Profile Monitoring via Sensor Fusion: the use of PCA Methods for Multi-Channel Data

M. Grasso[1a], B.M. Colosimo[a], M. 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

Profile Monitoring via Sensor Fusion: the use of PCA Methods for Multi-Channel Data

Abstract: Continuous advances of sensor technology and real-time computational capability is leading to data rich environments to improve industrial automation and machine intelligence. When multiple signals are acquired from different sources (i.e., multi-channel signal data), two main issues must be faced: i) reduce data dimensionality, to make the overall signal analysis system efficient and actually implementable in industrial environment, and ii) fuse together all the sensor outputs to achieve a better comprehension of the process. In this frame, Multi-way Principal Component Analysis (PCA) represents a multivariate technique to perform both the tasks. The paper investigates two main multi-way extensions of the traditional PCA to deal with multi-channel signals, one based on unfolding the original datasets, and one based on multi-linear analysis of data in their tensorial form. The approaches proposed for data modelling are combined with appropriate control charting to achieve multi-channel profile data monitoring. The developed methodologies are demonstrated with both simulated and real data. The real data come from an industrial sensor fusion application in waterjet cutting, where different signals are monitored to detect faults affecting the most critical machine components.

1   Introduction

The development of low-cost, non-intrusive and smart sensors on one hand, and the continuous improvement of real-time computational capability on the other hand, make a large amount of data potentially available in industry. In this frame, sensor signals acquired during the process provide a suitable source of information to develop an in-process quality control and to allow a faster implementation of corrective actions. In several applications, the acquired signals present cyclically repeating patterns; in those cases the suite of profile monitoring techniques (Woodall et al., 2004; Williams et al., 2007) provides the natural framework to evaluate the stability over time of process

quality. An overview of parametric and nonparametric approaches for profile data as well as application domains investigated at this time can be found in the recent book edited by Noorossana et al. (2012).

This paper focuses on the specific case of monitoring profiles that are signal data. On this topic, the first seminal paper on signal profile monitoring is due to Jin and Shi (1999), who suggested using wavelet analysis to monitor tonnage signals in stamping processes. Some years later, the same authors (Jin and Shi, 2001) proposed a feature extraction and classification approach based on wavelet analysis for force signals in welding processes. With a similar approach, Chang and Yadama (2010) combined wavelet decomposition and B-spline smoothing for quality control of tonnage signals. Zhou et al. (2005) studied a directionally variant multivariate control chart in forging processes.

The largest portion of profile monitoring literature focuses on single signal analysis, regardless the strong industrial interest for multi-signal applications. Data-rich environments in industry, in fact, are leading to an increasing demand for multi-sensor data fusion methods to solve quality-related problems. The most widely studied applications in literature include stability analysis and chatter detection (Kuljanic et al., 2009; Inasaki, 1999) and tool condition monitoring (Cho et al., 2010; Wang et al., 2007; Chen and Jen, 2000; Bahr et al., 1997; Bhattacharyya and Sengupta, 2009; Lezanski, 2001; Shi and Gindy, 2007; Aliustaoglu et al. 2009). However, only few authors studied profile monitoring approaches in the field of sensor fusion. Among them, Kim et al. (2006) proposed a multi-channel profile monitoring method based on Principal Curves to monitor multiple homogeneous signals in forging processes. Amiri et al. (2013) investigated the problem of integrated monitoring of mixed-type data, i.e. profile data and multivariate quality characteristics, possibly coming from multiple sources. Recently,

Paynabar et al. (2013) proposed a multi-way extension of the Principal Component Analysis (PCA) technique to classify multi-channel profile data.

In this study, we consider the use of multi-way extensions of the PCA to deal with information fusion of multi-channel signals. The goal consists in transforming a set of profile data from multiple sources into a synthetic feature set that explains the correlation structure of original data.

Two different multi-way PCA control chart formulations are proposed: one based on unfolding the multi-way dataset into a matrix, in order to apply the traditional PCA on the transformed matrix (hereafter called Vectorized PCA - VPCA), and one based on applying the PCA directly on the multi-way dataset, preserving its higher-order tensor representation (Multi-linear PCA - MPCA).

The two techniques have different features that make them suitable to multi-channel signal analysis and monitoring problems. In some cases, the VPCA may provide reasonable performances, and it may result more flexible than the MPCA. In some other cases, however, the MPCA may be preferred because of its higher computational efficiency, and because it provides a better interpretability of results.

This study extends the work of Paynabar et al. (2013) in two directions: by introducing a multi-way generalization of PCA-based control charts, whereas Paynabar et al. (2013) focused on clustering techniques, and by considering the general case of heterogeneous signals. Differently from Paynabar et al. (2013), an unconstrained and more general formulation of the MPCA has been exploited in our study, in which the principal components obtained by the method may be correlated, contrary to regular PCA, but they are not limited by the number of channels considered in the specific test case. This study is also an extension of the previous studies of Colosimo and Pacella (2007; 2010), which presented the use of the PCA for profile monitoring of single profiles.

The proposed methods are tested both on simulated and real industrial data. Monte Carlo simulations based on multi-channel combinations of benchmark signals were used to evaluate the performances in a number of controlled scenarios. Then, real data are used to demonstrate the main features of the proposed methods on a real test case of industrial interest. Multi-sensor signals were acquired in a waterjet cutting process, and multi-way PCA methods were applied to detect faults affecting the most critical machine components. In waterjet processes the aggressiveness of abrasive particles and the challenging operative pressure conditions affect the reliability of machine tool components, which are subject to different types of faults and performance degradation. The lifetime of most stressed components is difficult to predict and the nature of different types of faults makes them almost impossible to prevent. Therefore, there is the need for a reliable health monitoring equipment able to provide a continuous automated assessment of machine conditions. These capabilities are required to cope with unattended processes, to implement remote monitoring services and to enhance maintenance and production management strategies.

In Section 2, a review of the theoretical background of multi-way PCA methods is provided; Section 3 reports the results achieved by comparing the two methods in the frame of Monte Carlo simulations; Section 4 briefly describes the real test case in waterjet machining; Section 5 discusses the comparative analysis results; and Section 6 concludes the paper.

2   Theoretical Background

Multi-way data analysis is the extension of two-way methods to higher-order datasets (Acar and Yener, 2009; De Latheuwer et al., 2000). A 2-way dataset may be represented in terms of a $N \times P$ matrix, where N is the number of samples and P is the number of variables: in this frame the PCA is a well understood and used multivariate

technique to explain the variance-covariance structure through a few linear combinations of the original variables (Jolliffe, 2002). The first proposals for 3-way and higher-way generalizations of the PCA date back to the 1960s and early 1970s (Kiers, 2000).

One possible approach to deal with multi-way arrays involves the 'matricization' operation (Kiers, 2000), which consists of unfolding the multi-dimensional dataset into a bidimensional one. As far as the PCA for process monitoring is concerned, this is the Multi-way PCA approach proposed by Nomikos and MacGregor (1995) for batch processes. The application of the PCA to unfolded datasets is the VPCA technique, which is a commonly exploited approach in a number of applications. A different technique consists in performing the PCA directly on the original tensorial data representation, without pre-processing the data by the unfolding procedure. This second approach is referred to as MPCA (Lu et al., 2008, 2009).

Different multi-linear extensions of the PCA have been proposed in literature: some of them are limited to the case of 2D data (and are especially used in image analysis), like 2D-PCA (Yang et al., 2004) or the Generalized PCA (Ye et al., 2004), and some other may be applied to tensors of any order (Jolliffe, 2002).

Several authors pointed out different advantages of the multi-linear approach over the one based on matricization, concerning the higher efficiency in terms of computational costs and memory demands, the easer interpretation of retained Principal Components (PCs), and the possibility to better characterize the actual multi-linear correlation structure (Acar and Yener, 2009; Lu et al., 2009; Paynabar et al., 2013). Regardless the respective pros and contras, the two methods may lead to different interpretation of results, and are both suitable to deal with multi-channel data.

The VPCA and MPCA for multi-channel profile data are described hereafter. In both cases, a multivariate control chart approach is combined to the multi-way PCA for

statistical process control. The application of Multi-way PCA techniques to synthetic indexes extracted from multi-channel signals is discussed at the end of the Section.

2.1 The VPCA Approach

A Q-way array χ is a tensor object $\chi \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_Q}$ such that I_q represents the dimension of the q-mode, $q = 1, \ldots, Q$, where the term '*mode*' refers to a generic set of entities (Kiers, 2000). In the frame of multi-channel profile data, the simplest Q-way dataset is a $(N \times P \times J)$ 3-way array such that N is the number of channels, P is the number of data points collected on each profile, and J is the number of multi-channel profiles. Note that more articulated datasets may be generated by introducing additional modes, e.g. by adding a further mode to group together different families of profiles.

The VPCA approach consists of unfolding the array χ slice by slice, rearranging the slices into a large two-dimensional matrix **X**, and then performing the regular PCA on **X** (Nomikos and MacGregor, 1995).

Notice that there are multiple possible rearrangements of the original array χ into a matrix, and each of them corresponds to looking at a different type of variability.

INCLUDE FIG. 1 ABOUT HERE

As far as the aforementioned multi-channel 3-way array is concerned, the most meaningful unfolding approach consists in concatenating the N (channel) P-dimensional profiles in the jth sample ($j = 1, 2, \ldots, J$) into a single 1-dimensional vector of length PN as shown in Fig. 1.

In this case a $(J \times PN)$ matrix **X** is obtained. Although an equal number P of data points in each channel is required to generate the 3-way array χ, such a constraint is not necessary with VPCA (even though samples from the same channel must be of equal

length). If the number of data points in the nth channel ($n = 1, 2, \ldots, N$) is P_n, then the transformed data matrix will have dimensions $J \times (\sum_{n=1}^{N} P_n)$. This may lead to a more flexible approach than Multi-linear PCA in some applications.

The control chart approach based on regular PCA (Colosimo and Pacella, 2007) may be applied to the matricized data. Let P' be the number of columns obtained by unfolding the original multi-way dataset, and M the number of sample to be used to estimate the PCA model (each sample is a realization of N profiles, one from each channel). Then, the PCA-based method consists of performing a spectral decomposition of the sample variance-covariance matrix $\mathbf{S}_{1:M}$ of the ($M \times P'$) data matrix $\mathbf{X}_{1:M}$, i.e. finding the matrices $\mathbf{L}$ and $\mathbf{U}$ that satisfy the relationship:

$$\mathbf{U}^T \mathbf{S}_{1:M} \mathbf{U} = \mathbf{L}$$

1)

Where $\mathbf{L}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{S}_{1:M}$ (λ_i; $i = 1, \ldots, P'$), while $\mathbf{U}$ is an orthonormal matrix whose i^{th} column $\mathbf{u}_i$ is the ith eigenvector of $\mathbf{S}_{1:M}$.

When the profiles refer to heterogeneous quantities, data standardization is required before computing the sample variance-covariance matrix $\mathbf{S}_{1:M}$. Standardization consists of subtracting to each column of $\mathbf{X}_{1:M}$ the corresponding sample mean value computed on the M samples, and dividing the result by the corresponding sample standard deviation. The projection of the jth sample onto the K-dimensional Principal Component (PC) orthogonal space is defined as follows:

$$\mathbf{z}_j = \mathbf{U}^T(\mathbf{x}_j - \bar{\mathbf{x}}) = \left[z_{j,1}, \dots, z_{j,K}\right]^T \qquad (j = 1,2,\dots) \qquad (2)$$

Where $\mathbf{x}_j$ is the jth row of the data matrix $\mathbf{X}_{1:M}$ and $\bar{\mathbf{x}} = (1/M)\sum_{j=1}^{M}\mathbf{x}_j$ is the average profile among the M ones used to estimate the PCA model. K is the maximum number of PCs that can be extracted, i.e. the maximum number of non-zero eigenvalues. K is upper-bounded by $\min\{P', M\}$.

The ith eigenvector $\mathbf{u}_i$ contains the weights (*loadings*) associated with the ith PC, and hence it weights the contribution of each profile data point to the corresponding linear combination.

The first PC is the maximum variance linear combination; the second PC is the maximum variance linear combination having zero-correlation with the first one; and so on. The relative importance of each PC, i.e. the amount of explained variance, is represented by the value of the corresponding eigenvalue. Therefore, the relevant information content may be captured by a reduced number of PCs, providing the dimensionality reduction at the origin of the PCA popularity. Different methods have been proposed to automatically select a number m of PCs to be retained. A very effective one was proposed by Wold (1978) and it is based on a cross-validation algorithm. For a comparison of methods see Valle *et al.* (1999).

By retaining the first m PCs, each sample – i.e. each row of the matrix $\mathbf{X}_{1:M}$ – may be reconstructed as follows:

$$\hat{\mathbf{x}}_j(m) = \bar{\mathbf{x}} + \sum_{i=1}^{m} z_{j,i}\mathbf{u}_i \qquad (j = 1,2,\dots) \qquad (3)$$

The process monitoring strategy requires the computation of two statistics (Colosimo and Pacella, 2007): one is the Hotelling's T^2 statistics, used to detect possible deviations along the directions of the first m PCs:

$$T_j^2(m) = \sum_{i=1}^{m} \frac{z_{j,i}^2}{\lambda_i} \qquad (j = 1,2, \ldots) \tag{4}$$

The second is the Sum of Squared Errors (SSE) statistics, used to detect possible deviations in directions orthogonal to the ones associated to the first m PCs, given by:

$$SSE_j(m) = (\hat{\mathbf{x}}_j(m) - \bar{\mathbf{x}})^T (\hat{\mathbf{x}}_j(m) - \bar{\mathbf{x}}) \qquad (j = 1,2, \ldots) \tag{5}$$

For the design procedure of control charts based on the T^2 and SSE statistics see Sec. 2.3.

