A Comparison Study of Distribution-Free Multivariate SPC Methods for Multimode Data

Marco Grasso[1a], Bianca Maria Colosimo[a], Quirico Semeraro[a], Massimo Pacella[b]

[a]*Dipartimento di Meccanica, Politecnico di Milano, Via La Masa 1, 20156 Milano, Italy*

[b]*Dipartimento di Ingegneria dell'Innovazione, Università del Salento, 73100 Lecce, Italy*

[1]Corresponding author. E-mail: marco.grasso@musp.it

Abstract — The data-rich environments of industrial applications lead to large amounts of correlated quality characteristics that are monitored using Multivariate Statistical Process Control (MSPC) tools. These variables usually represent heterogeneous quantities that originate from one or multiple sensors and are acquired with different sampling parameters. In this framework, any assumptions relative to the underlying statistical distribution may not be appropriate, and conventional MSPC methods may deliver unacceptable performances. In addition, in many practical applications, the process switches from one operating mode to a different one, leading to a stream of multimode data.

Various non-parametric approaches have been proposed for the design of multivariate control charts, but the monitoring of multimode processes remains a challenge for most of them. In this study, we investigate the use of distribution-free MSPC methods based on statistical learning tools. In this work, we compared the kernel distance-based control chart (***K***-chart) based on a one-class-classification variant of Support Vector Machines and a fuzzy neural network method based on the Adaptive Resonance Theory. The performances of the two methods were evaluated using both Monte Carlo simulations and real industrial data. The simulated scenarios include different types of out-of-control conditions to highlight the advantages and disadvantages of the two methods. Real data acquired during a roll grinding process provide a framework for assessment of the practical applicability of these methods in multimode industrial applications.

Index Terms — Multimode Processes, Support Vector Machine, Artificial Neural Networks, Adaptive Resonance Theory, Statistical Process Control

1 Introduction

The technological advances in sensor systems together with the continuous improvement of real-time data processing capabilities have increased the use of in-process sensing to enhance process quality control outcomes. The signals from one or multiple sensors are analyzed to extract synthetic features used to characterize the stability of the ongoing process. Therefore, the quality characteristics represent multiple variables and consist of a collection of features that represent heterogeneous quantities resulting from different pre-processing procedures. In this framework, the conventional assumptions relative to the underlying distribution may not be appropriate for the design of statistical process monitoring tools. Moreover, in many practical applications, the natural process behavior switches from an operating mode to the following one, producing streams of data from different distributions that follow one another over time, without information about the duration of each mode

and their temporal allocation. Motivational examples are qualitatively depicted in Fig. 1. Fig. 1a) depicts a milling process performed on a workpiece whose geometry produces different levels of cutting depth along the cutting trajectory. The monitored variables consist of features extracted from a cutting force signal (e.g., the mean force level and the standard deviation). This process is multimode in nature, as it switches from a data distribution to another one as the tool moves along its path. Fig. 1b) shows a drilling process applied to a multi-layer (hybrid material) workpiece for which different materials produce different thrust force levels, and hence, the monitored features extracted from the force signal are multimode in nature, with different distributions corresponding to different material layers.

INSERT FIGURE 1 ABOUT HERE
INSERT FIGURE 2 ABOUT HERE

A further example from a real industrial case study is shown in Fig. 2 and refers to a roll grinding process that consists of grinding of large cylindrical rolls for use in a subsequent rolling process on metal sheets. An accelerometer is mounted on the wheel head (Fig. 2 a), and the signal along the X-axis is used to detect any out-of-control departure from a nominal and stable cutting condition during the process itself. The bivariate monitored statistics $\{\boldsymbol{x}_j \in \mathbb{R}^2, j = 1,2,\dots\}$ consists of two synthetic indices, nominally the rms index and the $kurtosis$ index (see Section 7), such that $\boldsymbol{x}_j = [rms_j, kurt_j]^T$ is computed by segmenting the signal into sliding time windows of fixed duration.

In common operating conditions, a grinding cycle is composed of different passes, each with cutting parameters that vary within given ranges. This situation yields a multimode process in which the in-control (IC) distribution of the monitored indices is characterized by temporally consecutive distributions, which correspond to different combinations of wheel speed n_s and infeed a_e parameters.

Fig. 2b) shows the multimode distribution of $\boldsymbol{x}_j$ under IC conditions for different combinations of cutting parameters (n_s, a_e) and its distribution in the presence of an out-of-control (OOC) chatter onset (this real case study is discussed in additional detail in Section 7).

If the process exhibits multimode behavior, clustered data represent the natural pattern that characterizes the IC condition. Any transition between one mode and another should not be signaled by the control chart because such a transition is the consequence of the natural process variability. The result is one of the most challenging violations of the traditional SPC distributional assumptions.

Despite of a great industrial interest for multimode process monitoring in the field of discrete part manufacturing, this problem attracted very limited attention in the literature. Certain authors proposed Multivariate Statistical Process Control (MSPC) schemes for multimode processes in chemometrics applications, but most of them are based on distributional assumptions within each mode[1-3] or assumptions about the covariance structure of multiple modes[4-5]. The proposed methods include either the use of multiple models[6-8] or the use of local models to be iteratively updated[9]. Some authors assumed the availability of external knowledge to filter out the effect of operating conditions[10-11]. However, there is a lack of non-parametric methods in this area that motivates the present study.

For in-process MSPC industrial applications, parametric methods may be of limited practical interest because single modes frequently exhibit deviations from multivariate normality and they may be difficult to classify in a reliable way. Furthermore, assumptions about the covariance structure of the data and the availability of external knowledge are rarely applicable in practice.

Various authors have discussed the need for distribution-free MSPC methods[12-18] and the limits of traditional control charts in the presence of non-normal data[19], but not in the frame of multimode processes. Our study represents a first attempt to assess the applicability of selected non-parametric MSPC methods to multimode process monitoring in the field of industrial production. Our analysis is aimed at demonstrating and comparing their performances and limitations by means of both simulated and real industrial data.

In particular, we focus on a category of methods that is often referred to as *one-class-classification*[20-21] or *novelty detection*[22]. These approaches are statistical learning methods that can be trained on a dataset (Phase I) consisting of only natural process data, i.e., data collected under IC process conditions. During the actual process-monitoring phase (Phase II), every observation that cannot be classified within the IC class is signaled. This approach allows implementation of traditional classification techniques in the SPC frame in which no information on the nature of possible departures from the natural condition is available, at least in the preliminary implementation stage.

One interesting feature of this category of methods consists of their distribution-free properties, and hence, they can be applied regardless of the single-mode or multimode nature of the monitored data. A rationale for the choice of such a family of techniques for multimode process monitoring is discussed in Section 3.

Our comparison study involves two one-class-classification methods, which are respectively based on the Support Vector Data Description (SVDD) methodology and on unsupervised Artificial Neural Networks (ANNs). Their performances are evaluated using Monte Carlo simulations in the presence of multimode data and a real dataset acquired in roll grinding operations. The improved K-chart

design proposed by Ning and Tsung[23] and the Fuzzy-ART-based scheme proposed by Pacella and Semeraro[24-27] are reviewed and compared. The previous work of Pacella and Semeraro[24-27] focused on process monitoring of univariate time series or streams of profile data. In this study, we extend this Fuzzy-ART-based scheme to monitoring of a multivariate multimode process. We also build on the study by Ning and Tsung[23], which focused on mixed type data; in the current paper, we extend the previous analysis to characterize the Phase II performances of this method under different out-of-control scenarios in a multimode process framework, and we further study the role played by the kernel function and the design parameters in Phase I of control charting.

In the remainder of the paper, the terms *Phase I dataset* and *training dataset* are used interchangeably under the assumption that the IC state of the process has been proven, which ensures that the collected samples are representative of its natural condition.

The performances of all competing methods are compared using a real case study that addresses chatter detection in roll grinding. In this case, in-process signal monitoring is aimed at detecting anomalous vibration onset. The real case study allows us to evaluate the practical applicability of the proposed methods for actual industrial scenarios in which streams of multimode data are acquired via sensors installed on the machine tool.

The paper is organized as follows. Section 2 lists the nomenclature used in the paper, Section 3 presents a rationale for the use of one-class-classification methods, Section 4 describes the K-chart approach, Section 5 details the Fuzzy ART-based approach, Section 6 presents the simulated scenarios and discusses the results of the comparative analysis, Section 7 presents the real case study in roll grinding, and Section 8 concludes the paper.

2 Nomenclature

$\#SV(S)$	Number of support vectors (K-chart approach)
b_{out}	Output signal of the Fuzzy ART network
C	Penalty coefficient (K -chart approach)
$f_{o_+}(S)$	Proportion of artificial outliers classified as in-boundary data (K-chart approach)
F_0, F_1, F_2	Layers of the Fuzzy ART architecture
h	Control limit (K-chart approach)
IC	In-control
L	Number of mixtures
M	Overall number of samples collected under natural process conditions

M_o	Number of artificial outliers
M_1	Number of training samples (Fuzzy-ART-based approach)
$MSPC$	Multivariate Statistical Process Control
N	Number of tuning samples (Fuzzy-ART-based approach)
$\boldsymbol{o}$	Center of the irregular shaped region in the K-chart approach
OOC	Out-of-control
p	Number of variables
Q	Number of nodes in the F_2 layer
q_{max}	Index of the node with maximum bottom-up input (Fuzzy-ART-based approach)
R	Radius of the irregular shaped region in the K-chart approach
S, S^*	Kernel width parameter (K-chart approach)
$[S_0, S_U]$	Range for the selection of the kernel width parameter
$T_q(\boldsymbol{x}_j)$	Bottom-up inputs (Fuzzy ART based approach), $j = 1, 2, \ldots$.
ν	Weight used in Eq. 10
$\boldsymbol{x}_j$	j^{th} multivariate sample; elements are denoted by $x_{1,j}, \ldots, x_{p,j}$, $j = 1, 2, \ldots$.
$\boldsymbol{x}_j^c$	Input vector after the complement coding step (Fuzzy-ART-based approach), $j = 1, 2, \ldots$.
$\boldsymbol{w}_q$	Weight vector (Fuzzy-ART-based approach)
$\boldsymbol{z}$	Generic new multivariate observation
α	Targeted Type I error
α_j and γ_j	Lagrangian coefficients (K-chart approach), $j = 1, 2, \ldots$.
β	Choice parameter (Fuzzy-ART-based approach)
$\gamma(S)$	Objective function for the kernel width parameter selection
δ_1, δ_2, δ_3	Shift parameters
η	Vigilance step (Fuzzy-ART-based approach)
$\boldsymbol{\mu}_l$	Multivariate mean of the Gaussian mixture model, $l = 1, \ldots, L$
ξ_j	Slack variables (K-chart approach)
π_l	Prior probabilities in Gaussian mixture model, $l = 1, \ldots, L$
ρ	Vigilance parameter (Fuzzy-ART-based approach)
ρ_u	Maximum value of the vigilance parameter that induces one class
$\boldsymbol{\Sigma}_l$	Variance-covariance matrix of the Gaussian mixture model, $l = 1, \ldots, L$

3 A Rationale for the Use of One-Class-Classification Methods

A commonly used approach for the design of non-parametric multivariate control charts involves certain generalizations of rank-based methods and/or a transformation of the original data into a categorical form[15-17;28-31]. However, these methods usually rely on the assumption of identically distributed data and are not designed to deal with processes that jump from one operating mode to another, without being known the duration of each mode and the temporal allocation of multiple modes. In multimode processes, the i.i.d. assumption may hold only within each mode, and the transitions between consecutive natural modes should not be signaled by the control chart. In addition, rank-based methods may be affected by limited efficiency in the presence of large shifts[31]. For a brief review of non-parametric MSPC methods, see the recent book of Qiu[18] and the work of Bersimis *et al.*[32].

