Plant-Wide Oscillation Detection using Multivariate Empirical Mode Decomposition

Muhammad Faisal Aftab[a,*], Morten Hovd[a], Selvanathan Sivalingam[b]

[a]*Department of Engineering Cybernetics, NTNU, Trondheim Norway*
[b]*Siemens AS, Trondheim, Norway*

Abstract

Plant-wide oscillation detection is an important task in the maintenance of large-scale industrial control systems, owing to the fact that in an interactive multi-loop environment oscillation generated in one loop may propagate to the different parts of the plant. In such a scenario, its is required that different loops oscillating due to a common cause and hence similar frequency may be grouped together. In this paper an adaptive method for plant-wide oscillation detection based on multivariate empirical mode decomposition (MEMD) along with a grouping algorithm is proposed. The method can identify multiple oscillation groups among different variables as well as variables with random noise only. The proposed method is also applicable to both non-linear and non-stationary time series where the techniques based on the conventional Fourier analysis are prone to errors. Within each group that oscillate due to a common cause, the method can also indicate the location of the probable root cause of oscillations. The efficacy of the proposed method is established with the help of both simulation and industrial case studies.

Keywords: Multivariate empirical mode decomposition, mode alignment, grouping algorithm, dyadic filter bank property.

*Corresponding author
Email addresses: muhammad.faisal.aftab@itk.ntnu.no (Muhammad Faisal Aftab), morten.hovd@itk.ntnu.no (Morten Hovd)

1. Introduction

The observation that only a third of all industrial controllers give acceptable performance and 30% of all industrial control loops oscillate (Srinivasan et al. (2007)) led to increased research in the field of oscillation detection and root cause diagnosis. The topic is important because oscillations give rise to product variability and greatly impact the profitability of plant operations. Sustained oscillations are also harmful for plant equipment, leading to increased wear and possibly premature failure. There can be thousands of loops running in a plant, and oscillations in any of them can propagate to different parts of the plant owing to underlying interactions and coupling effects. It is therefore important to find and group the variables that oscillate due to a common cause. This will be very helpful in finding a root cause of a specific plant-wide oscillation, as there is little reason for searching for a root cause among variables not affected by the oscillation in question.

Oscillations can be caused due to a variety of reasons, like poor controller tuning, external disturbances and process non-linearities, to name a few Choudhury et al. (2008). Oscillation detection can be subdivided into two categories, namely oscillation detection in individual loops and plant wide oscillation detection. The present work deals with the later subcategory only. The presence of multiple oscillations, noisy measurements along with non-stationary and non-linear effects make plant wide oscillation detection a challenging problem. The salient features of a reliable and robust plant wide oscillation detection method can be outlined as:

- The ability to group different variables oscillating with similar frequency.
- The capacity to detect multiple oscillations (oscillations due to multiple sources) in a variable and to assign each oscillation to corresponding group. This also includes the ability to separate out variables void of any oscillation, thus containing random noise only.
- Applicability to signals with both non-stationary and non-linear effects.

- Adaptability i.e. the minimum possible dependence on any *a priori* assumptions, pre-processing or underlying process dynamics.

- Ability to identify the set of variables/nodes that are closest to the source of the oscillations.

Different attempts have been made in recent years by researchers to address the problem of plant wide oscillation detection using both time and frequency domain methods. Time domain methods require the knowledge of process order and time delays among measurements. Gathering this information typically requires significant effort. Frequency domain methods are insensitive to phase lags, but suffer from the restrictive condition of signal stationarity.

The use of principal component analysis (PCA) for the detection of plant wide oscillations is proposed by Thornhill et al. (2002). In this method, normalized power spectra of different time trends are subjected to dimensionality reduction using PCA and each principal component represents a group oscillating with similar characteristics. Apart from the power spectrum, PCA of the Auto Covariance Function (ACF), which is independent of phase lags, is also shown to give similar performance Thornhill et al. (2002).

Another method based on similarity among the power spectra of different variables is put forward by Tangirala et al. (2005). In this method a correlation index among the power spectra is used to construct the power spectral correlation map (PSCMAP).

Xia et al. (2005) and Xia and Howell (2005) proposed a spectral independent component analysis (ICA) based method where dimensionality reduction and grouping is accomplished by employing ICA, but with a condition that the sources of oscillation are statistically independent.

The limitations of the PCA and ICA based method are highlighted by Tangirala et al. (2007), and another method based on Non-Negative Matrix factorization (NMF) is proposed instead. This method is aimed at overcoming the limitations of the PCA and ICA based methods. In an analogy to the singular value decomposition performed in standard PCA, the proposed method call for the pseudo singular value decomposition (PSVD) to

determine the dimensionality of basis space. The main issues in NMF as highlighted in Tangirala et al. (2007) are initialization of basis functions and the absence of any solid method for basis size determination, though singular value decomposition (SVD) of the spectral matrix is used as a tool for the basis initialization in the work by Tangirala et al. (2007). The NMF based method is powerful tool yet the emphasis is more on determining the basis shape of the linearly independent oscillations in a plant-wide setting and at times it fails to group the variables oscillating with similar frequencies. This aspect will be illustrated by the simulation example later in the paper.

Moreover, A genetic algorithm (GA) based matrix factorisation technique has been adopted by El-Ferik et al. (2012). A method based on intrinsic time-scale decomposition has been proposed recently by Lang et al. (2018).

Almost all the methods used for the plant wide oscillation detection, except Lang et al. (2018), are based on conventional Fourier analysis, in one way or the other. Therefore they are vulnerable to errors when it comes to time trends that are outcome of non-linear and non-stationary processes. In order to overcome these problems a fully adaptive and automated plant wide oscillation detection method based on Multivariate Empirical Mode Decomposition (MEMD) has been proposed in this work.

The MEMD is the extension of the standard univariate Empirical Mode Decomposition (EMD) to higher dimensions. The MEMD likes its predecessor EMD is applicable to both non-linear and non-stationary time series and can adaptively decompose the n-dimensional signal into functions called Intrinsic Mode Functions (IMFs), without any *a priori* assumptions or knowledge about the underlying process. The MEMD has also been used by the authors for detecting non-linearity induced oscillations in individual control loops in the presence of non-stationary effects (Aftab et al., 2017a,b), but the present work is focussed on the plant-wide oscillation detection problem.

An important consequence of MEMD is its mode alignment property where common modes in data are extracted in the same indexed IMFs. This mode alignment property along with a grouping algorithm is exploited to give a robust plant wide oscillation detection method in this work. The

proposed method does not require any pre-processing or assumptions regarding the signal generating process and is both applicable to non-linear and non-stationary time series. The method can handle multiple oscillations and group the variables oscillating due to the same cause. An added advantage is the automatic determination of the oscillation frequency of each group via ACF (Thornhill et al., 2003; Aftab et al., 2017a), that is not possible for existing Fourier based methods. This fact also helps in identifying the presence of harmonics for non-linearity induced oscillations. Furthermore the method is also helpful in identifying the variables associated with the root cause of oscillations.