2.2 The MPCA Approach

The Multi-linear methodology allows applying the PCA technique without unfolding the original dataset.

The basic MPCA approach (as all the other multi-linear extensions of regular PCA) produces correlated PCs, contrary to PCA (Lu et al. 2008). An extension of the MPCA named Uncorrelated Multi-linear PCA (UMPCA), which introduces the zero-correlation constraint among PCs derived from an iterative procedure aimed at finding directions capturing maximum variance, has been also proposed in the literature (Lu et al. 2009). The zero-correlation constraints introduces a limitation on the maximum number of PCs that may be extracted: such a number is upper-bounded by $\min\{\min_q I_q, M\}$, whereas the remaining portion of data variability may be captured by

removing the zero-correlation constraint. In the frame of process monitoring applications, the existence of correlation among the extracted PCs does not affect the performances, since the information content is summarized by the Hotelling's T^2 statistics. Because of this, the MPCA approach is here adopted instead of the UMPCA one. In particular, the method proposed by Lu et al. (2008) is hereafter reviewed, and a novel control chart approach is proposed.

Given a Q-way array $\chi \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_Q}$, we assume, without loss of generality, that the last mode represents the sample replicates (i.e., I_Q is the number of samples). Thus, $\chi_{\cdot,j}$ is the jth sample of $Q - 1$ dimensional tensor objects $\chi_{\cdot,j} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_{Q-1}}$. The MPCA objective is to determine a multilinear transformation $\{\widetilde{\mathbf{U}}^{(q)} \in \mathbb{R}^{I_q \times P_q}, q = 1, \dots, Q - 1\}$ that maps the original tensor space $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_{Q-1}}$ into a tensor subspace $\mathbb{R}^{P_1 \times P_2 \times \dots \times P_{Q-1}}$ with $P_q < I_q$ for $q = 1, \dots, Q - 1$:

$$Y_j = (\chi_{\cdot,j} - \bar{\chi}_{\cdot}) \times_{q=1}^{Q-1} \left\{ \widetilde{\mathbf{U}}_i^{(q)^T}, q = 1, \dots, Q - 1 \right\} \quad (j = 1, 2, \dots) \tag{6}$$

where $\bar{\chi}_{\cdot} = (1/M) \sum_{j=1}^{M} \chi_{\cdot,j}$ is the average $(Q - 1)$-way array among the M ones used to estimate the MPCA model. The symbol $\times_q$ denotes the *mode q multiplication*, i.e., the multiplication of an array by a matrix along the q-mode of the array.

The variability captured by the projected tensor $Y_j \in \mathbb{R}^{P_1 \times P_2 \times \dots \times P_{Q-1}}$ is measured by the total tensor scatter defined as $\psi_Y = \sum_{j=1}^{M} \|Y_j\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm defined by $\|\cdot\|_F = \sqrt{\langle \cdot, \cdot \rangle}$.

Thus, the goal is to compute the $Q - 1$ projection matrices $\{\widetilde{\mathbf{U}}^{(q)} \in \mathbb{R}^{I_q \times P_q}, q = 1, \dots, Q - 1\}$ that maximize the total tensor scatter ψ_Y. As far as the full projection is concerned, i.e. the projection with $P_q = I_q$ for $q = 1, \dots, Q - 1$, $\widetilde{\mathbf{U}}^{(q)} \in \mathbb{R}^{I_q \times I_q}$ is the

matrix comprised of the eigenvectors of $\mathbf{\Phi}^{(q)} = \sum_{j=1}^{M}(\mathbf{X}_{j(q)} - \overline{\mathbf{X}}_{(q)})(\mathbf{X}_{j(q)} - \overline{\mathbf{X}}_{(q)})^{T}$,

where $\mathbf{X}_{j(q)}$ is the equivalent matrix representation of $\chi_{.j}$ by unfolding the qth mode,

and $\overline{\mathbf{X}}_{(q)} = (1/M)\sum_{j=1}^{M}\mathbf{X}_{j(q)}$.

Let $T_S \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_{Q-1}}$ be the total scatter tensor of the full projection such that

each entry is defined as:

$$T_S(i_1, i_2, ..., i_{Q-1}) = \sum_{j=1}^{M}[Y_j(i_1, i_2, ..., i_{Q-1})]^2$$

$$(i_q = 1, ..., I_q, \text{ for } q = 1, ..., Q - 1) \tag{7}$$

where $Y_j \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_{Q-1}}$, then the $(i_q)^{th}$ eigenvalue $\lambda_{i_q}^{(q)}$ is the sum of all the

entries of the $(i_q)^{th}$ q-mode slide of T_S:

$$\lambda_{i_q}^{(q)} = \sum_{i_1=1}^{I_1} ... \sum_{i_{q-1}=1}^{I_{q-1}} \sum_{i_{q+1}=1}^{I_{q+1}} ... \sum_{i_{Q-1}=1}^{I_{Q-1}} T_S(i_1, ..., i_{q-1}, i_q, i_{q+1}, ..., i_{Q-1})$$

$$(i_q = 1, ..., I_q, \text{ for } q = 1, ..., Q - 1) \tag{8}$$

An approximate approach for dimensionality reduction consists of truncating the

q-mode eigenvectors beyond the $(P_q)^{th}$ such that the retained fraction of the total scatter

in the q-mode is about $S^q = \left(\sum_{i_q=1}^{P_q} \lambda_{i_q}^{(q)} / \sum_{i_q=1}^{I_q} \lambda_{i_q}^{(q)}\right)$, being $S^1 \cong S^2 \cong ... \cong S^{Q-1} \cong S$.

This approach allows retaining a reduced number of PCs such that a given percentage of

the overall variability is captured in each mode. This method is an extension of the

dimensionality reduction strategy of the regular PCA to the multi-linear case.

In the 3-way case of multi-channel data above mentioned, given the tensor $\chi_{.j} \in$

$\mathbb{R}^{N \times P}$, the full projection consists of two matrices $\widetilde{\mathbf{U}}^{(1)} \in \mathbb{R}^{N \times N}$ and $\widetilde{\mathbf{U}}^{(2)} \in \mathbb{R}^{P \times P}$, such

that the $(i_1)^{th}$ column of $\widetilde{\mathbf{U}}^{(1)}$ includes the loadings associated to the $(i_1)^{th}$ 1-mode PC ($i_1 = n = 1, \dots, N$), i.e., one weight per channel, and the $(i_2)^{th}$ column of $\widetilde{\mathbf{U}}^{(2)}$ includes the loadings of associated to the $(i_2)^{th}$ 2-mode PC ($i_2 = i = 1, \dots, P$), i.e., one weight per data-point.

The distinction between weights associated to channels and weights associated to data-points is expected to improve the interpretability of extracted PCs with respect to the VPCA method.

A different approach for dimensionality reduction has been proposed by Lu et al. (2008): it is based on an iterative procedure called Sequential Mode Truncation (SMT). The optimization of $\widetilde{\mathbf{U}}^{(q)} \in \mathbb{R}^{I_q \times P_q}$ with $P_q < I_q$ depends on the projections in other modes, since $\widetilde{\mathbf{U}}^{(q)}$ consists of the P_q eigenvectors corresponding to the largest P_q eigenvalues of the following matrix:

$$\mathbf{\Phi}^{(q)} = \sum_{j=1}^{m}\left(\mathbf{X}_{j(q)} - \overline{\mathbf{X}}_{(q)}\right) \cdot \widetilde{\mathbf{U}}_{\Phi^{(q)}} \cdot \widetilde{\mathbf{U}}_{\Phi^{(q)}}^{T} \cdot \left(\mathbf{X}_{j(q)} - \overline{\mathbf{X}}_{(q)}\right)^{T}$$
$$(q = 1, \dots, Q-1)$$
$$(9)$$

where:

$$\widetilde{\mathbf{U}}_{\Phi^{(q)}} = \left(\widetilde{\mathbf{U}}^{(q+1)} \otimes \widetilde{\mathbf{U}}^{(q+1)} \otimes \dots \otimes \widetilde{\mathbf{U}}^{(Q-1)} \otimes \widetilde{\mathbf{U}}^{(1)} \otimes \widetilde{\mathbf{U}}^{(2)} \otimes \dots \otimes \widetilde{\mathbf{U}}^{(q-1)}\right)$$
$$(q = 1, \dots, Q-1)$$
$$(10)$$

being $\otimes$ the Kronecker product. The proof of (9) can be found in Lu et al. (2008).

Because of this, the dimensionality reduction in one mode cannot be determined independently from the others. The MPCA algorithm proposed by Lu et al. (2008) allows solving the optimization problem by an iterative truncation of q-mode eigenvectors,

subject to a targeted dimensionality reduction specified by the user. However, Lu *et al.* (2008) demonstrated that the results provided by the approximated approach are very close to the ones provided by the SMT procedure, and they concluded that the former method could be safely used instead of the more computationally extensive SMT alternative. Because of this, the former approach is used in this study.

Once a reduced number $m = \prod_q P_q$ of PCs is retained, for $q = 1, \dots, Q-1$, by capturing an approaximately equal percentage of total scatter S in every mode, the jth sample can be reconstructed as follows:

$$\hat{\chi}_{\cdot,j}(m) = \bar{\chi}_{\cdot} + \sum_{i_1=1}^{P_1} \dots \sum_{i_{Q-1}=1}^{P_{Q-1}} Y_j(i_1, \dots, i_{Q-1}) \mathbf{u}_{i_1}^{(1)} \circ \mathbf{u}_{i_2}^{(2)} \circ \dots \circ \mathbf{u}_{i_{Q-1}}^{(Q-1)}$$

$$(j = 1, 2, \dots) \tag{11}$$

where $\mathbf{u}_{i_q}^{(q)}$ is the $(i_q)^{th}$ column of the projection matrix $\widetilde{\mathbf{U}}^{(q)}$.

In analogy with the regular PCA approach, the Hotelling's T^2 statistics can be used to detect possible deviations along the directions of the m extracted PCs. Since the PCs may result correlated, the general formulation of Hotelling's T^2 statistics should be used as follows:

$$T_j^2(m) = (Y_j^* - \overline{Y^*}) \mathbf{S}_{Y^*}^{-1} (Y_j^* - \overline{Y^*})^T$$

$$(j = 1, 2, \dots) \tag{12}$$

where $\mathbf{Y}^*$ is the matricized version of the projected tensor Y after q-mode eigenvectors truncation, such that Y_j^* is a $(1 \times m)$ vector and Phase I $\mathbf{Y}^*$ is a $(M \times m)$

matrix; $\overline{\mathbf{Y}^*} = (1/M)\sum_{j=1}^{M}\mathbf{Y}_j^*$; and $\mathbf{S}_{Y^*}$ is the $(m \times m)$ estimated covariance matrix of $\mathbf{Y}^*$ in Phase I. The estimated covariance matrix $\mathbf{S}_{Y^*}$ is computed as follows:

$$\mathbf{S}_{Y^*} = \mathbf{V}'\mathbf{V}/2(M-1) \tag{13}$$

where $\mathbf{V}$ is the $(M-1 \times m)$ matrix that consists of row vectors of the differences $\mathbf{v}_j = \mathbf{Y}_{j+1}^* - \mathbf{Y}_j^*$, for $j = 1, \dots, M-1$. The definition in Eq. 13 based on successive differences is an unbiased estimator of the variance-covariance matrix if the observations are i.i.d. in Phase I, and it provides better control chart performances than the common estimator, which pools all the Phase I observations (Williams et al., 2006).

In addition, the SSE statistics can be used to detect possible deviations in directions orthogonal to the ones associated to the m extracted PCs. The SSE statistics may be computed as follows:

$$SSE_j(m) = (\hat{\chi}_{\cdot,j}(m) - \bar{\chi}_{\cdot})^T(\hat{\chi}_{\cdot,j}(m) - \bar{\chi}_{\cdot}) \qquad (j = 1,2,\dots) \tag{14}$$

Finally, since the MPCA requires an equal number P of data points in all the channels, some synchronization step is required. In this study, a synchronous re-sampling procedure is implemented. The time reference is converted into a new reference expressed in terms of a percentage of a complete pumping cycle. The new reference step is selected in order to avoid any pattern distortion or relevant information loss. Then, a periodic cubic spline interpolation is used to compute the signal values in the new reference points. A discussion about the synchronization issue can be found in Gao (2012).

2.3 *Control Chart Design and Utilization*

The VPCA and MPCA techniques involve the same profile monitoring procedure, which consists of a control chart design phase (Phase I), and a monitoring phase (Phase II).