A more widely used approach applied to address unknown distributions consists of adjusting the control limits of conventional control charts by estimating the empirical percentiles of the monitored statistic. If the number of Phase I samples is small, the bootstrap re-sampling technique can be used[33-35]. However, in presence of multimode processes, conventional control statistics may be not adequate and the bootstrap approach may produce unreliable estimates of the tail probabilities[36].

Alternatively, a different paradigm consists of using one-class-classification methods that are designed to adapt classical data-mining and machine-learning techniques to MSPC.

These methods are aimed at estimating a multivariate region that envelops the Phase I data such that a target Type I error is achieved. Next, the contour of this region plays the role of the control limit, and any observation that falls outside of this region is signaled. Thus, when a multimode historical dataset is available, one-class-classification methods provide a non-parametric framework to monitor the process, regardless of the jumps from one mode to another. It is worth to notice that this kind of methods may be applied without the need for prior knowledge about the number or the temporal allocation of the modes, and without the need for a preliminary data clustering step.

Selected one-class-classification variants of statistical learning techniques have been proposed in the literature, including Support Vector Machines (SVMs)[20] and Artificial Neural Networks (ANNs)[24].

With respect to the use of ANNs for SPC applications, few authors have considered the one-class-classification case in which the training dataset is composed of natural process data only. Among these, Al-Ghanim[37] presented an Adaptive Resonance Theory (ART) neural network to distinguish natural from unnatural variations in the outcomes of a manufacturing process. Pacella and

Semeraro[24-26] extended the study of Al-Ghanim[37] by proposing a Fuzzy ART neural network to address arbitrary sequences of input patterns, whereas the ANN model proposed by Al-Ghanim[37] was limited to binary inputs. Pacella and Semeraro[27] derived a variant of the previously proposed Fuzzy ART-based scheme to monitor the stability over time of profile data. A different type of unsupervised ANN method, i.e., the Self Organizing Map (SOM), was discussed by Tax[38] for one-class-classification applications. For a survey of ANN methods in SPC applications, the interested reader should refer to the papers by Guh[39] and Psarakis[40].

A different statistical learning tool for which one-class-classification variants were proposed in the literature is the SVM method. A one-class-classification variant known as the Support Vector Data Description (SVDD) was proposed by Tax[38] and Tax and Duin[41] and was used by Sun and Tsung[42] to design a kernel distance-based control chart referred to as the K-chart. Other authors studied SVDD-based MSPC approaches. Camci *et al.*[43] studied a version of the K-chart believed to be robust to Phase I contamination. Sukchotrat *et al.*[44] compared the SVDD-based approach with a method based on the k-Nearest Neighbors (k-NN) algorithm. Ge *et al.*[45] applied an SVDD-based control chart to batch process monitoring. Gani *et al.*[46] eventually applied the K-chart to monitor the quality characteristics in a refrigerator metal sheet manufacturing process.

The K-chart approach shares certain common points with data-depth based methods[29-30], but it involves neither data-depth measures nor ranking operations. The control statistic consists of the kernel distance of any observation from a common multivariate center estimated using the one-class variant of the SVM optimization procedure. The K-chart also differs from control charting approaches whose control region consists of a percentile of an estimated multivariate density function. Indeed, this approach does not require the estimation of the complete density but only a boundary around a data set. Furthermore, Tax and Duin[41] demonstrated that the SVDD technique outperforms basic control charting methods based on density estimation.

A critical issue affecting any kernel distance-based control chart involves the proper selection of kernel parameters and the estimation of the control limits. Recently, Ning and Tsung[23] proposed an improved design of the K-chart, including an effective strategy for kernel parameter selection, which will be used as a reference in this study.

Both the K-chart approach and the Fuzzy ART-based approach are suitable for monitoring of the stability over time of multivariate data, regardless of the single-mode or multimode nature of the process. These approaches do not require any assumption on the underlying distribution of the natural process data, their covariance structure, or the number of modes. Thus, these methods can be

used to design completely distribution-free MSPC tools and to address such complex signal data as those that characterize actual industrial applications.

To the best of the authors' knowledge, no comparison studies for monitoring of multimode data via distribution-free MSPC have been presented in the literature. Our paper aims to fill this gap and address practical issues that arise in real industrial scenarios.

4 K-chart Based on Support Vector Data Description

The SVDD method was presented by Tax and Duin[20] to extend the SVM classification technique to problems characterized by single-class training sets. Given a multivariate Phase I (or training) dataset $\{\boldsymbol{x}_j \in \mathbb{R}^p, j = 1, \dots, M\}$, where $\boldsymbol{x}_j = [x_{1,j}, x_{2,j}, \dots, x_{p,j}]^T$, the SVDD method consists of finding a minimal volume control region characterized by a center $\boldsymbol{o} \in \mathbb{R}^p$, and a radius R, that can envelop a given percentage of the original data. The K-chart[42] is a multivariate control chart whose control statistic consists of the kernel distance of any observation $\boldsymbol{z} \in \mathbb{R}^p$ from the center $\boldsymbol{o} \in \mathbb{R}^p$ of that region. The control limit is estimated to guarantee a target Type I error with the available dataset. A kernel distance, hereafter denoted by $kd(\boldsymbol{z})$, replaces the traditional Euclidean and statistical distance notions to adapt the control region boundary to the actual spread of the data, whereas using the Hotelling's T^2 distance, the control region would become a p-dimensional ellipsoid, as an example.

The SVDD methodology is briefly reviewed in Sub-section 4.1 to explain how the kernel distance $kd(\boldsymbol{z})$ is computed. Sub-section 4.2 is devoted to the selection of the kernel parameter and the K-chart design procedure.

4.1 The SVDD Methodology

The SVDD works by estimating a minimal volume control region that adapts to the actual spread of the data. The estimation of such a region, centered in $\boldsymbol{o} \in \mathbb{R}^p$ and with radius R, requires the solution of the following data-driven optimization problem:

$$\begin{aligned} &\min(R^2 + C\sum_{j=1}^{M}\xi_j) \\ &\text{s.t. } (\boldsymbol{x}_j - \boldsymbol{o})^T(\boldsymbol{x}_j - \boldsymbol{o}) \leq R^2 + \xi_j \text{ and } \xi_j \geq 0, j = 1, \dots, M \end{aligned} \tag{1}$$

where ξ_j, $j = 1, \dots, M$, are slack variables, and C is a penalty coefficient used to weight the trade-off between the volume of the region and the percentage of enclosed data ($C > 0$). By introducing the Lagrangian function:

$$L(R, \boldsymbol{o}, \xi_j; \alpha_j, \gamma_j) = R^2 + C\sum_{j=1}^{M}\xi_j - \sum_{j=1}^{M}\alpha_j(R^2 + \xi_j - (\boldsymbol{x}_j - \boldsymbol{o})^T(\boldsymbol{x}_j - \boldsymbol{o})) - \sum_{j=1}^{M}\gamma_j\xi_j \quad (2)$$

and by setting the partial derivatives w.r.t. R, $\boldsymbol{o}$, and ξ_j, $j = 1, \dots, M$, to zero, the problem (1) can be simplified as follows[23]:

$$\begin{aligned} &\max\left(\sum_{j=1}^{M}\alpha_j \boldsymbol{x}_j{}^T\boldsymbol{x}_j - \sum_{j,k=1}^{M}\alpha_j\alpha_k \boldsymbol{x}_j{}^T\boldsymbol{x}_k\right) \\ &\text{s.t. } \sum_{j=1}^{M}\alpha_j = 1 \text{ and } 0 \le \alpha_j \le C, j = 1, \dots, M \end{aligned} \quad (3)$$

The points whose Lagrangian coefficients are larger than zero are known as *support vectors*. It can be demonstrated that the shape of region is determined by those points only[38].

By introducing the kernel trick, it is possible to replace the inner product $\boldsymbol{a}^T\boldsymbol{b}$ by a kernel function $K(\boldsymbol{a} \times \boldsymbol{b})$ that allows generation of a more flexible and data-adaptive control region. The K-chart is aimed at monitoring the stability over time of the kernel distance $kd(\boldsymbol{z})$ of any new observation $\boldsymbol{z} \in \mathbb{R}^p$ from the center $\boldsymbol{o}$:

$$kd(\boldsymbol{z}) = K(\boldsymbol{z} \times \boldsymbol{z}) - 2\sum_{j=1}^{M}\alpha_j K(\boldsymbol{x}_j \times \boldsymbol{z}) + \sum_{j,k=1}^{M}\alpha_j\alpha_k K(\boldsymbol{x}_j \times \boldsymbol{x}_k) \quad (4)$$

Ning and Tsung[23] showed that there are different possible approaches to the design of the K-chart because there are three major parameters to set: the kernel width parameter denoted by S, the penalty coefficient C, and the control limit denoted by h. By comparing different design solutions, Ning and Tsung[16] showed that the best performances might be achieved by reducing the number of parameters to two (i.e., S and h). In fact, by assuming $C > 1$, the constraint $0 \le \alpha_j \le C$ is replaced by $\alpha_j \ge 0$, and problem (3) can be solved by introducing the kernel function $K(\boldsymbol{x}. \times \boldsymbol{x}.)$.

In this case, no penalty is applied, and hence, the kernel-based boundary is estimated by enclosing all of the training data. The false alarm rate is controlled by setting a proper value for the control limit h. Thus, only the S and h parameters remain to be determined. The next sub-section reviews the procedure used to automatically select those two parameters and to design the K-chart.

With respect to the kernel function, the most common choices include the Gaussian Radial Basis (GRB) and the polynomial and sigmoidal functions[47].

Tax[38] demonstrated that the GRB function is more appropriate than other kernel functions in classification problems. In the framework of K-chart-based monitoring, this approach has been used in previous studies[23,42], although its benefits over other kernels were not fully investigated. Thus, the

GRB function is used as the default choice in this study, but we discuss the use of other kernel functions in Section 6.

