Multivariate detection of transient disturbances for uni- and multi-rate systems

Inês M. Cecílio, James R. Ottewill, Harald Fretheim, Nina F. Thornhill

Abstract

This paper presents a method to detect transient disturbances in a multivariate context, and an extension of that method to handle multi-rate systems. Both methods are based on a time series analysis technique known as nearest neighbors, and on multivariate statistics implemented as a singular value decomposition. The motivation for these developments is that there is an increasing industrial requirement for the analysis of data sets comprising measurements from industrial processes together with their associated electrical and mechanical equipment. These systems are increasingly affected by transient disturbances, and their measurements are commonly sampled at different rates. The paper demonstrates superior results with the multivariate method in comparison to the univariate approach, and with the multi-rate method in comparison to a uni-rate method, for which the fast-sampled measurements had to be downsampled. The method is demonstrated on experimental and industrial case studies.

Index Terms

Industrial process, electric machines, fault detection, multivariate, multirate, nearest neighbors, SVD.

I. INTRODUCTION

There is increasing use of electromechanical equipment in industrial processes as electric motors replace other rotating machinery, such as gas turbines, to drive pumps and compressors [1], [2]. The motivations for the replacement are the high energy-efficiency and easy maintenance of electric motors. However, the electromechanical equipment and electrical utilities can be sources of disturbance to the process. One reason is that disturbances in the electrical supply, such as voltage deviations, can enter the process via the operation of the electromechanical equipment. Another reason is that equipment has its own fault modes. In addition, a recent report [3] indicates that disturbances in the electrical supply are becoming more frequent.

Techniques to analyse plant-wide disturbances have been effective in the analysis of large scale industrial plants, where process disturbances can propagate due to interactions between components of the plant [4], [5]. Recent contributions [6] have highlighted the importance of extending such analysis techniques to include the measurements from the electrical and mechanical equipment which service the process. Examples of these measurements are

I. M. Cecílio and N. F. Thornhill are with the Centre for Process System Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK e-mail:i.cecilio09@imperial.ac.uk. J. R. Ottewill is with ABB Corporate Research Center, Kraków, Poland. H. Fretheim is with ABB Oil, Gas, and Petrochemicals, Oslo, Norway.

voltage levels in the motor power supply, and motor speed. An example of a process measurement is flow rate. Transient disturbances are particularly relevant when considering electromechanical measurements because disturbances related to the electrical utility are mostly of a transient nature. Examples are voltage dips and spikes, which are caused by power imbalances in the grid that lead to momentary frequency and voltage instabilities [7].

Transient disturbances are defined as short-lasting deviations of a measurement from its previous and subsequent trend. In addition, the deviation seldom repeats within the time horizon of analysis. The reason to use this definition is that the time scale and timing of the disturbance are more important than its shape. An important task in process disturbance analysis is the detection of transient disturbances [8], [9]. Recently, Cecílio et al. [10] proposed a method which framed the detection of transients as an anomaly detection problem and solved it with nearest neighbors techniques. That paper explored the performance of the nearest neighbors method in depth and proposed default parameters which optimized it for a wide range of cases. The method, however, took a univariate approach, which means that the measurements from each variable are analysed separately.

The first contribution of this paper is to extend the optimized Nearest Neighbor method by Cecílio et al. [10] into a multivariate analysis. A multivariate extension allows the detection of a transient disturbance which influences several measurements. As this paper later shows, the identification of a transient disturbance can be difficult in measurements with strong oscillatory trends or noise. In such cases, the presence of the same disturbance in other measurements can be exploited for an improved outcome.

The second contribution of this paper is to extend the multivariate approach to the case of a multi-rate system. This is defined as a discrete-time system with two or more operating sampling rates [11]. The system comprising process, electrical, and mechanical measurements is a multi-rate system because, in industry, process measurements are typically sampled 100 to 1000 times slower than electromechanical measurements. Therefore, without the multi-rate extension, electromechanical measurements could only be included if downsampled to the process rate. Other authors have addressed multi-rate systems, however the majority focused on problems other than the detection of disturbances [8], [12].

In brief, the multivariate extension is solved using singular value decomposition on a multivariate set of features known as anomaly index vectors [10]. Periods with transient disturbances lead to high anomaly index values. The multi-rate extension is solved by expanding the anomaly index vectors of slow-sampled measurements to match those of fast-sampled measurements.

These methods are illustrated with experimental data from a gas compression set-up, and with industrial data from routine operation of a gas processing plant. The results demonstrate that the multivariate method improves the detection outcome in comparison to the univariate approach. Likewise, the multi-rate method successfully incorporates measurements with different sampling rates, and improves the detection results by comparison to a uni-rate method, for which the fast-sampled measurements had to be downsampled.

The paper is organized as follows. Section II discusses related work in the analysis of transient disturbances, multivariate analysis, and multi-rate systems. It also gives the theoretical background on the concepts of nearest neighbors, and singular value decomposition. Section III recalls the method for univariate detection of transient

disturbances proposed in [10]. It also presents a case study, which is used throughout the paper to illustrate the various methods. Section IV then explains the method to detect transient distubances in a multivariate set of process and electromechanical measurements, sampled at equal rate. This method involves three classes of parameters, so section V analyses their influence in the detection results and gives clear recommendations for their values. Section VI-A presents the extension of the multivariate detection method to multi-rate systems. The two proposed methods are tested in section VII on two industrial case studies. The paper ends with comments and conclusions.

II. Background

A. Analysis of transient disturbances

Detection of transient disturbances in process measurements has previously been addressed in two main ways: qualitative trend analysis and numerical data-driven methods. The review paper [13] by Venkatasubramanian et al. presents a comprehensive overview of qualitative and quantitative methods to analyse process measurements in order to detect transients and other types of disturbances.

Qualitative trend analysis [9] is based on projecting the trends of process time series into basic shapes, such as spikes, ramps and wavelets [14], [15]. Recent contributions have extended the approach to adapt to disturbances at different time scales [16], to work on-line [17], and to find the optimal start and end points of a shape in the time series [18]. However, this approach requires the creation of a complete database with the possible shapes of a transient disturbance. The reason is that the transient is detected by comparison to these shapes.

A different strand of research applies numerical data-driven methods. Most of these assume that the transient disturbance is distinct from the normal trend of the measurement by its amplitude in the time-domain or by its scale. Traditional methods of statistical process monitoring use differences in time-domain amplitude [19]. However, this approach is not appropriate if the system generating the measurement has oscillatory or cyclical dynamics.

Differences in scale have been exploited with wavelet decomposition [8] in order to detect transient disturbances which map to wavelet coefficients of high amplitude in the lower scales. To detect the transient-related coefficients, the non-transient parts of the signal must not produce other high amplitude coefficients at those same scales. However, this may not always be the case, as Figure 1 exemplifies. The top panel of the figure shows a temperature measurement from an industrial gas processing plant. The measurement has an oscillatory trend and a pulse at hour 4 caused by a transient disturbance. The bottom panel of Figure 1 shows the wavelet coefficients for the measurement at different times (horizontal axis) and scales (vertical axis). The light tones represent high amplitude coefficients. The plot shows that the transient disturbance maps to high amplitude coefficients from scales one to four. However, the measurement shows various steep increases and decreases which are not part of transient disturbances. The steep variations also map to high amplitude coefficients at the same scales as the transient, and hence the transient cannot be distinguished.

This paper addresses the detection of transients with the nearest neighbors method recently proposed in [10]. This method is more generic than the approaches discussed in the previous paragraph. It does not assume the transient to have a specific shape, nor does it require the wavelet coefficients or the relative amplitude of the transient to
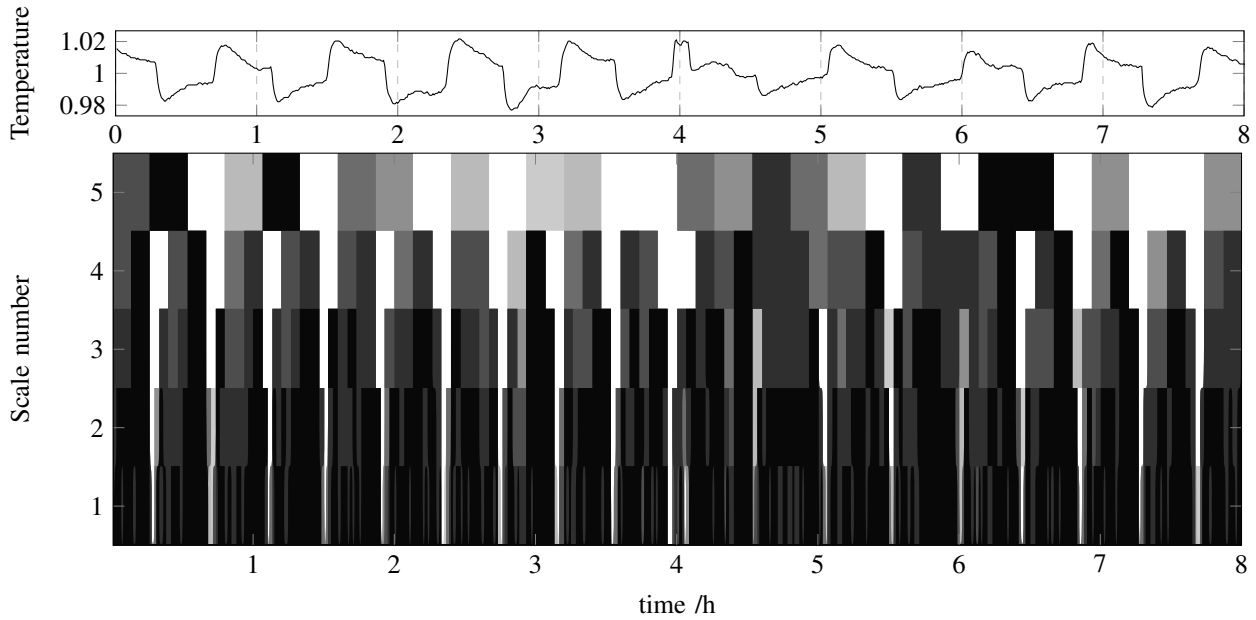
Fig. 1: Top panel: measurement with normalized units. Bottom panel: corresponding coefficients of Haar wavelet at different times (horizontal axis) and scales (vertical axis). Lighter tones represent higher amplitude coefficients.
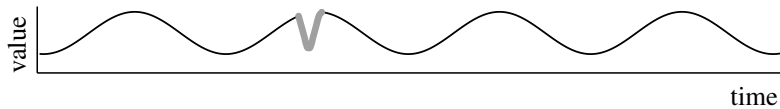

Fig. 2: Time series with an anomalous segment, drawn in a thick grey line.

be distinct from the overall measurement trend. The method is also independent of models of routine or healthy operation. Reference [10] presents detailed background on this topic, and a brief summary is given here in the interests of a self-contained paper.

1) Nearest neighbors for detection of transient disturbances: Transient disturbances can be detected by considering them as anomalous segments in a time series. A segment $\mathbf{x} = [x_1 \ x_2 \ \cdots x_m]$ is a sequence of samples x_i from a time series X, which are ordered in time. Anomalous segments are those different from the underlying trend of the time series, as illustrated in Figure 2 where the anomalous segment is drawn in a thick grey line.

Nearest neighbors techniques detect anomalous segments by measuring their similarity to all other segments in the time series. The most similar segments are known as nearest neighbors. As illustrated in Figure 2, the similarity between an anomalous segment and any of the other segments is low because its sequence of samples is significantly different. Therefore, this segment will not have near neighbors.

2) Implementation: The Euclidean distance is a common metric to assess similarity between segments of time series. Zero distance indicates maximum similarity, and occurs only between segments which are equal in all samples. Equation (1) defines the Euclidean distance between segments $\mathbf{x}_1$ and $\mathbf{x}_2$, with m samples in each segment.

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^{m}(x_{1,i} - x_{2,i})^2} \tag{1}$$

An anomalous segment is detected by its anomaly index. This is a single value defined as the distance between the segment and its kth nearest neighbor [10], [20], [21]. The distance of an anomalous segment to its kth nearest neighbor will be significantly higher than the distances of normal segments to their k^{th} nearest neighbors. The reason for using the kth nearest neighbour and the choice of the optimum value of k are explained in [10]. The authors discuss also other distance metrics and definitions of anomaly index, along with their adequacy to detect transient disturbances.

B. Multivariate analysis with SVD

In this paper, the objective of a multivariate analysis is to identify representative features of a set of variables, where representative features are those which capture the largest proportion of variance of the variables. This happens if a feature has itself a large variance or is present in many of the variables. These features can be identified through a singular value decomposition of a matrix $\mathbf{A}$, containing the variables arranged as row vectors.

Singular Value Decomposition (SVD) is a mathematical technique for factorizing any data matrix $\mathbf{A}$ into a product of three other matrices, with specific properties [22], [23]. Equation (2) shows this decomposition.

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\top} \tag{2}$$

Matrices $\mathbf{U}$, $\mathbf{S}$, and $\mathbf{V}^{\top}$ show the following properties:

- The columns of matrix $\mathbf{U}$ are orthonormal vectors which are basis functions for the columns of $\mathbf{A}$. In other words, the columns of matrix $\mathbf{U}$ are linearly independent and, by linear combination, can represent every column of $\mathbf{A}$.
- Similarly, the rows of matrix $\mathbf{V}^{\top}$ are orthonormal basis functions for the rows of $\mathbf{A}$.
- Matrix $\mathbf{S}$ is a diagonal matrix of singular values of $\mathbf{A}$. A singular value s_j gives the total projection of the rows, or columns, of $\mathbf{A}$ along the direction of the associated orthonormal basis, namely, row vector $\mathbf{v}_j^{\top}$, or column vector $\mathbf{u}_j$. Furthermore, singular values are ordered by decreasing value.

When the rows of matrix $\mathbf{A}$ are mean-centred, SVD has also a statistical interpretation: the square of a singular value is directly related to the total variance of $\mathbf{A}$ along the direction defined by the corresponding row. Specifically, it equals the variance times $n - 1$, where n is the number of elements in the row vector. This means that the first few row vectors $\mathbf{v}_j^{\top}$ are the directions which capture the largest proportion of variance of the rows of $\mathbf{A}$. This is the reason why SVD can be used for multivariate analysis.

