^T$ denotes this window after the reception of observation x_N. The vector Y_N is modeled with the following normal distribution:

$$\mathbf{Y}_N \sim \mathcal{N}(\mu_N, \sigma^2 \mathbf{I}_L), \tag{7}$$

where μ_N is the expectation in this window, $\mathbf{I}_L$ is the identity matrix of size L, and σ^2 is the variance which is assumed constant for all windows $Y_N, \forall N \geq L$.

A linear parametric model is proposed to represent the expectation μ_N. It essentially consists in representing all the observations in the window Y_N as a weighted sum of q basis vectors that represent the columns of a matrix $\mathbf{H}$ of size $L \times q$. The weight of this sum represents the vector of q parameters $\mathbf{d}_N$. Hence, the expectation μ_N can be written as:

$$\mu_N = \mathbf{H}\mathbf{d}_N. \tag{8}$$

In this paper, the model of $\mathbf{H}$ is based on the following algebraic polynomial:

$$h(x) = \sum_{j=0}^{q-1} d_j x^j, \tag{9}$$

with $q-1$ the degree of the algebraic polynomial. The use of a linear parametric model in statistical testing theory has been widely exploited [13]. One can note that we have used such approach of polynomial image modeling in some of our prior work on image processing [14]–[16] and especially for the detection of defects on wheels' surface [17], [18]. However, here it is used in a simplistic manner within a sequential detection method to remove the possible slight "natural" intensity changes that are not abnormal and should thus be removed.

It follows from Eqs. (7) and (8) that in the absence of any anomaly, the vector of observations Y_N is modeled by:

$$\mathbf{Y}_N \sim \mathcal{N}(\mathbf{H}\mathbf{d}_N, \boldsymbol{\sigma}^2 \mathbf{I}_L). \tag{10}$$

On the opposite, when a defect happens in the process, a change occurs in the mean value which will affect all the observations after the change-point. Consequently, when the change occurs, the observations Y_N can be modeled as:

$$\mathbf{Y}_N \sim \mathcal{N}(\mathbf{H}\mathbf{d}_N + a\mathbf{K}_M, \boldsymbol{\sigma}^2 \mathbf{I}_L), \tag{11}$$

where the sudden shift in the mean value is described by the vector K_M, of size L, containing $L - M$ zeros before the change occurs and minus ones M times after, and the constant $a > 0$ represents the amplitude of the change. Here, M is the number of maximal acceptable observations with defects. For example, the change vector $K_1 = (0, 0, \ldots, 0, -1)$ describes a change that only affects the last observation in the window of size L.

It is important to note that the "acceptable" variation of mean value, modeled by $\mathbf{H}\mathbf{d}_N$, is a nuisance parameter as it is of no use for the considered detection problem. To deal with this nuisance parameter, it is proposed to use the maximum likelihood (ML) estimation method to perform a rejection of this nuisance parameter as follows:

$$\mathbf{r}_N = \frac{1}{\sigma} \mathbf{W}\mathbf{Y}_N. \tag{12}$$

Here $\mathbf{W}$ is the orthogonal projection of size $L - q \times L$, onto the null space of $\mathbf{H}$, whose columns correspond to the eigenvectors of the matrix $\mathbf{I}_L - \mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}$ associated with eigenvalues equal to 1. The vector $\mathbf{r}_N$ represents the projection of the observations onto the null space of $\mathbf{H}$.

B. 2FLW-SEQ PROCEDURE

Among others, the matrix $\mathbf{W}$ has the following useful properties: $\mathbf{W}\mathbf{W}^T = \mathbf{I}_{L-q}$; it thus follows from Eqs. (10)-(12), that the residuals $\mathbf{r}_N$ can be modeled under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ by the following statistical distribution:

$$\begin{cases} \mathcal{H}_0 : \left\{\mathbf{r}_N \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{L-q})\right\} \\ \mathcal{H}_1 : \left\{\mathbf{r}_N \sim \mathcal{N}\left(\dfrac{a}{\sigma}\boldsymbol{\theta}_M, \mathbf{I}_{L-q}\right)\right\}, \end{cases} \tag{13}$$

where $\boldsymbol{\theta}_M$ represents the shift of expectation, due to the process failure, projected onto the null space of $\mathbf{H}$: $\boldsymbol{\theta}_M = \mathbf{W}\mathbf{K}_M$.

From the definition of the hypotheses in Eq. (13), after the rejection of the nuisance parameter $\mathbf{H}\mathbf{d}_N$, it is obvious that the considered detection problem essentially consists in the detection of a specific signal in noise. In this paper, it is proposed to use a sequential method with a fixed window of length M which also corresponds to a pre-defined fixed maximal detection delay. Similar approaches have been studied in the context of sequential detection in [19] and [20]. They proposed to use the well-known match space detection which is given in our case by:

$$\delta_N = \begin{cases} 0 & \textit{if } \widetilde{S}_{N-L+1}^{N} = \boldsymbol{\theta}_M^T \mathbf{r}_N < \tau \\ 1 & \textit{if } \widetilde{S}_{N-L+1}^{N} = \boldsymbol{\theta}_M^T \mathbf{r}_N \geq \tau. \end{cases} \tag{14}$$

From Eq. (13) it is straightforward to establish the statistical distribution of results $\widetilde{S}_{N-L+1}^{N}$ of the proposed 2FLW-SEQ:

$$\begin{cases} \mathcal{H}_0 : \left\{ \widetilde{S}_{N-L+1}^{N} \sim \mathcal{N}(\mathbf{0}, \|\boldsymbol{\theta}_M\|_2^2) \right\} \\ \mathcal{H}_1 : \left\{ \widetilde{S}_{N-L+1}^{N} \sim \mathcal{N}\left(\frac{a}{\sigma}\|\boldsymbol{\theta}_M\|_2^2, \|\boldsymbol{\theta}_M\|_2^2\right) \right\}. \end{cases} \tag{15}$$

which can be normalized, for the sake of clarity, as follows:

$$\begin{cases} \mathcal{H}_0 : \left\{ \dfrac{\widetilde{S}_{N-L+1}^{N}}{\|\boldsymbol{\theta}_M\|_2} \sim \mathcal{N}(\mathbf{0}, 1) \right\} \\ \mathcal{H}_1 : \left\{ \dfrac{\widetilde{S}_{N-L+1}^{N}}{\|\boldsymbol{\theta}_M\|_2} \sim \mathcal{N}(\frac{a}{\sigma}\|\boldsymbol{\theta}_M\|_2, 1) \right\}. \end{cases} \tag{16}$$

It is important to note that the choice of the first window size L and the polynomial degrees $q-1$ is crucial and essentially depends on the observations. First, L must be much greater than M in order to avoid any significant impact of the abrupt change on the estimate of the linear model parameters $\mathbf{d}_N$. On the opposite, L must remain reasonably small such that the linear model will well model the observations' expectation and to ensure that the residuals $\mathbf{r}_N$ follow a standard normal distribution under $\mathcal{H}_0$.