During Phase I, a number M of multi-channel samples representative of the normal working condition of the process must be collected. The VPCA or the MPCA procedure is then applied to those M samples, and two control charts are designed, for the T^2 and SSE statistics, respectively. Control limits may be estimated as $(1 - \alpha')\%$ percentiles of the empirical distributions of $T_j^2(m)$ and $SSE_j(m)$, $j = 1, \dots, M$, where α is the overall Type I error, and $\alpha = 1 - (1 - \alpha')^{(1/2)}$ is the Type I error associated to each chart, computed by using the Sidak correction (Montgomery, 2008). When few profiles are available during Phase I, the empirical percentiles can be estimated by using a bootstrap-based procedure (Liu and Tang, 1996), consisting of drawing B bootstrap samples of size M from the original one, computing the PCA model and the $T_j^2(m)$ and $SSE_j(m)$ statistics for each sample, and then using the collection of BM realizations to estimate the empirical cumulative distribution function. During Phase II, i.e., the actual monitoring Phase, $T_j^2(m)$ and $SSE_j(m)$ values are estimated for each new observed multi-channel sample. When the VPCA approach is used, the computation of $T_j^2(m)$ and $SSE_j(m)$ statistics for the jth observed sample are based on Phase I estimates of $\mathbf{L}$, $\mathbf{U}$, and $\bar{\mathbf{x}}$. When the MPCA approach is used, those statistics are based on Phase I estimates of $\widetilde{\mathbf{U}}^{(q)}$, S_{Y^*} and $\bar{\chi}$. In both cases, a violation of at least one control limit leads to an alarm activation.

3   Simulation Analysis

The performances of the VPCA and MPCA methodologies are first compared by means of Monte Carlo simulations. The benchmark signals proposed by Donoho and Johnstone (1994) are used to generate a multi-channel test case. Those signals have been used by different authors to test wavelet-based algorithms, but also in the frame of statistical models and machine learning literature (e.g., see Koo and Kil, 2008; Ko et al., 2009; Fan et al., 2012).

Three signals proposed by Donoho and Johnstone (1994) are shown below; they are called respectively 'blocks', 'heavysine', and 'bumps'.

INCLUDE FIG. 2 ABOUT HERE

Let $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ be respectively the 'blocks', 'heavysine', and 'bumps' profiles shown in Fig. 2 (the number of data points is set to $P = 128$ for all the signals). A 4-channel profile dataset consisting of a three-way array $\chi \in \mathbb{R}^{N \times P \times J}$ was generated, such that $N = 4$ and J is the overall number of simulated profiles. χ is generated as follows:

$$\chi_{1,\cdot,j} = b_{1,j}\boldsymbol{x}_1 + b_{2,j}\boldsymbol{x}_2 + \varepsilon_{1,j}$$

$$\chi_{2,\cdot,j} = b_{3,j}\boldsymbol{x}_1{}^2 + b_{4,j}\boldsymbol{x}_3 + \varepsilon_{2,j}$$

$$(j = 1,2,\dots) \tag{15}$$

$$\chi_{3,\cdot,j} = b_{5,j}\boldsymbol{x}_2{}^2 + b_{6,j}\boldsymbol{x}_3{}^2 + \varepsilon_{3,j}$$

$$\chi_{4,\cdot,j} = b_{7,j}\boldsymbol{x}_1\boldsymbol{x}_2 + \varepsilon_{4,j}$$

where $\varepsilon_{n,j}$ is a random term ($\varepsilon_{n,j} \sim N(0,0.5)$, $n = 1, \dots, 4$ and $j = 1,2, \dots$), and $\boldsymbol{b}_j = [b_{1,j}, \dots, b_{7,j}]^T$ is the model parameter vector, such that $\boldsymbol{b}_j \sim MN(\boldsymbol{\mu_b}, \Sigma_b)$, for $j = 1,2, \dots$

The following settings were used to generate the dataset:

$$\boldsymbol{\mu_b} = [0.2, 1, 1.5, 0.5, 1, 0.7, 0.8]^T$$

$$\Sigma_b = diag(\sigma_{b_1}, \ldots, \sigma_{b_7})$$

$$= diag(0.08, 0.015, 0.05, 0.01, 0.09, 0.03, 0.06)$$

(16)

The benchmark signals proposed by Donoho and Johnstone (1994) were chosen because their complex pattern features lead to profile modeling difficulties when a parametric modeling approach is used. PCA-based methods, instead, allows capturing the main pattern features without any further profile modeling or smoothing step.

Different out-of-control scenarios were generated to simulate different kinds of deviations from the natural multi-channel pattern. The following out-of-control scenarios were considered.

Mean shift of the reference signals:

$$\boldsymbol{x}_u = \boldsymbol{x}_u + \delta_a \quad (u = 1,2,3)$$

(17)

where $\delta_a \in \{0.01, 0.025, 0.05, 0.075, 0.1\}\sigma_{x_u}$, and σ_{x_u} is the standard deviation of $\boldsymbol{x}_u$ signal, $u = 1,2,3$.

Superimposition of a sinusoid term on the reference signals:

$$\boldsymbol{x}_u = \boldsymbol{x}_u + \delta_b \boldsymbol{y}_s \quad (u = 1,2,3)$$

(18)

where $\delta_b \in \{0.025, 0.05, 0.075, 0.1, 0.125\}\sigma_{x_u}$, $u = 1,2,3$, and $\boldsymbol{y}_s$ is the sine function over the domain $[0, p]$, with period p and peak-to-peak amplitude equal to 1.

Standard deviation increase of the error term:

$$\sigma_{\varepsilon_{n,j}} = \delta_c \sigma_{\varepsilon_{n,j}} \quad (n = 1, \dots, 4 \text{ and } j = 1,2, \dots) \tag{19}$$

where $\delta_c \in \{1.1, 1.5, 2, 2.5, 3\}$ and $\sigma_{\varepsilon_{n,j}}$ is the standard deviation of the error term.

Mean shift of the model parameters:

$$\mu_{b,w} = \mu_{b,w} + \delta_d \quad (w = 1, \dots, 7) \tag{20}$$

where $\delta_d \in \{1, 2, 3, 4, 5\}\sigma_{b_w}$; $\mu_{b,w}$ and σ_{b_w} are respectively the mean value and the standard deviation of the wth model parameter, $w = 1, \dots, 7$.

Standard deviation increase of the model parameters:

$$\sigma_{b_w} = \delta_e \sigma_{b_w} \quad (w = 1, \dots, 7) \tag{21}$$

where $\delta_e \in \{1.5, 2, 2.5, 3, 4\}$.

3.1   *Simulation Results*

The explained variance associated to the first PCs and the corresponding cumulative explained variance resulting from the VPCA and the MPCA applied to a set of 5000 in-control profile samples are shown in Fig. 3.

INCLUDE FIG. 3 ABOUT HERE

Fig. 3 shows that the variance explained by the first five PCs in the VPCA case is considerably higher than the one captured by the higher order PCs, even though the

cumulative percentage of variance associated to the first five PCs is relatively low (about 33%). This is because a large contribution to the overall variability is due to the noise term. In this case, the first five PCs are suitable to capture the variability associated to the systematic pattern of the signals, and to filter out the noise effect.

In order to guarantee a comparison analysis under the same conditions, the VPCA and MPCA methods are compared being about equal the total percentage of explained variance (about 33%). With regard to the MPCA, this leads to retaining the first 1-mode PC and the first three 2-mode PCs, where 1-mode corresponds to channels and 2-mode corresponds to profile data points. Notice that the first 1-mode PC accounts for about 25% of variability, whereas the first two 1-mode PCs account for about 50%; however, we observed that, for the considered cases, by adding the second 1-mode PC no significant performance improvement was achieved, and hence only the first 1-mode PC was retained.

The performances were compared in terms of the Average Run Length (ARL), for a targeted Type I error $\alpha = 0.01$. In each scenario, 1000 runs were performed. In each run, a set of 10000 randomly generated 4-channel profiles was used in Phase I. The 10000 samples were divided into two sets of 5000 samples: the former set was used to estimate the VPCA or MPCA model, and the latter one was used to estimate the empirical control limits in order to guarantee an in-control ARL equal to 100.

Table 1 summarizes the ARL results achieved under in-control conditions, and in out-of-control scenarios a), b), and c). Table 2 summarizes the ARL results in scenarios d) and e).

INCLUDE TABLE 1 ABOUT HERE

INCLUDE TABLE 2 ABOUT HERE

The batch means method was used to estimate the 99% confidence intervals of ARL estimates, by dividing the 1000 ARL values into 20 batches of 50 observations.

Table 1 shows that the VPCA outperforms the MPCA in scenarios a), b), and c), i.e. in presence of deviations that involves the generating signals x_u ($u = 1, 2, 3$) and the standard deviation of the error terms. Table 2, instead, shows that the MPCA approach becomes a feasible competitor in presence of more complex out-of-control scenarios, where the deviations involve the distribution of the model parameters. In this case, the MPCA performs better than the VPCA in three on six cases.

Notice that the results of scenarios d) and e) for the model parameter $b_{4,j}$ ($j = 1,2, ...$) are not included in Table 2 because no effect was observed at the considered severity levels for both the methods.

The different behavior of the two methods is due to the different nature of the extracted PCs. The MPCA is thought to capture the correlation structure among different modes, and hence it is more sensitive to deviations that involve such a structure, e.g., model parameter modifications. The VPCA, instead, is more suitable to detect pattern changes that involve one or more profiles, including a mean shift or a variance increase.

INCLUDE FIG. 4 ABOUT HERE

INCLUDE FIG. 5 ABOUT HERE

INCLUDE FIG. 6 ABOUT HERE

INCLUDE FIG. 7 ABOUT HERE

Fig. 4 to Fig. 7 graphically depict the ARL performances and the corresponding 99% confidence intervals for the two methods in each simulated scenario. When the VPCA performs better than the MPCA, e.g., in scenarios b) and c), the performance improvement is considerable; otherwise, when the MPCA performs better than the VPCA the margin of improvement is reduced. This leads to an overall preference for the VPCA approach, at least for applications characterized by a limited number of profiles, as the one considered in this study.

4   A Real Case Study in Waterjet Cutting

Waterjet/abrasive waterjet (WJ/AWJ) is a flexible technology that can be profitably exploited for different operations on a wide range of materials (Kovacevic et al., 1997).

Due to challenging pressure conditions, cyclic pressure loadings, aggressiveness of abrasives and other factors, most of the components of the Ultra High Pressure (UHP) pump and the cutting head are subject to wear and unpredictable faults. Therefore, the continuous monitoring of machine health conditions is of great industrial interest, as it allows implementing condition-based maintenance strategies, and providing automatic reaction to critical faults as far as unattended processes are concerned.

Different authors studied WJ/AWJ process monitoring solutions, aimed at assessing the cutting stability (Perzel et al., 2012; Krenicky and Miroslav, 2012) and detecting process malfunctions, including non-correct jet penetration (Axinte and Kong, 2009; Rabani et al., 2012) and workpiece crack detection (Choi and Choi, 1997). A number of studies has been also focused on condition monitoring of cutting head components – orifice and mixing tube – (e.g., see Annoni et al. 2008; Annoni et al. 2009; Jurisevic et al. 2004). However, limited attention has been devoted to the condition

monitoring of UHP pump components, which is extremely relevant for condition-based maintenance purposes.

In this study, we consider a multi-sensor data fusion approach for health condition monitoring of some of the most stressed machine components, including both the UHP pump and the cutting head.

The study refers to the most common pump configuration, characterized by two circuits – an oil circuit and a water circuit –, where water pressure intensification is provided by a positive-displacement pump including three single-acting pistons (see Annoni et al., 2008 for a detailed description of the plant).

The pressure signal, acquired on the high pressure water duct, is a suitable source of information for monitoring purposes, as it is characterized by fluctuations that are influenced by both upstream and downstream flow rate modifications. A different type of signal that is strongly influenced by any variation in the pumping regime conditions is the plunger displacement signal, one for each pumping plunger. Since the two kinds of sensors provide correlated and partially complementary information, and different responses to process changes, we consider a multi-sensor fusion approach based on pressure and plunger displacement signals.

The cutting process is characterized by repeating pressure and plunger displacement profiles, one for each pumping cycle. Fig. 8 shows the dynamic pressure profiles (i.e. pressure fluctuations around the static level) and plunger displacement profiles corresponding to three complete pumping cycles. The plunger displacement profile is the result of consecutive pumping steps indicated in Fig. 5: a pre-compression step, a compression step and a suction step (see Grasso et al. 2013 for details). The signals were acquired on a 45 kW pump with a water pressure set value of 350 MPa and a 0.25 mm orifice with a sampling frequency of 2 kHz.

INCLUDE FIG. 8 ABOUT HERE

Real data were acquired both under normal health conditions and in presence of actual faults. The following faults scenarios were considered, as they involve the most critical components and refer to common contingencies in WJ shop floors:

- *Fault A:* cracked high pressure cylinder

- *Fault B*: cracked discharge check valve

- *Fault C*: worn discharge check valve seat

- *Fault D*: broken orifice

INCLUDE FIG. 6 ABOUT HERE

The experimental settings and the Multi-way PCA procedures are shown in Fig. 9. The availability of dead centre digital triggers allows performing the signal segmentation directly on-line. The signals under normal and faulty conditions were acquired by replicating the same cutting process on an aluminum plate. Regarding fault scenarios A, B and C, different faulty components were made available by the machine tool builders, and they came from actual faults. Dye penetrant analysis and visual inspection were applied to rank the components based on the fault severity. In case of fault D, instead, the effect of a broken orifice was simulated by installing orifices with a larger diameter (0.33 mm). For details about the working principle of the machine tool and the designed experiments used to collect the data, the interested reader may refer to Grasso et al. (2013).

5 Real Test Case Results

5.1 *Phase I*

The dataset to be used in Phase I includes $M = 130$ profiles acquired under normal working conditions.

The different methods are compared being equal the percentage of data variability explained by the retained PCs (a target value of 80% was set). The relative importance of each PC in the two considered methods is shown in Fig. 10.