If $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$, the GRB function with kernel width parameter $S \in \mathbb{R}^+$ is defined as follows:

$$K(\boldsymbol{a} \times \boldsymbol{b}) = \exp\left\{-\frac{\|\boldsymbol{a} - \boldsymbol{b}\|^2}{S^2}\right\} \tag{5}$$

4.2 Automated Selection of the Kernel Parameter and Control Chart Design

In most cases, the selection of the kernel width parameter involves trial and error. When in-process monitoring is considered, an automated data-driven procedure is required. To this aim, Tax and Duin[48] proposed a method that was further improved by Ning and Tsung[23]. The method is derived from multi-class SVM problems in which the classification errors can be used as a standard to select S. In a one-class-classification problem, a similar approach might be applied by generating artificial outliers. Tax and Duin[48] proposed drawing of those outliers from a block-shaped or a hyper-spherical uniform distribution that encloses the training data in $\mathbb{R}^p$.

Given $f_{o_+}(S)$, the proportion of artificial outliers that are classified as in-boundary data for a given choice of S, and $\#SV(S)$, the number of support vectors, S can be selected by minimizing:

$$\gamma(S) = (1 - \nu)\frac{\#SV(S)}{M} + \nu f_{o_+}(S), \tag{6}$$

because $\#SV(S)/M$ is a counterpart of the Type I error, and $f_{o_+}(S)$ is the counterpart of the Type II error, where $0 < \nu < 1$ is a weight.

The procedure for the selection of the kernel width parameter is applied as follows:

1. Given a training set of M observations, generate a number M_o of artificial outliers;
2. Set S equal to an initial value S_0,, and solve problem (9) for the $M + M_o$ available data;
3. Compute $f_{o_+}(S_0)$ and $\#SV(S_0)$;
4. Set S equal to a new value $S_0 + s$, where s is a step value, and repeat steps 3 and 4 until S equals a pre-fixed upper limit S_U;
5. Find the value of S (known as S^*) such that $\#SV(S^*)/M$ is nearest to the targeted Type I error;
6. Calculate the weight ν as follows: $\nu(S^*) = \left(1 + \frac{f_{o_+}(S^*)}{(\#SV(S^*)/M)}\right)^{-1}$;
7. Calculate the $\gamma(S)$ value in Equation (10), where $\nu = \nu(S^*)$, for S values in the range $[S_0, S_U]$;

8. Eventually, S is determined by the minimal $\gamma(S)$.

Once the optimal value of the kernel width parameter is determined, the control region can be estimated. The control limit h can be estimated as the $100(1-\alpha)\%$ empirical percentile of the kernel distance $kd(\mathbf{z}_j)$, $j = 1,2,\dots,M$[23], where α is the targeted Type I error.
The same procedure can be applied using other kernel functions, provided that the control region geometry depends on a single kernel parameter. A discussion on the use of different kernel functions for the design of the K-chart is reported in Section 6.

5 A Quality Control Scheme Based on Fuzzy ART Networks

The ART methodology was introduced by Grossberg in 1976, which led to a number of ART-based neural network models widely used in different applications[49-50].
An ART-based neural network allows clustering of data into groups characterized by similar features in a self-organizing manner. The Fuzzy ART belongs to the class of unsupervised ART architectures; it is based on fuzzy set theory operations and allows clustering of arbitrarily complex analog input patterns. The method proposed by Pacella and Semeraro[24] adapts this learning paradigm to SPC applications by training the network on a Phase I dataset and associating a control region to natural process data. Similar to the SVDD approach, the control region is estimated to guarantee that the target Type I error is achieved on the IC data. When a new sample is presented to the Fuzzy ART network, an output signal b_{out} is generated. If the new sample is internal to the control region, it is judged as in-control ($b_{out} = 1$),, and no alarm is signaled. Otherwise, the output signal is $b_{out} = -1$, and an alarm is signaled. Therefore, the Fuzzy ART can operate as a non-parametric statistical process control tool. Sub-section 5.1 discusses the Fuzzy ART-based approach for multivariate data, and Sub-section 5.2 reviews a procedure used to estimate the so-called *vigilance parameter* ρ used to control the actual false alarm rate.

5.1 The Fuzzy ART Methodology

A block diagram of the Fuzzy ART architecture is shown in Fig. 3.

INSERT FIGURE 3 ABOUT HERE

The multivariate vector $\boldsymbol{x}_j = [x_{1,j}, x_{2,j}, \dots, x_{p,j}]^T \in \mathbb{R}^p$, $j = 1, 2, \dots$, represents the jth input vector of the Fuzzy ART network. Note that the values of input data can range only between 0 and 1, and hence, a re-scaling operation might be required in the pre-processing phase. The re-scaling operation is carried out by guaranteeing that any departure from the IC condition within a reasonable number of standard deviation units falls into the new $[0,1]$ range. Notably large shifts that fall outside the range can be directly signaled as out-of-control observations using a simple check rule.

The Fuzzy ART architecture consists of two subsystems, the *attentional subsystem* and the *orienting subsystem*[49]. The attentional subsystem consists of three layers of nodes denoted as F_0, F_1, and F_2. The F_0 layer consists of p nodes and is charged with applying a pre-processing operation known as *complement coding*. Given the input sample $\boldsymbol{x}_j$, the complement coding produces an output sample $\boldsymbol{x}_j^c$, such that:

$$\boldsymbol{x}_j^c = [x_{1,j}, x_{2,j}, \dots, x_{p,j}, 1 - x_{1,j}, 1 - x_{2,j}, \dots, 1 - x_{p,j}]^T \tag{7}$$

The F_1 layer is the *comparison layer* and consists of $2p$ nodes. The F_2 layer is the *recognition layer* and consists of a number Q of nodes equal to (or greater than) the number of clusters formed during the training phase. Every node in the F_2 layer is connected with every node in the F_1 layer via a weight vector $\boldsymbol{w}_q = [w_{q,1}, w_{q,2}, \dots, w_{q,2p}]^T$, $q = 1, \dots, Q$, which subsumes both the bottom-up and top-down weight vectors of the Fuzzy ART.

The orienting subsystem consists of a single node referred to as the *reset node*. The output of the reset node depends on the vigilance parameter ρ and affects the nodes in the F_2 layer. The vigilance parameter ρ determines the required degree of similarity among input samples. The reset signal to the recognition layer causes either a different class to be selected or, if no more classes are available, it indicates the end of the training process. We refer the reader to Appendix A for a brief description of the training procedure (see also Carpenter *et al.* [51]).

After training the network, a number Q of classes is generated to cluster all of the training data. Small values of the vigilance parameter ρ result in coarse clustering, whereas large values of ρ result in fine clustering. Thus, it is evident that the performances are strongly affected by the choice of the vigilance parameter ρ, both in terms of false alarm rates and actual disturbance detection rates.

Pacella and Semeraro[24-27] proposed a method used to select the proper value of the ρ parameter given a targeted Type I error. In the presence of multimode data, the Fuzzy ART network could be theoretically trained to find the most appropriate clustering configuration of IC variables. However, it is not possible to control both the number of classes and the false alarm rate by acting only on the

vigilance parameter. The procedure proposed by Pacella and Semeraro[24-27] assumes that a single control region is suitable for monitoring the process, regardless of the distribution and including mixture distributions. To the best of the authors' knowledge, this method is the only training approach that allows control of the false alarm rate, and hence, it is used as a reference in this study.

5.2 Automated Selection of the Vigilance Parameter and Design Procedure

When no more than one class is formed during the training process, a monotonic relationship exists between the false alarm rate and the vigilance parameter ρ. Moreover, by decreasing the vigilance parameter and fixing any other parameter, the number of classes generated during the training process decreases and converges to one. These two Fuzzy ART properties are useful in developing an iterative procedure for the selection of the proper value of ρ, given a targeted Type I error. The procedure requires the data acquired under natural process conditions to be divided into two sets: a *training set* used to train the network as discussed above and a *tuning set* used to determine the vigilance parameter value that provides the desired false alarm rate.

Given M training samples and N tuning samples, the procedure is applied as follows:

1. Let i be an iteration index and η the vigilance step; initialize $i = 1$, and set a small value for η (e.g., $\eta = 1.0e^{-4}$);
2. Set the vigilance parameter to $\rho_i = 1 - i\eta$, and train the network on the M training samples;
3. Repeat step 2 by setting $i = i + 1$ until only one class is formed. Let ρ_u be the maximum value that induces one class, and from this step on, only values in $[0, \rho_u]$ will be considered for the vigilance parameter;
4. Re-initialize the iteration index $i = 0$, and set a smaller vigilance step (e.g., $\eta = 1.0e^{-5}$);
5. Train the Fuzzy ART on the M training samples by setting $\rho_i = \rho_u - i\eta$;
6. Disengage learning, and calculate the actual false alarm rate on the N tuning samples;
7. Repeat steps 5 and 6 by setting $i = i + 1$ until the actual false alarm rate is equal to the targeted one within a given tolerance.

When the iterative procedure is completed, the final network and the resulting vigilance parameter ρ are saved and will be used to monitor any newly acquired sample. The procedure used to design and implement the MSPC approach based on the Fuzzy ART technique consists of a Phase I and a Phase II. The main steps in Phase I include (i) collection of a training dataset and a tuning dataset that are representative of the natural process conditions; (ii) re-scaling of original data (if required); (iii)

complement coding of original data; (iv) Fuzzy ART network training (see Appendix A); and (v) Fuzzy ART network tuning for the selection of the vigilance parameter (see above). Next, Phase II simply consists of applying the same re-scaling and complement coding operations to any new observations and submission of the resulting data vector to the Fuzzy ART network with the parameters estimated during Phase I. The observation is eventually classified as either IC or OOC.
For further details on the vigilance parameter selection and the neural network implementation, see Pacella and Semeraro[24-27].

6 Comparison of Methods

6.1 Simulated Scenarios

A condition often encountered in industrial practice consists of a multimode process characterized by a mixture of natural patterns. In this study, two multimode reference scenarios were considered to test and compare the performances provided by the different methods: (i) *Scenario A* is representative of a two-mode process, with a limited displacement between the modes; (ii) *Scenario B* is representative of a three-mode process, with clusters that are easily separable and located far away from one another. In both cases, without loss of generality, the data in each mode are randomly drawn from a bivariate Gaussian distribution with equal prior probability. The two reference scenarios are shown in Fig. 4 ($M = 2000$ samples). The same analysis were performed for different kinds of distributions, and the results confirmed the main findings and conclusions discussed in this study. Additional results are available from the authors upon request.