As for q, it is the opposite scenario. Indeed, high polynomial degrees may lead to the shift being eliminated with the projection (12), and thus removed from the residuals. On the other hand, very small polynomial degrees may not be sufficient to properly model the process, and thus putting parts of the healthy observations among the residuals, and probably losing the standard normal distribution under $\mathcal{H}_0$.

IV. ASSESSMENT OF 2FLW-SEQ STATISTICAL PROPERTIES

A sequential change-point detection procedure stops as soon as its decision rule δ_n becomes 1. Then, the stopping time T is defined as the smallest observation index n for which $\delta_n = 1$. A correct change detection consists in stopping the sequential procedure after the change has occurred, which means $T \geq v$ where v is the change point index. A false alarm is raised in case where $T < v$, i.e. the process has been stopped before the change occurred. A usual criteria for a sequential procedure is to detect the change as soon as it occurs, thus minimizing the detection delay $T-v$. Many criteria have been used to investigate the optimality of change point detection algorithms concerning the detection delay, as the "mean delay", the "conditional mean delay", the "worst mean delay", etc.... [10], [21]–[23]. In that context, the CUSUM algorithm has been proven to be optimal in [23]–[25]. However, in the proposed detection scheme, the goal is to fix a detection delay after which the change detection is considered too late. In fact, minimizing the detection delay does not necessarily lead to a higher detection power, or to a small probability of missed detection. Therefore, the aim of the proposed sequential method is to minimize the worst-case probability of missed detection under constraint on the worst-case probability of false alarm for a given run length.

A. MINIMIZING THE PROBABILITY OF MISSED DETECTION

The stopping time of the classical CUSUM procedure is given by:

$$T_c = \inf_{n\geq 1}\{n : \max_{1\leq k\leq n} \widetilde{S}_k^n \geq \tau\} \tag{17}$$

In this context, the CUSUM procedure takes into account all previous observations. However, for the proposed sequential method, after collecting the first L observations, the stopping time can be defined as:

$$T_{2FLW} = \inf_{n\geq L}\{n : \widetilde{S}_{n-L+1}^n \geq \tau\} \tag{18}$$

The probability of missed detection can be considered as the probability that the detection delay is higher than the acceptable one defined as M, knowing that the detection is made after the change has occurred with $T \geq v$. Then, to the purpose of minimizing the probability of missed detection, the following criteria can be applied:

$$\mathbb{P}_{md}(M) = \sup_{v\geq L} \mathbb{P}(T - v + 1 > M \mid T \geq v) \tag{19}$$

where $\mathbb{P}_{md}(M)$ is the worst-case probability of missed detection. Minimizing this probability will lead to maximizing the detection probability denoted as $\beta(M) = 1 - \mathbb{P}_{md}(M)$.

Eq. (19) can be developed to:

$$\mathbb{P}_{md}(M) = \sup_{v\geq L} \frac{\mathbb{P}\left(\bigcap_{n=L}^{M+v-1} \{\widetilde{S}_{n-L+1}^n < \tau\}\right)}{\mathbb{P}\left(\bigcap_{n=L}^{v-1} \{\widetilde{S}_{n-L+1}^n < \tau\}\right)} \tag{20}$$

It is complicated to calculate the exact value of $\mathbb{P}_{md}(M)$, instead it is proposed to calculate an upper bound. It can be seen that:

$$\mathbb{P}\left(\bigcap_{n=L}^{M+v-1} \{\widetilde{S}_{n-L+1}^n < \tau\}\right) \leq \mathbb{P}\left(\left\{\bigcap_{n=L}^{v-1} \{\widetilde{S}_{n-L+1}^n < \tau\}\right\} \bigcap \left\{\widetilde{S}_{M+v-L}^{M+v-1} < \tau\right\}\right) \tag{21}$$

Note that in Eq. (21), the two events have common observations of indexes $(M+v-L, \ldots, v-1)$. In order to calculate the result $\widetilde{S}_{M+v-L}^{M+v-1}$, observations of indexes $(M+v-L, \ldots, M+v-1)$ have been projected onto the null space of the model matrix $\mathbf{H}$, and then the resulting residuals have been multiplied by $\boldsymbol{\theta}_M$ which represents the shift of expectation, due to the process failure, projected onto the null space of $\mathbf{H}$. Because all the common observations are healthy observations, as they are acquired before the change v, their effect is neglected when multiplied by $\boldsymbol{\theta}_M$. Following that, the two events can be considered as independent, and Eq. (21) can be written as:

$$\mathbb{P}\left(\bigcap_{n=L}^{M+v-1}\{\widetilde{S}_{n-L+1}^{n}<\tau\}\right) \leq \mathbb{P}\left(\left\{\bigcap_{n=L}^{v-1}\{\widetilde{S}_{n-L+1}^{n}<\tau\}\right\}\right)\cdot\mathbb{P}\left(\left\{\widetilde{S}_{M+v-L}^{M+v-1}<\tau\right\}\right) \quad (22)$$

Then, from Eq. (20), we get:

$$\mathbb{P}_{md}(M) \leq \mathbb{P}\left(\left\{\widetilde{S}_{M+v-L}^{M+v-1}<\tau\right\}\right) = \mathbb{P}_{\mathcal{H}_1}\left(\left\{\widetilde{S}_{1}^{L}<\tau\right\}\right) \quad (23)$$

where $\mathbb{P}_{\mathcal{H}_1}$ is the probability under $\mathcal{H}_1$. Based on (15), under $\mathcal{H}_1$, the result $\widetilde{S}_1^L$ is a Gaussian random variable with mean $\frac{a}{\sigma}\|\boldsymbol{\theta}_M\|_2^2$ and variance $\|\boldsymbol{\theta}_M\|_2^2$. As a result, the worst-case probability of missed detection can be upper bounded as:

$$\mathbb{P}_{md}(M) \leq \Phi\left(\frac{\tau}{\|\boldsymbol{\theta}_M\|_2} - \frac{a}{\sigma}\|\boldsymbol{\theta}_M\|_2\right) \quad (24)$$

with Φ the standard normal cumulative distribution function.

Finally, the power function $\beta(M)$ of the proposed test (15), that is the probability of detecting a failure after at most M observations, is bounded by:

$$\beta(M) \geq 1 - \Phi\left(\frac{\tau}{\|\boldsymbol{\theta}_M\|_2} - \frac{a}{\sigma}\|\boldsymbol{\theta}_M\|_2\right). \quad (25)$$

In what follows, this lower bound will be referred to as $\widetilde{\beta}(M)$.

B. WORST-CASE PROBABILITY OF FALSE ALARM

On the other hand, for a given run length R and at a given time ℓ, the false alarm probability is given by:

$$\mathbb{P}_0(\ell \leq T \leq \ell + R) \quad (26)$$

Hence, the worst-case probability of false alarm for all $\ell \geq L$ can be defined as:

$$\mathbb{P}_{fa}(R) = \sup_{\ell \geq L} \mathbb{P}_0(\ell \leq T \leq \ell + R) \quad (27)$$

The calculation of the exact value of $\mathbb{P}_{fa}(R)$ is absurd, instead it is proposed to calculate an upper bound only. In this way, it is possible to guarantee a false alarm rate lower than that bound for all $\ell \geq L$.