INCLUDE FIG. 10 ABOUT HERE

The number of retained PCs to capture at least the 80% of overall data variability is $m = 16$ for the VPCA approach and $P_1 = 3$ (1-mode) and $P_2 = 7$ (2-mode) for the MPCA approach.

The loadings obtained when VPCA is applied directly to profile data are shown in Fig. 11. For sake of space only a subset of retained PCs is shown (notice that the first 4 PCs explain about 50% of the overall variability). The weights are associated to data points, and hence the loadings are profiles in the time domain. Regarding the displacements signals, the first PCs associate different weight levels to different pumping steps, with largest weight to the steps characterized by largest variability. Regarding the pressure signal, instead, the first PCs mainly capture the transient features, the low frequency ripples or a combination of them.

INCLUDE FIG. 11 ABOUT HERE

Fig. 12 shows the loadings obtained by applying the MPCA approach, i.e. two sets of loadings, one associated to channels (1-mode loadings) and one associated to data points (2-mode loadings).

The first 1-mode PC averages the contribution of the three displacement signals, with a lower weight given to the pressure signal. The second 1-mode PC is mainly influenced by the pressure signal and the third PC contrasts the third displacement signal against the other two. The 2-mode loadings of the seven retained 2-mode PCs capture both the different steps of the pumping cycle that characterize the plunger displacement signals (e.g. see PC 1, 2, 4, 5, and 6), and the fluctuations of the pressure signal (PC 3 and 7).

INCLUDE FIG. 12 ABOUT HERE

5.2   Testing Data (Phase II)

The effect of different faults on signal patterns are shown in Fig. 13, where the average Phase I profiles are compared with the average profiles in each fault scenarios, corresponding to the highest severity level. As far as Fault A, B, and C are concerned, the faulty components were installed into the plunger/cylinder group number 1. Fault A increases the compression speed of plunger 1, with a consequent impact on the duration of the pumping steps of other signals. Faults B and C mainly impact the pre-compression step of plunger 1 stroke, since the discharge check valve is closed during that step, and any leakage influences the pre-compression equilibrium state. Fault D is a downstream fault (it involves the orifice in the cutting head), and hence it has the same effect on all the plungers.

INCLUDE FIG. 13 ABOUT HERE

With regard to the dynamic pressure signal, a broken orifice results in a considerable reduction of the 6x harmonic component.

The fault detection percentage for the three different approaches is shown in Table 3. In this case the bootstrap-based approach was used to estimate the empirical control limits (Liu and Tang, 1996). A number $B = 1000$ of bootstrap samples were generated, with a targeted Type I error $\alpha = 0.01$.

INCLUDE TABLE 3 ABOUT HERE

The VPCA method provides a 100% detection capability. The MPCA method approaches the 99% detection rate: the only small fraction of missed detection occurs with Fault C, when medium and low wear levels are considered. Thus, the two methods provide very similar results, even though the interpretation of extracted PCs is different. Since the MPCA loadings associate different sets of weights to each mode, the quantification of the role played by each channel in the final PCA model may be simpler. However, as shown by the simulations discussed in previous Sections, the interpretability improvement provided by the MPCA approach is not necessarily associated to an improvement of monitoring performances.

6   Conclusions

The multi-way analysis provides a framework to extend the PCA technique to multi-dimensional datasets, like those encountered in multi-sensor data fusion problems.

Two Multi-way PCA extensions are considered in this study: the VPCA approach, consisting of applying regular PCA to a matrix generated by unfolding the original multi-

way dataset, and the MPCA approach, based on applying the PCA directly to the multi-way dataset, preserving its higher-order tensor representation. We reviewed the theoretical background of the two methods, and proposed the corresponding extensions of PCA-based control charts.

The interpretability of results provided by Multi-way PCA methods is expected to be an important advantage with respect to black box data fusion techniques widely exploited in mainstream literature.

The Monte Carlo simulations demonstrated that the VPCA may provide better performances than the MPCA with regard to simple out-of-control scenarios, including mean shifts and noise variance increase. The MPCA may be an effective competitor in presence of some departures from the natural pattern that affect the correlation structure among different channels. However, in terms of overall ARL performances, our simulation results suggest that the VPCA should be preferred, at least in applications characterized by a small number of channels, as the one considered in this study.

The application to real multi-channel profile data acquired in waterjet cutting showed the different nature of extracted features. In particular, MPCA loadings associate different sets of weights to each mode, which may improve the quantification of the contribution of each channel on the final PCA model. However, the MPCA requires an equal number of data points in each channel. From this last point of view, the VPCA is more flexible, since it can be applied to generic vectors of features with different length.

As highlighted by different authors, the MPCA could be a more efficient method from a computational and memory saving point of view, especially when high numbers of channels and variables/data points are involved (Lu et al. 2008). One could be interested in evaluating both the methods during the design phase, since they lead to different PC interpretations and to different subspace projections. The choice between the

two approaches should take into account the various features mentioned above, including the fact that their performances depend on the nature of the out-of-control condition. Further research and simulation efforts are expected to further clarify the benefits and limitations of the two proposed approaches in different scenarios, and in presence of a larger number of channels.

Nomenclature

B	Number of bootstrap samples
$\boldsymbol{b}_j$	j^{th} model parameter vector used in simulated scenarios
	$(\boldsymbol{b}_j \sim MN(\boldsymbol{\mu_b}, \Sigma_{\boldsymbol{b}}))$
I_q	Dimension of the q-mode of a Q-way array
J	Number of samples
$\mathbf{L}$	Eigenvalue matrix in VPCA (diagonal elements are denoted by λ_i)
m	Number of retained PCs
M	Number of samples used in Phase I
N	Number of channels
P	Number of data points (or generic number of variables) in each sample
P'	Number of matrix columns after unfolding
P_q	Number of retained PCs in the qth mode
Q	Number of modes
$\mathbf{S}_{1:M}$	Sample variance-covariance matrix of $\mathbf{X}_{1:M}$
S^q	Fraction of the total scatter captured by the retained PCs in the qth mode
$SSE_j(m)$	j^{th} sample of Sum of Squared Error statistics based on the first m PCs
$T_j^2(m)$	j^{th} sample of Hotelling T^2 statistics based on the first m PCs
T_S	Total scatter tensor

$\mathbf{U}$	Eigenvector matrix in VPCA (column elements are denoted by $\mathbf{u}_i$)
$\widetilde{\mathbf{U}}^{(q)}$	q^{th} projection matrix in MPCA (column elements are denoted by $\mathbf{u}_{i_q}^{(q)}$)
$\widetilde{\mathbf{U}}_{\Phi^{(q)}}$	Kronecker product of projection matrices – see Eq. 10
$\mathbf{V}$	Consecutive differences matrix used in Eq. 13
$\mathbf{X}$	Resulting matrix after unfolding; the jth sample is denoted by $\mathbf{x}_j$
$\bar{\mathbf{x}}$	Average profile (or multivariate vector) among the M samples used in Phase I
$\mathbf{X}_{1:M}$	Unfolded Phase I matrix (it includes only the M samples used in Phase I)
$\mathbf{X}_{j(q)}$	Equivalent matrix representation of $\chi_{\cdot,j}$ by unfolding the qth mode
$\overline{\mathbf{X}}_{(q)}$	Average of the matrix obtained by q^{th} mode unfolding of $\chi_{\cdot,j}$
$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$	Blocks, heavysine, and bumps profiles
$\hat{\mathbf{x}}_j(m)$	Reconstructed profile (or multivariate vector) by using the first m PCs
$\boldsymbol{y}_s$	Sine function, used in simulated scenarios
$\mathbf{z}_j$	Projection of the jth sample onto orthogonal space spanned by retained PCs
α, α'	Type I error
$\delta_a, \dots, \delta_e$	Shifts applied in simulated out-of-control scenarios
$\varepsilon_{n,j}$	Random error term added to the nth channel of the jth sample in simulated scenarios
$\lambda_{i_q}^{(q)}$	$(i_q)^{th}$ eigenvalue in the qth mode (MPCA)
σ_{x_u}	Standard deviation of the uth benchmark signal (blocks, heavysine, or bumps)
γ_j	j^{th} projected tensor

$\mathbf{Y}^*$	Matricized version of the projected tensor Y, with Phase I estimates of mean vector and covariance matrix respectively denoted by $\overline{Y^*}$ and $\mathbf{S}_{Y^*}$
$\mathbf{\Phi}^{(q)}$	q-mode matrix in MPCA
χ	Q-way array
$\chi_{\cdot,j}$	j^{th} sample of $Q-1$ dimensional tensor objects
$\bar{\chi}_{\cdot}$	Average $(Q-1)$-way array among the M ones used to estimate the MPCA model
$\hat{\chi}_{\cdot,j}(m)$	Reconstruction of jth $(Q-1)$-way sample by using m PCs
ψ_Y	Total tensor scatter

Subscripts

i	Data point or variable index $(i = 1, \dots, P)$
i_q	q-mode index $(i_q = 1, \dots, I_q)$
j	Sample index $(j = 1, \dots, J)$
n	Channel index $(n = 1, \dots, N)$
q	Mode index $(q = 1, \dots, Q)$
u	Benchmark signal index used in simulated scenarios $(u = 1, 2, 3)$
w	Model parameter index used in simulated scenarios $(w = 1, \dots, 3)$

References

Acar E., Yener B. 2009, Unsupervised Multiway Data Analysis: a Literature Survey, IEEE Transactions on Knowledge and Data Engineering, 21:1, 6 - 20

Aliustaoglu C., Ertunc H. M., Ocak H. 2009, Tool Wear Condition Monitoring Using a Sensor Fusion Model Based on Fuzzy Inference System, Mechanical Systems and Signal Processing, 23, 539 - 546

Amiri A., Zou C., Doroudyan M. H. 2013, Monitoring Correlated Profile and Multivariate Quality Characteristics, Quality and Reliability Engineering International, doi: 10.1002/qre.1483

Annoni M., Cristaldi L., Lazzaroni M. 2008, Measurements, Analysis and Interpretation Of The Signals From A High-Pressure Waterjet Pump, IEEE Transactions on Instrumentation and Measurement, 57:1, 34-47

Annoni, M., Cristaldi, L., Lazzaroni, M., Ferrari, S. 2009, Nozzles Classification in a High-Pressure Water Jet System, IEEE Transactions on Instrumentation and Measurement, 58:10, 3739 – 3745

Axinte, D. A., Kong, M. C. 2009, An Integrated Monitoring Method to Supervise Waterjet Machining, In: CIRP Annals – Manufacturing Technology, 58, 303 – 306

Bahr, B., Motavalli, S. and Arfi, T. 1997, Sensor Fusion For Monitoring Machine Tool Conditions, International Journal of Computer Integrated Manufacturing, 10:5, 314 – 323

Bhattacharyya P., Sengupta D. 2009, Estimation of Tool Wear Based on Adaptive Sensor Fusion and Power in Face Milling, International Journal of Production Research, 47:3, 817-833

Chang S. I., Yadama S. 2010, Statistical Process Control for Monitoring Non-Linear Profiles Using Wavelet Filtering and B-Spline Approximation, International Journal of Production Research, 48:4, 1049-1068

Chen S-L., Jen Y. W. 2000, Data Fusion Neural Network for Tool Condition Monitoring in CNC Milling Machining, International Journal of Machine Tools & Manufacture, 40, 381 - 400

Cho S., Binsaeid S., Asfour S. 2010, Design of Multisensor Fusion based Tool Condition Monitoring System in End Milling, The International Journal of Advanced Manufacturing Technology, 46:5-8, 681 - 694

Choi G. S., Choi G. H. 1997, Process Analysis And Monitoring In Abrasive Waterjet Machining Of Alumina Ceramics, International Journal of Machine Tools and Manufacture, 37:3, 295-307

Colosimo, B.M., Pacella, M. 2007, On the Use of Principal Component Analysis to Identify Systematic Patterns in Roundness Profiles, Quality and Reliability Engineering International, 23, 925 - 941

Colosimo, B.M., Pacella, M. 2010, A Comparison Study of Control Charts for Statistical Monitoring of Functional Data, International Journal of Production Research, 23, 707 – 725

De Latheuwer L., De Moor B., Vandewalle J. 2000, A Multilinear Singular Value Decomposition, SIAM Journal on Matrix Analysis and Applications, 21:4, 1253 – 1278

Donoho D. L., Johnstone I. M. 1994, Ideal Spatial Adaptation by Wavelet Shrinkage, Biometrika, 81, 425–55

Fan Z., Cai M., Wang H. 2012, An Improved Denoising Algorithm Based On Wavelet Transform Modulus Maxima For Non-Instrusive Measurement Signals, Measurement Science and Technology, 23, 1-11

Gao X. 2012, On-line Monitoring of Batch Process with Multiway PCA/ICA, Chapter 13 in Principal Component Analysis – Multidisciplinary Application, edited by Sanguansat P., Intech Ed.