SVD is closely related to the statistical technique of Principal Component Analysis (PCA). PCA is formulated as $\mathbf{A} = \mathbf{T}\mathbf{V}^{\top}$, where $\mathbf{T} = \mathbf{U}\mathbf{S}$, from equation (2). In this context, matrices $\mathbf{T}$ and $\mathbf{V}^{\top}$ are known as score and loading matrix, respectively. PCA has been extensively used in process monitoring [24], [25]. For example, in

multivariate statistical process monitoring the interest is to identify representative features across plant profiles [26]. These features are retrieved as the rows of matrix $\mathbf{V}^\top$. A plant profile is the ensemble of the values of all process measurements at a particular instant in time.

C. Multi-rate systems

Multi-rate systems have been identified as a challenge in areas related to process disturbance analysis.

Review of the literature shows that a significant part of research has been in the fields of control [12] and soft sensors [27], and the aim is usually to estimate the outputs of the system at the faster samping rate. Significant contributions have addressed the development of state space models at the faster sampling rate [28], and parameter and state estimation [29]. Multi-rate data has also been used to calibrate models by applying Bayesian methods [30]. These techniques have been mostly applied on petroleum, chemical and biological systems, in which the measurements with lower sampling rate derive from quality variables which require laboratory analysis.

Multi-rate systems have also been addressed in the field of process monitoring. In this case, the objective was to develop multivariate models that describe process data during normal operation, so that new incoming measurements can be compared to these models. To that end, Bakshi [25] and Misra et al. [8] combined wavelet analysis and principal component analysis. The authors suggested that these methods could handle multi-rate systems because the integration of the multi-rate variables is done over different time scales, into which the measurements are previously decomposed by the wavelet analysis.

As discussed in the Introduction, the current paper has a focus on methods which are independent of models of routine or healthy operation, a problem for which multi-rate systems remain an open question.

III. UNIVARIATE DETECTION OF TRANSIENT DISTURBANCES

The first step of the method proposed in the current paper follows the algorithm for univariate detection of transient disturbances proposed in reference [10]. This section recalls that algorithm. It first presents a reference example, which is used throughout the paper to illustrate the different methods.

A. Reference example

Figure 3 shows a schematic of a gas compression rig in ABB Corporate Research Center, Kraków, Poland. A particular set of experiments with this rig lead to the seven time series plotted in Figure 4, which correspond to the measurement tags in Figure 3. All time series are available at a rate of 1 kHz and consist of 25000 samples.

Two step changes were imposed in the set-point of the compressor speed around 6 and 18 s. These resulted in transient disturbances, which are visible in Figure 4 around these times in the first five measurements. Additionally, the first five measurements are also affected by oscillations, which are due to unstable operation of the compressor close to the surge limit. Of course, it is not desirable to operate an industrial compressor so close to its surge limit, but the experimental rig in question was able to explore such operation. The measurements of flow, F1, and valve
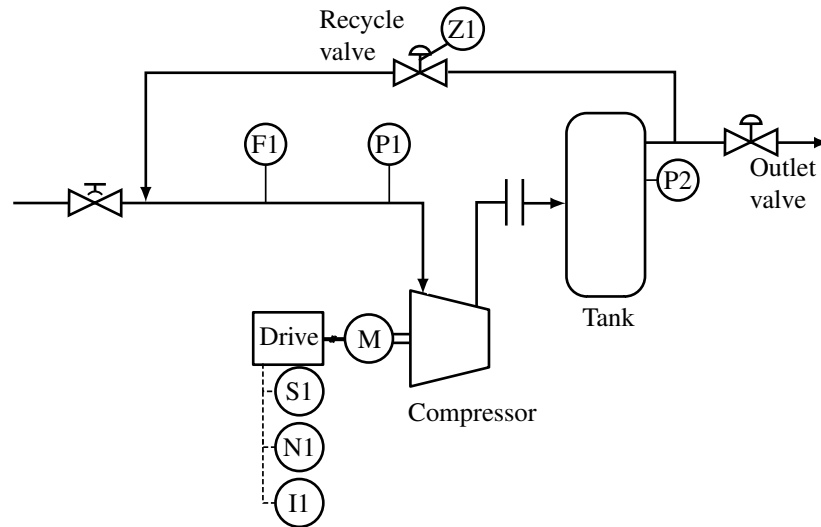
Fig. 3: Simplified schematic of the gas compression experimental rig.
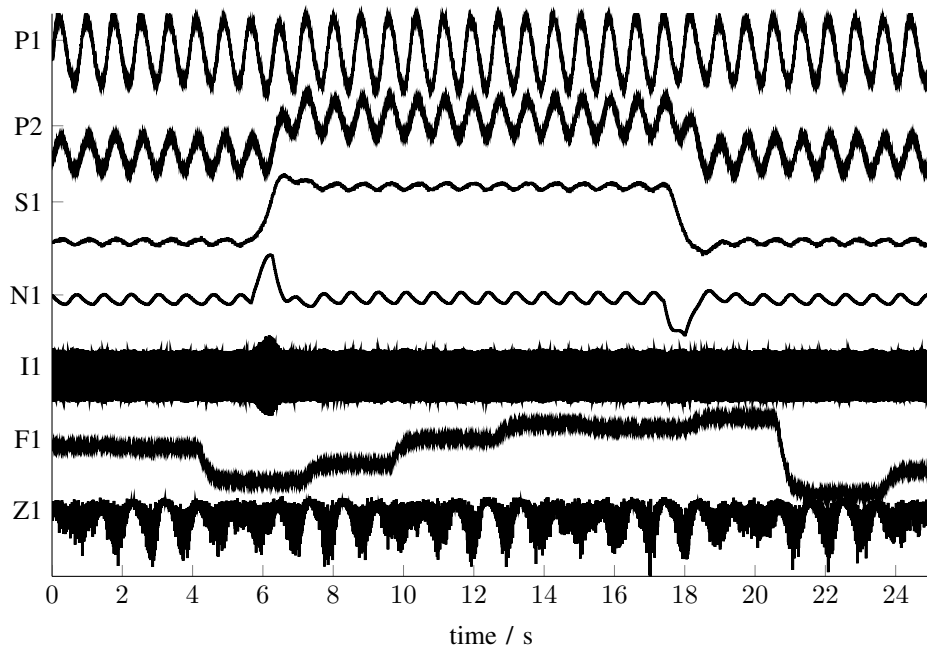

Fig. 4: Reference example: time series of measurements from a gas compression rig.

position, Z1, do not show these transients. These seven measurements were chosen because they reflect four distinct situations with regards to the presence of transients.

In measurements of speed, S1, and torque, N1, the transient disturbances can be seen clearly, whereas in measurements of suction pressure, P1, and discharge pressure, P2, the transients are present but masked by the

oscillations. The measurement of current, I1, is also affected by the two transients, but only the first is clearly visible in its time series.

The interest of F1 is that it does not show the transient disturbances at 6 and 18 s. The reason is that the compressor was operating unstably in this experiment, and the flow rate was low and pulsating. However, the flow instrument is a vortex-shedding flow meter, not suited to disturbed flows especially at the lower end of its measurement range. Even though the flow measurement shows artefacts, it is included in the data set because it does show abrupt decays around 4 and 20 s, as well as several smaller positive steps which are not synchronized with the transient features of the other measurements in the data set. As a result, F1 poses a useful challenge to the development of the method.

The interest of Z1 is being free of transient disturbances. The valve position, Z1, was fixed during the experiment. The persistent variations visible in its measurement were an electronic artifact, and thought to be produced by interference between sensors or in the communication between the sensors and the data aquisition module. The trend of Z1 is normalized and the persistent variations are only a small percentage (2%) of the mean value. Although Z1 has dips of short duration, these are not considered as transients because they are frequent and repeating.

B. Univariate anomaly index vector

The basis of detecting transient disturbances is the identification of anomalous segments in the time series of a measurement. This section recalls the method proposed in reference [10] to identify such segments in an univariate approach, that is, in the time series of a single measurement.

1) Embedding matrix: As explained in reference [10], the samples of a time series X are arranged into an embedding matrix $\mathbf{X}$ according to equation (3). Row vectors $\mathbf{x}_i$ are known as embedded vectors, and are m-sample long. Parameters m, τ and δ are called the embedding dimension, the embedding granularity, and the embedding step. The number of embedded vectors, N_E, depends on these parameters and on the total number of samples n.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{N_E} \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_{1+\delta} & x_{1+\delta+\tau} & \cdots & x_{1+\delta+(m-1)\tau} \\ \vdots & \vdots & & \vdots \\ x_{1+(N_E-1)\delta} & x_{1+(N_E-1)\delta+\tau} & \cdots & x_{1+(N_E-1)\delta+(m-1)\tau} \end{bmatrix} \tag{3}$$

Figure 5 illustrates the concepts of embedded vector, m, τ and δ. It shows a symbolical time series represented by dots, and the first three embedded vectors given the parameter values indicated in the top left corner.

Each embedded vector in the matrix is then mean-centered. This means that embedded vectors which had identical trends but different ranges of their numerical values will now become similar. For instance, the three-sample segments $\begin{Bmatrix} 1 & 2 & 3 \end{Bmatrix}$ and $\begin{Bmatrix} 11 & 12 & 13 \end{Bmatrix}$ will be identical and both equal to $\begin{Bmatrix} -1 & 0 & 1 \end{Bmatrix}$ after mean centering.

2) Similarity: The purpose of embedded vectors is to capture the trend of every m-sample segment in the time series. If any of these segments is anomalous, it will be significantly different from all others. This dissimilarity is evaluated using the Euclidean distance metric to compare each embedded vector to every other. This comparison
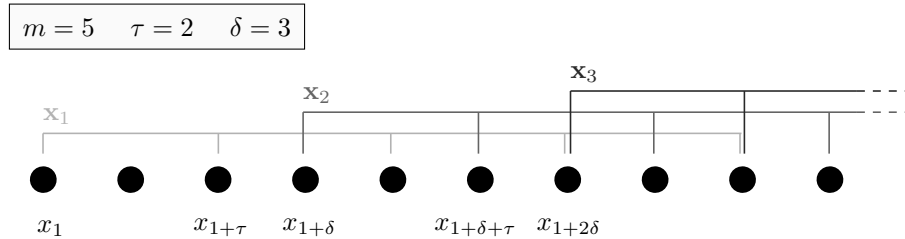
Fig. 5: Symbolical time series of samples x_i and its first three embedded vectors, given parameters m, τ and δ.

can be implemented with two nested loops, running down the rows of the embedding matrix, which will generate a $N_E \times N_E$ matrix of distances.

3) Anomaly index vector: The anomaly index vector, **ai**, is the sequence of anomaly indices, ai_i, of each of the N_E embedded vectors. The anomaly index ai_i is defined as the kth smallest distance between embedded vector $\mathbf{x}_i$ and all others. As explained in reference [10], distances to near-in-time embedded vectors are excluded. Near-in-time embedded vectors to $\mathbf{x}_i$ are its adjacent segments with overlapping samples [31].

Figure 6a shows the anomaly index vectors for the measurements in the reference example. These anomaly index vectors were computed using $\tau = 1$, $\delta = 1$, $m = 1001$, and $k = 5$, which are the values suggested in reference [10], after optimization. The dashed lines represent the detection threshold, as defined in section IV-B.

Each anomaly index vector in Figure 6a is normalized. As suggested in reference [10], every anomaly index ai_i is divided by the median value of all the anomaly indices in the vector. The median approximates the average anomaly index of non anomalous embedded vectors [10].

The univariate detection method yields the desired results for all measurements except P1, that is, periods with transient disturbances generate anomaly indices above the detection threhold, while transient-free periods have anomaly indices below the detection threshold. The detection is noteworthy in measurements P2 and I1 because in these measurements the transients are not clearly distinct from the rest of the time series. In measurement P1 the transients are even less distinct from the rest of the time series, and hence the corresponding anomaly indices did not protrude above the threshold. This means that in the univariate approach the presence of the transients in P1 is not recognized. The next section of the paper describes a multivariate method which strengthens the evidence that the transients occur in P1 by exploiting the presence of the same disturbances in the other measurements.

IV. MULTIVARIATE DETECTION OF TRANSIENT DISTURBANCES

The previous section showed it can be difficult to detect transient disturbances in measurements with strong oscillatory trends if the detection is done in a univariate context. This section presents the first contribution of the paper, which is a method to detect transient distubances in a multivariate approach. The multivariate method exploits the presence of the same disturbance in several measurements for an improved outcome. The measurements can be from process and electromechanical variables if they are sampled at equal rate, such as with the reference example.
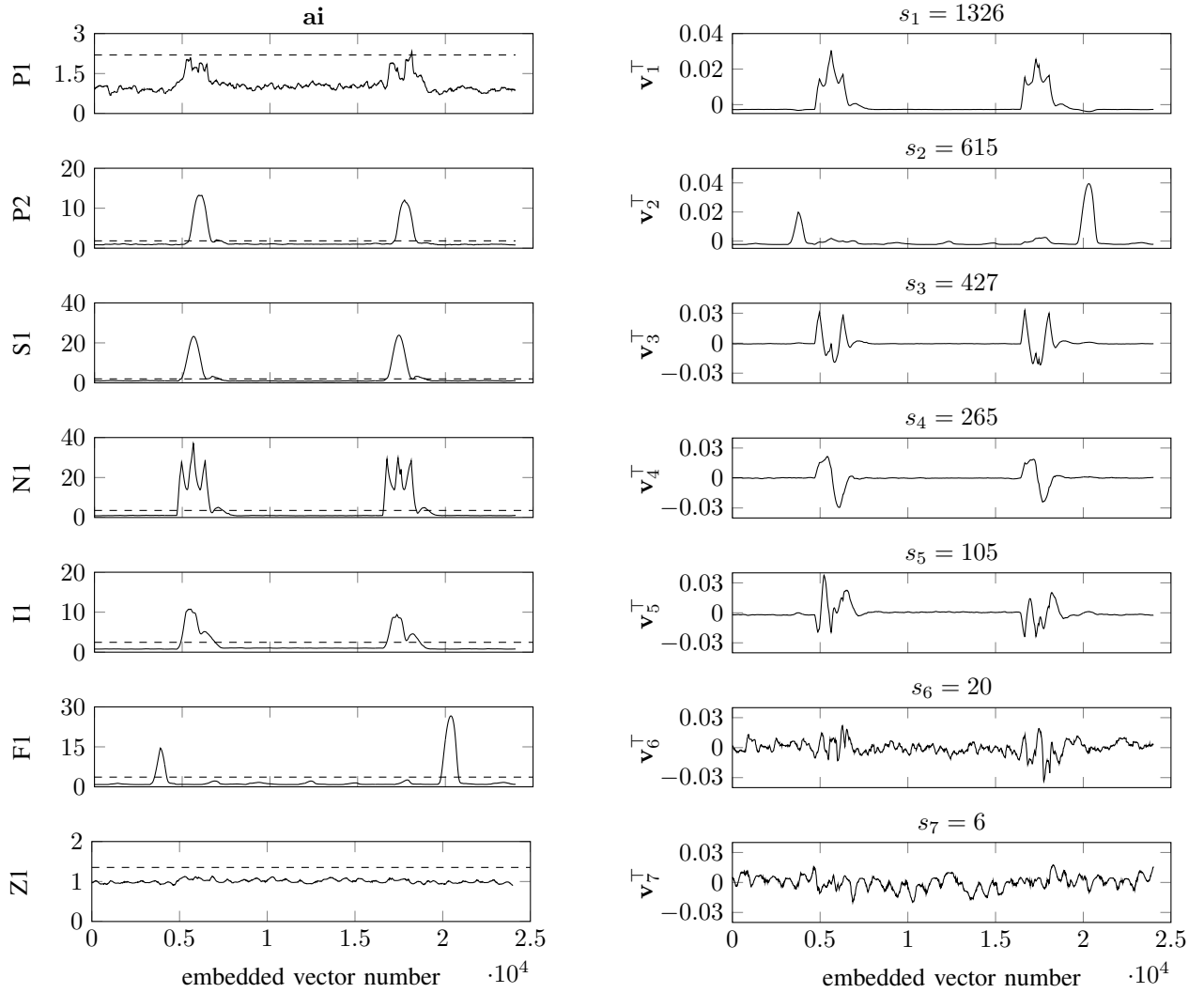
(a) Anomaly index vectors.