The calculation will be done in two steps. First, the proof that the worst-case probability of false alarm is indeed the probability of false alarm at the starting point L. And then, the second step is to determine the upper bound.

First, let us start the proof of the following equality:

$$\mathbb{P}_{fa}(R) = \sup_{\ell \geq L} \mathbb{P}_0(\ell \leq T \leq \ell + R) = \mathbb{P}_0(L \leq T \leq L + R) \quad (28)$$

Let us denote $U_\ell = \mathbb{P}_0(T = \ell)$ for all $\ell \geq L$. For the first point L, it can be clearly seen that:

$$U_L = \mathbb{P}_0(\widetilde{S}_1^L \geq \tau) \quad (29)$$

and that:

$$U_{L+1} = \mathbb{P}_0\left(\left\{\widetilde{S}_1^L < \tau\right\}\bigcap\{\widetilde{S}_2^{L+1} \geq \tau\}\right) \leq \mathbb{P}_0\left(\left\{\widetilde{S}_2^{L+1} \geq \tau\right\}\right) \quad (30)$$

As all the observations of indexes $(1, \ldots, L+1)$ follow the same distribution under $\mathcal{H}_0$, then the inequality in Eq. (30) can be rewritten as:

$$U_{L+1} \leq \mathbb{P}_0\left(\left\{\widetilde{S}_1^L \geq \tau\right\}\right) = U_L \quad (31)$$

In a similar manner, for $\ell > L$, we can verify that:

$$U_\ell = \mathbb{P}_0\left(\bigcap_{n=L}^{\ell-1}\{\widetilde{S}_{n-L+1}^{n} < \tau\}\bigcap\left\{\widetilde{S}_{\ell-L+1}^{\ell} \geq \tau\right\}\right) \quad (32)$$

and that:

$$\begin{aligned} U_{\ell+1} &= \mathbb{P}_0\left(\bigcap_{n=L}^{\ell}\{\widetilde{S}_{n-L+1}^{n} < \tau\}\bigcap\left\{\widetilde{S}_{\ell-L+2}^{\ell+1} \geq \tau\right\}\right) \\ &\leq \mathbb{P}_0\left(\bigcap_{n=L+1}^{\ell}\{\widetilde{S}_{n-L+1}^{n} < \tau\}\bigcap\left\{\widetilde{S}_{\ell-L+2}^{\ell+1} \geq \tau\right\}\right) \\ &\leq \mathbb{P}_0\left(\bigcap_{n=L}^{\ell-1}\{\widetilde{S}_{n-L+1}^{n} < \tau\}\bigcap\left\{\widetilde{S}_{\ell-L+1}^{\ell} \geq \tau\right\}\right) = U_\ell \end{aligned} \quad (33)$$

Therefore, it is concluded that $(U_\ell)_{\ell \geq L}$ is a decreasing sequence. Now let us define $V_\ell = \mathbb{P}_0(\ell \leq T \leq \ell + R)$ for all $\ell \geq L$. It can be seen that:

$$V_\ell = \sum_{n=\ell}^{\ell+R-1} \mathbb{P}_0(T = n) = \sum_{n=\ell}^{\ell+R-1} U_n \quad (34)$$

Then:

$$V_\ell - V_{\ell+1} = \sum_{n=\ell}^{\ell+R-1} U_n - \sum_{n=\ell+1}^{\ell+R} U_n = U_\ell - U_{\ell+R} \geq 0 \quad (35)$$

Consequently, $(V_\ell)_{\ell \geq L}$ is also a decreasing sequence. As a result, the equality in Eq. (28) is proven to be correct:

$$\sup_{\ell \geq L} V_\ell = V_L = \mathbb{P}_0(L \leq T \leq L + R) = \mathbb{P}_{fa}(R) \quad (36)$$

The second step consists in calculating the upper bound of V_L. From Eq. (29), U_L can be rewritten as:

$$U_L = 1 - \mathbb{P}_0(\widetilde{S}_1^L < \tau) \quad (37)$$

Similarly for all $\ell > L$, Eq. (32) can be rewritten as:

$$\begin{aligned} U_\ell &= \mathbb{P}_0\left(\bigcap_{n=L}^{\ell-1}\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \\ &\quad - \mathbb{P}_0\left(\bigcap_{n=L}^{\ell-1}\{\widetilde{S}^n_{n-L+1} < \tau\}\bigcap\left\{\widetilde{S}^\ell_{\ell-L+1} < \tau\right\}\right) \\ &= \mathbb{P}_0\left(\bigcap_{n=L}^{\ell-1}\{\widetilde{S}^n_{n-L+1} < \tau\}\right) - \mathbb{P}_0\left(\bigcap_{n=L}^{\ell}\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \end{aligned} \tag{38}$$

It follows from Eqs. (37), (38), and (34), that the worst-case probability of false detection V_L is:

$$V_L = 1 - \mathbb{P}_0\left(\bigcap_{n=L}^{L+R-1}\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \tag{39}$$

For any two positive integers $n \neq n'$, it is possible to prove that the covariance of the two Gaussian variables $\widetilde{S}^n_{n-L+1}$ and $\widetilde{S}^{n'}_{n'-L+1}$ is non-negative $cov\left(\widetilde{S}^n_{n-L+1}, \widetilde{S}^{n'}_{n'-L+1}\right) \geq 0$. As a consequence, one can immediately get:

$$\mathbb{P}_0\left(\bigcap_{n=L}^{L+R-1}\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \geq \prod_{n=L}^{L+R-1}\mathbb{P}_0\left(\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \tag{40}$$

Thus, V_L is upper bounded by:

$$V_L \leq 1 - \prod_{n=L}^{L+R-1}\mathbb{P}_0\left(\{\widetilde{S}^n_{n-L+1} < \tau\}\right) \tag{41}$$

Finally, based on (15), under $\mathcal{H}_0$, the results $\widetilde{S}^n_{n-L+1}$ $\forall n \geq L$ are Gaussian random variables with zero mean and variance $\|\boldsymbol{\theta}_M\|_2^2$. As a result, the probability of having a false alarm $\alpha(R)$ after R observations is bounded by:

$$\alpha(R) \leq 1 - \Phi\left(\frac{\tau}{\|\boldsymbol{\theta}_M\|_2}\right)^R, \tag{42}$$

In what follows, this upper bound will be referred to as $\widetilde{\alpha}(R)$.

Equations (25) and (42) emphasize the main advantages of the proposed approach. First, the statistical performance of the proposed test is bounded. The false alarm probability $\alpha(R)$ is upper bounded which will enable to calculate a detection threshold τ using a pre-defined false alarm rate knowing that the application is guaranteed not to exceed. On the other hand, the detection power $\beta(M)$ of the test is lower bounded which will allow to guarantee, for a pre-defined false alarm rate, a minimal detection power that the application will not decrease bellow. Second, the false alarm probability $\alpha(R)$ only depends on the prescribed run-length R and the maximal acceptable detection delay M. Last, the power function (25) shows that the accuracy of the proposed method essentially depends on the ''change-to-noise ratio'' a/σ, along with the maximal acceptable detection delay M.