Gardner M. M., Lu J-C., Gyurcsik R. S. 1997, Equipment Fault Detection Using Spatial Signatures, IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part C, 20:4, 295-304

Grasso M., Goletti M., Annoni M., Colosimo B.M. 2013, A New Approach for Online Health Assessment of Abrasive Waterjet Cutting Systems, accepted for publication on the International Journal of Abrasive Technology

Inasaki I. 1999, Sensor Fusion for Monitoring and Controlling Grinding Processes, The International Journal of Advanced Manufacturing Technology, 15, 730-736

Jin J., and Shi, J. 1999, Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets, Technometrics, 41:4, 327 – 339

Jin, J., Shi, J. 2001, Automatic Feature Extraction of Waveform Signals for In-process Diagnostic Performance Improvement, Journal of Intelligent Manufacturing, 12, 257 – 268

Jolliffe I. T. 2002, Principal Component Analysis, 2nd Edition, Springer Series in Statistics

Jurisevic B., Brissaud D., Junkar M. 2004, Monitoring of Abrasive Water Jet (AWJ) Cutting Using Sound Detection, International Journal of Advanced Manufacturing Technology, 24, 733 – 737

Kiers H. A. L. 2000, Towards a Standardized Notation and Terminology in Multiway Analysis, Journal of Chemometrics, 14, 105 - 122

Kim J., Huang Q., Shi J., Chang T-S. 2006, Online Multichannel Forging Tonnage Monitoring and Fault Pattern Discrimination Using Principal Curve, Transactions of the AMSE, 128, 944 – 950

Ko K., Qu L., Vannucci M. 2009, Wavelet-Based Bayesian Estimation Of Partially Linear Regression Models With Long Memory Errors, Satistica Sinica, 19, 1463-1478

Koo I., Kil R. M. 2008, Model Selection for Regression with Continuous Kernel Functions Using the Modulus of Continuity, Journal of Machine Learning Research, 9, 2607-2633

Kovacevic R., Hashish M., R. Mohan, Ramulu M., Kim T.J., Geskin E.S. 1997, State Of The Art of Research and Development In Abrasive Waterjet Machining, Journal of Manufacturing Science and Engineering, 119:4, 776 – 785

Krenicky, T., Mikoslav R. 2012, Monitoring of Vibrations in the Technology of AWJ, Key Engineering Materials, 96, 229 - 234

Kuljanic E., Totis G., Sortino M. 2009, Development of an Intelligent Multisensor Chatter Detection System in Milling, Mechanical Systems and Signal Processing, 23, 1704 – 1718

Lezanski P. 2001, An Intelligent System, for Grinding Wheel Condition Monitoring, Journal of Materials Processing Technology, 109, 258 - 263

Liu R. Y., Tang J. 1996, Control Charts for Dependent and Independent Measurements Based on Bootstrap Methods, Journal of the American Statistical Association, 91:436, 1694 - 1700

Lu H., Konstantinos N., Venetsanopoulos A. N. 2008, MPCA: Multilinear Principal Component Analysis of Tensor Objects, IEEE Transactions on Neural Networks, 19:1, 18 - 39

Lu H., Konstantinos N., Venetsanopoulos A. N. 2009, Uncorrelated Multilinear Principal Component Analysis for Unsupervised Multilinear Subspace Learning, IEEE Transactions on Neural Networks, 20:11, 1820 - 1836

Luo R. C., Yih C-C., Su K. L. 2002, Multisensor Fusion and Integration: Approaches, Applications, and Future Research Directions, IEEE Sensors Journal, 2:2, 107 - 119

Montgomery D. C. 2008, Introduction to Statistical Quality Control, John Wiley & Sons, 6th Ed

Nomikos P., MacGregor J. F. 1995, Multivariate SPC Charts for Monitoring Batch Processes, Technometrics, 37:1, 41 – 59

Noorossana R., Saghaei A., Amiri A. 2012, Statistical Analysis of Profile Monitoring, John Wiley & Sons

Paynabar K., Jin J., Pacella M. 2013, Analysis of Multichannel Nonlinear Profiles Using Uncorrelated Multilinear Principal Component Analysis with Applications in Fault Detection and Diagnosis, IEE Transactions, 45:11, 1235 – 1247

Perzel V., Hreha P., Hloch S., Tozan H., Valicek J. 2012, Vibration Emission as a Potential Source Of Information for Abrasive Waterjet Quality Process Control, The International Journal of Advanced Manufacturing Technology, 61:1-4, 285 – 294

Rabani A., Marinescu I., Axinte D. 2012, Acoustic Emission Energy Transfer Rate: a Method for Monitoring Abrasive Waterjet Milling, International Journal of Machine Tools & Manufacture, 61, 80 – 89

Rao B. K. N. 1996, Handbook of Condition Monitoring, Elsevier Science Ltd

Shi D., Gindy N. N. 2007, Tool Wear Predictive Model Based on Least Squares Support Vector Machines, Mechanical Systems and Signal Processing, 21, 1799 - 1814

Sick, B. 2002, On-Line and Indirect Tool Wear Monitoring in Turning with Artificial Neural Networks: A review of more than a decade of research, Mechanical Systems and Signal Processing, 16:4, 487 - 546

Valle S., Li W., Qin S. J. 1999, Selection of the Number of Principal Components: the Variance of Reconstruction Error Criterion with a Comparison to Other Methods, Industrial & Engineering Chemistry Research, 38, 4389-4401

Wang, W. H. , Hong, G. S. , Wong, Y. S. and Zhu, K. P. 2007, Sensor Fusion for Online Tool Condition Monitoring in Milling, International Journal of Production Research, 45: 21, 5095 – 5116

Williams, J. D., Woodall, W. H., Birch, J. B., and Sullivan, J. H. 2006, On the Distribution of Hotelling's T^2 Statistic based on the Successive Differences Covariance Matrix Estimator, Journal of Quality Technology, 38, 217-229.

Williams, J.D., Woodall, W.H., Birch, J.B. 2007, Statistical Monitoring of Nonlinear Product and Process Quality Profiles, Quality and Reliability Engineering International, 23, 925 - 941

Wold S. 1978, Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models, Technometrics, 20:4, 397-405

Woodall, W.H., Spitzner, D.J., Montgomery, D.C., Gupta, S. 2004, Using Control Charts to Monitor Process and Product Quality Profiles, Journal of Quality Technology, 36:3, 309 - 320

Yang J., Zhang D., Frangi A. F., Yang J. 2004, Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26:1, 131 - 137

Ye J., Janardan R., Li Q. 2004, GPCA: An Efficient Dimension Reduction Scheme for Image Compression and Retrieval, Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 354 – 363

Zhou S., Jin N., Jin J. 2005, Cycle-based Signal Monitoring Using a Directionally Variant Multivariate Control Chart System, IEE Transactions, 37, 971 - 982

Table 1 – ARLs and 99% Confidence Intervals for VPCA and MPCA – in-control and out-of-control scenarios a), b), and c)

Scenario	Affected signal	Severity	VPCA ARL	VPCA CI 99%	MPCA ARL	MPCA CI 99%
In-control			*100.80*	*[98.03, 103.58]*	*101.97*	*[98.52, 105.43]*
a) Signal mean shift	x_1	0.01	**84.61**	**[82.09, 87.13]**	**86.32**	**[83.00, 89.64]**
		0.025	**38.23**	**[37.48, 38.98]**	45.18	[43.52, 46.84]
		0.05	**5.67**	**[5.62, 5.73]**	11.45	[11.01, 11.89]
		0.075	**1.34**	**[1.34, 1.35]**	3.61	[3.53, 3.69]
		0.1	**1.00**	**[1.00, 1.00]**	1.65	[1.61, 1.69]
	x_2	0.01	**77.04**	**[74.87, 79.21]**	92.33	[88.98, 95.68]
		0.025	**13.83**	**[13.64, 14.03]**	58.06	[54.24, 61.87]
		0.05	**1.10**	**[1.10, 1.11]**	19.51	[16.69, 22.33]
		0.075	**1.00**	**[1.00, 1.00]**	5.57	[4.80, 6.35]
		0.1	**1.00**	**[1.00, 1.00]**	1.68	[1.53, 1.83]
	x_3	0.01	99.04	[97.26, 100.83]	**100.47**	**[96.38, 104.56]**
		0.025	87.11	[84.92, 89.29]	100.96	[97.95, 103.96]
		0.05	57.73	[56.28, 59.17]	94.97	[91.46, 98.47]
		0.075	32.61	[31.86, 33.35]	86.10	[81.96, 90.24]
		0.1	16.70	[16.35, 17.04]	72.75	[69.68, 75.82]
b) Sinusoid term superimposition	x_1	0.025	**21.99**	**[21.47, 22.52]**	86.72	[83.36, 89.07]
		0.05	**1.41**	**[1.41, 1.42]**	49.74	[48.32, 51.16]
		0.075	**1.00**	**[1.00, 1.00]**	17.72	[17.16, 18.29]
		0.1	**1.00**	**[1.00, 1.00]**	4.19	[4.01, 4.37]
		0.125	**1.00**	**[1.00, 1.00]**	1.20	[1.17, 1.24]
	x_2	0.025	**31.90**	**[31.26, 32.55]**	94.12	[90.95, 97.29]
		0.05	**2.25**	**[2.23, 2.27]**	63.94	[61.66, 66.23]
		0.075	**1.01**	**[1.01, 1.01]**	30.49	[29.69, 31.30]
		0.1	**1.00**	**[1.00, 1.00]**	10.06	[9.66, 10.45]
		0.125	**1.00**	**[1.00, 1.00]**	2.52	[2.38, 2.67]
	x_3	0.025	95.06	[92.56, 97.56]	**99.89**	**[96.23, 103.55]**
		0.05	76.74	[74.92, 78.57]	100.12	[96.96, 103.29]
		0.075	51.17	[50.26, 52.08]	95.94	[93.95, 97.92]
		0.1	28.29	[27.81, 28.77]	91.62	[88.76, 94.48]
		0.125	13.51	[13.27, 13.74]	84.10	[81.93, 86.27]
c) Error term variance increase	$\chi_{1,\cdot,j}$	1.1	**16.90**	**[16.55, 17.26]**	88.24	[85.56, 90.92]
		1.5	**1.02**	**[1.02, 1.02]**	32.73	[31.52, 33.93]
		2	**1.00**	**[1.00, 1.00]**	4.28	[3.98, 4.57]
		2.5	**1.00**	**[1.00, 1.00]**	1.08	[1.06, 1.10]
		3	**1.00**	**[1.00, 1.00]**	**1.00**	**[1.00, 1.00]**
	$\chi_{2,\cdot,j}$	1.1	**23.91**	**[23.56, 24.26]**	92.14	[88.40, 95.89]
		1.5	**1.10**	**[1.10, 1.10]**	41.93	[39.95, 43.91]
		2	**1.00**	**[1.00, 1.00]**	8.12	[7.62, 8.62]
		2.5	**1.00**	**[1.00, 1.00]**	1.49	[1.43, 1.55]
		3	**1.00**	**[1.00, 1.00]**	1.01	[1.01, 1.01]
	$\chi_{3,\cdot,j}$	1.1	**37.16**	**[36.41, 37.92]**	96.26	[93.43, 99.09]
		1.5	**1.55**	**[1.54, 1.56]**	56.59	[54.32, 58.85]
		2	**1.00**	**[1.00, 1.00]**	18.37	[17.64, 19.11]
		2.5	**1.00**	**[1.00, 1.00]**	4.16	[3.85, 4.48]
		3	**1.00**	**[1.00, 1.00]**	1.29	[1.24, 1.34]
	$\chi_{4,\cdot,j}$	1.1	**22.84**	**[22.50, 23.17]**	91.76	[88.01, 95.50]
		1.5	**1.09**	**[1.09, 1.09]**	41.15	[39.62, 42.68]
		2	**1.00**	**[1.00, 1.00]**	7.48	[7.04, 7.93]
		2.5	**1.00**	**[1.00, 1.00]**	1.40	[1.34, 1.45]
		3	**1.00**	**[1.00, 1.00]**	**1.01**	**[1.00, 1.01]**

Table 2 – ARLs and 99% Confidence Intervals for VPCA and MPCA – out-of-control scenarios d) and e)