INSERT FIGURE 4 ABOUT HERE

Let $\{\boldsymbol{x}_j \in \mathbb{R}^2, j = 1, \dots, M\}$, where $\boldsymbol{x}_j = [x_{j1}, x_{j2}]^T$, be a bivariate dataset generated by the underlying process $p(\boldsymbol{x}_j) = \sum_{l=1}^{L} \pi_l MN(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, where L is the number of Gaussian distributions, and π_l is the kth prior probability, such that $\sum_{l=1}^{L} \pi_l = 1$. The following parameters were used to simulate the training data (natural process condition):

Scenario A:
$L = 2, \pi_l = 1/L$, for $l = 1,2$,
$\boldsymbol{\mu}_1 = [0.3, 0.3]^T, \boldsymbol{\mu}_2 = [0.5, 0.4]^T$,

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.25 & 0.125 \\ 0.125 & 0.3 \end{bmatrix} e^{-2}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.125 & -0.0625 \\ -0.0625 & 0.15 \end{bmatrix} e^{-2};$

Scenario B:

$L = 3, \pi_l = 1/L$, for $l = 1,2,3,$

$\boldsymbol{\mu}_1 = [0.33, 0.45]^T, \boldsymbol{\mu}_2 = [0.55, 0.3]^T, \boldsymbol{\mu}_3 = [0.70, 0.55]^T,$

$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.11 & -0.08 \\ -0.08 & 0.13 \end{bmatrix} e^{-2}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.42 & 0 \\ 0 & 0.04 \end{bmatrix} e^{-2}, \boldsymbol{\Sigma}_3 = \begin{bmatrix} 0.56 & 0.28 \\ 0.28 & 0.56 \end{bmatrix} e^{-3}.$

The departure from multivariate normality caused by the multimode nature of the process strongly affects the SPC performances when a traditional control chart with theoretical control limits is applied. This effect is shown in Table 1, where the Average Run Length (ARL) of a T^2 control chart under IC conditions is computed in both scenarios. A total of $M = 10000$ samples were used to design the control chart, and the other 10000 samples were used to test its performances. The actual ARL values were computed in 1000 runs. Table 1 shows that the actual ARL is much larger than the target, and the more broadly the clusters are spread, the larger the gap between the target value and the actual value. In Scenario B, the ARL is even larger than the number of samples used in the test set. With respect to Scenario A, the 99% confidence intervals were computed using the batch means approach by dividing the 1000 ARL values into 20 batches of 50 observations.

INSERT TABLE 1 ABOUT HERE

The results shown in Table 1 are due to an inflation of the sample variance-covariance matrix caused by the clustered nature of the multimode data. The result is an overestimated control limit that yields a reduced Type I error, and consequently, a larger Type II error. The use of empirical control limits for the T^2 statistics allows for improved performances, but it is not a sufficiently reliable approach, as demonstrated in the remainder of the paper.

With respect to the simulation of unnatural departures from the IC conditions, different types of deviation were considered and are listed in Table 2 and Table 3 for Scenario A and Scenario B, respectively, where δ_1, δ_2 and δ_3 are the shift parameters. The out-of-control conditions were chosen because they are representative of the most typical unnatural shifts that may occur in multimode processes. These shifts include (i) location shifts of only one cluster, (ii) location shifts of all clusters, (iii) variability increase in only one cluster, and (iv) variability increase in all clusters.

Furthermore, different directions of location shift were considered to highlight the need for a flexible and adaptive control region. Five severity levels were simulated for each disturbance, namely:

$$\begin{aligned} &\delta_1 \in \{0.025, 0.0375, 0.05, 0.0625, 0.075\} \\ &\delta_2 \in \{1.25, 1.5, 2, 2.5, 3\} \\ &\delta_3 \in \{0.01, 0.025, 0.05, 0.075, 0.1\} \end{aligned} \tag{8}$$

INSERT TABLE 2 ABOUT HERE
INSERT TABLE 3 ABOUT HERE

The Phase II performances were compared in terms of the ARL for a targeted Type I error with $\alpha = 0.01$. Without loss of generality, the proposed results are based on setting $ARL_0 = 100$ to ensure a reasonable duration of the overall simulation tests. As a matter of fact, larger values of ARL_0 require a higher computational effort in tuning the proposed methods without affecting the conclusion of the comparison analysis. Nevertheless, the control regions for both $ARL_0 = 100$ and $ARL_0 = 370$ are depicted in the graphics from here on.

In each simulation scenario, 1000 runs were performed. In each run, a set of $M = 2000$ randomly generated bivariate samples was used as the training set. Such a large number of training samples is compatible with in-process monitoring applications in which the observed variables are synthetic indices automatically extracted from sensor signals. In typical data-rich manufacturing operations, thousands of signal-based observations can be acquired in a few minutes.

With respect to the Fuzzy ART-based approach, the M samples were divided into $M_1 = 150$ training samples and N tuning samples ($M_1 + N = M = 2000$). The choice of the ratio M_1/N represents a tradeoff between the ability to train the network on a sufficient number of representative samples and the ability to estimate the actual false alarm rate on a sufficient number of tuning samples.

For all of the considered methods, a set of 10000 randomly generated samples was used as the test set. The batch means approach was used to estimate the 99% confidence intervals of the ARL estimates by dividing the 1000 ARL values into 20 batches of 50 observations.

6.2 Analysis of Results

Both the K-chart and the Fuzzy-ART-based methods assume that the Phase I dataset is representative of the natural process conditions, regardless of the number of clusters associated with different

operating modes. Thus, no prior information on the actual data distribution is used for process monitoring purposes, and no cluster identification or separation is involved in the training procedures, i.e., the two methods are completely distribution-free in nature. Moreover, in both cases, an automated procedure is used to tune the control region with respect to the Phase I data by guaranteeing compliance with the targeted Type I error. These factors make the two methods realistically comparable on a fair basis.

The K-chart and the Fuzzy ART-based method were also compared with the traditional T^2 chart with an empirical limit. In this case, the empirical limit was estimated using the approach proposed by Phaladiganon *et al.*[35].

INSERT FIGURE 5 ABOUT HERE

INSERT FIGURE 6 ABOUT HERE

With respect to the K-chart approach, Fig. 5 and Fig. 6 show the contour patterns of the irregular shaped region that encloses 100% of the training data for different values of the GRB kernel width parameter S. A discussion on the effects of using a different kernel function is reported in Sub-section 6.3.

The values of the kernel width parameter S resulting from 1000 runs under natural process conditions for different choices of the number M_o of artificial outliers are shown in Fig. 7 (the 95% confidence intervals are depicted). In particular, four different choices of M_o are considered: $M_o = M/4$, $M_o = M$, $M_o = 2M$, and $M_o = 3M$, with $M = 2000$.

INSERT FIGURE 7 ABOUT HERE

Fig. 7 shows that if a sufficient number of outliers are used in the kernel width selection procedure, such a parameter has no statistically significant effect on the estimated mean of S. In Scenario A, if $M_o \geq M$, there is no significant difference at level $\alpha = 0.05$ in the mean of S at different values of M_o, whereas in Scenario B, no significant difference in the result is also observed at $M_o = M/4$. Moreover, the automatically selected value of S in Scenario A oscillates near $S = 0.31$, which may be slightly overestimated. In Scenario B, instead, the kernel width selection procedure yields an average value of approximately $S = 0.15$, which seems to be more appropriate (see Fig. 6).

In this study, a number of artificial outliers $M_o = M = 2000$ was used for the design of the K-chart.

The boundaries of the control regions resulting from the T^2 chart, the K-chart, and the Fuzzy-ART-based approach for a training set of $M = 2000$ samples are shown in Fig. 8.

INSERT FIGURE 8 ABOUT HERE

Fig. 8 shows that the Fuzzy-ART-based approach yields a rectangular control region with possibly blunted corners. Therefore, the K-chart approach is the only one that provides a control region that adapts to the actual spread of the data. As shown in the final portion of the paper, this result leads to improved performances with respect to the T^2 chart and the Fuzzy ART method, especially in the presence of strong departures from an elliptic spread of the data.

The resulting ARLs and the corresponding 99% confidence intervals for the different disturbances simulated in Scenario A are shown in Table 4 and are also depicted in Fig. 9. The severity levels are ranked from 1 to 5, where 1 refers to the lowest level of the corresponding shift parameter and 5 refers to the highest level. Bold fonts identify the approach that provides the best performances.

INSERT TABLE 4 ABOUT HERE
INSERT TABLE 5 ABOUT HERE
INSERT TABLE 6 ABOUT HERE
INSERT TABLE 7 ABOUT HERE

For Scenario A, Table 4 and Fig. 9 show that the K-chart outperforms the T^2 chart in the presence of Disturbance 2 (shift of cluster A2), Disturbance 5 (variance increase of cluster A2), and Disturbance 6 (variance increase of both clusters). In the case of Disturbance 4 (variance increase of cluster A1), the K-chart provides lower ARLs than the T^2 chart only for large shifts because the control region in the K-chart approach better adapts to the actual shape of cluster A2, leading to a faster reaction to a shift of its centroid or an increase of its variance. However, the tightness of the T^2 elliptic control region along its semi-minor axis allows achievement of lower ARLs than the K-chart in the presence of Disturbance 1 (shift of cluster A1) and Disturbance 3 (shift of both clusters).

In Scenario A, the Fuzzy-ART-based approach outperforms the competitor methods only in the presence of Disturbance 1 (shift of cluster A1). In the case of Disturbance 2, it provides better performances than the T^2 chart, but for Disturbances 3 and 4, the T^2 chart is more reactive than the Fuzzy ART in detecting the occurrence of a shift.

The resulting ARLs in Scenario B are shown in Tables 5, 6, and 7.

Table 5 summarizes the results for the IC case and the four out-of-control scenarios that consist of rigid translations of one or more centroids towards the outside of the ellipse that encloses the data (see Fig. 7). Table 6 summarizes the results for the scenarios that consist of rigid translations towards the inside of the ellipse that encloses the data. Eventually, Table 7 shows the results for scenarios involving a variance increase of at least one cluster.

INSERT FIGURE 9 ABOUT HERE

Figs. 10, 11, and 12 graphically depict the results listed in Table 5, Table 6, and Table 7, respectively. Table 5 and Fig. 10 show that the Fuzzy-ART-based approach outperforms the T^2 chart if outward shifts of one or more centroids are present in Scenario B, with the only exception of Disturbance 2 (leftward shift of cluster B2) in which the improvement occurs for large shifts only. The Fuzzy-ART-based approach performs better than the K-chart for Disturbance 6 (up- and right-ward shift of cluster B3) and Disturbance 7 (outward shift of all clusters), whereas the two methods provide analogous results for Disturbance 4 (downward shift of cluster B2). In case of Disturbance 2 (left-ward shift of cluster B1), the K-chart performs slightly better for large shifts only.
The K-chart approach performs better than the T^2 chart for all of the disturbances reported in Table 3, at least for shifts of medium and large severity.