(b) Basis functions. The values on top of the plots indicate the singular value associated with each basis function.

Fig. 6: Univariate anomaly index vectors and orthonormal basis functions obtained from SVD of the anomaly index matrix, for the reference example.

A. Multivariate method

The multivariate detection method starts from the univariate anomaly index vectors **ai**. These are generated as in section III, for each of N_V measurements.

1) Pre-treatment and anomaly index matrix: The multivariate analysis requires the normalization and mean-centering of each vector **ai**. Normalization makes the univariate anomaly index vectors independent of their engineering units, and of the number of samples in the embedded vectors. Otherwise, it would be meaningless to compare the numerical values. This can be done by dividing the anomaly indices in vector **ai** by the median

value of all the indices in the vector, as done in section III-B3. The normalized anomaly index vectors are then mean-centered. The reason is to take advantage of the statistical interpretation of SVD, as discussed in section II-B.

The anomaly index vector of measurement r is now treated as a row vector, $\mathbf{ai}_r^\top$, and arranged in row r of an anomaly index matrix, $\mathbf{A}$, of size $N_V \times N_E$. This arrangement is shown in the left-hand side of equation (4).

2) SVD of the anomaly index matrix: The anomaly index matrix $\mathbf{A}$ is factorized with SVD as $\mathbf{A} = \mathbf{USV}^\top$. Equation (4) shows the SVD, and puts the orthonormal basis functions $\mathbf{v}_j^\top$ in evidence. The aim of SVD is to identify representative features across the various anomaly index vectors. These vectors are in the rows of $\mathbf{A}$, hence the features of interest are captured by the rows of matrix $\mathbf{V}^\top$, as discussed in section II-B.

In the equation, $u_{r,j}$ are elements of $\mathbf{U}$, and s_j are entries in the diagonal of $\mathbf{S}$.

$$\mathbf{A} = \begin{bmatrix} \mathbf{ai}_1^\top \\ \mathbf{ai}_2^\top \\ \vdots \\ \mathbf{ai}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{N_V,1} \end{bmatrix} s_1 \mathbf{v}_1^\top + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{N_V,2} \end{bmatrix} s_2 \mathbf{v}_2^\top + \cdots + \begin{bmatrix} u_{1,N_V} \\ u_{2,N_V} \\ \vdots \\ u_{N_V,N_V} \end{bmatrix} s_{N_V} \mathbf{v}_{N_V}^\top \tag{4}$$

Figure 6b shows the basis functions $\mathbf{v}_j^\top$ derived from the $N_V = 7$ measurements in the reference example. They are ordered by the magnitude of the associated singular value, s_j, which is indicated on top of each plot.

As desired, the first few basis functions show entries with values which are distinctively higher than the rest of $\mathbf{v}_j^\top$. In the first and third basis functions, the higher entries correspond to embedded vectors which are associated with the transient disturbances in measurements P1, P2, S1, N1, and I1. In the second basis function, the higher entries are associated with the transient disturbances in measurement F1.

Some of the basis functions $\mathbf{v}_j^\top$, along with the corresponding singular values s_j and column vectors $\mathbf{u}_j$, will be discarded from the univariate anomaly index vectors $\mathbf{ai}_r$. The two following selection steps determine which terms to discard. The detection of transients will be done on the final anomaly index vectors $\tilde{\mathbf{ai}}_r$.

3) Selection of basis vectors relevant on a multivariate level: Transient disturbances generate anomaly indices that account for most of the variance of the rows of matrix $\mathbf{A}$. These representative features are captured by the first few basis functions $\mathbf{v}_j^\top$. For instance, if there are two independent transient events, one expects that these are captured by the first two basis functions. The rest of the basis functions should capture details of comparatively less importance to the description of the transients.

For the above reason, a first selection step discards basis functions $\mathbf{v}_j^\top$ which are not representative of the anomaly index matrix. This step also discards the corresponding singular values s_j and column vectors $\mathbf{u}_j$. Specifically, a basis function is retained if its associated singular value, s_j, satisfies the inequality below.

$$s_j^2 \geq \alpha \cdot \sum_{r=1}^{N_V} s_r^2 \tag{5}$$

As discussed in section II-B, s_j^2 directly relates to the total variance of the rows of $\mathbf{A}$ along the direction of the basis function $\mathbf{v}_j^\top$. Accordingly, $\sum_{r=1}^{N_V} s_r^2$ relates to the variance of $\mathbf{A}$ along all directions $\mathbf{v}_r^\top$.

The selection criterion (5) is based on Kaiser's rule [32] for selecting components in principal component analysis. This was shown to be a reliable method in the study by Valle et al. [33].

Parameter α should be such as to discard basis functions $\mathbf{v}_j^\top$ which are unrelated to the description of the transients. The reference example uses $\alpha = 0.05$. As a result, the selection retains the first three basis function shown in Figure 6b, along with the corresponding singular values and column vectors. These are the basis functions that characterize the two independent transient events, along with details which accomodate different time durations of the transients in different variables. Section V will provide guidelines for the setting parameter α.

The anomaly index vectors which result after this step, $\hat{\mathbf{ai}}_r$, are formed by the retained terms $u_{r,j}s_j\mathbf{v}_j^\top$, as shown in equation (6) for the reference example.

$$
\hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{ai}}_1^\top \\ \hat{\mathbf{ai}}_2^\top \\ \vdots \\ \hat{\mathbf{ai}}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{N_V,1} \end{bmatrix} s_1\mathbf{v}_1^\top + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{N_V,2} \end{bmatrix} s_2\mathbf{v}_2^\top + \begin{bmatrix} u_{1,3} \\ u_{2,3} \\ \vdots \\ u_{N_V,3} \end{bmatrix} s_3\mathbf{v}_3^\top \tag{6}
$$

Figure 7a shows the intermediate anomaly index vectors $\hat{\mathbf{ai}}_r$ after this selection step. The new detection thresholds, marked with dashed lines, are based on the intermediate anomaly index vectors $\hat{\mathbf{ai}}_r$. The numerical values of $\hat{\mathbf{ai}}$ can be negative as well as positive, and this is due to the mean-centering step at the start of the method.

The first plot, for P1, is evidence that the first selection step facilitates the detection of transients in measurements with strong oscillatory trends or noise. Figure 6a had shown that the univariate anomaly index $\mathbf{ai}$ for P1 did not recognize the presence of the transients. On the other hand, the intermediate anomaly index vector $\hat{\mathbf{ai}}$ of P1 already shows values above the detection threshold. This is because the first selection step discarded terms of the univariate anomaly index vector which were associated with the last four basis functions.

4) Selection of terms relevant on an individual level: The last plot of Figure 7a also shows that the anomaly index vector $\hat{\mathbf{ai}}$ of Z1 after the first selection step erroneously presents values above the detection threshold. The reason is that the basis functions $\mathbf{v}_j^\top$ retained in the previous step show little similarity to the behavior of the univariate anomaly index vector $\mathbf{ai}$ of Z1. Therefore, a second selection step discards from each anomaly index vector $\hat{\mathbf{ai}}_r$ the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are not relevant to the univariate anomaly index vector $\mathbf{ai}_r$. Specifically, a term $u_{r,j}s_j\mathbf{v}_j^\top$ from anomaly index vector $\hat{\mathbf{ai}}_r$ in equation (6) is only retained if its variance, given by $\dfrac{(u_{r,j}s_j)^2}{(N_E-1)}$, satisfies

$$
\frac{(u_{r,j}s_j)^2}{N_E-1} \geq \beta \cdot \mathrm{var}(\mathbf{ai}_r) \ . \tag{7}
$$

As a parallel with PCA, β compares the contribution of each score $t_{r,j} = u_{r,j}s_j$ of anomaly index vector $\mathbf{ai}_r$ with all its scores.

The second selection step can be visualized in Figure 8. Each plot in this figure compares the values of $(u_{r,j}s_j)^2$ for each variable r. The dashed lines are the selection thresholds dependent on β, that is, $(N_E-1)\beta \cdot \mathrm{var}(\mathbf{ai}_r)$.
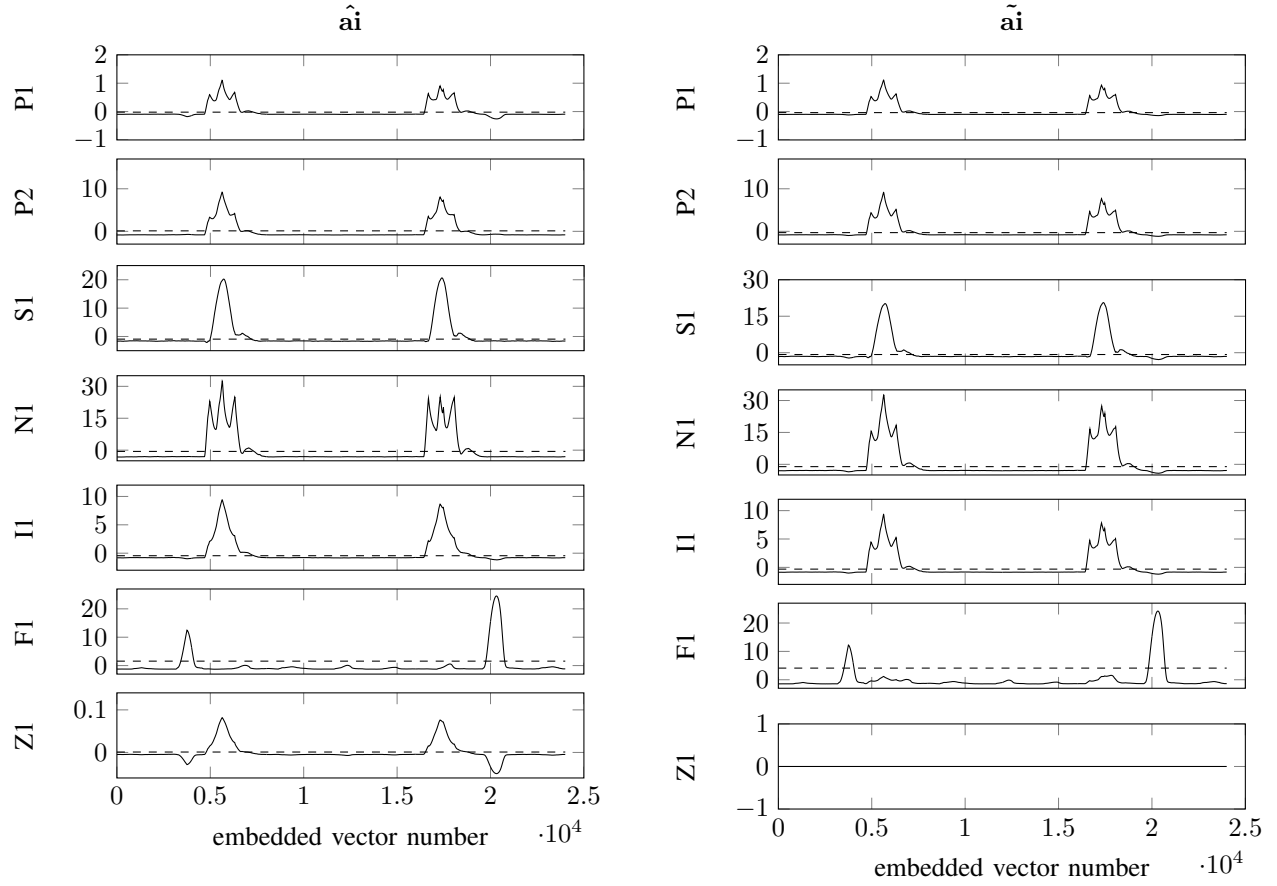
(a) Anomaly index vectors after the first selection step, $\hat{\mathbf{ai}}_r$.

(b) Final anomaly index vectors $\tilde{\mathbf{ai}}_r$.

Fig. 7: Intermediate and final anomaly index vectors for the reference example.

These thresholds were calculated using $\beta = 0.2$. Only terms $u_{r,j}s_j\mathbf{v}_j^\top$ such that $(u_{r,j}s_j)^2$ is above the threshold are retained in the final anomaly index vector.

Equation (8) shows that the final anomaly index vectors $\tilde{\mathbf{ai}}_r$ are formed by the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are retained after this second selection step. Terms are discarded by setting $u_{r,j}$ to zero. The zeros in equation (8) follow the reference example for illustration.