Scenario	Affected parameter	Delta	VPCA		MPCA	
			ARL	CI 99%	ARL	CI 99%
d) Mean shift of model parameters	$b_{1,j}$	1	**52.40**	**[51.11, 53.70]**	68.72	[65.90, 71.54]
		2	**12.64**	**[12.43, 12.85]**	25.74	[23.60, 27.89]
		3	**3.55**	**[3.51, 3.59]**	8.63	[8.21, 9.04]
		4	**1.60**	**[1.59, 1.61]**	3.30	[3.20, 3.40]
		5	**1.12**	**[1.11, 1.12]**	1.68	[1.64, 1.73]
	$b_{2,j}$	1	**94.78**	**[92.61, 96.95]**	100.33	[96.01, 104.65]
		2	**75.86**	**[74.01, 77.70]**	99.97	[95.77, 104.17]
		3	**51.42**	**[50.47, 52.38]**	90.17	[86.71, 93.62]
		4	**30.55**	**[29.90, 31.19]**	84.07	[80.30, 87.84]
		5	**16.74**	**[16.47, 17.00]**	74.62	[70.91, 78.34]
	$b_{3,j}$	1	51.88	[50.88, 52.87]	**36.76**	**[35.49, 38.02]**
		2	11.91	[11.71, 12.12]	**8.26**	**[8.00, 8.52]**
		3	3.34	[3.32, 3.37]	**2.63**	**[2.56, 2.71]**
		4	1.54	[1.53, 1.54]	**1.39**	**[1.37, 1.41]**
		5	1.10	[1.10, 1.10]	**1.07**	**[1.07, 1.08]**
	$b_{5,j}$	1	51.61	[50.48, 52.75]	**27.76**	**[26.80, 28.73]**
		2	11.88	[11.66, 12.09]	**5.55**	**[5.46, 5.65]**
		3	3.29	[3.26, 3.33]	**2.00**	**[1.97, 2.03]**
		4	1.53	[1.52, 1.53]	**1.21**	**[1.20, 1.21]**
		5	1.10	[1.10, 1.10]	**1.03**	**[1.03, 1.03]**
	$b_{6,j}$	1	**55.48**	**[54.16, 56.79]**	94.04	[90.41, 97.67]
		2	**14.23**	**[14.01, 14.45]**	79.17	[76.86, 81.49]
		3	**3.94**	**[3.90, 3.98]**	56.51	[53.91, 59.11]
		4	**1.71**	**[1.70, 1.72]**	37.69	[35.90, 39.47]
		5	**1.15**	**[1.15, 1.15]**	21.64	[20.54, 22.74]
	$b_{7,j}$	1	52.15	[50.95, 53.35]	**42.39**	**[40.94, 43.85]**
		2	11.90	[11.70, 12.11]	**9.21**	**[8.90, 9.53]**
		3	3.32	[3.29, 3.36]	**2.80**	**[2.72, 2.89]**
		4	1.54	[1.53, 1.54]	**1.41**	**[1.39, 1.44]**
		5	1.10	[1.10, 1.10]	**1.07**	**[1.06, 1.07]**
e) Variance increase of model parameters	$b_{1,j}$	1.5	**33.17**	**[32.62, 33.71]**	54.14	[51.24, 57.04]
		2	**11.76**	**[11.64, 11.87]**	22.58	[21.65, 23.51]
		2.5	**6.24**	**[6.20, 6.30]**	11.11	[10.76, 11.45]
		3	**4.22**	**[4.19, 4.24]**	6.77	[6.60, 6.93]
		4	**2.70**	**[2.69, 2.71]**	3.80	[3.73, 3.87]
	$b_{2,j}$	1.5	**94.10**	**[91.84, 96.36]**	100.75	[96.08, 105.41]
		2	**75.90**	**[74.37, 77.42]**	99.97	[96.70, 103.25]
		2.5	**55.79**	**[54.36, 57.21]**	95.94	[92.68, 99.20]
		3	**37.84**	**[37.29, 38.39]**	92.39	[89.74, 95.04]
		4	**16.60**	**[16.40, 16.80]**	83.08	[79.11, 87.05]
	$b_{3,j}$	1.5	31.68	[31.14, 32.23]	**23.49**	**[23.01, 23.96]**
		2	11.13	[10.97, 11.29]	**8.92**	**[8.80, 9.05]**
		2.5	6.00	[5.97, 6.04]	**5.11**	**[5.04, 5.18]**
		3	4.09	[4.07, 4.12]	**3.61**	**[3.59, 3.64]**
		4	2.63	[2.62, 2.64]	**2.42**	**[2.41, 2.44]**
	$b_{5,j}$	1.5	31.64	[31.02, 32.26]	**17.46**	**[17.22, 17.69]**
		2	11.12	[11.00, 11.25]	**6.88**	**[6.82, 6.94]**
		2.5	5.96	[5.91, 6.00]	**4.16**	**[4.13, 4.19]**
		3	4.05	[4.03, 4.07]	**3.07**	**[3.05, 3.09]**
		4	2.63	[2.61, 2.64]	**2.18**	**[2.17, 2.19]**
	$b_{6,j}$	1.5	**36.30**	**[35.78, 36.83]**	94.08	[91.16, 96.99]
		2	**12.80**	**[12.60, 12.99]**	81.10	[78.68, 83.53]
		2.5	**6.72**	**[6.67, 6.78]**	62.43	[60.07, 64.80]
		3	**4.46**	**[4.44, 4.49]**	41.65	[40.31, 42.99]
		4	**2.79**	**[2.78, 2.81]**	17.12	[16.79, 17.46]
	$b_{7,j}$	1.5	32.21	[31.61, 32.81]	**26.42**	**[25.75, 27.10]**
		2	11.32	[11.20, 11.44]	**9.54**	**[9.28, 9.79]**
		2.5	6.04	[5.99, 6.09]	**5.33**	**[5.26, 5.40]**
		3	4.09	[4.06, 4.11]	**3.71**	**[3.67, 3.74]**
		4	2.63	[2.62, 2.65]	**2.47**	**[2.45, 2.48]**