INSERT FIGURE 10 ABOUT HERE

Table 6 and Fig. 11 show that when the cluster shift is directed toward the inside of the ellipse (or the rectangle) that encloses the data, both the T^2 chart and the Fuzzy-ART-based approach completely fail in detecting the disturbance. In the case of inward translations, an increasing trend of the ARL is observed because the probability of observing data outside the elliptical or rectangular control regions decreases. If the natural process condition is described by a mixture distribution, the operator may be interested in detecting any deviation from that condition, regardless of the direction of the shift. In this frame, a rectangular or an elliptical control region is not adequate to detect a translation of one or more clusters in the multivariate variable space, and the methods based on those types of control regions lack actual flexibility.

INSERT FIGURE 11 ABOUT HERE

In contrast, the K-chart is able to detect the shift regardless of the direction of the centroid translation due to the irregular-shaped control region, which adapts to the actual spread of the data. Therefore, the K-chart is the most flexible non-parametric method among the methods considered in this work. Table 7 and Fig. 12 show that the K-chart outperforms both the T^2 chart and the Fuzzy-ART-based approach in detecting any variance increase that involves one or more clusters. Additionally, in this case, the particular nature of the control region provides a faster reaction with respect to disturbances that affect the data dispersion.

INSERT FIGURE 12 ABOUT HERE

6.3 Use of Different Kernel Functions

As stated in Section 4, the two most popular alternatives to the RGB kernel function in the SVM literature[47] are the polynomial function $K_{poly}(\boldsymbol{a} \times \boldsymbol{b})$, and the sigmoidal function $K_{sigm}(\boldsymbol{a} \times \boldsymbol{b})$:

$$\begin{aligned} K_{poly}(\boldsymbol{a} \times \boldsymbol{b}) &= (1 + \mathbf{a}^T \boldsymbol{b})^d \\ K_{sigm}(\boldsymbol{a} \times \boldsymbol{b}) &= \tanh(d_1 \mathbf{a}^T \boldsymbol{b} + d_2), \end{aligned} \tag{9}$$

where $d \in \mathrm{N}^+$ is the polynomial kernel parameter and $(d_1, d_2) \in \mathbb{R}$ are two sigmoidal kernel parameters. Different combinations of values for d_1 and d_2 were compared by Lin and Lin[52], who demonstrated that the choices of $d_1 > 0$ and $d_2 < 0$ are the most suitable, but they concluded that the RGB function is preferable in general. Furthermore, the dependency on two parameters renders the kernel optimization procedure more expensive from a computational viewpoint. For these reasons, only the polynomial function is considered and compared with the GRB function in this work.

The automatic procedure for the selection of the kernel parameter described in Section 4 can be applied without modifications to the polynomial function. The only difference is represented by the discrete nature of the parameter $d \in \mathrm{N}^+$. Fig. 13 shows the geometry of the control regions for $ARL_0 = 100$ and $M = 2000$ in Scenario A and Scenario B using the polynomial kernel and the RGB kernel, where the values of S and d result from the automatic selection procedure.

One limitation of the polynomial kernel[38] is that it is strongly influenced by observations with a large norm. The result is an undesired inflation of the control-region volume, which can be reduced by centering and re-scaling the original observations. As an example, Fig. 13 shows the boundary of the

bivariate regions estimated for the data transformed to a zero mean and unit standard deviation. Notice that by rescaling the data, the range of suitable values of the RGB kernel parameter S is changed as well. The optimal values estimated in this case are $S = 2.2$ in Scenario A and $S = 1.3$ in Scenario B.

INSERT FIGURE 13 ABOUT HERE

Fig. 13 shows that the control region based on the GRB function fits the data better than the control region based on the polynomial function. The difference is larger in Scenario B in which the clusters are more widely spread. The optimal polynomial degree is $d = 17$ in Scenario A and $d = 15$ in Scenario B. At lower degrees, the contour of the region approaches an ellipse, whereas at higher degrees, the contour becomes more irregular without enhancing the data fitting. However, in the overall explored domain of function degrees, the control region generated by the polynomial kernel is always much less representative of the actual data distribution than the region generated by the GRB kernel. This empirical outcome shows that in the presence of mixed distributions, the GRB is preferred over other common kernels in accordance with the results of previous studies devoted to single-mode non-normal distributions[38,52].

7 A Real Test Case

In this section, a real case study that addresses chatter detection in a roll grinding process, as mentioned in Section 1, is discussed and used to highlight the applicability of the proposed methods in an actual industrial operation. The in-process acquisition of sensor signals has particular industrial relevance because it allows the detection of undesired process phenomena that affect product quality and implementation of adaptive control actions. However, signal data might present a multimode pattern caused by frequent changes of the cutting parameters during each grinding cycle[53-54], and this situation makes chatter detection[55] a troublesome task using traditional control charting methods. Despite a body of literature devoted to the chatter detection problem[56-62], few automatic methods have been considered for actual industrial implementation. In this framework, multimode SPC techniques may represent a valuable alternative to common approaches. For a review of chatter vibration fundamentals in grinding processes, see Altintas and Weck[63] and Inasaki *et al.*[64].

The scheme of the experimental setup used to collect real signal data during a roll grinding process is presented in Fig. 2a). The grinding process was performed on a special alloyed steel roll with an

initial diameter of $500\ mm$ and an axial length of $1700\ mm$. The grinding wheel was constructed of an aluminum oxide material with a nominal diameter of $700\ mm$ and a width of $75\ mm$. The accelerometer signal along the X axis was sampled at $2\ kHz$ and segmented into sliding time windows of duration $T = 1\ s$, with an overlap ratio of 90%. The signal was processed online to compute two synthetic indices denoted by rms_j and $kurt_j$, $j = 1,2, \ldots$, which consist of the root mean square index of the vibration signal and the kurtosis of its time-domain distribution within the jth time window. The rms index was chosen because it represents the most basic choice for vibration monitoring in industrial applications[53]. Our previous experimental studies showed that use of the $kurt$ index in combination with the rms index enhances the capability of chatter detection.

The result is a bivariate quality characteristic $\{\boldsymbol{x}_j \in \mathbb{R}^2, j = 1,2, \ldots\}$, where $\boldsymbol{x}_j = [rms_j, kurt_j]^T$.

The experiments were performed as follows. A total of M grinding passes were carried out under chatter-free conditions with different combinations of the cutting parameters. These combinations of parameters are representative of the different operative conditions adopted by the operator during each grinding cycle. Therefore, the data collected during this phase are expected to be representative of the natural process behavior and are hence used as the Phase I dataset.

To simulate the OOC Phase II data, a new set of grinding passes was performed to induce chatter vibrations that grow over time as the waviness on the workpiece and the wheel becomes increasingly severe. The waviness on the wheel was artificially induced to produce the chatter onset. At the end of each sub-set of the grinding passes, the surface condition of the roll was checked by visual inspection, and the presence or absence of chatter marks was used to qualify the passes as IC or OOC, respectively.

The different combinations of cutting parameters used in all the test runs are reported in Table 8. The roll speed was held constant during the experiments at $n_w = 30\ rpm$.

INSERT TABLE 8 ABOUT HERE

The data collected under chatter-free conditions (see Fig. 2b)) with the eight combinations of parameters shown in Table 8 were used as the Phase I dataset (approximately $M = 3000$ observations).

A traditional SPC approach would require the design of one control chart for each cutting condition (each mode of the multimode IC state), which could lead to a considerable amount of work. In addition, any departure from the multivariate normality assumption within each mode implies the need for a non-parametric approach. The major advantage provided by the use of a non-parametric

multimode technique consists of the design of a single monitoring tool that is able to cover multiple operative conditions regardless of the actual data distribution in each mode.

INSERT TABLE 9 ABOUT HERE

Table 9 shows the detection percentages of data acquired under the chattered conditions provided by the T^2 chart, the Fuzzy-ART-based approach and the K-chart. The corresponding control regions at $ARL_0 = 100$ and $ARL_0 = 370$ are shown in Fig. 14 in which the data are depicted after the re-scaling operation imposed by the Fuzzy ART-based approach.

INSERT FIGURE 14 ABOUT HERE

Due to its adaptive properties, the K–chart allows detection of greater than 98% of the data acquired under chattered conditions. Fig. 14 shows that the SVDD-based procedure yields a double control region that encloses the two separate clusters. In this case, the estimated kernel parameter is $S = 0.0425$. The Fuzzy-ART-based approach is the one that provides the lowest performance because a large percentage of the out-of-control observations are spread within the rectangular control region that encloses the IC data. The T^2 chart performs slightly better than the Fuzzy ART, but an elliptical control region (analogous to a rectangular region) is far from a good choice for the multimode grinding data.

The results in the real case study confirm the major conclusions drawn based on the simulation experiments. The K–chart outperforms the other two approaches due to its kernel-based procedure for estimation of a flexible and adaptive control region.

8 Conclusions

The implementation of traditional MSPC tools based on conventional assumptions on the underlying data distribution might not be appropriate for industrial applications of practical interest, especially if in-process sensor data are used. A challenging violation of traditional MSPC assumptions consists of a multimode process characterized by transitions from one operating mode to another. The development of distribution-free MSPC tools that are able to monitor multimode processes is of great practical interest, but it has attracted limited attention in the mainstream literature thus far, especially in the discrete part manufacturing field.

Our study is aimed at investigating the applicability of certain non-parametric MSPC methods to multimode processes and demonstrating their performances using a comparison study that includes both simulated and real industrial data.

The paper compares two distribution-free methods based on one-class classification variants of two well-known statistical learning techniques, i.e., the Fuzzy-ART-based approach and the so-called K-chart, which require no assumption on the data distribution and their variance-covariance structure. Our analysis showed that the Fuzzy-ART-based approach generates a rectangular-like control region, possibly with blunted corners. Such a control region is the result of the procedure adopted to select the vigilance parameter ρ, which assumes that all of the training data can be grouped into a single IC class. The single-class assumption allows design of a simple-to-implement procedure to control the false alarm rate, but it may lead to reduced performance in the presence of clustered data.

Further research efforts are required to design a Fuzzy-ART-based approach that allows control of the number of classes generated during the training step and the false alarm rate at the same time.

The K-chart provides a more flexible solution due to an irregularly shaped control region that adapts to the actual spread in the data. The simulation results showed that the K-chart is able to detect departures from the natural multimode distribution regardless of the direction of the shift within the multivariate variable space. The greater the departure from multi-normality and/or single-mode distributions, the greater the expected benefits provided by the K-chart over other methods will be.

A real case study that addresses chatter detection in roll grinding via in-process sensor signals was proposed to evaluate the implementation of the proposed methods in an actual industrial application. The results achieved in the real case study confirm the main conclusions drawn based on the simulation analysis. In particular, due to its adaptive properties, the K-chart is the more flexible approach among those considered in this work for monitoring of a multimode process regardless of the actual distribution of the acquired variables.