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{ai}}_1^\top \\ \tilde{\mathbf{ai}}_2^\top \\ \tilde{\mathbf{ai}}_3^\top \\ \vdots \\ \tilde{\mathbf{ai}}_{N_V}^\top \end{bmatrix} = \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ u_{3,1} \\ \vdots \\ 0 \end{bmatrix} s_1\mathbf{v}_1^\top + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} s_2\mathbf{v}_2^\top + \begin{bmatrix} 0 \\ 0 \\ u_{3,3} \\ \vdots \\ 0 \end{bmatrix} s_3\mathbf{v}_3^\top \tag{8}$$
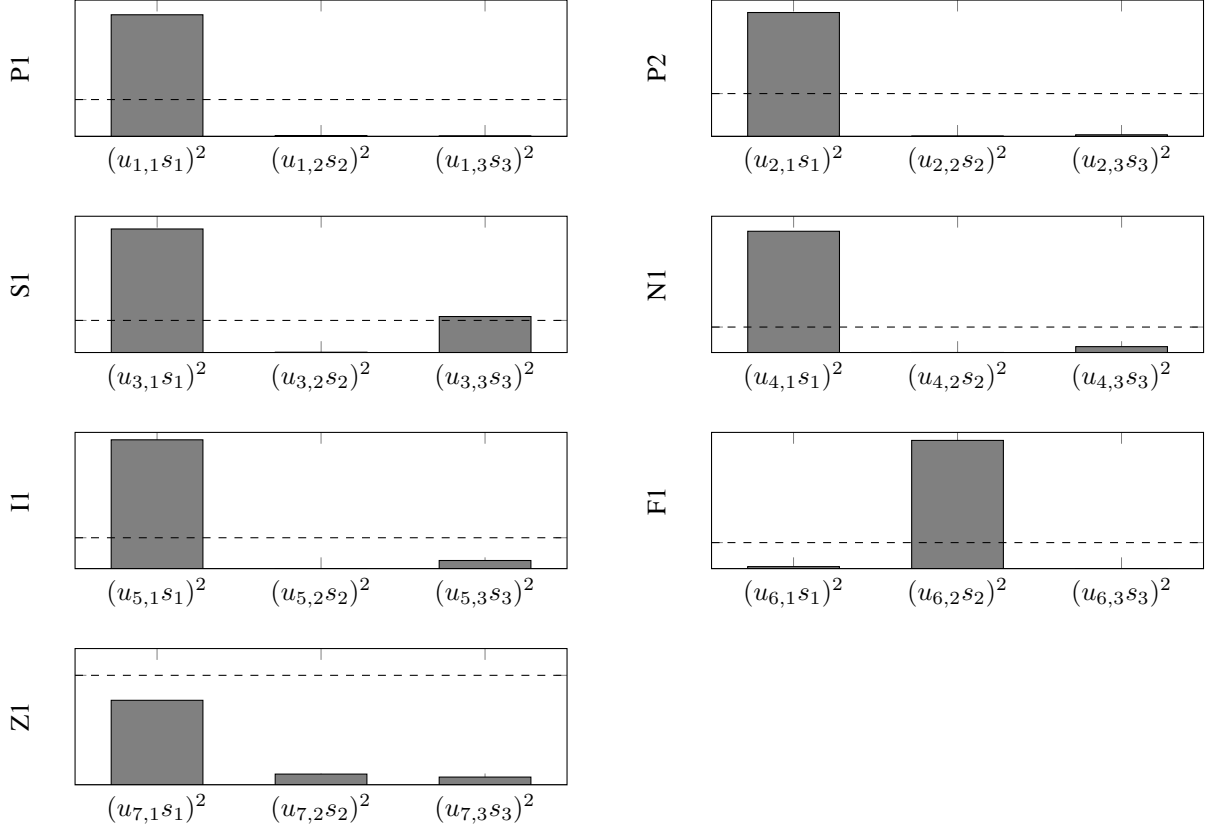
Fig. 8: Comparison of $(u_{r,j}s_j)^2$ for each variable r with the selection thresholds (dashed lines) derived from $\beta = 0.2$.

Figure 7b shows the final anomaly index vectors $\tilde{\mathbf{ai}}_r$ for the reference example. It should be recalled that the objective is to detect the presence of two transient disturbances in all measurements except Z1, which must be considered as transient-free. In addition, the pair of transients in measurement F1 must be recognized as occuring at different times than the other measurements. Figure 7b shows that the proposed method achieves these objectives. The dashed lines represent the detection thresholds, which are based on the final anomaly index vectors.

5) Plant-wide anomaly index vector: A global characterization of the group of variables with regards to transient disturbances can be given by a plant-wide anomaly index vector, $\tilde{\mathbf{ai}}_{PW}$. This is given by the arithmetic average of the N_V final anomaly index vectors, as shown in equation (9).

$$\tilde{\mathbf{ai}}_{PW} = \frac{1}{N_V} \left[\sum_{r=1}^{N_V} \tilde{ai}_{r,1} \quad \sum_{r=1}^{N_V} \tilde{ai}_{r,2} \quad \cdots \quad \sum_{r=1}^{N_V} \tilde{ai}_{r,N_E} \right] \tag{9}$$

Figure 9 shows the plant-wide anomaly index vector for the reference example. It clearly captures the occurence of the two groups of transient disturbances. The two protusions of larger magnitude correspond to the transients which are present in most measurements, while the two protusions of smaller magnitude correspond to the transients which are present only in measurement F1, and hence are less important plant-wide.
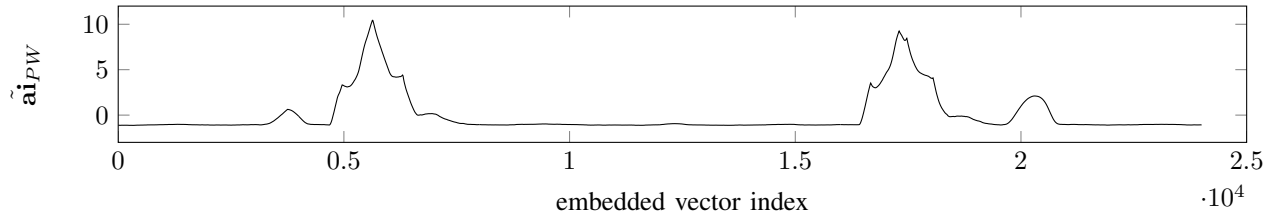
Fig. 9: Plant-wide anomaly index vector for the reference example.

B. Detection threshold and confidence level

A transient disturbance is detected when embedded vectors generated from the time series of a measurement are classified as anomalous. Classifying an embedded vector as anomalous happens when its anomaly index $\tilde{ai}$ is greater than a threshold.

The threshold used is indicated in the right-hand side of equation (10). It is the same as proposed in reference [10], but operating on the probability distribution of $\tilde{\mathbf{ai}}$, specifically on its median and interquartile range, IQR. The median is defined as the 50th percentile (Q_2), and IQR as the difference between the 75th and 25th percentiles. These are robust statistics because they are less influenced by outlier values than classical statistics [34].

$$ai > Q_2(\tilde{\mathbf{ai}}) + 6 \times \text{IQR}(\tilde{\mathbf{ai}}), \tag{10}$$

Appendix A analyzes the behavior of the threshold under the null hypothesis of a time series with no anomalies. The analysis uses three cases representative of time series with no anomalies: (i) steady state operation with only random noise, (ii) operation with non-random variability, and (iii) oscillatory operation. The conclusion is that the confidence level of this detection threshold depends on parameter β. With $\beta > 0.15$, the probability of false positives is less than one in a thousand.

V. PARAMETER SETTINGS

Multivariate detection of transient disturbances involves the selection of (i) the number of samples to characterize a transient, N_t, (ii) the parameters of the nearest neighbors method (NN in Table I), and (iii) the parameters for the SVD selection steps. Table I indicates all parameters. Symbols n and N_V were defined earlier.

The recommended values for the parameters in the nearest neighbors method are those suggested by Cecílio et al. [10]. These authors optimized the parameters to yield the best univariate detection results. The multivariate detection method proposed in the current paper builds on univariate anomaly index vectors, and aims to improve the univariate detection results. Thus, it should start from the optimal univariate anomaly index vectors.

This section optimizes the values for N_t and the SVD parameters, which are specific to the multivariate detection method. The results are summarized in Table I. Like many data-driven methods, the tuning is inexact and it is challenging to provide a single recommendation that will work for all cases. The paper has tried to point to

TABLE I: Parameters involved in the multivariate detection method, with recommended values.

Method	Parameters	Recommended value
	N_t, number of samples in a transient	$40 \leq N_t \leq 200$
NN	τ, embedding granularity	1
	m, embedding dimension	$0.5\, N_t$
	δ, embedding step	$\delta \leq 5$
	k, number of nearest neighbors	$3 \leq k \leq {}^n/_{10}$
SVD	α	${}^{0.3}/_{N_V}$
	β	0.2

some guidelines, which are generalizable because they relate the parameters of the method to the dynamics of the system.and the number of variables in the data set.

It should be noted that parameter N_t is set indirectly, while all other parameters are set explicitly. Specifically, parameter N_t is enforced by adjusting the sampling interval of the time series in a pre-processing step. This requires having an idea of the typical duration of a transient, which can be judged from past experience with the system or initial visual inspection of the data.

A. Influence of parameters in the detection results

A transient disturbance shoud be described by a number of samples, N_t, that is sufficient to characterize its dynamics. On the other hand, increasing N_t implies, in general, an increase in computational time with a power law relationship. The reason is that N_t is, in general, directly related to the number of samples in the time series, n, and the nearest neighbors step is $\mathcal{O}(n^2)$.

Parameter α determines which basis functions $\mathbf{v}_j^\top$ are retained. Thus, α influences the detection results for all variables in the set. Parameter β determines which terms $u_{r,j} s_j \mathbf{v}_j^\top$ are retained. Thus, β influences the detection results of each variable individually.

B. Parameter optimization

This section determines the best values for the parameters using the measurements in Figure 4. First, different parameter values are used to compute the final anomaly index vector $\tilde{\mathbf{ai}}_r$ from each measurement r. Each $\tilde{\mathbf{ai}}_r$ is then evaluated individually, using metrics based on false negatives and false positives [10].

Metric FN (false negatives) assesses whether embedded vectors that cover a transient are incorrectly classified as normal. In the comparison, N_E^{trans} is the total number of embedded vectors that include a transient, while N_{TP} is the number of embedded vectors correctly classified as anomalous. The metric is FN defined in equation (11), and has a value between zero and one.

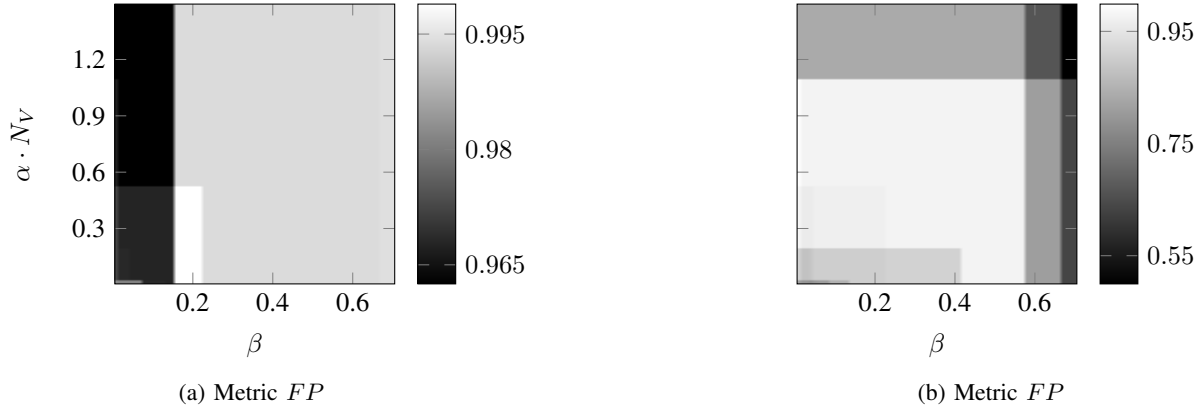$$FN = \frac{N_{TP}}{N_E^{trans}} \tag{11}$$

(a) Metric FP

(b) Metric FP

Fig. 10: Performance of the detection method as a function of parameters α and β. N_V denotes the number of measurements in the set. Paler colours denote better performance.

Metric FP (false positives) assesses whether embedded vectors not covering a transient were incorrectly classified as anomalous. In the comparison, N_E^{norm} is the total number of embedded vectors that do not include a transient, while N_{TN} is the number of embedded vectors correctly classified as normal. The metric is FP defined in equation (12), and has a value between zero and one.

$$FP = \frac{N_{TN}}{N_E^{norm}} \tag{12}$$

For parameter optimization, the correct classification of embedded vectors into normal and anomalous is defined visually beforehand. The actual classification is based on applying the threshold proposed in section IV-B to the final anomaly index vector. The measurement-specific metrics are then averaged over all measurements in order to obtain global FN and FP metrics for the whole set.

1) Selection of basis functions $\mathbf{v}_j^\top$ with α and terms $u_{r,j}s_j\mathbf{v}_j^\top$ with β: Figure 10 shows the influence of α and β on the detection results from the example in Figure 4. The performance is assessed by the global FP and FN metrics. The colors represent the magnitude of the metrics according to the scales shown in the figure. The vertical axis in the plots refers to $\alpha \cdot N_V$. The reason is that Kaiser [32] suggests that the best value for α varies inversely with N_V.

The highest value of FP is attained in the range $\alpha < {}^{0.5}\!/_{N_V}$ and $0.15 < \beta < 0.25$. The highest value of FN is attained in the range ${}^{0.2}\!/_{N_V} < \alpha < {}^{1.1}\!/_{N_V}$ and $\beta < 0.6$. These results suggest that optimal values for the parameters are $\alpha = {}^{0.3}\!/_{N_V}$ and $\beta = 0.2$.

The optimal values for the parameters are confirmed by additional analyses shown in appendix B. These analyses use five other sub-groups of measurements from the reference example. The analyses show that the influence of $\alpha \cdot N_V$ and β on the performance of the detection is generally consistent between the groups of measurements. The results also justify the use of $\alpha \cdot N_V$ as vertical axis. Each sub-group has one more measurement than the previous because the measurements and the size of the sub-group could affect how the performance responded to α and β.

The effects of α and β can be observed in more detail from Figure 10. In particular, the effects of α are:

- The increase in FN at $\alpha \cdot N_V = 0.2$ reflects the need for a lower limit for α. Above this limit, the selection step can discard basis functions $\mathbf{v}^\top$ which difficult the detection of transients in measurements such as P1.