Table 3 – Fault detection percentages in different fault scenarios

Fault	Fault Severity	Fault Detection (%)			
		VPCA		MPCA	
		T^2	$T^2 + SSE$	T^2	$T^2 + SSE$
A	Severe crack	100	100	100	100
	Medium crack	100	100	100	100
	Small crack	100	100	100	100
B	Severe crack	100	100	100	100
	Small crack	100	100	100	100
C	Severe wear	100	100	100	100
	Medium wear	100	100	96.15	96.15
	Low wear	100	100	79.17	91.67
D	Broken 1	100	100	100	100
	Broken 2	100	100	100	100
	Broken 3	100	100	100	100
Tot		100	100	97.95	98.98
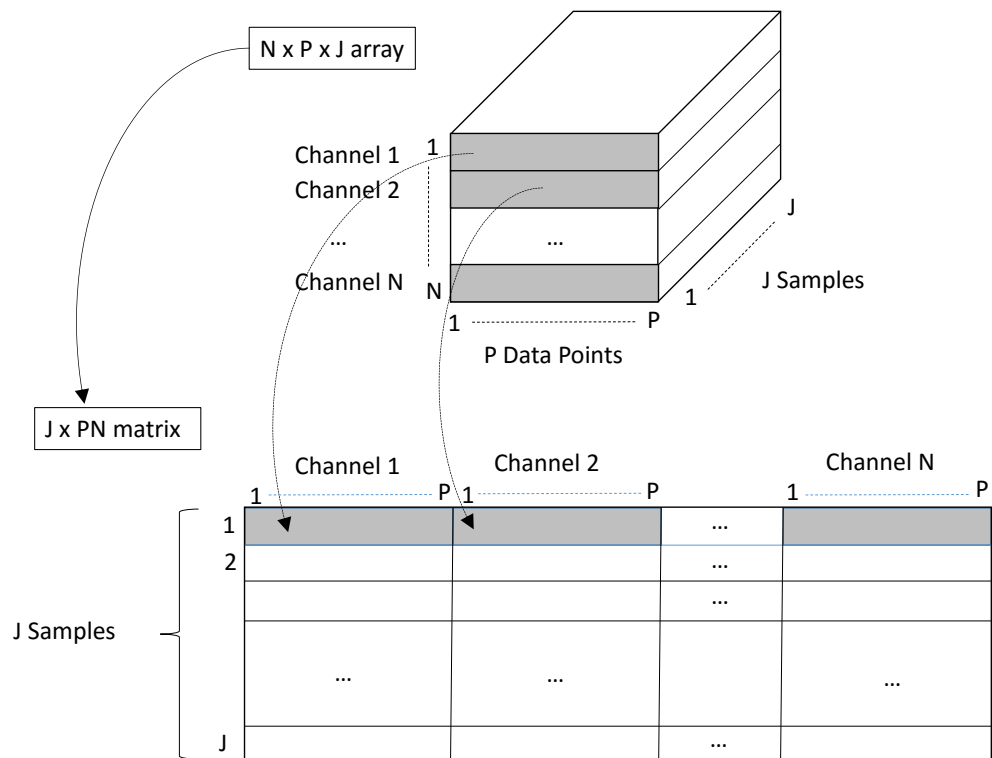
Fig. 1 – Unfolding of a 3-way array into a matrix
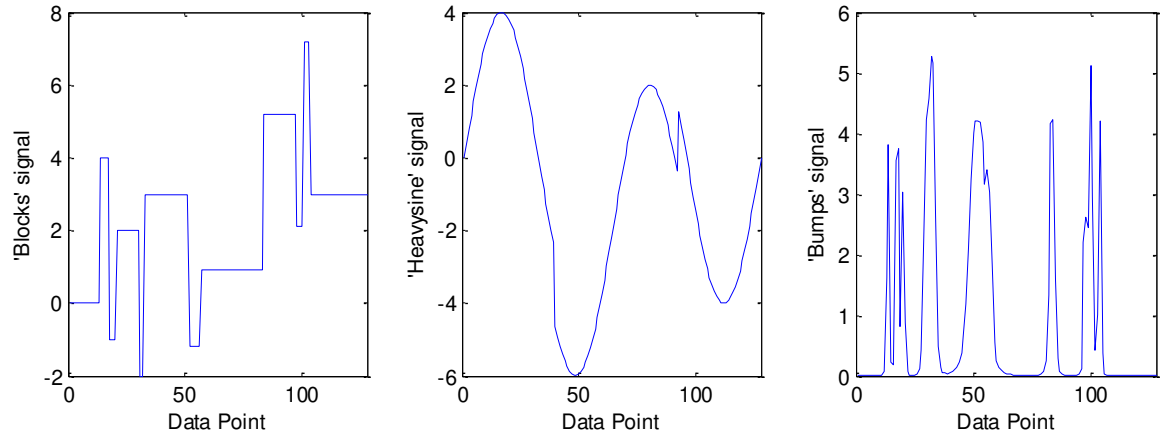
Fig. 2 - 'Blocks', 'heavysine', and 'bumps' profiles
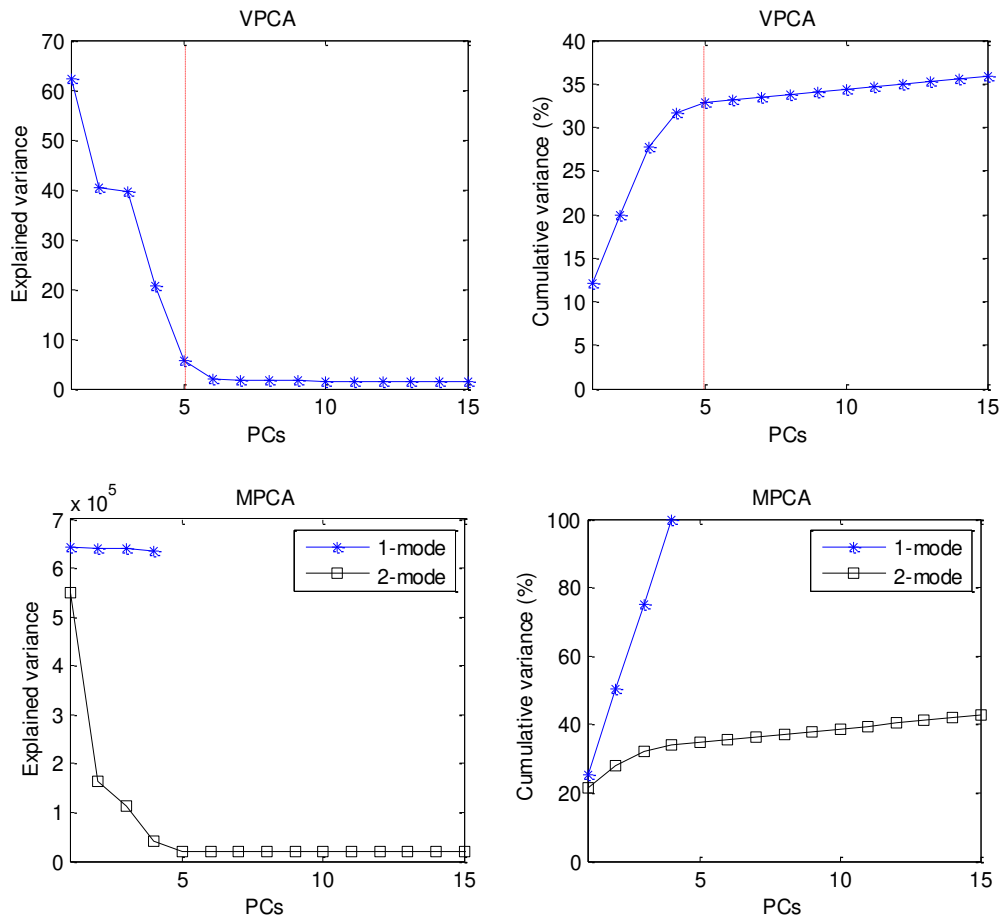
Fig. 3 – Explained variance and cumulative variance - VPCA and MPCA methods
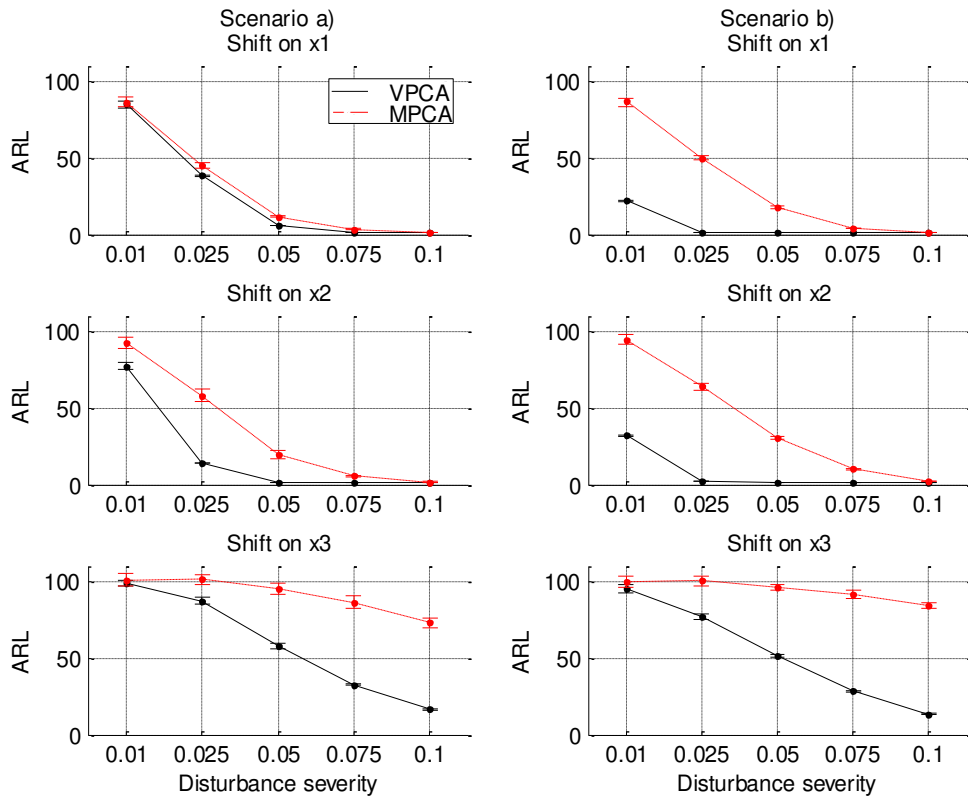
Fig. 4 – ARLs and 99% Confidence intervals in simulated scenarios a) and b) – VPCA and MPCA approaches
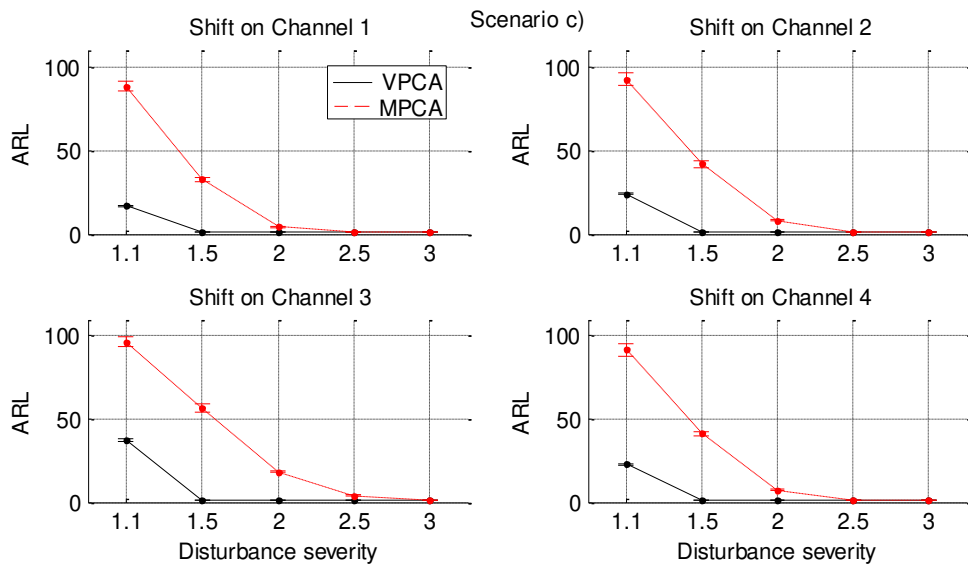
Fig. 5 – ARLs and 99% Confidence intervals in simulated scenario c) – VPCA and MPCA approaches
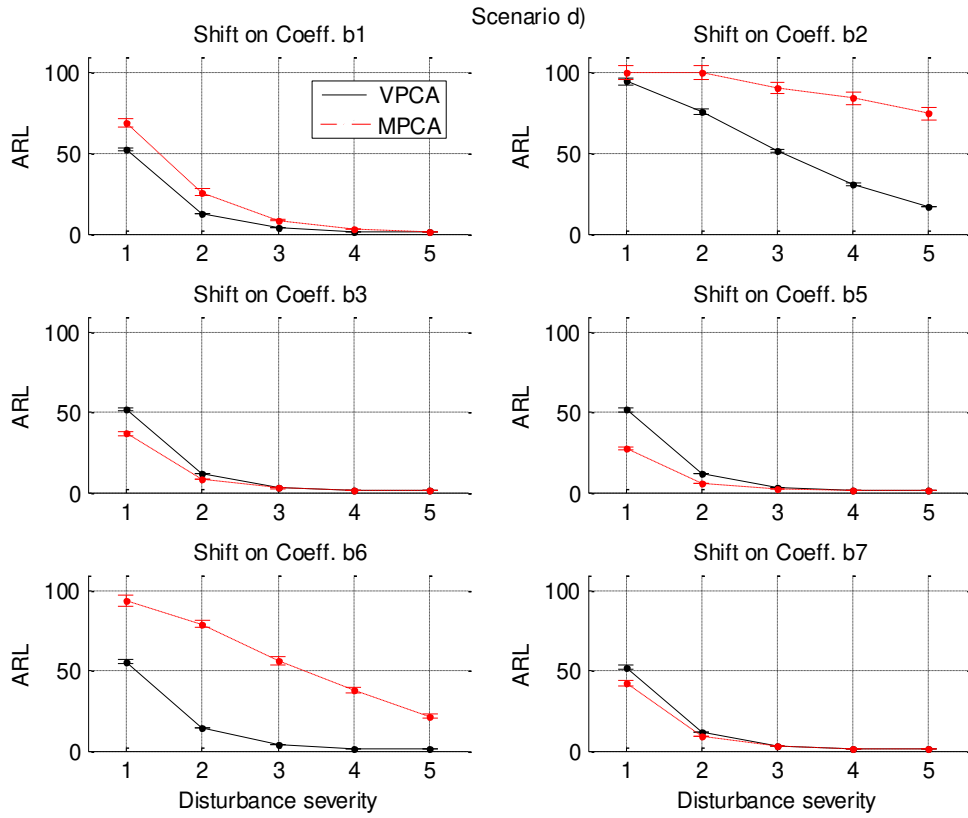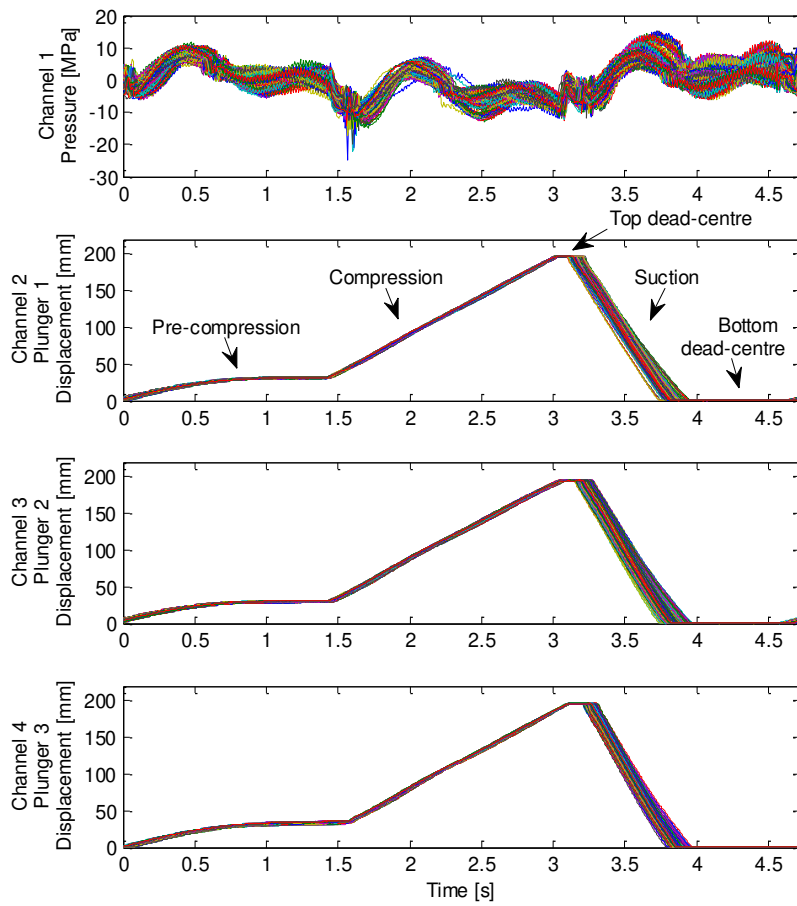
Fig. 6 - ARLs and 99% Confidence intervals in simulated scenario d) – VPCA and MPCA approaches
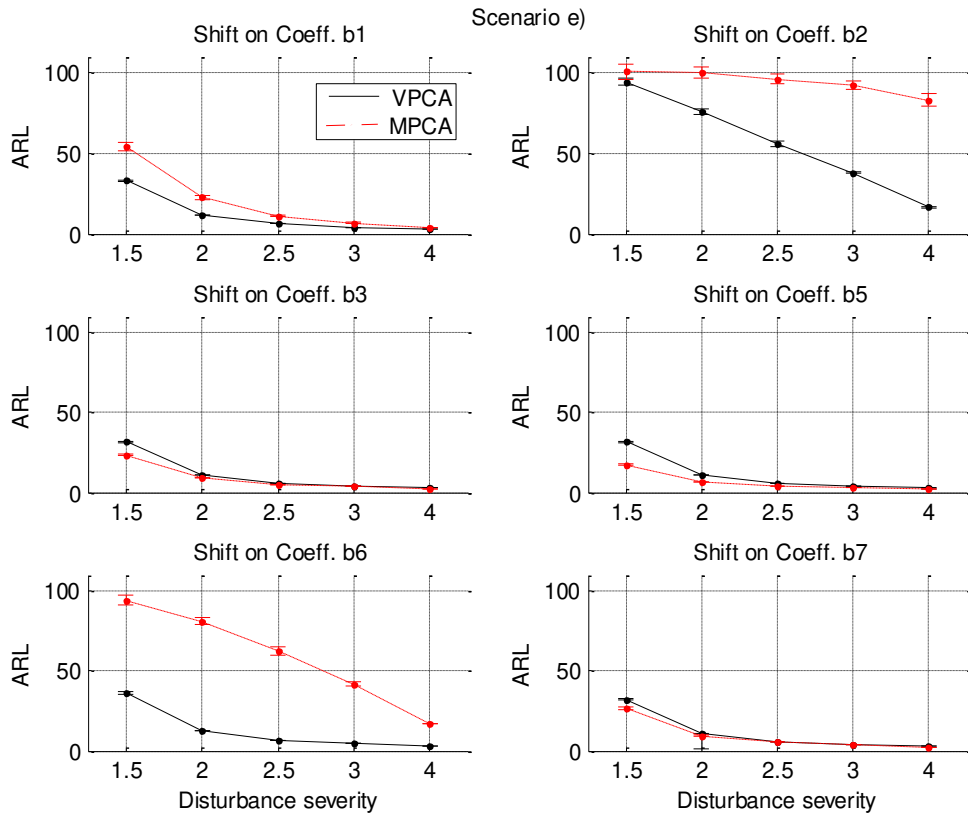
Fig. 7 - ARLs and 99% Confidence intervals in simulated scenario e) – VPCA and MPCA approaches