With respect to the K-chart design, we tested two different kernel functions, i.e., the GRB function and the polynomial function. We showed that the automated procedure for selection of the kernel parameter could be applied without modification to a different kernel without any modifications. However, the results show that the GRB function provides better adaptability to mixture distributions than the polynomial function, and hence, it should be generally preferred.

The automated procedure for selection of the kernel parameter involves simulation of artificial outliers during the training phase. Our simulation analysis showed that such a procedure is robust with respect to the number of artificial outliers, at least when such a number is reasonably large. Future studies might be aimed at further assessment of the sensitivity of such a procedure with

respect to different parameters and settings. Possible improvements of the procedure may be studied as well, e.g., avoiding the need for the inclusion of artificially generated outliers.

Acknowledgements

The real case study discussed in this work is the result of a joint project involving Politecnico di Milano and ITIA CNR. The authors especially thank Giacomo Bianchi, Marco Leonesio and Paolo Parenti. This research was partially supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 285075 – MuProD.

References

1 Yu, J., Qin, S. J., Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models, AIChE Journal, 2008, 54:7, 1811-1829

2 Choi S. W., Park J. H., Lee I. B., Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis, Computers and Chemical Engineering, 2004, 28, 1377-1387

3 Yoo C. K., Villez K., Lee I. B., Rosén C., Vanrolleghem P. A., Multimodel statistical process monitoring and diagnosis of a sequencing batch reactor, Biotechnology and Bioengineering, 2007, 96:4, 687-701

4 Lane S., Martin E. B., Kooijmans R., Morris A. J., Performance monitoring of a multiproduct semi-batch process, Journal of Process Control, 2001, 11:1, 1-11

5 Hwang D. H., Han C., Real-time monitoring for a process with multiple operating modes, Control Engineering Practice, 1999, 7:7, 891–902

6 Zhao S. J., Zhang J., Xu Y. M., Monitoring of processes with multiple operating modes through multiple principal component analysis models, Industrial and Engineering Chemistry Research, 2004, 43:22, 7025-7035

7 Xie, X., Shi, H., Dynamic multimode process modeling and monitoring using adaptive Gaussian mixture models, Industrial & Engineering Chemistry Research, 2012, 51(15), 5497-5505

8 Wang, F., Tan, S., Peng, J., Chang, Y., Process monitoring based on mode identification for multi-mode process with transitions, Chemometrics and Intelligent Laboratory Systems, 2012, 110(1), 144-155

9 Ge, Z., Song, Z., Online monitoring of nonlinear multiple mode processes based on adaptive local model approach, Control Engineering Practice, 2008,16(12), 1427-1437

10 Kano, M., Hasebe, S., Hashimoto, I., Ohno, H., Evolution of multivariate statistical process control: application of independent component analysis and external analysis, Computers & chemical engineering, 2004, 28(6), 1157-1166

11 Ge, Z., Yang, C., Song, Z., Wang, H., Robust online monitoring for multimode processes based on nonlinear external analysis, Industrial & Engineering Chemistry Research, 2008, 47(14), 4775-4783

12 Woodall W. H., Montgomery, D. C., Research Issues and Ideas in Statistical Process Control, Journal of Quality Technology, 1999, 31, pp. 376–386

13 Chakraborti S., Van der Laan P., Bakir S.T., Nonparametric Control Charts: an Overview and Some Results, Journal of Quality Technology, 2001, 33:3, 304–315

14 Graham M. A., Mukherjee A., Chakraborti S., Distribution-free exponentially weighted moving average control charts for monitoring unknown location, Computational Statistics & Data Analysis, 2012, 56:8, 2539-2561

15 Qiu, P., Hawkins, D.M., A Rank Based Multivariate CUSUM Procedure, Technometrics, 2001, 43, 120–132.

16 Qiu, P., Hawkins, D.M., A nonparametric multivariate CUSUM procedure for detecting shifts in all directions, JRSS-D (The Statistician), 2003, 52:2, 151-164

17 Qiu, P., Distribution-Free Multivariate Process Control Based On Log-Linear Modeling, IIE Transactions, 2008, 40, 664–677

18 Qiu, P., Introduction to Statistical Process Control, 2014, Boca Raton, FL: Chapman & Hall/CRC

19 Coleman, D. E., "Individual Contributions" in "A Discussion on Statistically-Based Process Monitoring and Control", edited by D. C. Montgomery and W. H. Woodall, Journal of Quality Technology, 1997, 29, 148–149

20 Tax, D.M.J., Duin, R.P.W., Data Domain Description Using Support Vectors, European Symposium on Artificial Neural Networks, 1999, 251–256

21 Kim S. B., Sukchotrat T., Park S-K., A Nonparametric Fault Isolation Approach Through One-Class Classification Algorithms, IIE Transactions, 2011, 43:7, 505-517

22 Zorriassatine F., Al-Habaibeh A., Parkin R. M., Jackson M. R., Coy J., Novelty Detection for Practical Pattern Recognition in Condition Monitoring of Multivariate Processes: A Case Study, International Journal of Advanced Manufacturing Technology, 2005, 25, 954-963

23 Ning X., Tsung F., Improved Design of Kernel Distance-Based Charts Using Support Vector Methods, IIE Transactions, 2013, 45:4, 464-476

24 Pacella M., Semeraro Q., Anglani A., Adaptive Resonance Theory based neural algorithms for manufacturing process quality control, International Journal Of Production Research, 2004, 42, 4581-4607

25 Pacella M., Semeraro Q., Anglani A., Manufacturing quality control by means of a Fuzzy ART network trained on natural process data, Engineering Applications Of Artificial Intelligence, 2004, 17, 83-96

26 Pacella M., Semeraro Q., Understanding ART-Based Neural Algorithms As Statistical Tools For Manufacturing Process Quality Control, Engineering Applications Of Artificial Intelligence, 2005, 18, 645-662

27 Pacella M., Semeraro Q., Monitoring Roundness Profiles based on an Unsupervised Neural Network Algorithm, Computers & Industrial Engineering, 2011, 60, 677-689

28 Liu R.Y., Control Charts For Multivariate Processes, Journal of the American Statistical Association, 1995, 90:432, 1380 – 1387

29 Liu R.Y., Singh K., Teng J.H., DDMA-charts: Nonparametric Multivariate Moving Average Control Charts Based On Data Depth, Allgemeines Statistisches Archiv, 2004, 88, 235-258

30 Messaoud A., Weihs C., Hering F., A Nonparametric Multivariate Control Chart Based On Data Depth, Technical Reports from Technische Universität Dortmund, Sonderforschungsbereich 475: Komplexitätsreduktion in multivariaten Datenstrukturen, 2004

31 Zou, C., Wang, Z., and Tsung, F., A Spatial Rank-based Multivariate EWMA Control Chart, Naval Research Logistics, 2012, 59:2, 91–110

32 Bersimis S., Psarakis S., Panaretos J., Multivariate Statistical Process Control Charts: An Overview, Quality and Reliability Engineering International, 2007, 23, 517-543

33 Martin E.B., Morris A.J., Non-Parametric Confidence Bond for Process Performance Monitoring Charts, Journal of Process Control, 1996, 6:6, 349 – 358

34 Chou Y-M., Mason R.L., Young J.C., The Control Chart For Individual Observations From A Multivariate Non-Normal Distribution, Communication in Statistics – Theory and Methods, 2001, 30:8-9, 1937 – 1949

35 Phaladiganon P., Kim S.B., Chen V.C.P., Baek J-G., Park S-K., Bootstrap-based T2 Multivariate Control Charts, Communications in Statistics – Simulation and Computation, 2011, 40:5, 645 – 662

36 Hall, P., DiCiccio, T.J., Romano, J.P., On Smoothing and the Bootstrap, The Annals of Statistics, 1989, 17, 692–704

37 Al-Ghanim A., An Unsupervised Learning Neural Algorithm For Identifying Process Behavior On Control Charts And A Comparison With Supervised Learning Approaches, Computers and Industrial Engineering, 1997, 32, 627–639

38 Tax D. M. J., One-Class Classification; Concept-Learning In The Absence Of Counter-Examples, Ph.D. thesis, Delft University of Technology, 2001

39 Guh R-S., Integrating Artificial Intelligence into On-line Statistical Process Control, Quality and Reliability Engineering International, 2003, 19, 1-20

40 Psarakis S., The Use of Neural Networks in Statistical Process Control Charts, Quality and Reliability Engineering International, 2011, 27, 641-650

41 Tax D.M.J., Duin R.P.W. (2004), Support Vector Data Description, Machine Learning, 54, 45-66

42 Sun R., Tsung F., A Kernel-Distance-Based Multivariate Control Chart Using Support Vector Methods, International Journal of Production Research, 2003, 41:13, 2975-2989

43 Camci F., Chinnam R. B., Ellis R. D., Robust Kernel Distance Multivariate Control Chart Using Support Vector Principles, International Journal of Production Research, 2008, 46:18, 5075-5095

44 Sukchotrat T., Kim S. B., Tsung F., One-Class Classification-based Control Charts for Multivariate Process Monitoring, IIE Transactions, 2009, 42:2, 107-120

45 Ge Z., Gao F., Song Z., Batch Process Monitoring Based on Support Vector Data Description Method, Journal of Process Control, 2011, 21, 949-959

46 Gani W., Taleb H., Limam M., An Assessment of the Kernel-Distance-Based Multivariate Control Chart Through an Industrial Application, Quality and Reliability Engineering International, 2011, 27, 391-401

47 Hastie T., Tibshirani R., Friedman J., The elements of statistical learning, 2009, New York: Springer.

48 Tax, D.M.J., Duin, R.P.W., Uniform Object Generation For Optimizing One-Class Classifiers, Journal of Machine Learning Research, 2002, 2, 155–173

49 Hagan M. T., Demuth H. B., Beale M., Neural Network Design (Boston, MA, USA: PWS Publishing Co.), 1996

50 Georgiopoulos M., Dagher I., Heileman G. L., Bebis G., Properties of Learning of a Fuzzy ART Variant, Neural Networks, 1999, 12, 837-850

51 Carpenter G. A., Grossberg S., Rosen D. B., Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System, Neural Networks, 1991, 4:6, 759–771

52 Lin H. T., Lin C. J., A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods, Technical report, Department of Computer Science, National Taiwan University, 2003

53 Maggioni M., Marzorati E., Grasso M., Colosimo B.M., Parenti P., In-Process Quality Characterization of Grinding Processes: a Sensor-Fusion Based Approach, to be presented at the ASME 12th Biennial Conference on Engineering Systems Design, ESDA 2014, June 25-27, 2014, Copenhagen, Denmark

54 Parenti P., Leonesio M., Cassinari A., Bianchi G., Monno M., Model-Based Identification of Chatter Marks during Cylindrical Grinding, International Conference on Advanced Manufacturing Engineering and Technologies, NEWTECH 2013, October 27-30, 2013 Stockholm, Sweden