- The small decrease in FP at $\alpha \cdot N_V = 0.5$ shows one of the disadvantages of increasing α beyond certain limits. This disadvantage is the removal of basis functions $\mathbf{v}^\top$ which capture details that define the precise start and end times of the transients.

- The decrease in FN at $\alpha \cdot N_V = 1.1$ shows the other disadvantage of increasing α beyond certain limits. This effect in performance is greater than in the previous paragraph. In this case, the disadvantage is the removal of basis functions $\mathbf{v}^\top$ which capture transients that are present in few measurements, such as the transients in measurement F1.

The effects of β are the following:

- The increase in FP at $\beta = 0.15$ reflects the need for a lower limit for β. Above this limit, the selection step can discard from the anomaly index vector of transient-free measurements, such as Z1, the terms $u_{r,j}s_j\mathbf{v}_j^\top$ which are associated with transients. This lower limit for β was expected from the discussion in section IV-B, and guarantee a probability of false positives of around one in a thousand.

- The small decrease in FP at $\beta = 0.5$ shows one of the disadvantages of increasing β beyond certain limits. Similarly to α, this disadvantage is the removal of terms $u_{r,j}s_j\mathbf{v}_j^\top$ which capture details that define the precise start and end times of the transients.

- The decrease in FN at $\beta = 0.6$ shows the other disadvantage of increasing β beyond certain limits. For measurements with masked transients, such as P1, transient-related terms $u_{r,j}s_j\mathbf{v}_j^\top$ capture a limited fraction of variance of the univariate anomaly index vectors. The value of β should be limited to retain such terms.

2) Number of samples in a transient, N_t: The number of samples in a transient was varied by downsampling the time series. The parameters of the nearest neighbors method were set as $\tau = 1$, $\delta = 1$, and $k = 5$, as recommended. Parameter m was reduced with the downsampling according to $m = 0.5 \cdot N_t + 1$. Parameters α and β were set to 0.05 and 0.2, respectively, following the discussion in section V-B1.

Figure 11 shows the influence of N_t on a performance index consisting of the sum of metrics FP and FN. As expected, the performance index increases when more samples characterize the transient. In particular, the results show that below $N_t = 40$ samples the performance decays below 85% of its maximum, and that above $N_t = 200$ samples the improvement in performance is only around 5%.

The first observation indicates that, whenever possible, the time series should be sampled so that a transient is described by at least 40 samples. The second observation indicates that if a transient is described by more than 200 samples then the time series is oversampled. This is important to know because the computational time of the algorithm grows as a power of N_t, as shown in Figure 12. Thus, if the sampling rate is high, and the computational time prohibitive, the time series can be downsampled until attaining 200 samples in the transient, with a limited loss in performance.
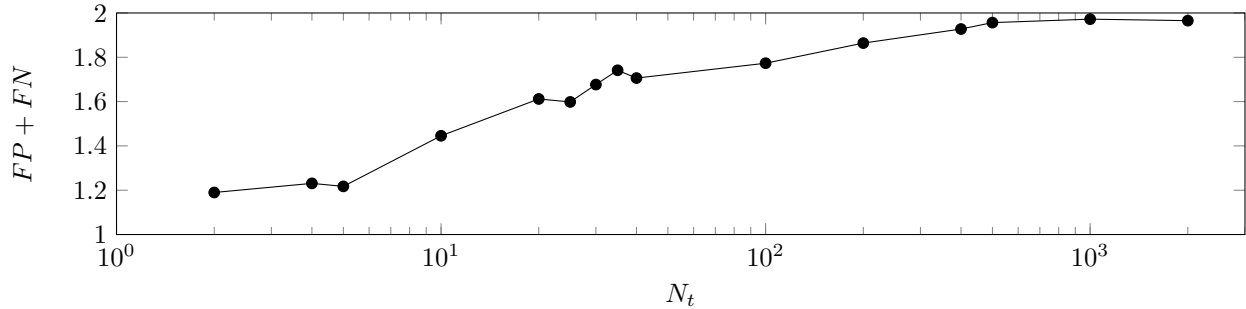
Fig. 11: Performance of the detection method as a function of the number of samples in a transient, N_t. The performance index can vary between zero and two.
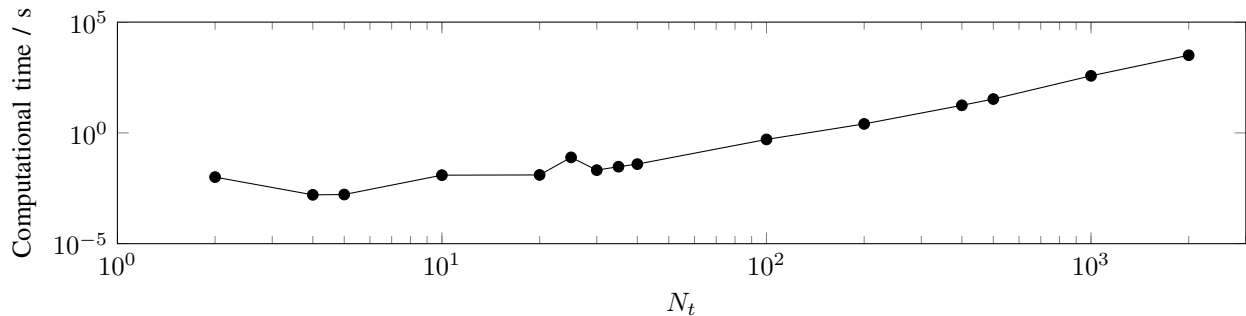

Fig. 12: Computational time for calculating one univariate anomaly index vector, shown as a function of the number of samples in a transient, N_t. The computational time is an average over the seven measurements in the set.

Users can resort to the recommended range of N_t to adjust the sampling interval of the time series in a pre-processing step. This requires having an idea of the typical duration of a transient, which can be judged from past experience with the system or initial visual inspection of the data. For instance, if the experience of the site is that electrical transients have a typical duration of 1 s, then the sampling interval should be shorter than 0.025 s to guarantee at least 40 samples in the transients.

VI. MULTIVARIATE DETECTION IN MULTI-RATE SYSTEMS

This section presents an extension of the multivariate detection method to multi-rate systems. The objective is to be able to use the advantages of the multivariate approach when some measurements are sampled at a slower rate than others, without having to downsample the fast-sampled measurements. Section VI-A shows the formulation of the method. Section VI-B demonstrates that the multi-rate method improves the detection of transients in a multi-rate data set compared to the alternative approach of downsampling to the lower sampling rate.

A. Extension of multivariate method to multi-rate systems

The multivariate method, as introduced in section IV, is not applicable to multi-rate systems. The reason is that SVD needs to be done on the anomaly index matrix $\mathbf{A}$. Each row in matrix $\mathbf{A}$ is the anomaly index vector $\mathbf{ai}$ of
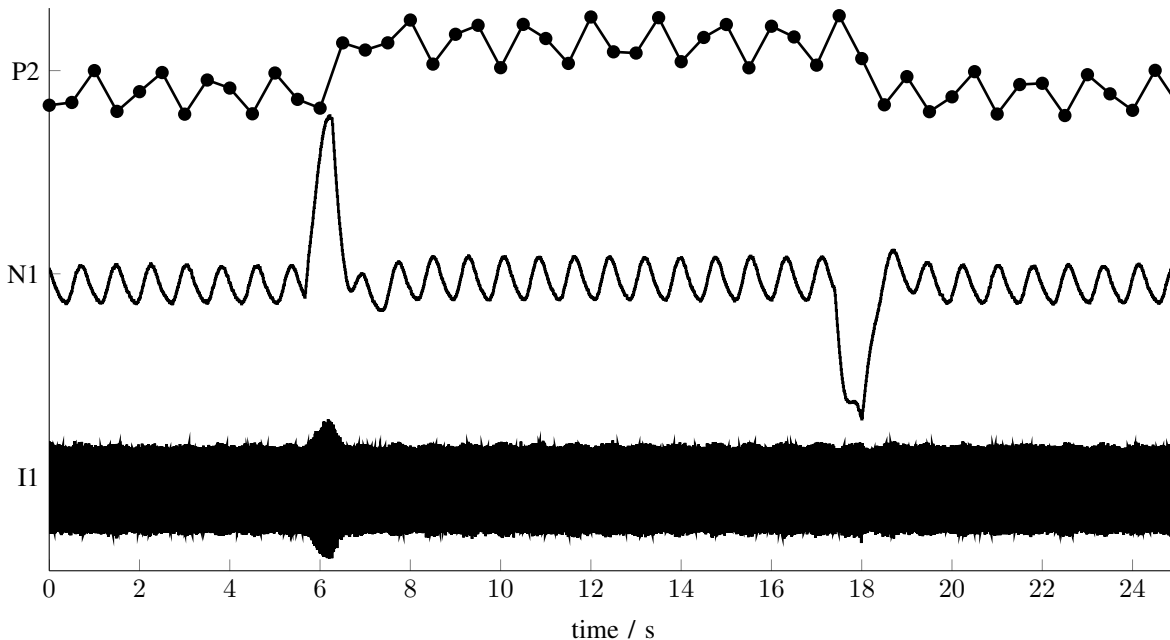
Fig. 13: Measurements from the reference example used to demonstrate the multi-rate detection method. Measurement P1 has been resampled with a sampling interval of 0.5 s to create a multi-rate data set.

one measurement, and each element in **ai** is the anomaly index for one embedded vector of that measurement. For the matrix implementation, the number of embedded vectors for each measurement must be the same. However, the slow-sampled measurements cannot have as many embedded vectors as fast-sampled measurements because of the longer intervals between samples. The alternative would, thus, be to downsample the fast-sampled measurements to the slower rate. This would lead to loss of information and, potentially, incorrect detection results.

This section shows how the multivariate detection method may be modified to handle multi-rate systems. The core modification is in the way the anomaly index vectors are constructed.

The method is illustrated on three measurements selected from the reference example in Figure 4. Figure 13 shows the time series of those three measurements. The reference example was not a multi-rate data set because all the measurements including the process measurements were sampled at 1 kHz. However, in practice the sampling rate for process measurements is slower than for electrical and mechanical measurements. Therefore the reference data set has to be manipulated to construct a realistic multi-rate data set. This is done by resampling the pressure measurement P1 with a sampling interval of 0.5 s, while leaving the sampling rate for torque and current measurements, N1 and I1, with a sampling interval of 1 ms.

This section has selected measurements P1, N1, and I1 to facilitate the visualization of the details of the method.

1) Embedding dimensions m^f and m^s: As with the uni-rate method, embedded vectors are defined as segments of m samples from a measurement, which are ordered in time and τ instants apart. However, the number of samples is different for fast-sampled and slow-sampled measurements. The requirement is that the embedded vectors of both must cover the same period of time. The period of time covered by an embedded vector is $(m-1)\tau\Delta t$, where Δt

is the sampling interval. The embedding dimension of the fast-sampled measurement is denoted as m^f, while the embedding dimension of the slow-sampled measurement is denoted as m^s. The relation between m^f and m^s must follow equation (13), where Δt^f and Δt^s denote the sampling intervals of fast and slow-sampled measurements, respectively.

$$\frac{(m^f - 1)\tau^f}{(m^s - 1)\tau^s} = \frac{\Delta t^s}{\Delta t^f} \tag{13}$$

The optimum period of time covered by an embedded vector follows the recommendations in Table I.

2) Numbers of embedded vectors N_E^f and N_E^s: According to [10], the number N_E of embedded vectors generated from a measurement is given by relation (14), where n indicates the number of samples in the measurement. The incomplete brackets indicate a floor function, which maps a real number to the largest previous integer.

$$1 + (N_E - 1)\delta + (m - 1)\tau \leq n \quad \Leftrightarrow \quad N_E = \left\lfloor \frac{n - (m - 1)\tau - 1}{\delta} + 1 \right\rfloor \tag{14}$$

In a fast-sampled measurement, the number of samples is denoted as n^f, while the number of samples in the slow-sampled measurements is denoted as n^s. If all measurements cover the same period of time, then n^f and n^s are related by equation (15).

$$\frac{n^f - 1}{n^s - 1} = \frac{\Delta t^s}{\Delta t^f} \tag{15}$$

This equation and equation (13) imply the relationship

$$\frac{(N_E^f - 1)\delta^f}{(N_E^s - 1)\delta^s} = \frac{\Delta t^s}{\Delta t^f} \tag{16}$$

where N_E^f is the number of embedded vectors in the fast-sampled measurements, and N_E^s is the number of embedded vectors in the slow-sampled measurements.

Parameters δ^f and δ^s in equation (16) are the embedding steps for fast and slow-sampled measurements, respectively. Their recommended values are the same as for δ, shown in Table I.

3) Expanded anomaly index vector for slow-sampled measurements: An anomaly index is computed for each embedded vector as in the uni-rate method (section IV). For the fast-sampled measurement, such anomaly indices are assembled in the anomaly index vector sequentially, as in section IV. This is illustrated in the top part of Figure 14. The black dots represent the initial samples of the fast-sampled measurement X, and each rectangle indicates the samples included in an embedded vector, for which $m^f = 4$ in the illustration. The anomaly index of embedded vector $\mathbf{x}_i$ is, thus, assigned to position i in the anomaly index vector.

The construction of the anomaly index vector for a slow-sampled measurement differs from the uni-rate method. The bottom part of Figure 14 illustrates this new procedure, where Y is the slow-sampled measurement and its embedded vectors have $m^s = 2$ samples. The anomaly index of embedded vector $\mathbf{y}_i$ is now assigned to position
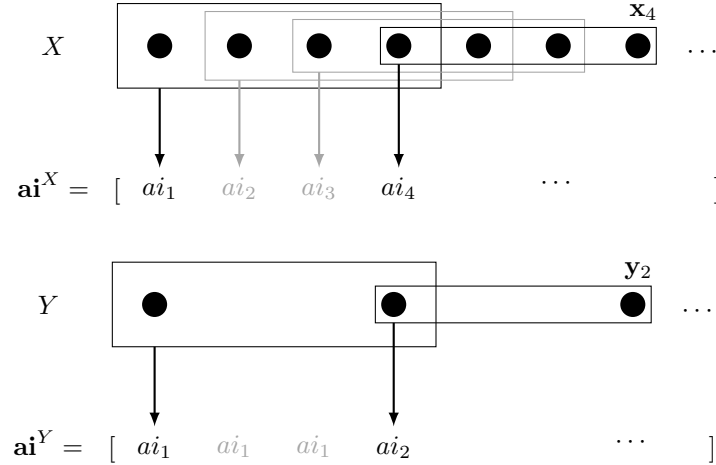
Fig. 14: Construction of the anomaly index vectors $\mathbf{ai}$ for a fast-, X, and a slow-sampled measurement, Y. The black dots represent samples in the time series of the measurements. The rectangles show the samples included in the embedded vectors $\mathbf{x}_i$ and $\mathbf{y}_i$. ai are the corresponding anomaly indices. The sampling intervals are related as $\Delta t^s / \Delta t^f = 3$, and the embedding dimensions are $m^f = 4$ and $m^s = 2$.