The paper is organized as follows. Section 2 gives an overview of the multivariate EMD. The mode alignment property and proposed grouping algorithm are discussed in sections 3 and 4, respectively. Section 5 and 6 outline the default parameter settings and the proposed algorithm. Simulation examples are discussed in section 7 followed by industrial case studies (section 8) and conclusions.

2. Multivariate Empirical Mode Decomposition (MEMD)

The multivariate EMD (MEMD) is an extension of the standard univariate EMD. EMD aims at decomposing the input signal into AM-FM modulated components called intrinsic mode functions or IMFs $c_i(t)$ and a bias term or residue $r(t)$.

$$x(t) = \sum_{i=1}^{N} c_i(t) + r(t) \tag{1}$$

where $x(t)$ is the input time series and N is the total number of IMFs. An IMF is defined as a signal having zero mean where the number of extrema and zero crossing must either be equal or at most differ by one (Huang et al., 1998). In order to process signals of higher dimensions, extended versions of the standard EMD, termed Bivariate EMD (Rilling et al. (2007)), Trivariate EMD (Rehman and Mandic (2010)) and Multivariate EMD (Rehman and Mandic (2009)) have been developed . Bivariate and Trivariate EMD

deals with two and three dimensional signals respectively, whereas MEMD is generalization of the same to n dimensions. The primary challenge in extending the univariate EMD to the multivariate case is the calculation of the envelope and the local mean in higher dimensions.

Rilling et al. (2007) proposed that, for the bivariate case, the envelope can be constructed by projecting the signal in multiple directions (represented by k uniformly distributed unit vectors in the complex plane) and by interpolating the extrema of projected signals via spline fitting. The same concept is extended to three and general n dimensions by considering uniformly sampled points on the unit three and n-dimensional spheres respectively. The multiple directions are then represented by the unit vectors (direction vectors) from the center of the n-sphere to these points. Mathematical details can be seen in Rehman and Mandic (2009) and Aftab et al. (2017a).

The multiple projections are used to generate the envelope and its mean. The projection onto each unit vector results in a one-dimensional variable for which it is easy to identify the extrema. Once the projections in each direction are calculated; the (one-dimensional) extrema are sampled and corresponding time indices are recorded. The corresponding time indices for each vector gives its extrema. That is, if an extremum is identified in any projection direction, the full n-dimensional point at that time is recorded as an extremum. The envelope is then generated for these extrema and the process is repeated for multiple directions. Thus multiple n-dimenisonal extrema curves are generated which are averaged to get the mean envelope.

The MEMD method to decompose n-dimensional input signal $x = [x_1, x_2, \ldots x_n]$ into IMFs is summarized below (Rehman and Mandic, 2009).

Step I Set up K direction vectors u^k with $k = 1 \ldots K$ by choosing uniformly distributed points on the n dimensional sphere.

Step II Find the projections $p^k(t)$ of the input signal $\mathbf{X}$ along the direction vectors u^k.

Step III Identify the maxima of projections $p^k(t)$ and corresponding time instants t^k.

Step IV Generate n-dimensional envelope curve $\mathbf{e}^k(t)$ by interpolating $[t^k, \mathbf{X}(t^k)]$.

Step V The n-dimensional mean of envelope curve is then given by

$$\mathbf{m} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{e}^k(t) \tag{2}$$

Step VI Extract n-dimensional detail $\mathbf{d(t)}$ using $\mathbf{d(t)} = \mathbf{X(t)} - \mathbf{m(t)}$.

Step VII Repeat the steps I-VI using $\mathbf{d(t)}$ as input till it fulfils the criteria for an IMF.

Step VIII Calculate the residue $\mathbf{r(t)} = \mathbf{X(t)} - \mathbf{d(t)}$ and iterate the procedure on $\mathbf{r(t)}$ till there are no more IMFs left to be extracted.

3. Mode Alignment Property of MEMD

An important consequence of MEMD is its mode alignment property. By virtue of this characteristics the different common oscillating modes in the multivariate signal are aligned in the same indexed IMFs Rehman and Mandic (2009). This can be explained by the use of following illustrative example.

3.1. Mode Alignment :: Illustrative Example

In order to accentuate the mode alignment property, we consider a multivariate signal of four variables A, B, C and D containing four oscillatory modes (f_1, f_2, f_3, f_4). The signals are generated in such a way that they have some modes common among them. In order to illustrate the robustness of the MEMD to phase difference between different data channels, different initial phase angles are also incorporated. A time varying bias $(b = 0.1t)$ is also added in variables A and B to make the signals non-stationary. The signals composition, variance of added noise and the associated phase angles are given in Table 1 (only phase angles with non-zero value are mentioned). The D variables also contains significantly higher noise level as well.

Table 1: Signal composition (mode alignment example)

Variables↓	f_1	f_2	f_3	f_4	Phase	Bias	Noise var
A	✓	–	✓	✓	$\angle f_4 = \pi/2$	+ve	0.2
B	–	✓	✓	–	$\angle f_3 = -\pi/2$	-ve	0.2
C	✓		✓	–	$\angle f_1 = \pi/4$	–	0.2
D	✓	✓	–	✓	–	–	1.0

The objective is to analyze this tetra-variate signal using MEMD and to observe whether it can align these modes in the same indexed IMFs. The signals and the resultant IMFs from MEMD process are shown in Figure 1.

In total nine (9) IMFs and a residual are extracted. The noise content has been extracted in the first three IMFs with variable D IMFs showing significant noise levels. It can be seen that oscillations common between A, B and C with frequency f_3 are present in the fifth IMF of A, B and C, whereas the f_2 mode, common between B and D, is extracted in the sixth IMF of these variables. Similarly f_1 is present in the eighth IMFs of A, C and D. The mode f_4, present in A and D, being the highest frequency oscillation is extracted in the fourth IMF of A and D variable i.e the first IMF after the noise content. The non-stationary effects arising due to the time varying bias are extracted in residue of variables A and B, as other two variables have no bias. Thus, this example illustrates the ability of the MEMD to align the common oscillatory modes in different signals that can be utilized for the plant wide oscillation detection.

The end effects are reduced using the mirror symmetry as proposed by Rilling et al. (2003). The parameters used for stopping criteria are $\sigma_1 = 0.075, \sigma_2 = 0.75, \alpha = 0.075$ (Rilling et al., 2003)[1].