V. PAINT COATING INTENSITY VARIATION

A. PAINTING PROCESS

Wheel paint has two purposes; to protect the underlying metal from the harsh environment to which it is exposed, and most importantly to improve the look of the wheel. Modern wheel coating methods consist of five main steps, starting with the pre-treatment which removes and cleans excess metal to form a smooth surface structure, and ending with the topcoats which provide the surface characteristics including color, appearance, gloss, smoothness, and weather resistance [11]. This paper focuses on the topcoats as they are the only visible layer.

Wheel topcoats are usually composed of several layers of paint coatings, with a precise thickness, spread on the whole surface of the wheel one after another [12]. They are generally applied in the form of liquid or powder using spray atomizers, also called spray gun nozzles [12]. The appearance (color, gloss, texture, etc...) of a coated surface greatly affects the perception on the product quality. In fact, every wheel manufacturer has a list of client requirements that defines every detail concerning the final product, including a ''top-coat requirements'' list that contains specifications about the color, the gloss level, and many other aspects of the topcoats. Given this set of specifications, any significant deviation from what is standard or normal to the product is considered an anomaly that has to be correctly detected. However, it is important to note that in this context a defective process will not only affect one wheel, but all of the following products. Therefore, a fast and accurate detection of any anomaly, as soon as it appears, is necessary in order to reduce the number of defective products, thus reducing the loss. Moreover, the deviation that is considered as anomalous is hardly distinguishable from other normal deviations, and hence may remain unnoticed by visual inspection.

All those points lead us to the necessity of an automatic inspection system that monitors the variations of the topcoat intensity, and signals the change-point with minimal delay time. The detection process has to be fast and sufficiently efficient in order to distinguish between a normal state and the anomalous state.

Technically speaking, many factors influence the quality of the coating, thus its appearance, such as temperature, paint viscosity, solvents, etc. [11], [12] .... Specifically for liquid painting, as time goes by, the viscosity of the paint in the paint bath decreases (the paint becomes more pasty) since the solvents are evaporating over time. This process may be faster or slower depending on the neighboring temperature [11], [12]. To rectify the effects of this process, usually the operators tend to increase the paint / airflow on the spray gun nozzle. These variations in the topcoats remain in the acceptable zone in accordance with the technical requirements. This paper focuses on a usual problem, that is when the spray gun nozzle partially clogs, or gets blocked, which will be translated in a sudden change in the intensity of the topcoats.

B. MONITORING SYSTEM

The monitoring system consists of an imaging system placed over the conveyor belt of a wheel factory, just after the painting process. It involves a camera that takes the image of each produced wheel, using a proper illumination setup, to uniformly brighten the whole surface of the wheel while reducing light reflection artifacts. An example of a wheel image acquired by the imaging system is shown in Figure 1.

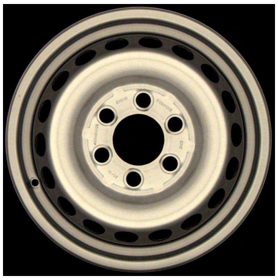

FIGURE 1. A typical example of a wheel image obtained from the monitoring system.

To the purpose of monitoring the variation of paint coating intensity on produced wheels, it is wished to consider a block containing s pixels in the image of the wheel, over which the mean value of all pixels is computed. The considered window maintains the same size and position on the surface of the wheel for all images. Then, for one image of a wheel, let $Z = \{z_w\}_{w=1}^{s}$ denote the window containing s pixels and $x = s^{-1}\sum_{w=1}^{s} z_w$ the mean value of pixels' intensity. Note that the behavior of the observations is independent from the window position on the surface of the wheel.

The variation of the mean value x from a wheel image to another describes the variation of the topcoat intensity. Indeed, the mean value is a sufficient parameter to detect coating failure as the change in pixel values that it causes affects the whole surface of the wheel. Figure 2 shows an example of series of mean pixels' value x_i for 1 000 images of consecutive wheels without change points, with i the image index. The observed variation in the mean values is considered to be normal, and it is due to the reasons detailed previously. It is shown that the mean value of observations x_i evolves smoothly.

Note that, the window Z has always the same position from the center of the wheel, but not exactly the same position on the wheel image. In fact, the wheels are not perfectly centered under the imaging system, which means that from an image to another, the position of the wheel may differ by few pixels. In addition, the illumination system is not ideal, meaning that the distribution of light over the whole wheel surface is not perfectly uniform, hence some locations on the wheel are slightly more or less illuminated than others. Therefore, it is concluded that the variance of the variable x_i is only related to the imaging system which is not modified during the acquisition, thus it remains constant for all observations, whether before or after the change. Based on these factors, and based on the behavior of the variable x_i observed

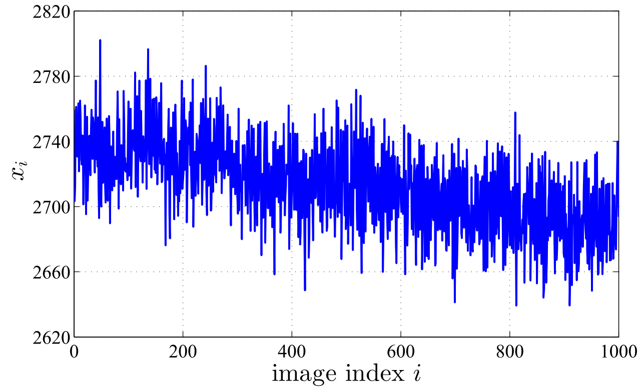

FIGURE 2. A typical example of variation of wheel images mean value.

in Figure 2, the process can be considered as a non-stationary process in the mean, with a constant variance over all the observations.

VI. EXPERIMENTS AND RESULTS

In this section, five types of results are presented. First, the proper choice of the first window length L and the degree of the polynomial $q-1$ is discussed with simulation results. The second experiment aims to study the effect of the second window length M on the performances of the proposed test. In the third part, it is wished to examine the efficiency of the bounds calculated in subsection IV and to study the detectability of the proposed test function of the abrupt change amplitude, given a set of requirements. Next, the fourth experiment is a study of a real case scenario with a real change point in the observations. Finally, a performance comparison is conducted to highlight the advantages of modeling the observations and examine the difference in the detection criteria between our approach and the CUSUM method.

To conduct these experiments, a data base of 500 000 successive healthy images has been acquired using an area scan camera installed over the production line of a wheel factory. The acquired images are made of 2046×2046 pixels of 12 bits depth. The procedure described in section V has been applied to obtain the observations x_i with $i = \{1, 2, \ldots, 500\,000\}$. The observed standard deviation, related to the imaging system, is $\sigma = 22$. As supposed in section V, this parameter is assumed to be constant during the monitoring process. However, the variance can be changed with the acquisition conditions, for instance, with the illumination intensity. To deal with the problem of imaging acquisition system drift, the variance is periodically computed (typically at the beginning of each week). The detection problem of abrupt changes in acquisition conditions is not addressed in this paper.