$1 + (i - 1)^{\Delta t^s}/\Delta t^f$. In this illustrative example, $\Delta t^s/\Delta t^f = 3$. This means that the anomaly indices of embedded vectors such as $\mathbf{x}_4$ and $\mathbf{y}_2$ in Figure 14 are assigned to the same position in the anomaly index vectors. Embedded vectors $\mathbf{x}_4$ and $\mathbf{y}_2$ cover the same period of time in X and Y.

The empty positions in the anomaly index vector of the slow-sampled measurement can be populated with the previous anomaly index, as illustrated in the bottom part of Figure 14 by the grey-colored anomaly indices. This is equivalent to assuming that the trend of the measurement stayed equally anomalous during those periods. This is an assumption at feature-level, as opposed to assuming values for the measurement samples, for example through interpolation.

The empty positions in the anomaly index vector of the slow-sampled measurement can also be populated in alternative manners. For example, the first half of the empty positions between two anomaly indices can be populated by the previous anomaly index and the second half by the subsequent anomaly index. Lowpass and cubic spline interpolations can also be used to interpolate between the available anomaly indices. All the alternatives yield similar results in the multivariate analysis of the next subsection.

Figure 15a shows the resulting anomaly index vectors $\mathbf{ai}$ for the selected measurements of the reference example. These are univariate anomaly index vectors, because they were built for each measurement separately. The first anomaly index vector corresponds to the slow-sampled measurement, and the step-wise shape is the result of expanding the anomaly indices as proposed.

If measurements X and Y originate from different sources, it may happen that their sampling instants are not synchronized, that is, each sample from the slow-sampled measurement falls in between two samples of the fast-sampled measurement. One strategy to handle asynchronous measurements is to resample the fast-sampled measurement X by interpolation. The new sampling instants of X retain the original sampling interval, but are
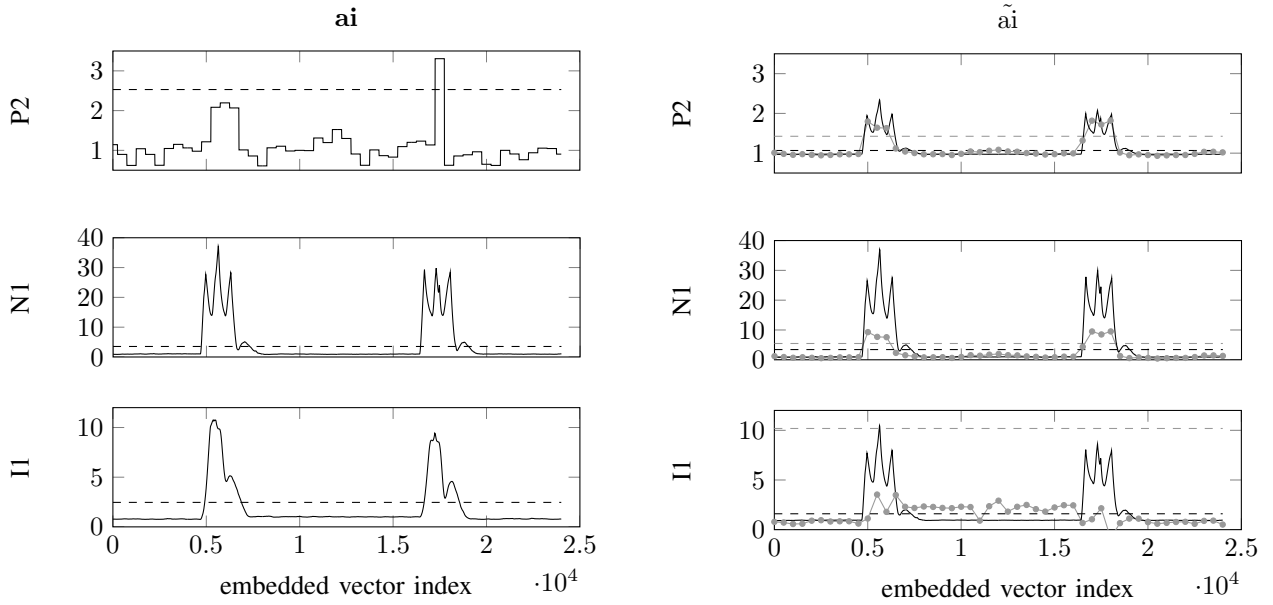
(a) Univariate anomaly index vectors. The first **ai** is the expanded anomaly index vector for the slow-sampled measurement.

(b) Final anomaly index vectors from the multi-rate method (black lines) and uni-rate method with the fast-sampled measurements downsampled to the slow rate (gey lines). Dashed lines represent the corresponding detection thresholds.

Fig. 15: Univariate and final anomaly index vectors for multi-rate measurements of the reference example.

time-shifted so that each sample from the slow-sampled measurement Y aligns with a sample of X. The values of the new samples are estimated using an interpolation function, such as quadratic and cubic polynomials, cubic splines, and Gaussian process [35]. In the problem of detection of transients, the error introduced by such interpolation should be small because electromechanical measurements have a large number of samples characterizing the events of interest.

4) Multivariate analysis: The anomaly index vectors constructed as above have the same number of elements, N_E^f. Therefore, it is now possible to build the anomaly index matrix as in the uni-rate method. The subsequent multivariate analysis and threshold-based detection follows as in section IV.

Figure 15b shows the final anomaly index vectors $\tilde{\mathbf{ai}}$ for the reference example. The black lines result from the multi-rate method, and the black dashed lines indicate the detection thresholds. In order to demonstrate the improvement obtained with the multi-rate method, Figure 15b also shows the results for the uni-rate method, with grey lines. To use the uni-rate method, the fast-sampled measurements had to be downsampled to the slow rate. As desired, the multi-rate method correctly identifies the transient disturbances in all measurements. On the other hand, the uni-rate method at slow rate misses the transients in the measurement I1, and estimates shorter durations for the transients in the other measurements, specially P2.

It should be noted that the sampling rate set up by automation engineers in a data historian normally reflects typical process time constants and the typical dynamic features of a process variable [36]. However, in some cases
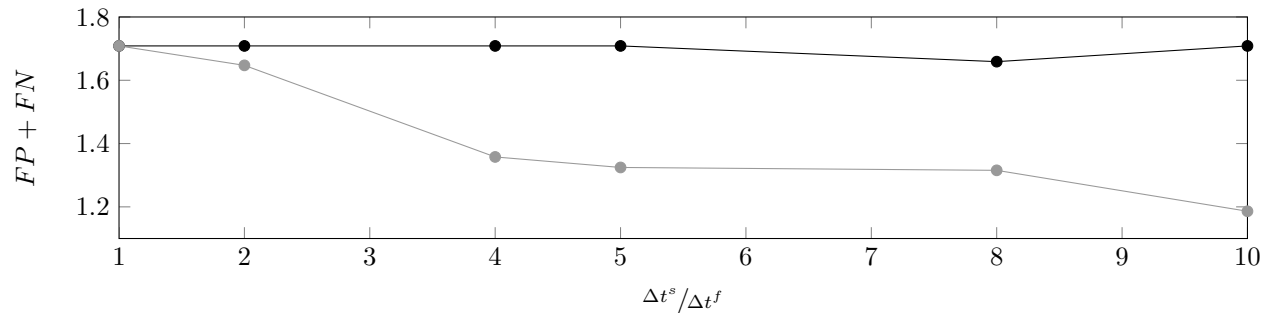
Fig. 16: Performance of the multi-rate (black line) and uni-rate methods (grey line) as a function of the sampling interval of the slow-sampled measurements Δt^s. The sampling interval of the fast-sampled measurements Δt^f is fixed. The performance index can vary between zero and two.

the sampling rate of process variables may be too low and completely miss a fast transient. In such cases, the anomaly index vector of that slow-sampled measurement will have no indication of the presence of the transient. Therefore, in the multivariate analysis the anomaly index vector will only contribute to the last few basis functions and the variable will be considered transient-free.

B. Comparison of the performance of the multi-rate and uni-rate methods

The purpose of this section is to quantify the improvement obtained when using the multi-rate method, in comparison to use of the uni-rate method with measurements downsampled to the lower rate. This is done using the measurements from the reference example.

The sampling interval Δt^s of the slow-sampled measurements is the limiting factor to the performance of both methods, since it affects the characterization of the trends of the slow-sampled measurements. Therefore, performance is assessed for different Δt^s.

In the multi-rate method, measurement P2 has the sampling interval Δt^s. Measurements N1 and I1 have a smaller sampling interval, Δt^f. The Δt^f used is that leading to $N_t = 40$ samples in the transients, which is the minimum recommended number of samples. The improvement obtained with the multi-rate method as determined in this section is, therefore, associated to $N_t = 40$. Any number of samples N_t greater than 40 should lead to improvements in performance equal to, or above, the improvements determined in this section.

Performance is measured by applying the metrics FP and FN presented in section V-B to the detection results of each measurement. The metrics are then averaged over the three measurements.

Figure 16 shows the influence of the sampling interval Δt^s on a performance index consisting of the sum of the global FP and FN metrics. The black line corresponds to the multi-rate method and the grey line to the uni-rate method. It is clear that the multi-rate method always performs better. At $\Delta t^s/\Delta t^f = 1$, the methods are equivalent. As expected, the improvement achieved by the multi-rate method increases as the sampling interval of the slow-sampled measurements increases.
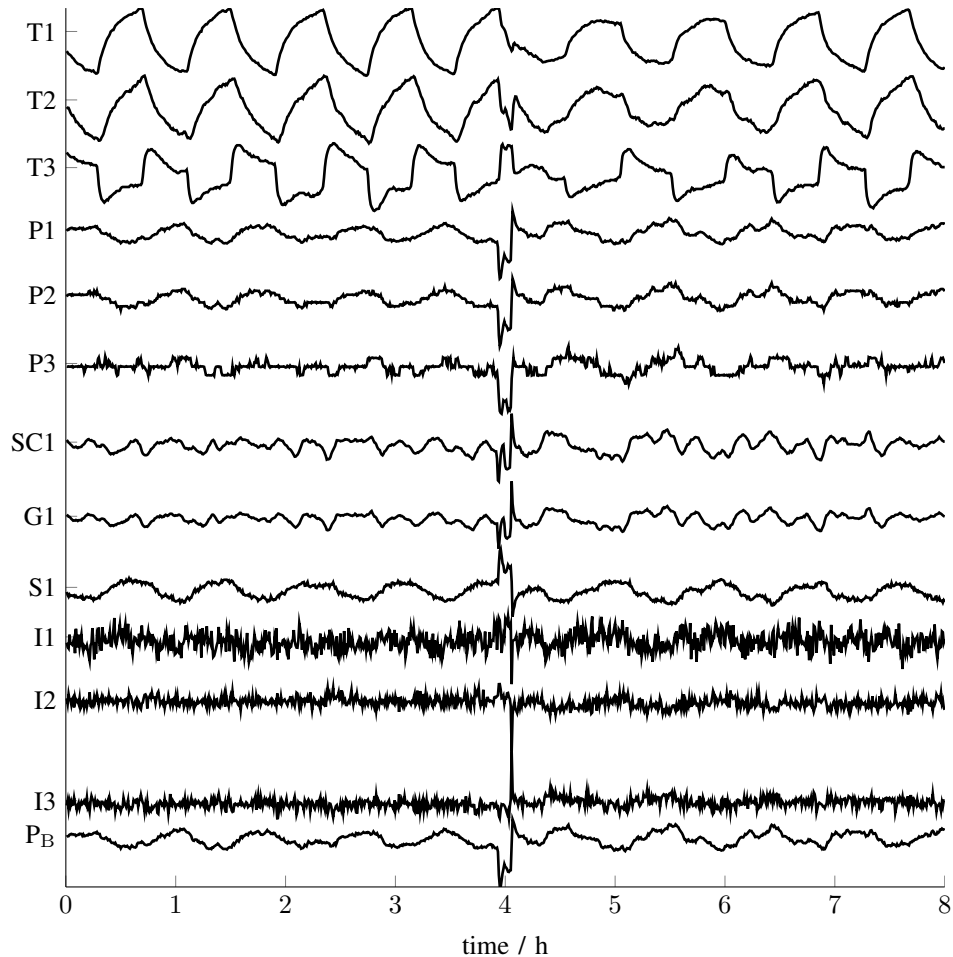
Fig. 17: Time series of the measurements in the gas processing plant case study.

Nonetheless, it should be noted that the time to compute the univariate anomaly index vectors is different for the two methods. Specifically, each fast-sampled measurement requires approximately $(\Delta t^s/\Delta t^f)^2$ times longer to process than the slow-sampled measurements.

VII. APPLICATION TO INDUSTRIAL CASE STUDIES

This section demonstrates the two methods proposed in the paper on two industrial case studies.

A. Multivariate detection in a uni-rate system

The first case study is part of an industrial process of gas processing, courtesy of ABB Oil, Gas and Petrochemicals, Oslo, Norway. Figure 17 shows the time series of 13 measurements from a gas expansion-recompression section of the process. These include gas temperatures (T1, T2, and T3), and pressures (P1, P2, and P3), control signals (SC1), valve position (G1), turbine speed (S1), drive currents (I1, I2, and I3), and pressure on the bearings of the turbine (P_B).

The transient affecting these measurements occurs around hour 4 in the figure. The transient manifests itself in different ways in each measurement. It must be noted that the underlying trends of most measurements are not constant. Several measurements show oscillations, namely T1, T2, T3, P1, P2, S1, and P_B, and the measurements of current I1, I2 and I3 are noisy.

The expectation for this case study is that the method combining nearest neighbors and SVD correctly detects the transient disturbance in all measurements. The multivariate step with SVD is expected to increase the robustness of the method to the other behavior in the trends, that is, the oscillations and the noise. All measurements are sampled synchronously, at 30 s intervals, thus the uni-rate method is applied.