4. Grouping Algorithm

It has been shown that MEMD is capable of aligning common oscillatory modes in multivariate signal in same indexed IMFs, but in order to have automatic plant wide detection and grouping of different oscillatory

[1]For details of $\sigma_1, \sigma_2, \alpha$ please refer to Rilling et al. (2003)

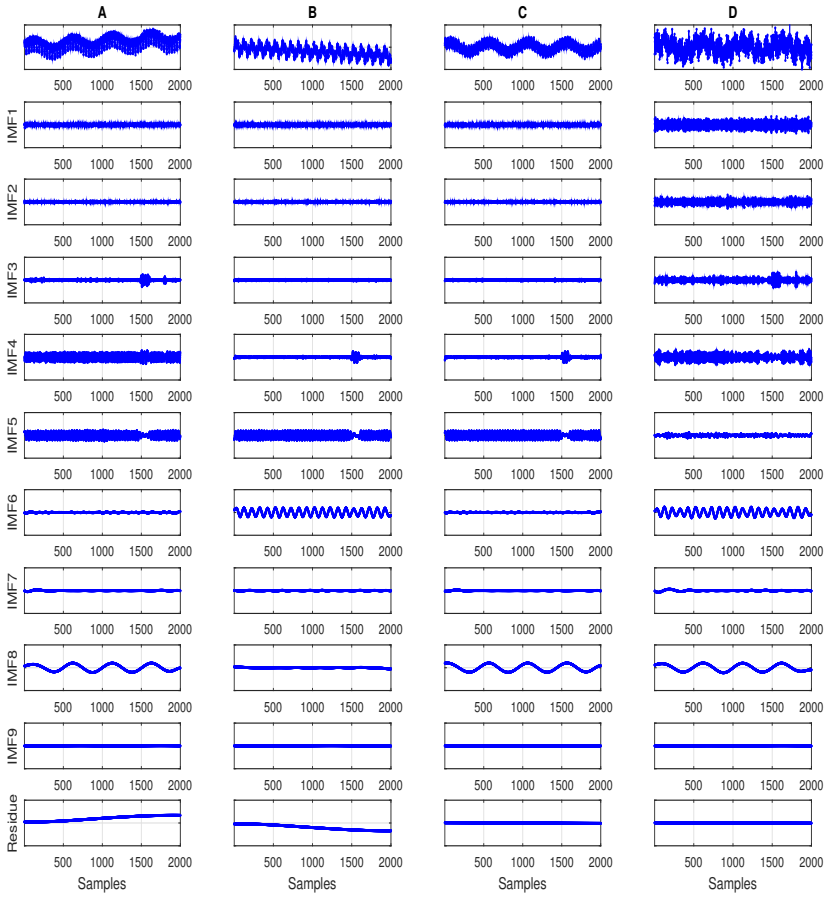

Figure 1: Mode alignment illustrative example

modes based on MEMD, some post processing is required. In this section we propose a grouping algorithm that is used for automatic plant wide oscillation detection.The major steps involved in grouping algorithm are explained next.

4.1. Normalized Correlation Coefficient Matrix

The first step in the algorithm is the calculation of a correlation coefficient matrix Λ. This is required because the MEMD, like the standard EMD process, may also generate spurious IMFs, due to spline fitting issues as highlighted in (Peng et al. (2005), Aftab et al. (2016), Aftab et al. (2017a), Srinivasan and Rengaswamy (2012)) and leakage effects where one principal mode may leak into other IMFs. As the IMFs so generated are nearly orthogonal, these pseudo-components will be poorly correlated with the original signal.

Therefore the correlation coefficient can be used to assess the significance of IMFs as reported in (Peng et al. (2005), Srinivasan and Rengaswamy (2012) and Aftab et al. (2016)). For the j^{th} variables of the multivariate signal $x_j(t)$, the correlation coefficient ρ_{ij} with corresponding i^{th} IMF is calculated from

$$\rho_{ij} = \frac{Cov(c_{i_j}, x_j)}{\sigma_{x_j}\sigma_{c_{i_j}}}, \quad i = 1, 2, 3 \ldots N \tag{3}$$

where Cov is the covariance, $c_{ij}(t)$ is the i^{th} IMF corresponding to j^{th} variable and $x_j(t)$ is the j^{th} variable in input signal; σ_{x_j} and $\sigma_{c_{i_j}}$ are the standard deviations of the j^{th} variable and corresponding IMF i, respectively. N is the total number of IMFs and J is the total number of variables in the multivariate signal.

The normalized correlation coefficient λ_{ij}, for each variable j, is calculated for all N IMFs using

$$\lambda_{ij} = \frac{\rho_{ij}}{max_i(\rho_{ij})}, \quad i = 1, 2, 3 \ldots N \tag{4}$$

The matrix Λ is then constructed from the individual λ_{ij}, in such a way that each row contains the correlation coefficient of same indexed IMF with

j^{th} variable, thus giving N by$J$ matrix of correlation coefficients.

$$\Lambda = \begin{Bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1(J-1)} & \lambda_{1J} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2(J-1)} & \lambda_{2J} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \lambda_{(N-1)1} & \lambda_{(N-1)2} & \dots & \lambda_{(N-1)(J-1)} & \lambda_{(N-1)J} \\ \lambda_{N1} & \lambda_{N2} & \dots & \lambda_{N(J-1)} & \lambda_{NJ} \end{Bmatrix} \tag{5}$$

4.2. Grouping Algorithm

Although owing to the mode alignment property of MEMD all the same indexed IMFs (each row in matrix Λ) represent similar oscillatory modes, yet due to presence of pseudo IMFs (as highlighted in previous section), a grouping algorithm is required, so that only the dominant modes in significant IMFs shall be considered for plant wide oscillation detection.

Thus the proposed grouping technique searches every row (same indexed IMFs for all J variables) for all the correlation coefficients greater than certain threshold, say η. The variables in a row, that are greater than the threshold, are grouped together representing same oscillatory mode. In this way all rows are searched one after another to group common oscillatory modes across all variables. The grouping is accomplished using the procedure outlined in Table 2

In case of multiple oscillations, one j^{th} variable may reside in more than one group, thereby giving the insight into the different oscillatory modes present in each variable.

4.3. Finding Noisy IMFs and Regrouping

Noisy IMFs may have significant correlation with the input signal, in particular for signals comprising of random noise only or with higher noise content. Therefore these noisy IMFs may be grouped together with high frequency oscillatory modes. In order to find the IMFs (identified in the previous step), void of any oscillatory mode, the method proposed by Hoyer Hoyer (2004) and used by Srinivasan and Rengaswamy (2012) is considered here.

Table 2: Grouping algorithm

1. For each row i of matrix Λ, find variables j with $\lambda_{ij} \geq \eta$.
2. Assign the variables with $\lambda_{ij} \geq \eta$ to group G_k.
3. The j variables within the same G_k represent oscillation group with similar frequency.

The process called sparseness index (SI) calculation uses the fact that the noisy signal will have power in a broad frequency range, whereas the oscillatory mode will have dominant magnitude at only one frequency. So the sparseness index of the signal x given by (6) will be approximately zero for the noise signal and will be nearly one for purely oscillatory signals.

$$SI(x) = \frac{\sqrt{N} - \left(\sum |X_i| / \sqrt{\sum |X_i|^2} \right)}{\sqrt{N} - 1} \tag{6}$$

where X is the frequency spectrum of times series $x(k)$ and N is the number of frequency bins.