First, let us start by discussing the choice of the first window length L, and the degree of the polynomial $q-1$. In fact, as mentioned in subsection III-B, the choice of parameters L and q has an important role, on the one hand, to increase the detection performances of the test, and, on the

TABLE 2. The empirical detection power $\beta(M)$ and the Hellinger distance H_D for different values of L and q.

		First window size L											
		50		100		150		200		250		1 000	
		$\beta(M)$	H_D	$\beta(M)$	H_D	$\beta(M)$	H_D	$\beta(M)$	H_D	$\beta(M)$	H_D	$\beta(M)$	H_D
Degrees of the polynomial $q-1$	2	0.4005	0.0405	0.8079	0.0426	0.8481	0.0433	0.9351	0.0443	0.8741	0.0451	0.5387	0.0564
	3	0.0559	0.0387	0.5135	0.0417	0.7549	0.0423	0.8020	0.0438	0.8775	0.0444	0.7049	0.0563
	4	0.0455	0.0373	0.2813	0.0409	0.4185	0.0415	0.6594	0.0429	0.8052	0.0438	0.8416	0.0463
	5	0.0441	0.0354	0.0601	0.0407	0.2378	0.0410	0.3493	0.0422	0.5582	0.0434	0.8706	0.0381
	10	0.0040	0.0264	0.0217	0.0364	0.0523	0.0395	0.0425	0.0414	0.0524	0.0420	0.6525	0.0350

other hand, to correctly model the paint coat intensity process as a Gaussian process. Hence, multiple Monte-Carlo simulations, with different values of L and q, have been performed to correctly tune these parameters to ensure the best performances. Two important factors are directly affected by the change in these parameters, those are the detection power $\beta(M)$ and the accuracy of the standard Gaussian distribution model for residuals' r_N distribution under $\mathcal{H}_0$. This accuracy can be expressed using the Hellinger distance H_D between the empirical residuals r_N and the theoretical standard Gaussian distribution. Table 2 contains the calculated values of $\beta(M)$ and H_D for values of L ranging from 50 to 1 000, and values of $q-1$ ranging from 2 to 10. These results have been obtained for a maximal detection delay $M = 5$, over a run length of $R = 5\,000$ which represents about half of a day's production, and for a pre-defined value of false alarm rate $\alpha(R) = 10^{-2}$. It can be observed from Table 2 that for a certain polynomial degree, increasing L will lead to a better detection performance as $\beta(M)$ increases, however, the Hellinger distance H_D increases alongside which indicates a decrease in accuracy. For large values of L, as $L = 1\,000$, small values of polynomial degree are not even sufficient to correctly represent the observations under $\mathcal{H}_0$, which can be seen by the increase in H_D and the decrease in $\beta(M)$. Thus it is necessary to increase the polynomial degree just to correctly model the observations. On the other side, for a certain value of L, increasing q will lead to an increase in the accuracy, in favor of a decreasing performance. For large values of q, as $q = 10$, the accuracy increases significantly, however, the test performance is low. To choose the optimal values of L and q, it is important to have the maximal detection power $\beta(M)$ alongside a sufficient accuracy so that the empirical performance matches at best the theoretical performance study. For the values $L = 200$ and $q-1 = 2$, we have the best detection power $\beta(M) = 0.9351$. Then, to better understand the relation between the Hellinger distance and the accuracy, figure 3 represents a comparison between the theoretical standard Gaussian cumulative distribution function (cdf) and the empirical cumulative distribution functions of the residuals r_N for $L = 200$ and $L = 1\,000$, and with $q-1 = 2$. It can be seen that for $L = 200$, the empirical distribution is accurate enough compared to the theoretical

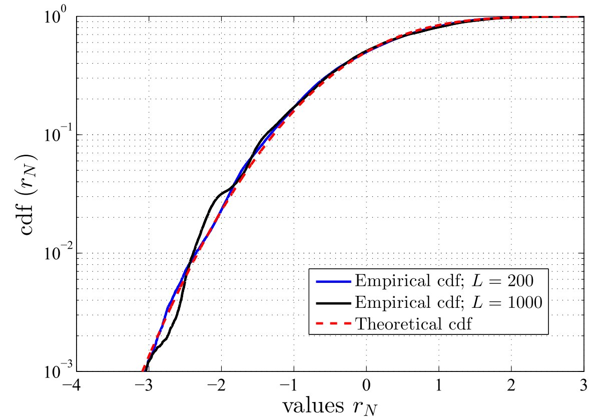

FIGURE 3. Empirical and theoretical cumulative distributions of the normalized residuals r_n with two different values of the first window size L with polynomial degrees $q-1=2$.

distribution, and that moving from $H_D = 0.0443$ for $L = 200$ to the highest distance values $H_D = 0.0564$ for $L = 1\,000$ will only have a small effect on the accuracy of the distribution under $\mathcal{H}_0$. Therefore, the choice of the parameters can be made on the basis of the highest detection power $\beta(M)$ for a Hellinger distance H_D lower than a certain value after which the accuracy is considered no longer acceptable. As a result, the correct choice of the parameters in our application is $L = 200$ and $q-1 = 2$, which will be considered in all following experiments.

Secondly, it is proposed to study the effect of the second window length M on the detection performances. The same data base has been used to perform a Monte-Carlo simulation, for which a simulated shift of amplitude $a = 60$ has been superimposed on some of the observations. Figure 4 represents the empirical false alarm probability $\alpha(R)$ and detection power $\beta(M)$ over a run length $R = 5\,000$ for 3 different values of the maximal allowed detection delay $M = \{1, 3, 5\}$, as a function of the decision threshold τ. It can be observed that when M increases, $\|\boldsymbol{\theta}_M\|_2$ increases, which affects both the false alarm rate $\alpha(R)$ and the detection power $\beta(M)$, as seen in (15). However, the increase rate of $\beta(M)$ is larger than the one of $\alpha(R)$. Hence, the shift between the detection power and the false alarm probability becomes larger which implies

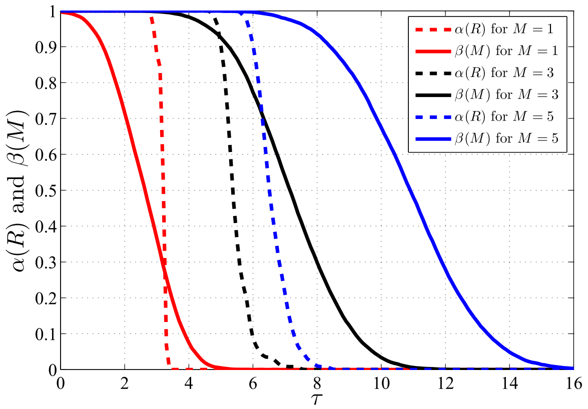

FIGURE 4. **Empirical false alarm probability $\alpha(R)$ and detection power $\beta(M)$ over a run length $R = 5000$ for 3 different values of M, plotted as a function of the decision threshold τ.**