The time scale of transient disturbance is approximately 15 min, which is common for process disturbances. The sampling interval of 30 s means these transients are described by 30 samples. The recommended value is 40, however this is not possible in this case study. Therefore the nearest feasible value is used. Other recommended parameter values for this case are $\tau = 1$, $m = 20$, $\delta = 1$ and $k = 3$, for the nearest neighbors step. For the multivariate step, the parameters are chosen according to section V-B as $\alpha = 0.03$ and $\beta = 0.2$.

Figure 18 shows, in black lines, the final anomaly index vectors, $\tilde{\mathbf{ai}}$, for each measurement. These can be compared to the anomaly index vectors $\mathbf{ai}$, obtained in the univariate analysis, which are represented in Figure 18 with grey lines. The comparison shows that the multivariate analysis is able to retrieve the feature between embedded vector numbers 400 and 500 which is common to all the anomaly index vectors $\mathbf{ai}$. The features of the individual $\mathbf{ai}$ which have been discarded include, most notably, the smaller protusions in the $\mathbf{ai}$ of measurements T1 and T3, and the noisy trend in the $\mathbf{ai}$ of measurement I1. By discarding such features, the anomaly indices of embedded vectors covering the transient become more distinct.

The outcome of the analysis is that a transient disturbance was detected in all measurements starting at 3.9 h and ending at 4.15 h. This result matches the expectation for this case study, hence it shows that the uni-rate method and recommended parameters are generalizable beyond the conditions of the reference example.

B. Multivariate detection in a multi-rate system

The second case study derives from tests conducted on a commercial ABB turbocharger compressor, courtesy of ABB Corporate Research Centers in Kraków, Poland and in Baden-Dättwil, Switzerland, and ABB Turbo Systems, Baden, Switzerland. The compressor is driven by an asynchronous motor, and is controlled by an a.c. measurement speed drive. The tests subjected the compressor to varying mass flow rates, at different rotational speeds.

Figure 19 shows the time series from six measurements monitored during these tests. The process measurements of flow (F1), inlet pressure (P1), outlet pressure (P2) and inlet temperature (T1) were collected by a process control module, and are given with a sampling interval of 5 s. The speed (S1) and torque (N1) of the motor shaft are measured by a monitoring module in the drive, and are given with a sampling interval of 0.1 s.

All measurements are affected by a sharp spike happening around minute 7. The transient is not very distinct in measurements N1 and T1, and the trends of most measurements are not constant. Specifically, measurements P1 and P2 show slow oscillations, and measurements T1, S1 and N1 are noisy. The expectation for this case study is
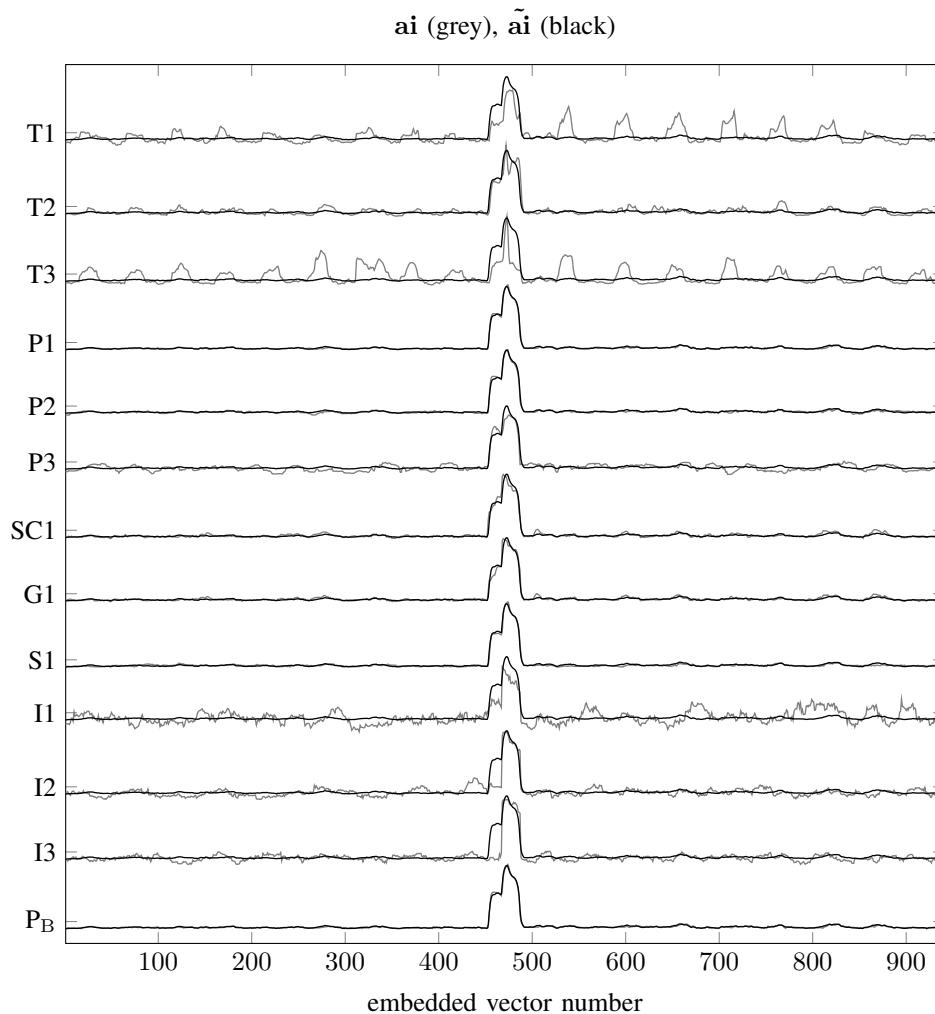
ai (grey), **ãi** (black)

Fig. 18: Anomaly index vectors for the measurements in the gas processing plant case study. Grey lines are the univariate anomaly index vectors **ai**. Black lines are the final anomaly index vectors **ãi** reconstructed after the multivariate step.

to detect the transient disturbance in all measurements. As in the previous example, the multivariate approach is used to improve the detection results in measurements in which the transient is less clear. In contrast to the previous example, this is a multi-rate system so the multi-rate version of the detection method is used.

The time scale of the transient disturbance is approximately 35 to 40 s. The transient in the measurements collected by the drive is characterized by 400 samples. This is above 40 samples, as recommended. The transient in the process measurements is characterized by 8 samples. As shown in section VI-B, the performance of the multi-rate method is similar whether the slow-sampled measurements have 40 samples or less.

Considering the sampling intervals, the recommended values for the embedding dimensions are $m^f = 250$, for the drive measurements, and $m^s = 5$, for the process measurements. The other parameters are $\tau = 1$, $\delta = 1$, $k = 3$, $\alpha = 0.05$ and $\beta = 0.15$.
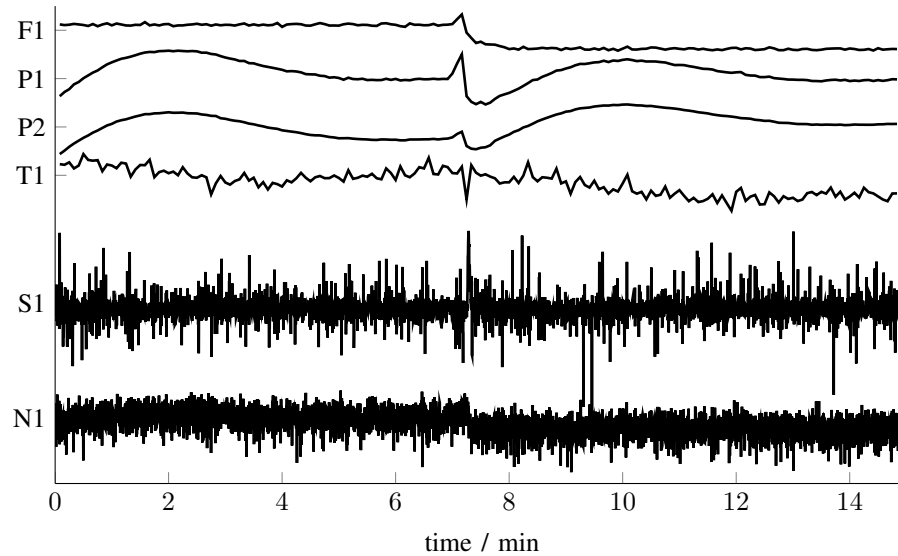
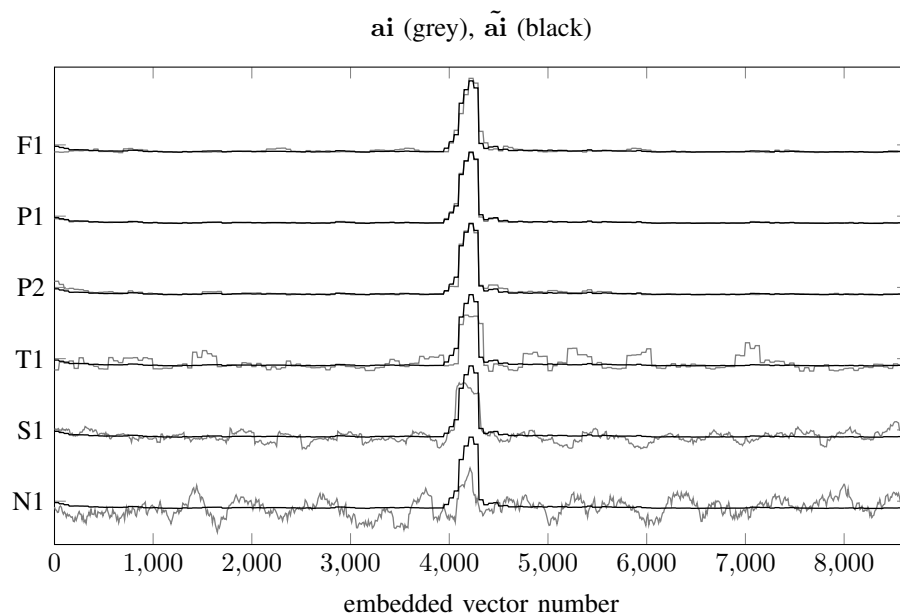Fig. 19: Time series of the measurements in the compressor tests case study.

Fig. 20: Anomaly index vectors for the measurements in the compressor tests case study. Grey lines are the univariate anomaly index vectors **ai**. Black lines are the final anomaly index vectors $\tilde{\textbf{ai}}$ after the multivariate step.

Figure 20 shows, in black lines, the reconstructed anomaly index vectors for each measurement. The grey lines represent the univariate anomaly index vectors. Those from process measurements are expanded as described in section VI-A3. This case study shows that the multi-rate method enables the multivariate analysis and, as desired, it retrieves the common features of the anomaly index vectors. The anomaly indices of embedded vectors covering the transient become more distinct particularly for T1, S1 and N1, leading to the correct detection results.

The outcome of the analysis is that a transient disturbance was detected in all measurements starting at 7.1 min and ending at 7.5 min. This result matches the expectation for this case study, hence it shows that the multi-rate method is generalizable beyond the conditions of the reference example.

VIII. Conclusion

This paper has presented two contributions to the detection of transient disturbances in a multivariate data set comprising process, electrical, and mechanical measurements. The extension of process disturbance analysis to electrical and mechanical measurements is justified by the increasing use of electrically-driven machinery in process industries, and the rising number of power quality incidents.

The first contribution was to extend the univariate transient detection method in [10] to a multivariate method. The results in this paper have demonstrated that the univariate method can fail in measurements with strong oscillatory trends or noise. In contrast, the multivariate method correctly detects the transients in such measurements because it exploits the presence of the same disturbance in other measurements. This method requires a detection threshold and parameters, thus the paper presented recommendations for both and analysed these recommendations.

The second contribution of this paper was to adapt the multivariate detection method to the case of a multi-rate system. This is justified because, in industry, the measurements from process, electrical, and mechanical systems are commonly available with different sampling rates. A uni-rate method would require the fast-sampled measurements to be downsampled to line up with the slow-sampled measurements. The paper has demonstrated that the multi-rate method achieves improved detection results in comparison to that scenario.

In summary, the multivariate problem was solved using singular value decomposition on a multivariate set of features known as anomaly index vectors. The values in these vectors are associated with segments of a measurement. Segments exhibiting a transient disturbance lead to high anomaly index values. The multi-rate extension was based on expanding the anomaly index vectors of slow-sampled measurements to match those of fast-sampled measurements.

The paper developed both methods on examples from experiments with a gas compression set-up. The applicability of the methods and recommended parameters to other examples was demonstrated in industrial routine data of a gas processing plant.

Appendix A
Confidence level of the detection threshold

This appendix analyzes the behavior of the detection threshold for the multivariate method, under the null hypothesis of a time series with no anomalies. Monte-Carlo simulations are used to generate three groups of N_{MC} time series with no anomalies. Each group is based on one time series, which is representative of an operation scenario with no transient disturbances. The three representative time series are: (i) steady state operation, (ii) operation with non-random variability, and (iii) oscillatory operation. The time series used for case (i) is synthetic while the time series for cases (ii) and (iii) belong to physical measurements monitored in a real gas processing
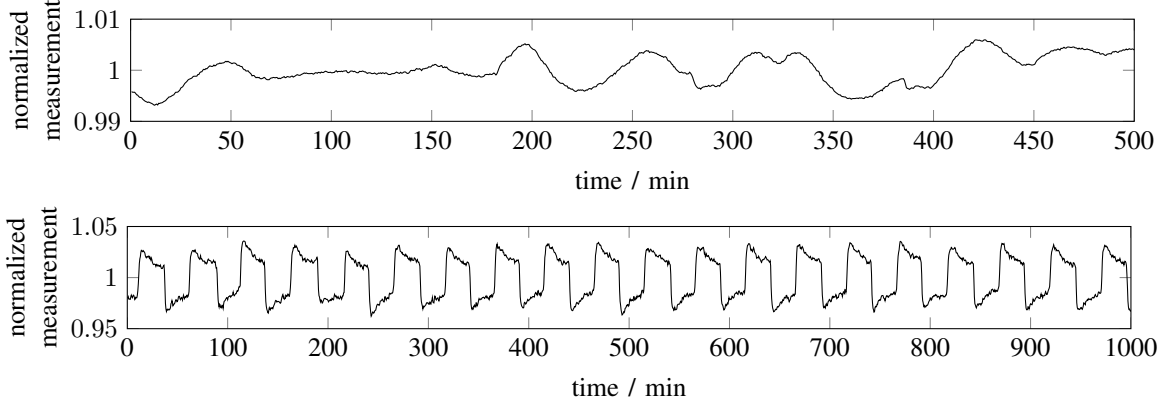
Fig. 21: Base time series for generation of cases representing normal operation scenarios. Top panel:operation with non-random variability. Bottom panel: oscillatory operation.
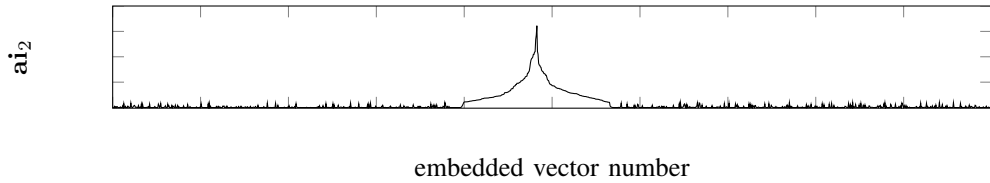

embedded vector number

Fig. 22: Univariate anomaly index vector $\mathbf{ai}_2$ used in the statistical study of the detection threshold.

plant and are shown in Figure 21. Each of the N_{MC} time series in a group is given by the representative time series added to one random time series, which represents Gaussian noise.