The SI of the IMFs grouped in each Group G_k is calculated from their frequency spectrum and only those IMFs are retained that have sparseness greater than some threshold γ, while rest are shifted to group G_0 representing noise only.

5. Default Parameter Settings

The grouping algorithm described in the previous section relies on tuning of parameters like number of direction vectors for MEMD and thresholds for the correlation coefficient and sparseness index. This section briefly outlines the motivation for choosing the default settings of these parameters. The

settings are also summarized in Table 3.

5.1. Number of Direction Vectors

The multidimensional envelope for MEMD process is generated using the signal projection in different directions. The number of direction vectors shall be sufficient to allow accurate envelopes to be generated and thus ensure that multivariate sifting process yields consistent results. A too high number of direction vectors gives rise to unnecessary computational load without improving the results much.

It has been observed that results are quite consistent if the number of direction vectors is about four times the number of control loops in question. Increasing the number of direction vectors further does not effect the results much. Therefore the default setting for number of direction vectors used in this work is four times the number of variables in the input signal.

5.2. Correlation coefficient (η)

The grouping algorithm requires the definition of the threshold η for the threshold normalized correlation coefficient. In the work by Srinivasan and Rengaswamy (2012) the threshold used to distinguish the dominant IMFs from the pseudo-components is taken as 0.5. The same threshold is used in this work as the default setting as it ensures that only dominant oscillatory modes are captured. The same threshold is also used in our recent work Andersson et al. (2017). This threshold can be lowered but this will increase the risk getting spurious groupings.

5.3. Sparseness Index Threshold (γ)

Sparseness index is used to distinguish between IMFs with significant noise content and ones with pure oscillations. The signal with higher SI will have a narrower frequency band and therefore the IMFs with SI greater than certain threshold, representing the pure oscillatory modes, are retained. The IMFs with SI lower than the threshold are classified as the ones with significant noise content. As the objective is to group variables having distinct oscillation the threshold used in this work is taken to be 0.58 (same threshold used in previous work Aftab et al. (2017a)). This threshold will ensure

that the signals with narrower frequency band, containing the primary oscillations and the related harmonics (if any) can be captured. The lower value is not desirable as only the dominant and distinct oscillations need to be considered. The threshold can be increased but it may risk some boundary line cases to be excluded.

More detailed explanation about the selection of the said threshold can be seen in Aftab et al. (2017a).

Table 3: Default Parameters Setting

Parameter	Default value
Direction Vectors	4 x number of loops
Correlation Threshold (η)	0.5
Sparseness Threshold (γ)	0.58

6. Proposed Algorithm

The proposed algorithm for plant wide oscillation detection can be divided into the following steps.

Step I Process the multivariate input signal using MEMD.

Step II Calculate the correlation coefficient ρ_{ij} of each i^{th} IMF with corresponding input variable. i.e j^{th} variable of multi-variable input as given in equation (3).

Step III Calculate corresponding normalized correlation coefficient λ_{ij} using relation in equation (4).

Step IV Construct the Correlation Coefficient Matrix Λ according to equation (5).

Step V Search each row of Λ to group similar oscillations using the grouping algorithm given in section 4.2.

Step VI Calculate the sparseness index all IMFs per criteria given in section 4.3.

Step VII IMFs with sparseness less than the given threshold are transferred to 0^{th} group G_0 .

Step VIII If transferring of IMFs as per step VII makes any group empty then re-number the group index k.

Step IX Find the characteristic power spectrum of each group using the member with maximum sparseness.

Step X Calculate the frequency/period of oscillations for each group using IMF with maximum sparseness via ACF method (Thornhill et al., 2003) .

Step XI In case of two or more groups show same frequency, the variables are merged into one group.

7. Simulation Example

In this section the proposed algorithm is applied on the simulation example consisting of tetra variate signal (same signal as discussed in Section (3.1)). The MEMD process generated 9 IMFs and a residue component and the normalized correlation coefficient matrix is given in Table 4.

The elements of Λ that fulfil the grouping criteria in each row are highlighted. The final grouping for the simulation example is shown in Table 6.

All the significant IMFs are checked for sparseness and it is found that first three IMFs of variable D are noisy ($SI < \gamma$) and hence transferred to G_0. In total four oscillatory groups are identified corresponding to four oscillatory modes ($f_1 \dots f_4$). Frequencies $f_4, f_3 \dots f_1$ are arranged in groups ($G_1 \dots G_4$) respectively with highest frequency residing in the first group (as the frequency of IMF decreases with IMF index). Furthermore, the proposed method also successfully highlighted the modes that are common among different variables. Apart from grouping similar modes significant noise component in D is also confirmed.

Table 4: Normalized correlation coefficient matrix (simulation example)

↓ IMF / Variable →	A	B	C	D
1	0.18	0.20	0.16	0.93
2	0.13	0.14	0.14	0.69
3	0.36	0.09	0.12	0.70
4	1.00	0.32	0.34	1.00
5	0.95	0.96	0.97	0.32
6	0.12	1.00	0.10	0.95
7	0.15	0.08	0.03	0.12
8	0.86	0.47	1.00	0.96
9	0.19	0.19	0.11	0.23
Residue	0.63	0.80	0.16	0.24

Table 5: Spectral shape associated with variables (NMF method)

↓ Basis Functions	Variables
w_1	A,C
w_2	B
w_3	D
w_4	A,D

7.1. Simulation Example with NMF Method

The same signals are analysed with NMF method for comparison purposes with the non-stationary bias terms removed from the data for this analysis (since it is known that the NMF method has problems with non-stationary signals). The NMF method, like principal component analysis (PCA) and independent component analysis (ICA), works on the matrix of power spectra and decomposes this power spectra matrix as a sum of p basis functions $w_1^{'} \ldots w_p^{'}$ (Thornhill and Horch, 2007). The basis functions determine spectrum of each cluster.

The results depicting the frequency spectrum of each basis function is shown in Figure 2. Variables associated with different spectral shapes (basis functions) are shown in Table 5. It can be seen that basis functions contains multiple frequency peaks, that are not too different from the indi-

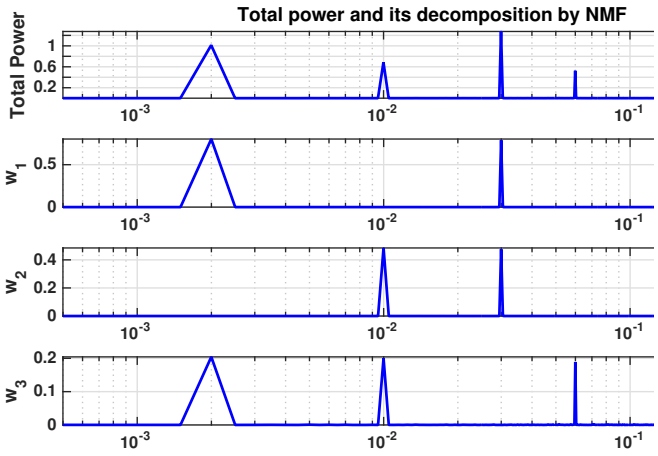

Figure 2: Basis functions from NMF method for simulation example

Table 6: Final grouping simulation example

Variable # ↓	G_0	G_1	G_2	G_3	G_4
A	–	✓	✓	–	✓
B	–	–	✓	✓	–
C	–	–	✓	–	✓
D	✓	✓	–	✓	✓

vidual power spectra of the variables. Moreover grouping based on common oscillations is not so obvious and needs further scrutiny of the basis spectra.