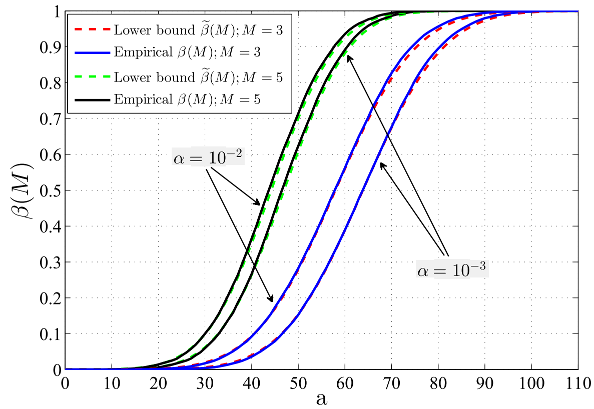

FIGURE 6. **Empirical and theoretical detection power $\beta(M)$ for 2 different values of M and 2 different values of false alarm rate $\alpha(R)$, plotted as a function of the change amplitude a.**

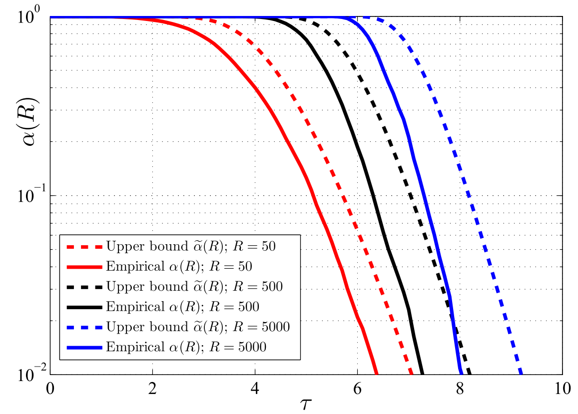

FIGURE 5. **Empirical and theoretical false alarm probability $\alpha(R)$ over three different values of run length R, plotted as a function of the decision threshold τ.**

a better detection performance, but at a larger delay M. As a result, it can be seen that the choice of M essentially depends on the application requirements. Depending on the application, this test allows to either increase the detection performance at a cost of a larger detection delay, or decrease the detection delay at a cost of a lower detection performance.

In the third part of the experiments, and because one of the main contributions of this paper is the design of a change-point detection method with bounded performance properties, it is wished first to examine the efficiency of the bounds calculated for $\alpha(R)$ and $\beta(M)$. Figure 5 shows the empirical false alarm probability $\alpha(R)$ and its theoretical upper bound $\widetilde{\alpha}(R)$ for three different values of the run length $R = \{50, 500, 5\,000\}$ as a function of the detection threshold τ. The maximal delay for detection is set to $M = 5$. It can be observed that the upper bound is accurate and relatively tight. However, as the run-length increases, one can notice that the upper bound is gradually losing its accuracy for smaller values of false alarm. At $\alpha(R) = 10^{-2}$, the distance between the empirical threshold and the theoretical one obtained by the upper bound is 0.7 for $R = 50$, but it increases to 1.2 for $R = 5\,000$. This is due mainly to the fact that the observations are not totally independent. In fact, the calculation of the upper bound of the false alarm probability is based on the inequality in equation (40) which is greatly affected by the independence of the observations. When R increases, the number of events in equation (40) increases, resulting in an increase in the difference between the probability of their intersection (the first term (40)) and the product of their individual probabilities (the second term in (40)). As a result, the sharpness of the upper bound for the false alarm probability decreases. In addition, a second factor can be the fact that the data base used to perform these experiments is rather small to be generally accurate in the empirical results for large values of run length as $R = 5\,000$. Then, in order to test the detectability of the proposed test and the sharpness of the detection power lower bound, figure 6 presents the empirical detection probability $\beta(M)$ and its theoretical lower bound $\widetilde{\beta}(M)$ for two different values of $M = \{3, 5\}$ and two different false alarm rates $\alpha(R) = \{10^{-2}, 10^{-3}\}$ over a run length $R = 5\,000$, as a function of the change amplitude a. First, it can be seen that the theoretical lower bound is precise and really tight for the different parameter values. Second, for a fixed value of the false alarm rate, when M increases, the detection power $\beta(M)$ increases accordingly. This result confirms the one obtained in the second experiment in Figure 4.

Next, it is wished to exemplify the efficiency of the proposed 2FLW-SEQ sequential detection method on a real case scenario with a real change-point in the observations. Figure 7 portrays a real case of observations when the spray gun nozzle got partially clogged. As a consequence, a sudden shift in the observations of amplitude $a = 55$ can be seen at exactly the image index 2434. The blue plot represents the real observations, while the red plot represents the expectation values (8) estimated using the polynomial model over a window of size $L = 200$ and a degree of $q - 1 = 2$. Because it is aimed to be as close as possible from the real practical requirement that corresponds to the specific application of paint coat

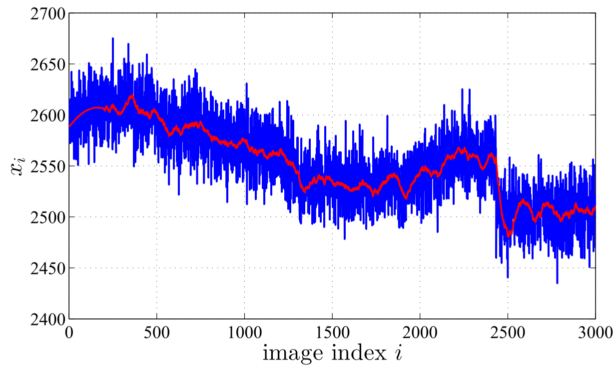

FIGURE 7. Real example of the variation of mean value with a change-point at index 2434.

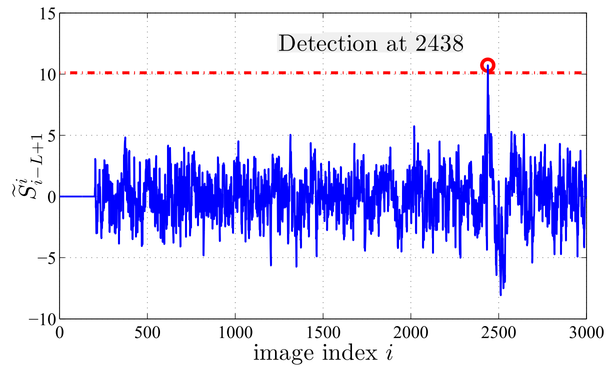

FIGURE 8. Result of the proposed 2FLW-SEQ detection method with $M = 5$.