55 Quintana G., Ciurana, J., Chatter in machining processes: a review, International Journal of Machine Tools and Manufacture, 2011, 51:5, 363-376

56 Karpushewski B., Wehmeier M., Inasaki I., Grinding monitoring system based on power and acoustic emission sensors, Annals of the CIRP, 2000, 49, 235 – 240

57 Govekar, E., Baus, A., Gradišek, J., Klocke, F., Grabec, I., A new method for chatter detection in grinding. CIRP Annals-Manufacturing Technology, 2002, 51:1, 267-270

58 Gradišek, J., Baus, A., Govekar, E., Klocke, F., Grabec, I., Automatic chatter detection in grinding. International Journal of Machine Tools and Manufacture, 2003, 43:14, 1397-1403

59 González-Brambila, O., Rubio, E., Jáuregui, J. C., Herrera-Ruiz, G., Chattering detection in cylindrical grinding processes using the wavelet transform. International Journal of Machine Tools and Manufacture, 2006, 46:15, 1934-1938

60 Tönshoff, H. K., Friemuth, T., Becker, J. C., Process monitoring in grinding. CIRP Annals-Manufacturing Technology, 2002, 51:2, 551-571

61 Inasaki I., Sensor fusion for monitoring and controlling grinding processes, International Journal of Advanced Manufacturing Technology, 1999, 15, 730-736

62 Aguiar, P. R., Serni, P. J., Dotto, F. R., & Bianchi, E. C., In-process grinding monitoring through acoustic emission, Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2006, 28:1, 118-124

63 Altintas Y., Weck M., Chatter stability of metal cutting and grinding, CIRP Annals-Manufacturing Technology, 2004, 53:2, 619-642

64 Inasaki I., Karpuschewski B., Lee H.S, Grinding Chatter - Origin and Suppression, Annals of the CIRP, 2001, 50:2, 515–534

Appendix A: The Fuzzy-ART Training Procedure

The weight initialization consists of setting $w_{q,1} = w_{q,2} = \cdots = w_{q,2p} = 1$. The F_2 node with these weights is referred to as an *uncommitted node*. After one class is chosen to represent an input pattern $\boldsymbol{x}_j^c$, the corresponding F_2 node is referred to as a *committed node*. Given a multivariate training set $\{\boldsymbol{x}_j\}, j = 1, \dots, M$, the training phase of Fuzzy ART is applied as follows:

1. Initialize the number of committed nodes to 0, and only one uncommitted node remains;
2. Set a choice parameter $\beta \in [0, \infty]$ (a small value is usually adopted, e.g., $\beta = 10^{-6}$ in this work) and a vigilance parameter $\rho \in [0,1]$;
3. Apply the complement coding to the new input sample $\boldsymbol{x}_j$;
4. Calculate the bottom-up inputs to the F_2 nodes as follows:

$$T_q(\boldsymbol{x}_j) = \begin{cases} \dfrac{p}{\beta + 2p} & \text{if the } q^{th} \text{ node is the uncommitted node} \\ \dfrac{|\boldsymbol{x}_j^c \wedge \boldsymbol{w}_q|}{\beta + |\boldsymbol{w}_q|} & \text{if the } q^{th} \text{ node is a committed node} \end{cases} \tag{A1}$$

   where operator '∧' gives the vector $\boldsymbol{u} \wedge \boldsymbol{v} = (\min\{u_1, v_1\}, \dots, \min\{u_i, v_i\} \dots)$, and operator '||' gives the scalar $|\boldsymbol{u}| = \sum_i \mathrm{abs}(u_i)$;
5. Choose the F_2 node that receives the maximum bottom-up input (assume that this node has index q_{max}). Three cases can now be distinguished:
   a) The q_{max} node is the uncommitted node, and in this case, the vigilance criterion is satisfied:

$$\frac{|\boldsymbol{x}_j^c \wedge \boldsymbol{w}_{q_{max}}|}{|\boldsymbol{x}_j^c|} \geq \rho \tag{A2}$$

   Increase the number of committed nodes by one. A new uncommitted node is introduced, and its weight vector is initialized as discussed above: Go to step 6;
   b) The q_{max} node is a committed node, and it satisfies the vigilance criterion (A2): Go to step 6;
   c) The q_{max} node is a committed node, but it does not satisfy the vigilance criterion (A2): Disqualify the node by setting $T_q(\boldsymbol{x}_j) = -1$ and repeat Step 5;
6. The weights associated with the q_{max} node are modified according to the following equation:

$$\boldsymbol{w}_{q_{max}} = \boldsymbol{w}_{q_{max}} \wedge \boldsymbol{x}_j^c \tag{A3}$$

If this is the last input sample, go to step 7; otherwise, go back to step 3;

7. After all M samples are presented, two cases are possible:
   a) In the previous training data presentation, at least one component of the weight vectors is changed: Go to step 3 by presenting each training sample again to the network;
   b) In the previous training data presentation, no weight change occurred: The learning process is complete.

The stop condition at step 7 is driven by a tolerance index (e.g., $tol = 1e - 6$) for comparison of the weight vector in consecutive training data presentations.

Note that by choosing a small value of the choice parameter β, the convergence of the learning process is guaranteed after one presentation of the input data to the network[50].

Table 1 – Actual IC ARL performances of T^2 chart with theoretical limit

Scenario	Target ARL	Actual ARL with 99% confidence intervals
A	100	198.48 [195.81, 201.15]
	370	874.76 [845.88, 903.64]
B	100	>10000
	370	>10000

Table 2 – Description of simulated disturbances in Scenario A

Disturbance	Description	Settings
1	Shift of cluster A1 centroid	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 - [\delta_1, \delta_1]^T$
2	Shift of cluster A2 centroid	$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2 + [\delta_1, \delta_1]^T$
3	Shift of both the cluster centroids	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 + [\delta_1, -\delta_1]^T$,
		$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2 + [\delta_1, -\delta_1]^T$
4	Variance increase of cluster A1	$\{\boldsymbol{\Sigma}_1\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_1\}_{i,i}, i = 1,2$
5	Variance increase of cluster A2	$\{\boldsymbol{\Sigma}_2\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_2\}_{i,i}, i = 1,2$
6	Variance increase of both the clusters	$\{\boldsymbol{\Sigma}_1\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_1\}_{i,i}, i = 1,2$
		$\{\boldsymbol{\Sigma}_2\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_2\}_{i,i}, i = 1,2$

Table 3 – Description of simulated disturbances in Scenario B

Disturbance	Description	Settings
1	Rightward shift of cluster B1 centroid	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 + [\delta_3, 0]^T$
2	Leftward shift of cluster B1 centroid	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 - [\delta_3, 0]^T$
3	Upward shift of cluster B2 centroid	$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2 + [0, \delta_3]^T$
4	Downward shift of cluster B2 centroid	$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2 - [0, \delta_3]^T$
5	Down-left-ward shift of cluster B3 centroid	$\boldsymbol{\mu}_3 = \boldsymbol{\mu}_3 - [\delta_3, \delta_3]^T$
6	Up-right-ward shift of cluster B3 centroid	$\boldsymbol{\mu}_3 = \boldsymbol{\mu}_3 + [\delta_3, \delta_3]^T$
7	Outward shift of all the clusters	$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1 - [\delta_3, 0]^T, \boldsymbol{\mu}_2 = \boldsymbol{\mu}_2 - [0, \delta_3]^T,$ $\boldsymbol{\mu}_3 = \boldsymbol{\mu}_3 + [\delta_3, \delta_3]^T$
8	Variance increase of cluster B1	$\{\boldsymbol{\Sigma}_1\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_1\}_{i,i}, i = 1,2$
9	Variance increase of cluster B2	$\{\boldsymbol{\Sigma}_2\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_2\}_{i,i}, i = 1,2$
10	Variance increase of cluster B3	$\{\boldsymbol{\Sigma}_3\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_3\}_{i,i}, i = 1,2$
11	Variance increase of all the clusters	$\{\boldsymbol{\Sigma}_1\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_1\}_{i,i}, i = 1,2$ $\{\boldsymbol{\Sigma}_2\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_2\}_{i,i}, i = 1,2$ $\{\boldsymbol{\Sigma}_3\}_{i,i} = \delta_2 \cdot \{\boldsymbol{\Sigma}_3\}_{i,i}, i = 1,2$

Table 4 – Comparison of ARLs and 99% confidence intervals (Scenario A)

Scenario A	Severity	T^2 chart		Fuzzy ART		K-chart	
In Control	-	*102.98*	*[98.06, 107.90]*	*99.37*	*[95.12, 103.63]*	*103.30*	*[97.79, 108.81]*
Disturbance 1 Shift of cluster A1	1	61.88	[60.15, 63.60]	**48.09**	**[46.40, 49.77]**	81.43	[77.28, 85.57]
	2	45.07	[44.01, 46.14]	**39.49**	**[34.88, 44.10]**	63.18	[58.97, 67.38]
	3	31.60	[30.92, 32.28]	**24.53**	**[22.53, 26.54]**	41.75	[39.46, 44.04]
	4	22.25	[21.73, 22.77]	**16.27**	**[14.34, 18.19]**	31.24	[28.42, 34.06]
	5	15.37	[15.03, 15.72]	**11.08**	**[10.00, 12.15]**	19.66	[18.19, 21.13]
Disturbance 2 Shift of cluster A2	1	85.09	[80.48, 89.70]	50.35	[44.44, 56.26]	**28.85**	**[27.34, 30.36]**
	2	71.75	[69.23, 74.28]	27.46	[23.70, 31.22]	**11.83**	**[11.08, 12.58]**
	3	54.90	[56.70, 57.11]	13.80	[12.05, 15.55]	**5.78**	**[5.42, 6.14]**
	4	40.92	[39.19, 42.65]	7.95	[7.26, 8.63]	**3.41**	**[3.27, 3.55]**
	5	27.59	[26.44, 28.74]	4.79	[4.50, 5.09]	**2.48**	**[2.44, 2.52]**
Disturbance 3 Shift of both the cluster centroids	1	**26.41**	**[25.58, 27.24]**	63.02	[53.15, 72.89]	34.61	[32.35, 36.87]
	2	**12.66**	**[12.41, 12.91]**	32.32	[28.64, 36.00]	15.54	[14.77, 16.30]
	3	**6.70**	**[6.62, 6.79]**	16.74	[15.03, 18.44]	8.30	[7.74, 8.85]
	4	**3.98**	**[3.91, 4.05]**	10.75	[9.36, 12.14]	4.53	[4.32, 4.74]
	5	**2.59**	**[2.57, 2.62]**	6.00	[5.40, 6.59]	2.92	[2.82, 3.01]
Disturbance 4 Variance increase of cluster A1	1	**54.58**	**[50.86, 54.58]**	68.68	[65.99, 71.37]	**52.10**	**[50.40, 53.81]**
	2	**33.46**	**[31.58, 33.46]**	48.93	[46.60, 51.25]	**31.27**	**[30.22, 32.33]**
	3	**17.13**	**[16.57, 17.13]**	27.29	[26.38, 28.21]	**16.14**	**[15.60, 16.70]**
	4	11.60	[11.30, 11.60]	18.60	[17.94, 19.26]	**11.05**	**[10.83, 11.26]**
	5	9.03	[8.78, 9.03]	14.12	[13.71, 14.53]	**8.55**	**[8.38, 8.73]**
Disturbance 5 Variance increase of cluster A2	1	73.98	[71.80, 76.15]	73.72	[68.20, 79.25]	**64.75**	**[61.29, 68.22]**
	2	53.79	[52.59, 55.00]	53.70	[51.09, 56.31]	**41.62**	**[39.87, 43.37]**
	3	33.11	[32.30, 33.92]	31.79	[30.30, 33.27]	**22.26**	**[21.74, 22.78]**
	4	23.15	[22.77, 23.52]	23.10	[21.84, 24.37]	**14.82**	**[14.40, 15.23]**
	5	17.34	[17.05, 17.63]	16.76	[15.99, 17.52]	**11.46**	**[11.23, 11.69]**
Disturbance 6 Variance increase of both the clusters	1	45.86	[43.60, 45.86]	50.93	[48.98, 52.89]	**39.48**	**[37.94, 41.02]**
	2	26.06	[24.61, 26.06]	30.58	[29.68, 31.49]	**21.34**	**[20.56, 22.12]**
	3	12.81	[12.34, 12.81]	15.96	[15.53, 16.44]	**10.28**	**[10.02, 10.55]**
	4	8.411	[8.16, 8.41]	10.63	[10.39, 10.87]	**6.77**	**[6.66, 6.88]**
	5	6.40	[6.25, 6.40]	7.88	[7.66, 8.10]	**5.14**	**[5.06, 5.23]**