The comparison shows that the proposed method points out different common modes in all loops quite effectively and in an automated manner, while catering for the the non stationary trends in the data as well.

8. Case Studies

8.1. Case Study-I

Time series data from 12 measurements (tags) of the challenge problem (1934 samples with sampling rate 1 minute) (Tangirala et al., 2005) (Thornhill et al., 2002) is analyzed using the proposed scheme. The data is from a pulp manufacturing process simulation where a stream of desired

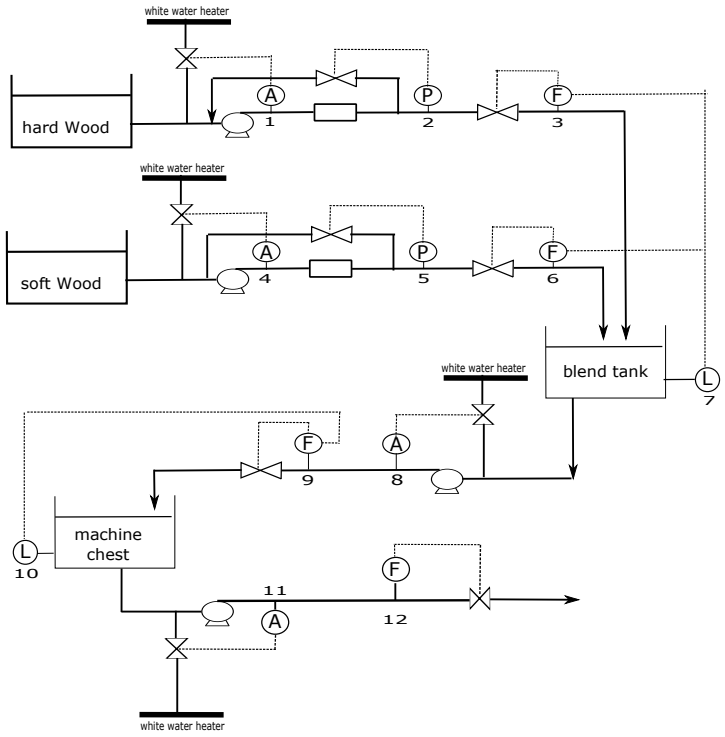

Figure 3: Process schematic case study-I

composition, from the combination of soft and hard wood pulp, is formed (process schematic shown in Figure 3). It is being reported that the trends exhibit two oscillatory modes, one at $0.002min^{-1}$ ($\approx$ 500 mins) and other with higher frequency $0.02min^{-1}$ ($\approx$ 50 mins) (Thornhill et al., 2002). The objective is to process this 12-variate signal (shown in Figure 4) using the proposed method, and to analyze how well the method can group the time trends with similar modes.

The correlation coefficient matrix (5) formed by the proposed method is given in Table 7. The entries highlighted in green are the ones that fulfilled the grouping criteria given in section 4.2.

In total four distinct groups are identified in the final arrangement (Table 8), G_0, with Tag 2 and 5, representing trends with noise and G_1, with Tag11 as the only member, represents presence of a mode with frequency 0.0424 min^{-1}.

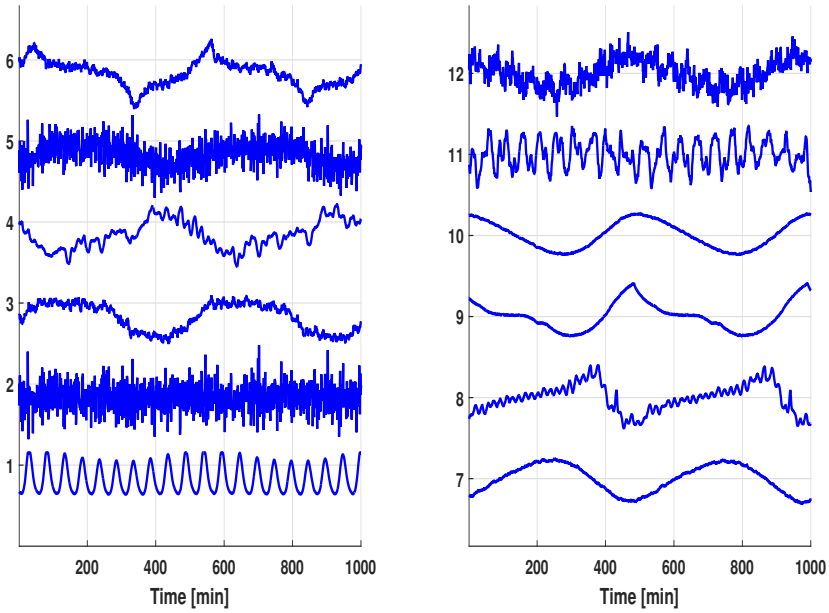

Figure 4: Time trends (normalized) case study-I

G_2 represents oscillation with frequency 0.02 min^{-1} and contains Tags 1 and 11. The group G_3 groups oscillations with frequency 0.004min^{-1} and tag 4 is the only member of this group. The primary oscillation with frequency 0.002min^{-1} is represented by G_4, and Tags 3-10 and 12 exhibit this oscillation. The frequencies and the members of each group are summarized in Table 9. The corresponding power spectra of the each group are shown in Figure 5.

8.1.1. Root Cause Analysis

The results (Table 8) shows that 9 out of 12 tags (Tags 3-10 and 12) are oscillating with frequency $0.002 min^{-1}$ and group G_4 represents this primary oscillatory mode. It is therefore, quite intuitive to search for the root cause of this oscillation.

In case this oscillation arises due to non-linearity some of the tags in G_4 (frequency 0.002min^{-1}) must exhibit higher order harmonics (Thornhill et al., 2001). G_3, coming from IMF 7, contains Tag 4 only. Figure 7 shows

that IMF7 contains the second and third harmonics of the oscillation in IMF8, which is present in G_4. Thus presence of higher order harmonics confirms the presence of non-linearity as the source of oscillation.