Fig. 8 – Superimposition of pressure and plunger displacement signals under normal working conditions
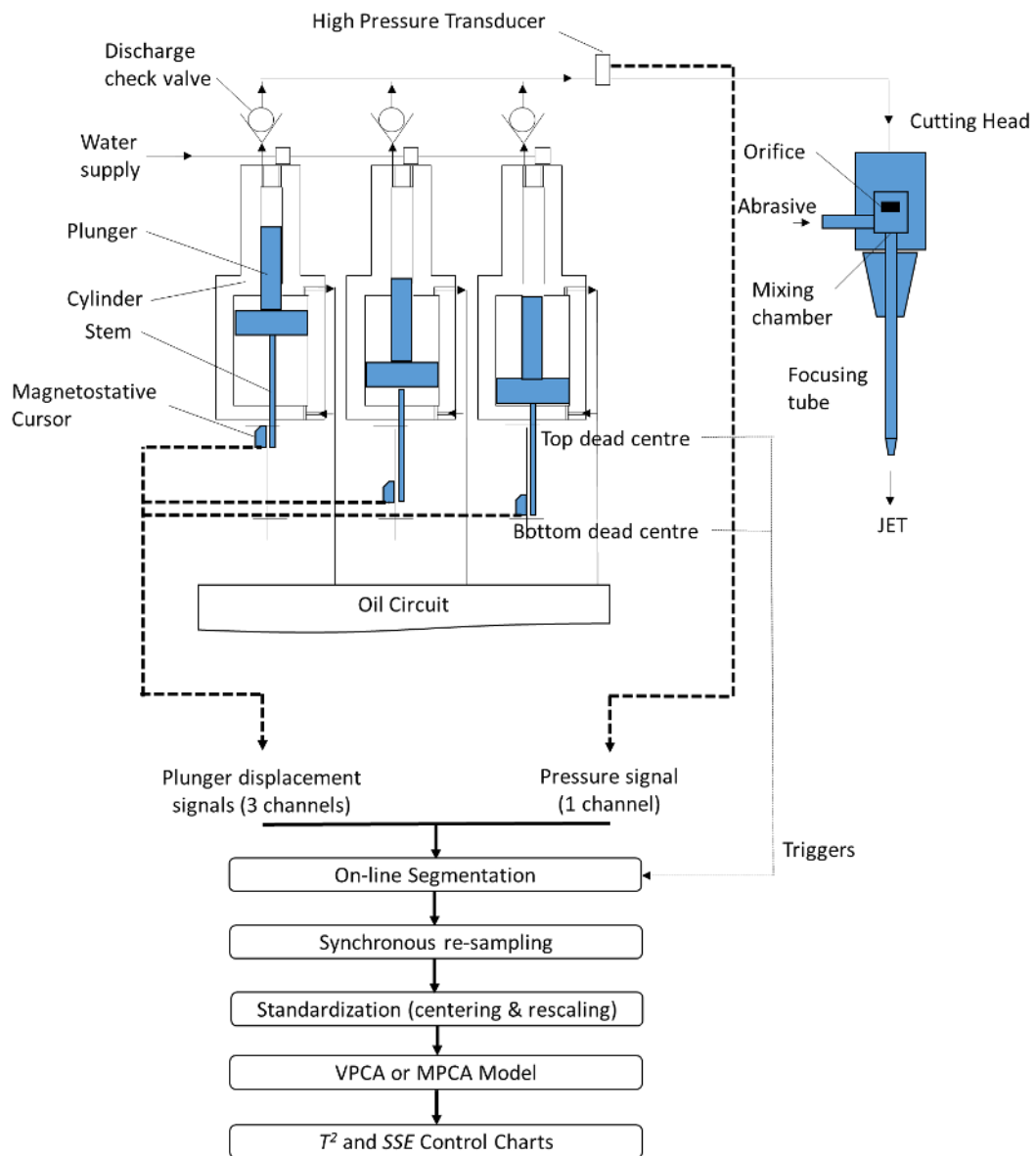
Fig. 9 – Sensor location and multi-channel signal analysis procedure in the WJ/AWJ real test case
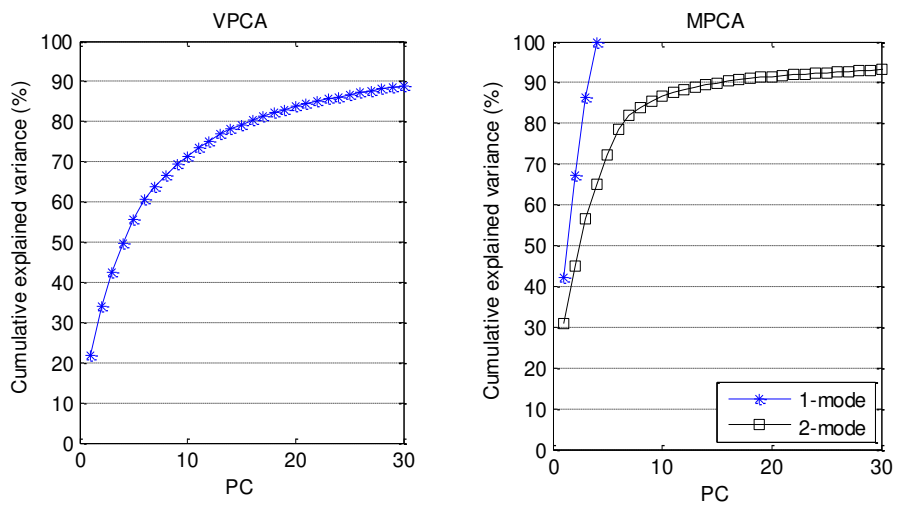
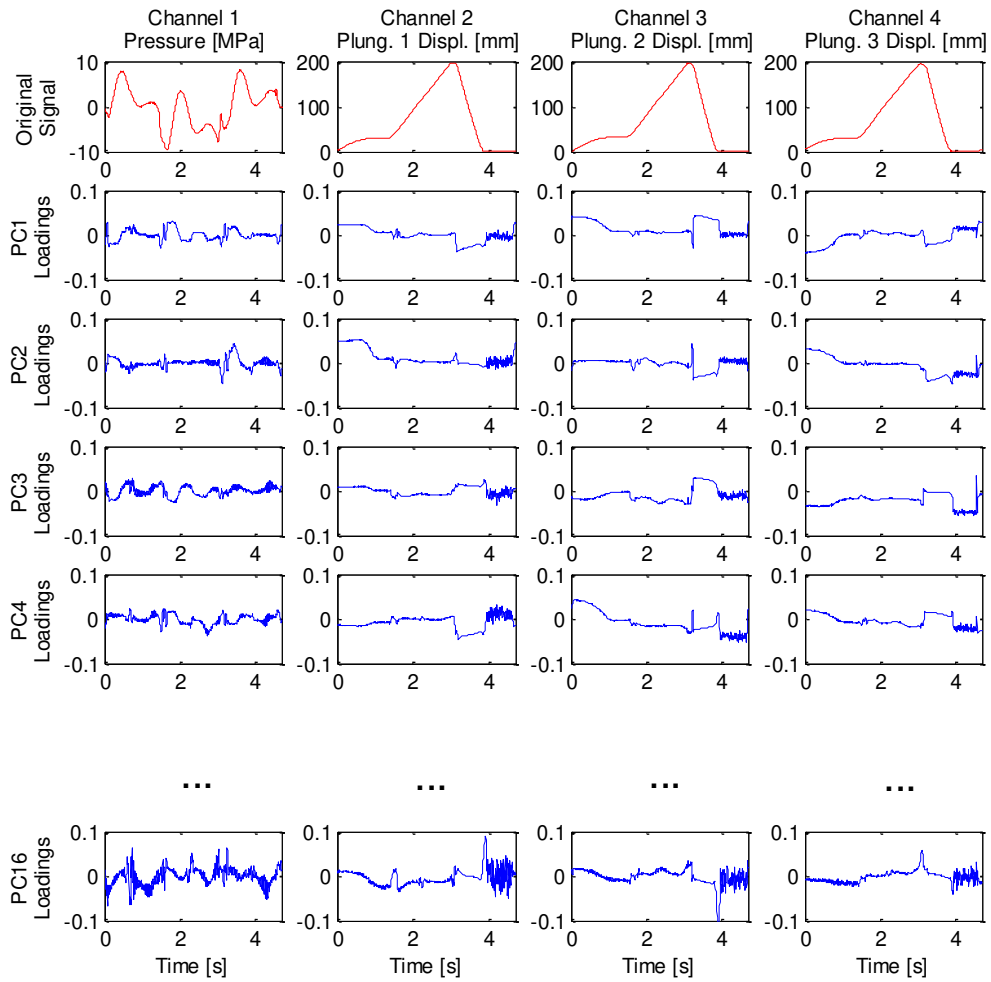Fig. 10 – Cumulative explained variance for the VPCA and the MPCA methods

Fig. 11 – Average Phase I profiles (first row) and loadings - VPCA approach
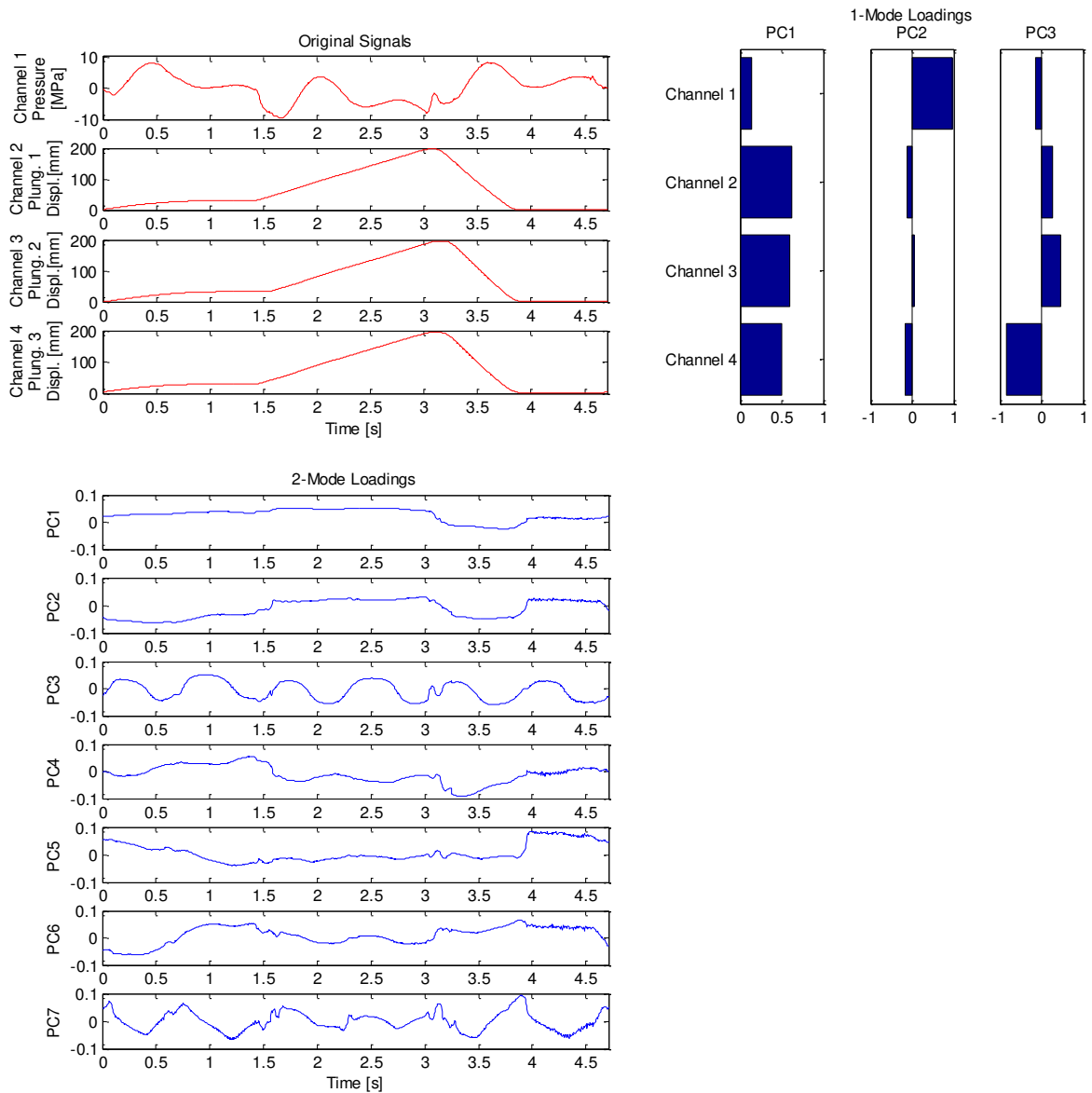
Fig. 12 – Average Phase I profiles and 1-Mode/2-Mode loadings - MPCA approach
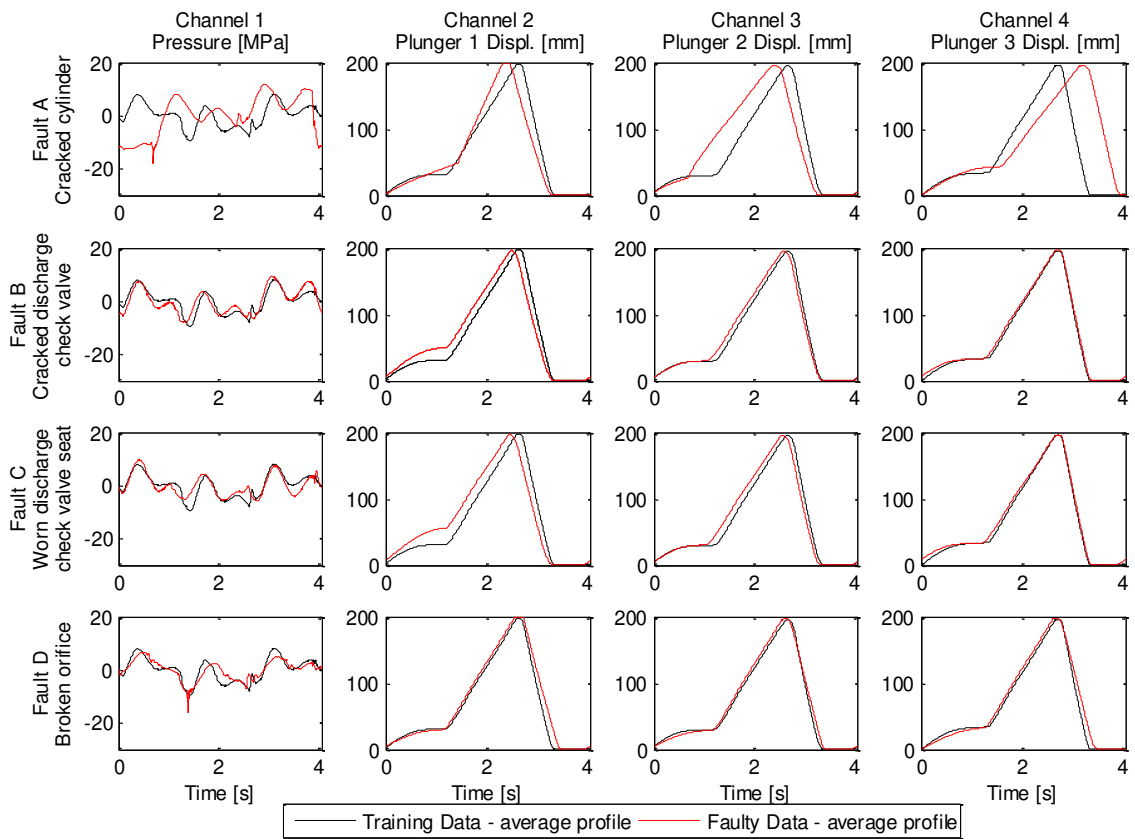
Fig. 13 – Effects of different faults on multi-channel signals (average profiles); faulty
data refer to the highest severity level in each fault scenario