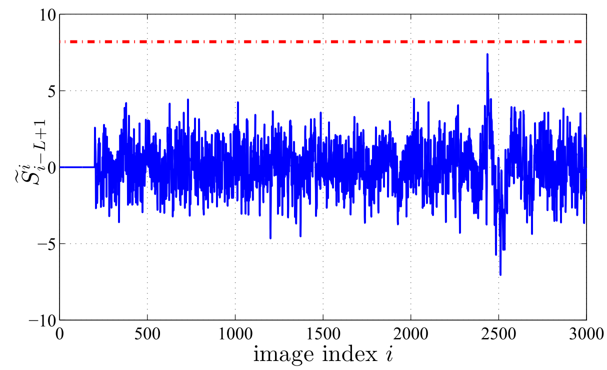

FIGURE 9. Result of the proposed 2FLW-SEQ detection method with $M = 3$.

monitoring, the false alarm rate is set to $\alpha(R) = 10^{-3}$ over a run length $R = 5\,000$. This will result in a detection threshold of $\tau = 10.12$ for $M = 5$ and a threshold of $\tau = 8.2$ for $M = 3$. Figure 8 illustrates the result of the proposed 2FLW-SEQ method with $M = 5$. It can be seen that the change point is detected at the index 2438 which means a delay of exactly 5 defective wheels.

Then, the same experiment has been performed for a maximal allowed delay of $M = 3$ where the detection power is much lower than the previous case of $M = 5$, as seen in Figure 6. Figure 9 illustrates the corresponding result where it can be seen that the change-point has been missed.

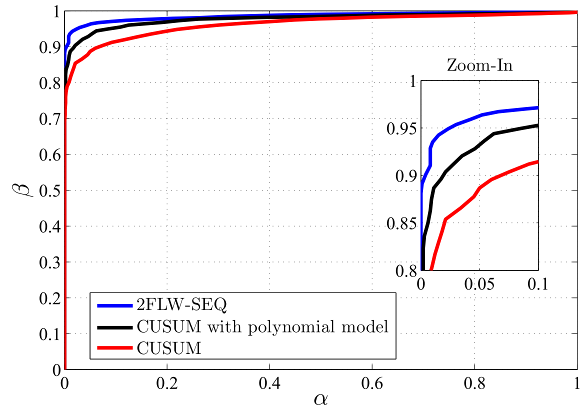

FIGURE 10. Empirical ROC curves for the proposed 2FLW-SEQ method and the CUSUM method with and without the polynomial model, computed over a run length $R = 5000$, with a maximal detection delay $M = 5$ and change amplitude $a = 60$.

Note that, usually when the change is detected, the sequential process stops. However to better illustrate the results of the test, the sequential procedure was allowed to continue. It is shown in Figure 7 that after the change occurs, the observations return to a state similar to the one just before the change occurred. Then, just after the change, the sequential procedure will re-operate under the hypothesis $\mathcal{H}_0$, and the results $\widetilde{S}_{i-L+1}^{i}$ will return to have a Gaussian distribution with zero mean and a variance $\|\boldsymbol{\theta}_M\|_2^2$, as it can be seen in Figures 8 and 9.

Last, but not least, the first goal is to investigate the advantages of modeling the paint coat intensity process to deal with its non-stationarity. To this purpose, it is proposed to compare the performance of the original 2FLW-SEQ method presented in this paper with a classical sequential detection method, more precisely the well-known CUSUM, in two different scenarios. In the first scenario the polynomial model is used to represent the expectation of the last L observations, while in the second scenario only the mean value of the last L observations is considered. In addition, the proposed 2FLW-SEQ is included in the comparison in order to show its efficiency. Note that when the polynomial model is used, the optimal parameters obtained from the first part of the experiments are considered, i.e. $L = 200$ and $q - 1 = 2$. However, for the CUSUM without a model, multiple simulations with different values of L have been conducted and lead to the choice of $L = 20$ which is the best in terms of detection power. Figure 10 presents the empirical ROC curves for the proposed 2FLW-SEQ method and the CUSUM method with and without the polynomial model, computed over a run length $R = 5000$, with a maximal detection delay $M = 5$ and change amplitude $a = 60$. It can be seen that using the polynomial model actually improves the performance of the CUSUM method. Figure 10 also shows that the proposed 2FLW-SEQ method outperforms the CUSUM method even when using the polynomial model. Indeed, the CUSUM method has proven many times to be optimal as mentioned in section IV, however, this optimality is related to the average detection delay. To better

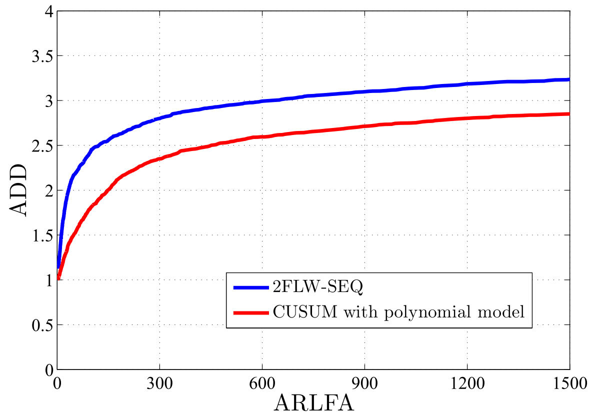

FIGURE 11. Average detection delay as a function of the average run length to false alarm for the proposed 2FLW-SEQ method and the CUSUM method with polynomial model with a maximal detection delay $M = 5$ and change amplitude $a = 60$.

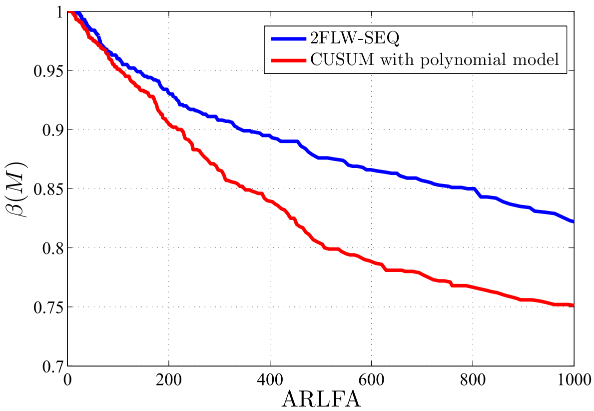

FIGURE 12. Empirical detection power $\beta(M)$ as a function of the average run length to false alarm for the proposed 2FLW-SEQ method and the CUSUM method with polynomial model with a maximal detection delay $M = 5$ and change amplitude $a = 40$.

understand the difference in the detection criteria under which each of the proposed 2FLW-SEQ method and the CUSUM method operates, two sets of simulations are conducted.

Figure 11 represents the average detection delay (ADD) as a function of the average run length to false alarm (ARLFA) for the proposed 2FLW-SEQ method, with a maximal detection delay set to $M = 5$, and for the CUSUM method with polynomial model with $q - 1 = 2$ and change amplitude $a = 60$. It can be seen that in this context, the CUSUM has proven to be optimal and, hence, outperforms the proposed 2FLW-SEQ method. On the other hand, as noted in the section IV, the aim of the proposed 2FLW-SEQ method is to minimize the worst-case probability of missed detection under constraint on the worst-case probability of false alarm for a given run length. To highlight this criteria, figure 12 represents the empirical detection power $\beta(M)$ as a function of the ARLFA for the same sequential methods and with a change amplitude $a = 40$. The smaller value of the change amplitude is considered to emphasize better the difference. Obviously, figure 12 shows that, in this context, with the increasing values of the ARLFA, the proposed 2FLW-SEQ method outperforms the CUSUM method in terms of detection power. Indeed, it is well known that minimizing the average detection delay does not necessarily lead to a higher detection power under a given maximal delay, or to a small probability of missed detection.