Table 7: Corelation coefficient matrix Λ for case study-I

IMF	Tags (j)											
(i)↓	1	2	3	4	5	6	7	8	9	10	11	12
1	0.01	1.00	0.12	0.03	1.00	0.13	0.03	0.05	0.02	0.02	0.11	0.47
2	-0.01	0.75	0.07	0.05	0.80	0.10	0.02	0.03	0.02	0.02	0.14	0.42
3	0.06	0.52	0.11	0.09	0.57	0.11	0.03	0.10	0.01	0.01	0.21	0.36
4	0.25	0.33	0.12	0.22	0.35	0.11	0.01	0.24	0.01	0.03	0.69	0.34
5	0.69	0.19	0.13	0.23	0.17	0.08	0.07	0.22	0.01	0.00	0.88	0.31
6	1.00	0.13	0.10	0.26	0.15	0.19	0.01	0.17	0.06	0.03	1.00	0.42
7	0.17	0.11	-0.02	0.61	0.29	-0.05	0.16	0.36	0.31	0.12	0.20	0.23
8	0.11	0.07	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.28	1.00
9	-0.01	0.07	0.12	-0.01	0.01	0.07	-0.02	-0.13	-0.07	-0.11	-0.03	0.06
10	-0.00	0.08	-0.09	-0.10	-0.04	-0.15	0.01	-0.15	-0.09	-0.03	-0.04	0.06

8.1.2. Advantages of the Proposed Scheme

Apart from the applicability to the non-stationary time series another foremost advantage offered by the proposed method is the level of automation where groups are identified readily and the frequency of oscillation is determined. The IMFs with their particular characteristics can be readily used to determine the frequency of oscillations via ACF method as discussed in Aftab et al. (2017a). This level of automation is hard to achieve in the

Table 8: Final grouping case study-I

Group # ↓	Tag 1	Tag 2	Tag 3	Tag 4	Tag 5	Tag 6	Tag 7	Tag 8	Tag 9	Tag 10	Tag 11	Tag 12
G_0		✓			✓							
G_1											✓	
G_2	✓										✓	
G_3				✓								
G_4			✓	✓	✓	✓	✓	✓	✓	✓		✓

Table 9: Groups and associated frequencies (case study-1)

Group # ↓	Frequency	Tags
G_0	Noise	$2, 5$
G_1	0.0424 min^{-1}	11
G_2	0.02 min^{-1}	$1, 11$
G_3	0.004 min^{-1}+ harmonics	4
G_4	0.002 min^{-1}	$3, 4, 5, 6, 7, 8, 9, 10, 12$

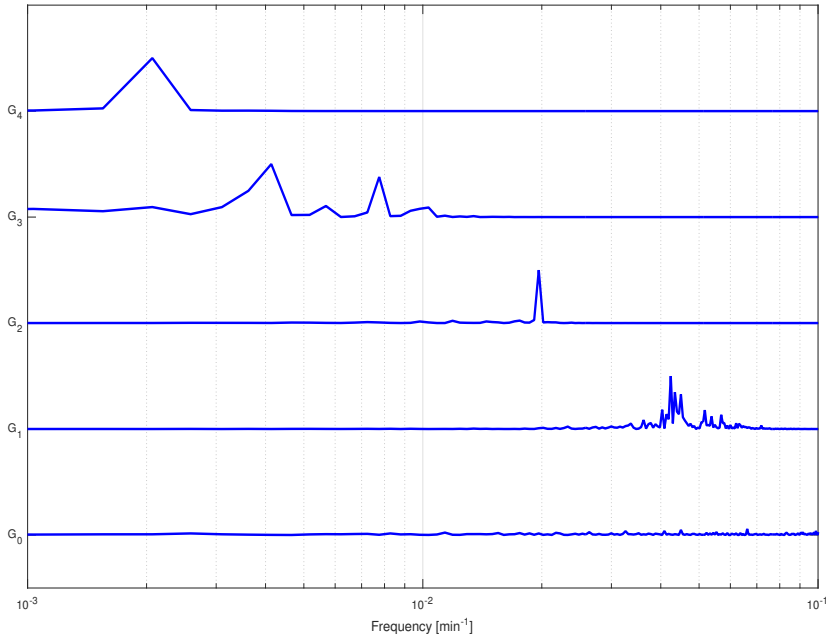

Figure 5: Frequency response of each group case study-I

conventional Fourier based methods like NMF, PCA and ICA. Moreover, this fact also helps in identifying the presence of harmonics, thereby indicating the presence of non-linearity as root cause and assisting in the further diagnosis.

Furthermore, the method is also able to identify the presence of multiple oscillation in Tag 11 with frequencies $0.0424 min^{-1}$ and $0.02 min^{-1}$. The faster mode cannot be captured by PCA (Thornhill et al., 2002) and GA (El-Ferik et al., 2012) methods.

8.2. Case Study-II

We now consider a more complex case study, where time trends from a refinery process are considered. The process is a hydrogen reformer and the same case study is being analyzed by Thornhill et al. (2002); Tangirala et al. (2005) and Tangirala et al. (2007) using PCA, PSCMAP and NMF based methods respectively. The schematic of refinery process is shown in Figure 6. The data consists of 37 Tags (Figure 7) each with 512 data points and sample rate 1 min.

The data trends from 37 tags are processed using the proposed method and the final grouping is given in Table 10. In total 9 groups (G_0-G_8) are identified with G_0 representing random noise. Different frequency modes associated with each group can also be seen in spectra plotted in figure 8.

Modes with low frequency ($< 0.0004 min^{-1}$) are captured in G_8 and contains Tags 5, 21, 22, 31 and 36. Tag 5 also finds its place in the slightly higher frequency group, i.e. G_7, thereby confirming the result reported by Thornhill et al. (2002). The next group G_6 gives the trends with oscillatory mode having frequency around $0.0008 min^{-1}$ and contains Tags 8, 12 and 35-37. Oscillatory mode with frequency around $0.001 min^{-1}$ are represented by G_5 and Tags 7-8, 12, 14-15, 31 and 35-37 are part of this group. Here again Tags 35-37 are part of both G_5 and G_6 correctly indicating presence of more than one modes in these tags. The frequencies around 0.002 are represented by G_4 with Tags 12, 14-18 and 26 as the members.

The biggest cluster is that at frequency $0.06 min^{-1}$ and the oscillations are attributed to valve non-linearity Thornhill (2005). The results show that Tags 1-4, 6-8, 10-11, 13, 15-20, 24-26, 28, 30, 33-34 and 37 exhibit