The univariate detection method, described in section III, generates $3N_{MC}$ univariate anomaly index vectors $\mathbf{ai}$ from the time series. The parameters used are $m = 50$, $\tau = \delta = 1$ and $k = 3$, as proposed in section V.

An anomaly index matrix $\mathbf{A}$ is formed for each univariate anomaly index vector, which is arranged in the first row of the matrix and hence is denoted $\mathbf{ai}_1$. A second row stores an additional anomaly index vector $\mathbf{ai}_2$, which is generated synthetically and is shown in Figure 22. This vector resembles the anomaly index vector from a time series with a clear transient disturbance.

The purpose of the anomaly index vector $\mathbf{ai}_2$ is two-fold. First, it allows a multivariate analysis. Second, it leads to a SVD basis function $\mathbf{v}_j^\top$ with a clear protusion. Such a basis function can potentially induce false positives in the anomaly index vectors of the time series with no anomalies.

With an anomaly index matrix $\mathbf{A}$ formed by two rows, the final anomaly index vector $\tilde{\mathbf{ai}}_1$ reconstructed after the multivariate step is formed by, at most, two terms. False positive detections on $\tilde{\mathbf{ai}}_1$ happen only if an included term leads itself to positive detections. Therefore, this section determines the probability that a term leading to positive detection is included in $\tilde{\mathbf{ai}}_1$. This gives the probability of $\tilde{\mathbf{ai}}_1$ leading to false positive detections.

A term j leading to positive detections is included in $\tilde{\mathbf{ai}}_1$ if its ratio

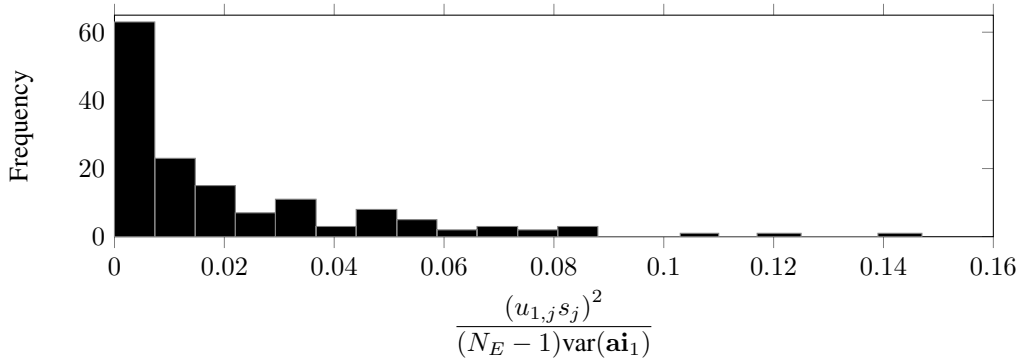$$\frac{(u_{1,j}s_j)^2}{(N_E-1)\text{var}(\mathbf{ai}_1)}$$

Fig. 23: Distribution of the values of relative variance of the SVD terms with positive detections.

TABLE II: Measurements included in each sub-group used for the optimization of α and β. N_V refers to the number of measurements in the group.

N_V	Tags					
3	N1	Z1	P1			
4	N1	Z1	P1	S1		
5	N1	Z1	P1	S1	I1	
6	N1	Z1	P1	S1	I1	P2

$$\frac{(u_{1,j}s_j)^2}{(N_E-1)\text{var}(\mathbf{ai}_1)} \tag{17}$$

is greater than parameter β (section IV-A). Figure 23 shows the frequency distribution of ratio (17) for terms j leading to positive detections. Positive detections are determined based on the threshold proposed in section IV-B.

The exponential distribution fits approximately to the distribution in Figure 23. The exponential distribution has a single parameter, estimated in this case as $\lambda = 48$. An exponential distribution with $\lambda = 48$ has a probability lower than one in a thousand that the ratio (17) is greater than 0.15. This means that, with $\beta > 0.15$, a term leading to positive detections has a probability lower than one in a thousand to be included in the anomaly index vector $\tilde{\mathbf{ai}}$ of a time series with no anomalies. As a result, with $\beta > 0.15$, the probability of false positives with the proposed threshold is less than one in a thousand.

APPENDIX B
OPTIMIZATION OF α AND β FOR ADDITIONAL GROUPS OF MEASUREMENTS

This appendix presents the optimization results for five sub-groups of measurements from the reference example. Table II indicates the measurements included in each sub-group.

Figure 24 shows the influence of α and β on the detection results from each of the five sub-groups, as measured by performance metrics FP, and FN. The colors represent the magnitude of the metrics according to the scales shown in the figure.
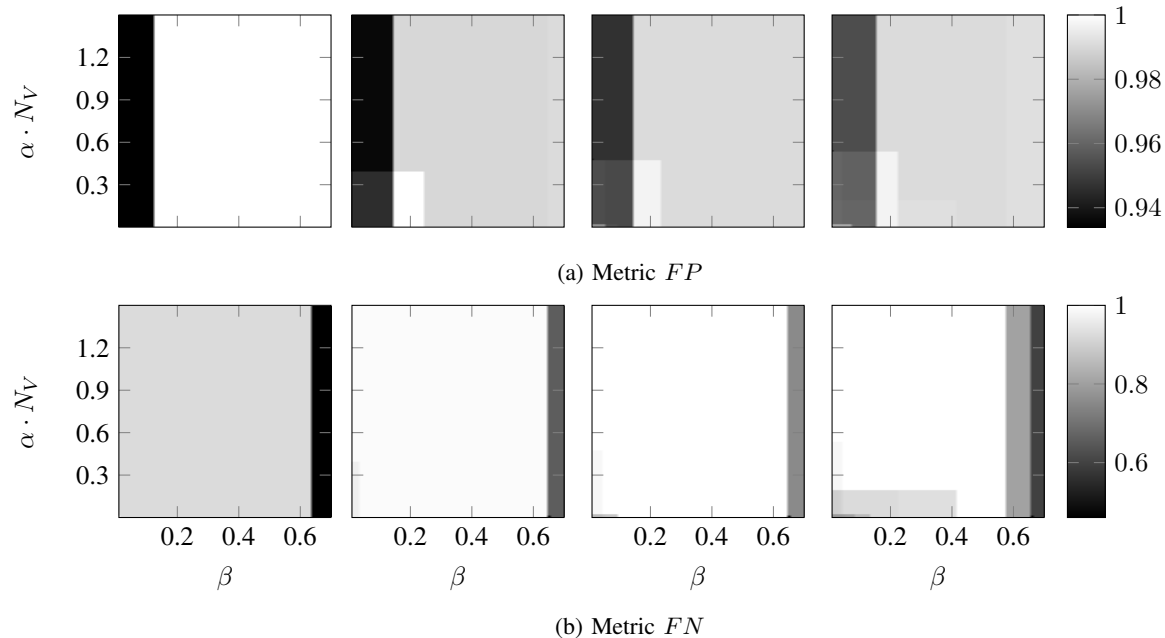
(a) Metric FP

(b) Metric FN

Fig. 24: Performance of the detection method as a function of parameters α and β for five different sub-groups of measurements. N_V denotes the number of measurements in the group. Paler colours denote better performance.

ACKNOWLEDGMENT

REFERENCES

[1] REAL-SMART. Marie Curie FP7-IAPP project. http://www3.imperial.ac.uk/realsmart. [Online; accessed 20/10/13].

[2] N. P. Directorate. (2012) Assessment of electrification of the utsira high. http://www.npd.no/Templates/OD/Article.aspx?id=4509&epslanguage=en. [Online; accessed 17/02/14].

[3] Statnett. (2012) Systemdrifts- og markedsutviklingsplan 2012. http://www.statnett.no/Documents/Kraftsystemet/Systemansvaret/Statnett_SMUP_24.05_lnk_Low.pdf. [Online; accessed 14/04/13].

[4] N. F. Thornhill and A. Horch, "Advances and new directions in plant-wide disturbance detection and diagnosis," *Control Engineering Practice*, vol. 15, no. 10, pp. 1196–1206, 2007.

[5] Y. A. Shardt, Y. Zhao, K. H. Lee, X. Yu, B. Huang, and S. L. Shah, "Determining the state of a process control system: Current trends and future challenges," *The Canadian Journal of Chemical Engineering*, vol. 90, no. 2, pp. 217–245, 2012.

[6] I. M. Cecílio, S.-L. Chen, and N. F. Thornhill, "Importance of auxiliary systems for process fault detection and diagnosis," in *Proceedings of the 19th Mediterranean Conference on Control & Automation (MED)*. Corfu: IEEE, 2011, pp. 952–957.

[7] H. Bevrani, *Robust Power System Frequency Control*. Springer, 2009.

[8] M. Misra, H. H. Yue, S. J. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Computers & Chemical Engineering*, vol. 26, no. 9, pp. 1281–1293, 2002.

[9] J. T. Cheung and G. Stephanopoulos, "Representation of process trends – Part I. A formal representation framework," *Computers & Chemical Engineering*, vol. 14, no. 4–5, pp. 495–510, 1990.

[10] I. M. Cecílio, J. R. Ottewill, J. Pretlove, and N. F. Thornhill, "Nearest neighbors method for detecting transient disturbances in process and electromechanical systems," *Journal of Process Control [Preprint]*, 2014.

[11] F. Ding and T. Chen, "Modeling and identification of multirate systems," *Acta Automatica Sinica*, vol. 31, pp. 105–122, 2005.

[12] D. Li, S. L. Shah, T. Chen, and K. Z. Qi, "Application of dual-rate modeling to CCR octane quality inferential control," *IEEE Transactions on Control Systems Technology*, vol. 11, no. 1, pp. 43–51, 2003.

[13] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis: Part III: process history based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003c.

[14] M. E. Janusz and V. Venkatasubramanian, "Automatic generation of qualitative descriptions of process trends for fault detection and diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 4, no. 5, pp. 329–339, 1991.

[15] B. R. Bakshi and G. Stephanopoulos, "Representation of process trends – III. Multiscale extraction of trends from process data," *Computers & Chemical Engineering*, vol. 18, no. 4, pp. 267–302, 1994.

[16] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 2, pp. 133–146, 2007.

[17] M. R. Maurya, P. K. Paritosh, R. Rengaswamy, and V. Venkatasubramanian, "A framework for on-line trend extraction and fault diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 6, pp. 950–960, 2010.

[18] K. Villez, V. Venkatasubramanian, and R. Rengaswamy, "Generalized shape constrained spline fitting for qualitative analysis of trends," *Computers & Chemical Engineering*, vol. 58, no. 0, pp. 116–134, 2013.

[19] D. C. Montgomery, *Introduction to Statistical Quality Control*. New York: John Wiley & Sons Inc., 2013.

[20] M. Chuah and F. Fu, "ECG anomaly detection via time series analysis," in *Proceedings of the ISPA International Workshops on Frontiers of High Performance Computing and Networking*. Niagara Falls: Springer Berlin Heidelberg, 2007, pp. 123–135.

[21] E. Keogh, J. Lin, and A. Fu, "HOT SAX: Efficiently finding the most unusual time series subsequence," in *Proceedings of the Fifth IEEE International Conference on Data Mining*. Houston: IEEE, 2005, pp. 226–233.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2012, vol. 3.

[23] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, 1997.

[24] A. AlGhazzawi and B. Lennox, "Monitoring a complex refining process using multivariate statistics," *Control Engineering Practice*, vol. 16, no. 3, pp. 294–307, 2008.

[25] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, 1998.

[26] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8-9, pp. 480–502, 2003.

[27] V. Prasad, M. Schley, L. P. Russo, and B. Wayne Bequette, "Product property and production rate control of styrene polymerization," *Journal of Process Control*, vol. 12, no. 3, pp. 353–372, 2002.

[28] D. Li, S. L. Shah, and T. Chen, "Identification of fast-rate models from multirate data," *International Journal of Control*, vol. 74, no. 7, pp. 680–689, 2001.

[29] F. Ding and T. Chen, "Hierarchical identification of lifted state-space models for general dual-rate systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 6, pp. 1179–1187, 2005.

[30] X. Shao, B. Huang, J. M. Lee, F. Xu, and A. Espejo, "Bayesian method for multirate data synthesis and model calibration," *AIChE Journal*, vol. 57, no. 6, pp. 1514–1525, 2011.

[31] N. F. Thornhill, "Finding the source of nonlinearity in a process with plant-wide oscillation," *IEEE Transactions on Control Systems Technology*, vol. 13, no. 3, pp. 434–443, 2005.

[32] H. F. Kaiser, "The application of electronic computers to factor analysis." *Educational and Psychological Measurement*, vol. 20, pp. 141–151, 1960.

[33] S. Valle, W. Li, and S. J. Qin, "Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods," *Industrial & Engineering Chemistry Research*, vol. 38, no. 11, pp. 4389–4401, 1999.

[34] P. J. Huber and E. M. Ronchetti, *Robust Statistics, Second Edition*.   John Wiley & Sons Inc., 2009.

[35] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1049–1075, 1999.

[36] N. F. Thornhill, M. Oettinger, and P. Fedenczuk, "Refinery-wide control loop performance assessment," *Journal of Process Control*, vol. 9, no. 2, pp. 109–124, 1999.