VII. CONCLUSION AND PERSPECTIVES

This paper proposes a method for online monitoring of a non-stationary process in the mean with a constant variance, to detect abrupt changes in the process mean value. Since the monitored process is non-stationary, in other words naturally changes over time, the proposed method must be able to distinguish those ''regular'' process changes from abrupt changes resulting from potential failures. In addition, since this method aims at being applied for industrial processes, it is required to detect the change within a given maximal detection delay and to control the false alarm probability over a fixed run length. Hence, in order to adapt to the operational requirements of the industrial context, the CUSUM method is modified to design a novel sequential method based on two fixed-length windows. Over the first window, the non-stationary process parameters are estimated using a linear parametric model, while the second window is used to execute the detection.

In this paper, the non-stationary process results from the variation of the paint quantity on inspected wheels surface, where the abrupt change corresponds to a sudden lack of paint. The mean value of pixels from all wheel images are used to measure the coating intensity. Numerical results on a large set of images show the accuracy of the proposed model, the efficiency of the proposed detection method, and the sharpness of the statistical performances theoretically established.

Since the proposed method is designed to monitor any non-stationary process in the mean with a constant variance, a first perspective is to test the efficiency of the proposed method for other industrial processes of such type. Furthermore, in this paper, modeling the non-stationary process over the first window was conducted using a polynomial model. Replacing this model with any other type of parametric models is feasible, and will not alter the detection procedure performed on the second window. Hence, a second interesting perspective is to test the accuracy of other types of parametric models to model the non-stationary process, such as the autoregressive integrated moving average (ARIMA) for example, and to compare it with the accuracy of the proposed polynomial model.

REFERENCES

[1] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, ''A review and comparison of changepoint detection techniques for climate data,'' *J. Appl. Meteorol. Climatol.*, vol. 46, no. 6, pp. 900–915, 2007.

[2] G. E. Evans, G. Y. Sofronov, J. M. Keith, and D. P. Kroese, ''Estimating change-points in biological sequences via the cross-entropy method,'' *Ann. Oper. Res.*, vol. 189, no. 1, pp. 155–165, 2011.

[3] T. Polushina and G. Sofronov, ''A cross-entropy method for change-point detection in four-letter dna sequences,'' in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2016, pp. 1–6.

[4] J. Chen and A. K. Gupta, "Testing and locating variance changepoints with application to stock prices," *J. Amer. Stat. Assoc.*, vol. 92, no. 438, pp. 739–747, 1997.

[5] A. Aprem and V. Krishnamurthy, "Utility change point detection in online social media: A revealed preference framework," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1869–1880, Apr. 2017.

[6] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer, 2001.

[7] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection* (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). New York, NY, USA: Taylor & Francis, 2014.

[8] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[9] K. Tout, F. Retraint, and R. Cogranne, "Wheels coating process monitoring in the presence of nuisance parameters using sequential change-point detection method," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 196–200.

[10] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, nos. 1–2, pp. 100–115, 1954.

[11] N. K. Akafuah, S. Poozesh, A. Salaimeh, G. Patrick, K. Lawler, and K. Saito, "Evolution of the automotive body coating process—A review," *Coatings*, vol. 6, no. 2, p. 24, 2016.

[12] N. K. Akafuah, "Automotive paint spray characterization and visualization," in *Automotive Painting Technology*, Dordrecht, The Netherlands: Springer, 2013, pp. 121–165.

[13] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Hoboken, NJ, USA: Wiley, 1973.

[14] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 221–234, Feb. 2016.

[15] R. Cogranne and F. Retraint, "An asymptotically uniformly most powerful test for LSB matching detection," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 464–476, Mar. 2013.

[16] R. Cogranne and F. Retraint, "Statistical detection of defects in radiographic images using an adaptive parametric model," *Signal Process.*, vol. 96, pp. 173–189, Mar. 2014.

[17] K. Tout, R. Cogranne, and F. Retraint, "Statistical decision methods in the presence of linear nuisance parameters and despite imaging system heteroscedastic noise: Application to wheel surface inspection," *Signal Process.*, vol. 144, pp. 430–443, Mar. 2017.

[18] K. Tout, R. Cogranne, and F. Retraint, "Fully automatic detection of anomalies on wheels surface using an adaptive accurate model and hypothesis testing theory," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 508–512.

[19] R. Cogranne, "A sequential method for online steganalysis," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, Nov. 2015, pp. 1–6.

[20] B. K. Guépié, L. Fillatre, and I. Nikiforov, "Sequential detection of transient changes," *Sequential Anal.*, vol. 31, no. 4, pp. 528–547, 2012.

[21] A. N. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process," *Sov. Math. Dokl.*, no. 2, pp. 795–799, 1961.

[22] K. W. Kemp, "The average run length of the cumulative sum chart when a V-mask is used," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 23, no. 1, pp. 149–153, 1961.

[23] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[24] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[25] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *Ann. Stat.*, vol. 18, no. 3, pp. 1464–1469, 1990.

KARIM TOUT received the Engineering degree in electrical engineering and the M.S. degree in industrial systems technology from Lebanese University in 2014. He is currently pursuing the Ph.D. degree in systems optimization and safety with the Troyes University of Technology. He is currently a Teaching and Research Temporary Attaché with the Troyes University Institutes of Technology. His research interests include statistical image processing, industrial system monitoring, hypothesis testing theory, anomaly detection, and computer vision applications.

FLORENT RETRAINT received the Engineering diploma in computer science from the Compiègne University of Technology in 1993, the M.S. degree in applied mathematics from the École nationale supérieure d'informatique et de mathématiques appliquées de Grenoble in 1994, and the Ph.D. degree in applied mathematics from the National Institute of Applied Sciences of Lyon, France, in 1998. Then, he held a post-doctoral position for one year at CEA Grenoble and was a Research Engineer for two years with Thales S.A. Since 2002, he has been with the Laboratory of system modeling and dependability, Troyes University of Technology. His research interests include image modeling, statistical image processing, hypothesis testing theory, anomaly detection, and localization.

RÉMI COGRANNE received the Engineering degree in computer science and telecommunication and the Ph.D. degree in systems optimization and safety from the Troyes University of Technology in 2008 and 2011, respectively. He was an Associate Professor with the University of Technology of Troyes. He was a Visiting Scholar with Binghamton University from 2014 to 2017. During his studies, he took a semester off to teach in a primary school in Ziguinchor, Senegal, and studied one semester at Jonkoping University, Sweden. His main research interests are in hypothesis testing for application in various fields, such as steganalysis, steganography, industrial system monitoring, computer network, and statistical image processing.