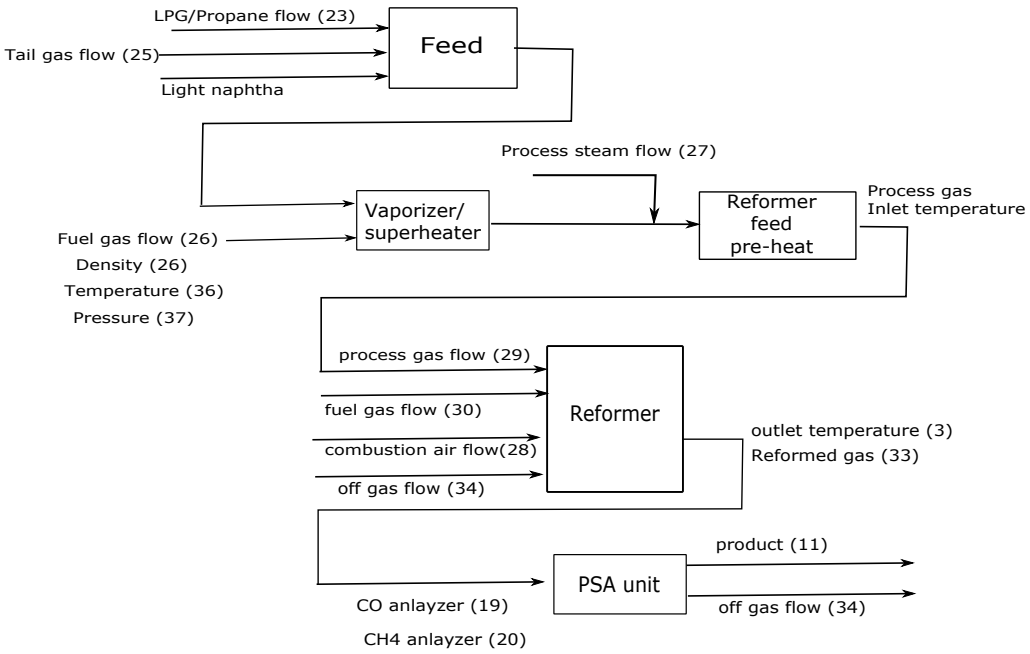

Figure 6: Process schematic case study-II

this oscillatory mode. Tags 2, 11 and 33-34 are shown to have oscillation of frequency $0.12min^{-1}$ and grouped in G_2. Tag 37 with frequency $0.3min^{-1}$ is the only member of group G_1. Group G_0 contains the tags that represent variables with only noise or significant noise contribution. Tag 23, 27 , and 29 are members of G_0 only thereby contain random noise, whereas other members 6-9, 13, 15-16, 24-26 and 30 have significant noise content in addition to the presence of oscillatory modes. A strange inclusion in G_0 is Tag 32, that contain high frequency mode at $0.3min^{-1}$ as reported in Thornhill et al. (2002). This is so because significant noise content in Tag 32 gives the sparseness index of 0.46 and is therefore not included in G_1 along side Tag 37.

8.2.1. Oscillation Diagnosis

The main oscillatory mode in this case study is the one with frequency $0.06min^{-1}$ and represented by group G_3. The majority of the control loops are affected by this oscillation and hence can be considered as the primary oscillation in the process. It is therefore interesting to see if the proposed

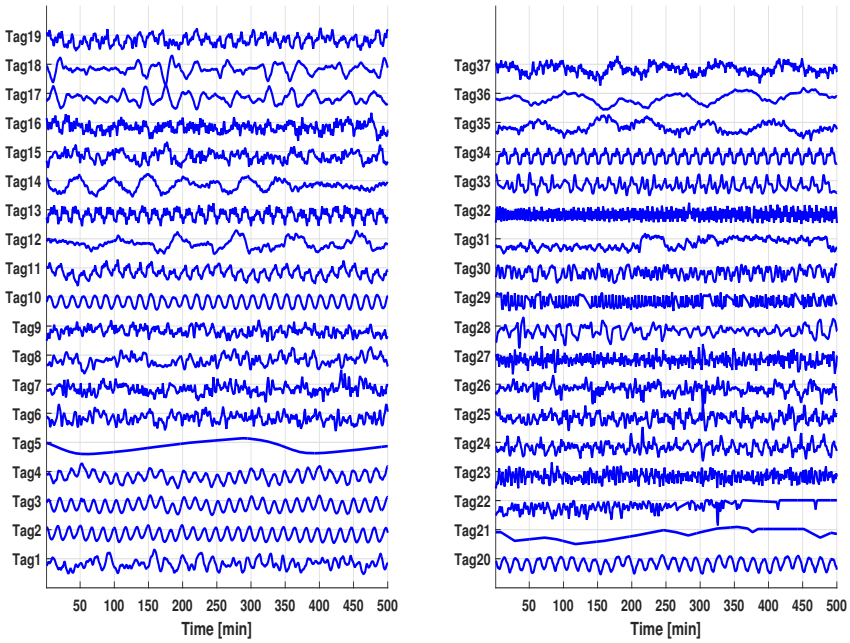

Figure 7: Time trends (normalized) case study-II

method can give some hint about the source of this oscillation. Table 10 shows that group G_2 is subset of G_3 and oscillation (frequency $0.12min^{-1}$) is twice as that of G_3 (Figure 8). Therefore G_2 can be considered as containing the second harmonic of oscillations captured by G_3 and the non-linearity is the source of this oscillation in the plant.

The tags included in G_2 can help in locating the source of this non-linearity induced oscillation. The analysis reported by Thornhill (2005), Choudhury (2006) and Zang and Howell (2007) also shows that Tags included in G_2 exhibit non-linearity signatures. Though Tags 33 and 34 are declared most probable source of this oscillation by Thornhill (2005), and Zang and Howell (2007) respectively.

Therefore the proposed method is able to detect different clusters in the plant data on the basis of dominant modes and at the same time identify the set of variables that are most likely to be the source of the primary oscillation. Clearly, it is recommended that the this analysis is complemented by additional investigations, for example non-linearity analysis (Aftab et al., 2017a; Thornhill, 2005; Choudhury, 2006; Zang and Howell, 2007) to accurately pinpoint the source of the primary oscillation. However, the preceding analysis has significantly reduced the number of tags that need to be analyzed in more detail.

9. Conclusions

In this paper a novel plant wide oscillation detection based on adaptive multivariate data analysis is presented. The method is shown to be robust in identifying dominant oscillations and grouping them together, as illustrated on case studies that have been frequently used in previous literature. The proposed method is capable of handling non-linear and non-stationary time series, where the standard Fourier-based methods are liable to be erroneous. The proposed method is also shown to be capable of indicating the variables that are most likely to be the root cause of oscillation. Furthermore, the method is adaptive in nature and doesn't require *a priori* assumption or pre-processing or any kind of filtering.