Table 5 – Comparison of ARLs and 99% confidence intervals (Scenario B) – part 1

Scenario B	Severity	T^2 chart		Fuzzy ART		K-chart	
In Control	-	*101.89*	*[97.77, 106.00]*	*102.39*	*[97.04, 107.75]*	*102.32*	*[98.57, 106.07]*
Disturbance 2 Leftward shift of cluster B1	1	**79.63**	**[77.25, 80.02]**	**82.98**	**[78.06, 87.90]**	93.10	[88.45, 97.74]
	2	**50.86**	**[49.15, 52.57]**	**51.96**	**[48.34, 55.57]**	57.26	[54.20, 60.32]
	3	24.64	[23.95, 25.32]	**19.29**	**[17.01, 21.56]**	**18.20**	**[16.04, 20.37]**
	4	12.90	[12.55, 13.25]	8.47	[7.83, 9.12]	**6.12**	**[5.71, 6.52]**
	5	7.42	[7.31, 7.53]	4.73	[4.41, 5.06]	**3.54**	**[3.46, 3.61]**
Disturbance 4 Downward shift of cluster B2	1	**79.56**	**[77.32, 81.80]**	**75.78**	**[70.47, 81.09]**	**77.30**	**[73.29, 81.30]**
	2	45.44	[44.12, 46.76]	**35.96**	**[32.77, 39.16]**	**34.59**	**[32.54, 36.64]**
	3	14.33	[13.87, 14.78]	**8.81**	**[7.87, 9.76]**	**8.23**	**[7.86, 8.60]**
	4	5.46	[5.40, 5.53]	**3.84**	**[3.72, 3.95]**	**3.80**	**[3.72, 3.87]**
	5	3.41	[3.38, 3.43]	**3.03**	**[3.01, 3.05]**	**3.04**	**[3.02, 3.05]**
Disturbance 6 Up-right-ward shift of cluster B3	1	96.07	[91.38, 100.75]	**61.73**	**[58.94, 64.52]**	90.39	[85.40, 95.37]
	2	70.23	[67.66, 72.80]	**23.26**	**[21.88, 24.65]**	44.50	[41.71, 47.29]
	3	16.37	[15.81, 16.93]	**6.19**	**[5.87, 6.51]**	9.71	[9.10, 10.33]
	4	4.93	[4.87, 4.99]	**3.43**	**[3.37, 3.49]**	3.96	[3.85, 4.08]
	5	3.14	[3.13, 3.16]	**3.00**	**[2.99, 3.00]**	3.05	[3.03, 3.07]
Disturbance 7 Outward shift of all the clusters	1	60.99	[59.09, 62.90]	**40.55**	**[38.66, 42.44]**	69.78	[67.20, 72.35]
	2	27.43	[26.74, 28.12]	**11.58**	**[11.19, 11.98]**	23.62	[22.28, 24.96]
	3	6.50	[6.34, 6.66]	**2.87**	**[2.81, 2.93]**	4.47	[4.31, 4.63]
	4	2.25	[2.23, 2.28]	**1.47**	**[1.45, 1.49]**	1.78	[1.75, 1.81]
	5	1.37	[1.37, 1.38]	**1.15**	**[1.14, 1.16]**	1.19	[1.18, 1.20]

Table 6 – Comparison of ARLs and 99% confidence intervals (Scenario B) – part 2

Scenario B	Severity	T^2 chart		Fuzzy ART		K-chart	
Disturbance 1 Rightward shift of cluster B1	1	132.96	[128.21, 137.71]	133.23	[126.37, 140.10]	**87.66**	**[83.55, 91.78]**
	2	185.16	[176.64, 193.67]	170.38	[151.96,188.79]	**50.89**	**[46.60, 55.18]**
	3	265.84	[251.86, 279.81]	186.42	[161.78, 211.07]	**14.07**	**[12.98, 15.15]**
	4	315.93	[303.19, 328.66]	183.44	[158.54, 208.35]	**5.18**	**[4.98, 5.38]**
	5	337.37	[318.44, 356.29]	202.70	[169.89, 235.51	**3.42**	**[3.37, 3.47]**
Disturbance 3 Upward shift of cluster B2	1	121.22	[118.15, 124.28]	135.26	[127.27, 143.26]	**94.42**	**[89.02, 99.83]**
	2	137.00	[132.83, 141.16]	155.14	[143.74, 166.54]	**51.86**	**[46.88, 56.85]**
	3	143.73	[138.98, 148.48]	143.76	[131.09, 156.43]	**13.16**	**[10.61, 15.72]**
	4	149.69	[145.24, 154.14]	166.18	[141.56, 190.79]	**5.27**	**[4.78, 5.74]**
	5	146.01	[141.51, 150.51]	150.34	[132.87, 167.81]	**3.52**	**[3.42, 3.61]**
Disturbance 5 Down-left-ward shift of cluster B3	1	102.89	[100.26, 105.52]	167.85	[159.02, 176.69]	**88.11**	**[84.72, 91.49]**
	2	107.46	[103.01, 111.91]	264.06	[226.10, 302.01]	**47.16**	**[44.27, 50.04]**
	3	102.96	[98.93, 106.98]	266.09	[232.21, 299.96]	**12.13**	**[10.95, 13.31]**
	4	103.98	[100.41, 107.57]	309.44	[259.93, 358.96]	**4.55**	**[4.20, 4.89]**
	5	102.46	[99.72, 105.20]	315.87	[234.46, 397.27]	**3.27**	**[3.10, 3.45]**

Table 7 – Comparison of ARLs and 99% confidence intervals (Scenario B) – part 3

Scenario B	Severity	T^2 chart		Fuzzy ART		K-chart	
Disturbance 8 Variance increase of cluster B1	1	79.88	[77.70, 82.07]	86.92	[81.95, 91.88]	**49.29**	**[47.52, 51.05]**
	2	64.41	[62.38, 66.43]	68.67	[65.41, 71.93]	**30.62**	**[29.45, 31.78]**
	3	45.93	[44.96, 46.90]	51.29	[48.65, 53.93]	**16.73**	**[16.28, 17.18]**
	4	35.36	[34.68, 36.05]	38.52	[36.80, 40.24]	**11.91**	**[11.61, 12.20]**
	5	29.40	[28.80, 30.01]	30.75	[29.76, 31.75]	**9.50**	**[9.34, 9.66]**
Disturbance 9 Variance increase of cluster B2	1	79.07	[76.58, 81.56]	86.99	[83.38, 90.59]	**70.54**	**[67.78, 73.31]**
	2	62.44	[60.49, 64.40]	71.00	[68.21, 73.78]	**48.94**	**[47.31, 50.56]**
	3	42.22	[41.41, 43.02]	46.01	[43.81, 48.20]	**28.73**	**[27.86, 29.60]**
	4	31.00	[30.34, 31.67]	33.62	[32.38, 34.86]	**19.53**	**[18.97, 20.08]**
	5	24.93	[24.45, 24.41]	26.12	[25.10, 27.15]	**15.10**	**[14.74, 15.47]**
Disturbance 10 Variance increase of cluster B3	1	100.42	[97.49, 103.34]	77.87	[75.23, 80.51]	**67.26**	**[64.83, 69.69]**
	2	95.26	[91.92, 98.60]	57.52	[55.65, 59.40]	**44.13**	**[42.07, 46.19]**
	3	86.41	[83.34, 89.48]	36.32	[34.60, 38.04]	**24.26**	**[23.40, 25.13]**
	4	74.02	[71.58, 76.47]	26.66	[25.23, 28.09]	**16.17**	**[15.79, 16.55]**
	5	63.07	[61.28, 64.86]	21.27	[20.23, 22.31]	**12.63**	**[12.31, 12.95]**
Disturbance 11 Variance increase of all the clusters	1	64.01	[62.22, 65.81]	56.13	[53.70, 58.55]	**33.53**	**[32.43, 34.64]**
	2	44.66	[43.50, 45.82]	34.87	[33.87, 35.86]	**17.68**	**[17.22, 18.14]**
	3	26.57	[25.89, 27.25]	19.04	[18.77, 19.31]	**8.40**	**[8.27, 8.53]**
	4	18.40	[18.14, 18.66]	13.00	[12.69, 13.30]	**5.65**	**[5.55, 5.74]**
	5	14.18	[13.98, 14.37]	9.63	[9.52, 9.75]	**4.31**	**[4.26, 4.36]**

Table 8 – Cutting parameters used in the real case study

Condition	Wheel speed [rpm] (n_s)	Infeed [mm] (a_e)
In-Control (IC) - Chatter-free	680	0.01
	680	0.02
	780	0.01
	780	0.02
	830	0.01
	830	0.02
	1100	0.01
	1100	0.02
Out-of-Control (OOC) - Chatter	1100	0.01

Table 9 – Chatter detection percentage in the real case study

ARL_0	T^2	Fuzzy ART	K-Chart ($S = 0.0425$)
100	63.73%	43.63%	98.04%
370	61.76%	42.65%	98.04%