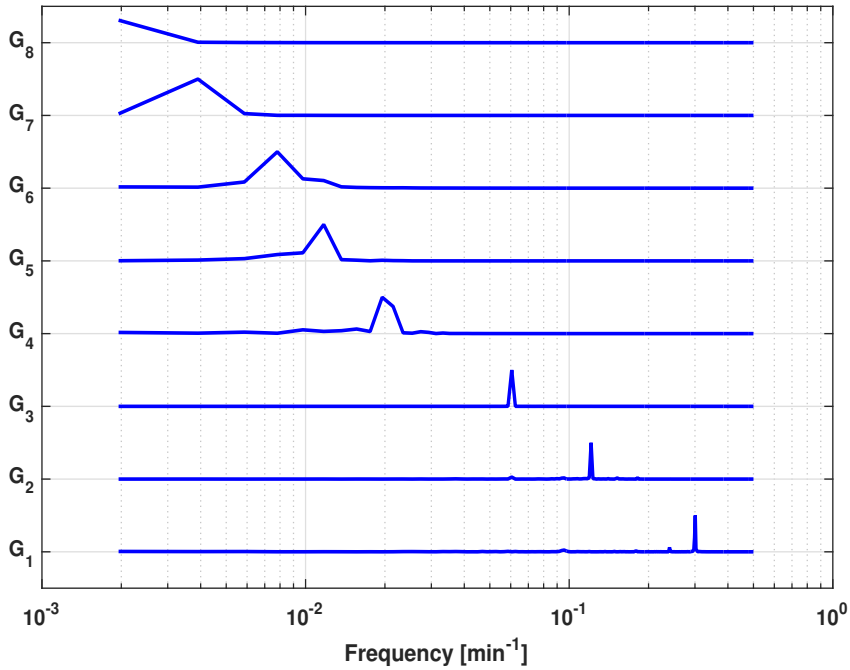

Figure 8: Frequency content of each group (case study-II)

Table 10: Final grouping case study-II

Tag # ↓	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
1				✓					
2			✓	✓					
3				✓					
4				✓					
5								✓	✓
6	✓			✓					
7	✓			✓		✓			
8	✓			✓		✓	✓		
9	✓								
10				✓					
11			✓	✓					
12					✓	✓	✓		
13	✓			✓					
14					✓	✓			
15	✓			✓	✓	✓			
16	✓			✓	✓				
17				✓	✓				
18				✓	✓				
19				✓					
20				✓					
21									✓
22	✓								✓
23	✓								
24	✓			✓					
25	✓			✓					
26	✓			✓	✓				
27	✓								
28	✓			✓					
29	✓								
30	✓			✓					
31									✓
32	✓								
33			✓	✓					
34			✓	✓					
35						✓	✓		
36						✓	✓		✓
37		✓		✓		✓	✓		

10. Acknowledgments

The financial support from Siemens AS,Trondheim is gratefully acknowledged. The authors would like to thank Prof A.K. Tangirala for providing the industrial data.

11. References

Aftab, M. F., Hovd, M., Huang, N. E., Sivalingam, S., 2016. An adaptive non-linearity detection algorithm for process control loops. IFAC-PapersOnLine 49 (7), 1020 – 1025.

Aftab, M. F., Hovd, M., Sivalingam, S., 2017a. Detecting non-linearity induced oscillations via the dyadic filter bank property of multivariate empirical mode decomposition. Journal of Process Control.
URL https://doi.org/10.1016/j.jprocont.2017.08.005

Aftab, M. F., Hovd, M., Sivalingam, S., 2017b. Diagnosis of plant wide oscillations by combining multivariate empirical mode decomposition and delay vector variance. Journal of Process Control(Submitted).

Andersson, L. E., Aftab, M. F., Scibilia, F., Imsland, L., Aug 2017. Forecasting using multivariate empirical mode decomposition; applied to iceberg drift forecast. In: 2017 IEEE Conference on Control Technology and Applications (CCTA). pp. 1097–1103.

Choudhury, A. A. S., Shah, S. L., Thornhill, N. F., 2008. Diagnosis of process nonlinearities and valve stiction: data driven approaches. Springer Science & Business Media.

Choudhury, M. S., 2006. Troubleshooting plantwide oscillations using nonlinearity information. Journal of Chemical Engineering, 50.

El-Ferik, S., Shareef, M. N., Ettaleb, L., 2012. Detection and diagnosis of plant-wide oscillations using ga based factorization. Journal of Process Control 22 (1), 321 – 329.

Hoyer, P. O., 2004. Non-negative matrix factorization with sparseness constraints. The Journal of Machine Learning Research 5, 1457–1469.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. R. Soc. Lond. A 454, 903–995.

Lang, X., Zhang, Z., Xie, L., Horch, A., Su, H., 2018. Time-frequency analysis of plant-wide oscillations using multivariate intrinsic time-scale decomposition. Industrial & Engineering Chemistry Research 57 (3), 954–966.

Peng, Z., Peter, W. T., Chu, F., 2005. A comparison study of improved Hilbert Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. Mechanical systems and signal processing 19 (5), 974–988.

Rehman, N., Mandic, D. P., 2009. Multivariate empirical mode decomposition. In: Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society, p. rspa20090502.

Rehman, N. U., Mandic, D. P., 2010. Empirical mode decomposition for trivariate signals. Signal Processing, IEEE Transactions on 58 (3), 1059–1068.

Rilling, G., Flandrin, P., Gonçalves, P., Lilly, J. M., 2007. Bivariate empirical mode decomposition. Signal Processing Letters, IEEE 14 (12), 936–939.

Rilling, G., Patrick, F., Paulo, G., 2003. On empirical mode decomposition and its algorithms. IEEE-EURASIP workshop on nonlinear signal and image processing 3, 8–11.

Srinivasan, B., Rengaswamy, R., 2012. Automatic oscillation detection and characterization in closed-loop systems. Control Engineering Practice 20 (8), 733–746.

Srinivasan, R., Rengaswamy, R., Miller, R., 2007. A modified empirical mode decomposition (emd) process for oscillation characterization in control loops. Control Engineering Practice 15 (9), 1135–1148.

Tangirala, A., Shah, S., Thornhill, N., 2005. PSCMAP: A new tool for plant-wide oscillation detection. Journal of Process Control 15 (8), 931–941.

Tangirala, A. K., Kanodia, J., Shah, S. L., 2007. Non-negative matrix factorization for detection and diagnosis of plantwide oscillations. Industrial & engineering chemistry research 46 (3), 801–817.

Thornhill, N., Shah, S., Huang, B., 2001. Detection of distributed oscillations and root-cause diagnosis.

Thornhill, N. F., 2005. Finding the source of nonlinearity in a process with plant-wide oscillation. Control Systems Technology, IEEE Transactions on 13 (3), 434–443.

Thornhill, N. F., Horch, A., 2007. Advances and new directions in plant-wide disturbance detection and diagnosis. Control Engineering Practice 15 (10), 1196–1206.

Thornhill, N. F., Huang, B., Zhang, H., 2003. Detection of multiple oscillations in control loops. Journal of Process Control 13 (1), 91–100.

Thornhill, N. F., Shah, S. L., Huang, B., Vishnubhotla, A., 2002. Spectral principal component analysis of dynamic process data. Control Engineering Practice 10 (8), 833–846.

Xia, C., Howell, J., 2005. Isolating multiple sources of plant-wide oscillations via independent component analysis. Control Engineering Practice 13 (8), 1027–1035.

Xia, C., Howell, J., Thornhill, N. F., 2005. Detecting and isolating multiple plant-wide oscillations via spectral independent component analysis. Automatica 41 (12), 2067–2075.

Zang, X., Howell, J., 2007. Isolating the source of whole-plant oscillations through bi-amplitude ratio analysis. Control Engineering Practice 15 (1), 69–76.