

Non-Negative Matrix Factorization for Detection and Diagnosis of Plant-Wide Oscillations

Arun K. Tangirala¹, Jitendria Kanodia¹ and Sirish L. Shah^{*2}

¹*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai 600036 India*

²*Department of Chemical & Materials Engineering, University of Alberta, Edmonton AB T6G 2G6 Canada*

Abstract

In this paper, we propose the use of Non-negative Matrix Factorization (NMF) of multivariate spectra for plant-wide oscillation detection. One of the key features of NMF is that it provides a parts-based representation which allows us to retain the causal basis spectral shapes or parts that constitute the spectra of measurements unlike the popular PCA-based methods. The contributions of this paper are as follows: (i) A novel measure known as the Pseudo-Singular Value (PSV) to assess the order of the basis space. The PSV is also useful in determining the most dominant features of a data set. (ii) A power decomposition plot which contains the total power (defined in this work) and its decomposition by NMF. The power plot is a useful and compact visual tool that provides overall spectral characteristics of the plant and shows the decomposition of these characteristics into well-localized frequency components. (iii) A novel measure defined as the Strength Factor (SF) to assess the strength of the localized features in the variables. It can be also used in isolating the root cause. Finally, it is shown that the proposed implementation of NMF is powerful and sensitive enough to capture small oscillations in the measurements. As a result, it largely eliminates the need for filtering the data. Industrial case studies are presented to illustrate the applications of NMF and to demonstrate the utility and practicality of the proposed measures.

Keywords: non-negative matrix factorization, pseudo-singular value, strength factor, total power, oscillations, multivariate spectral analysis, information-fusion.

1 Introduction

Detection of oscillatory signals on a plant-wide nature has generated widespread interest in the academic community over the last decade.^{1,2,3,4,5,6,7} The presence of oscillations in control loop measurements significantly affect control loop performance; oscillatory disturbances can propagate plant-wide and often force plants to back off from optimal operating conditions.⁸ Moreover, oscillations can result in off-spec products that directly impact the productivity of the plant.

Since oscillations can be conveniently characterized in the frequency domain, spectrum-based methods^{4,1,9} have emerged as popular tools in this area.

The objectives of multivariate spectral analysis in this context are primarily, (i) to find the dominant oscillations in the measured data (ii) find set(s) of variables which share these features

^{2*} Author to whom all correspondence should be addressed. Email: sirish.shah@ualberta.ca

and (iii) assess the relative strengths of these oscillations in those variables. The two important properties of spectra are that they contain only non-negative entries and that spectra of linearly related variables are represented by additive combinations of spectra. *Thus, the fact that the basis vectors and weights both have to be non-negative is a natural requirement for any multivariate spectral analysis tool.*

In the literature, these issues have been addressed to varying degrees using PCA of spectral data,¹ analyzing univariate and/or multivariate auto-covariance functions^{10,9} and Independent Component Analysis (ICA) of the spectral data.¹¹ However, each of these methods have their limitations and do not contain a full exploitation of the multivariate nature of the data. Some of the key limitations are:

- PCA of spectral data results in orthogonal basis shapes that can contain a mixture of positive and negative orientations. The orthogonality of basis vectors in PCA is achieved by allowing cancellations and additions of the features in the basis with the effect that there is no control either over the sign of the weights (combination coefficients) or the elements (entries) of basis vectors. Thus, the basis shapes attained with PCA of spectral data are not physically meaningful. In other words, PCA of spectral data cannot be expected to give basis vectors with multiple peaks lying in the same quadrant¹. An unfortunate consequence of this fact is that one cannot interpret the individual basis shapes but instead look at them in the p -dimensional space (p is the number of principal components), which is extremely restrictive in its use and also places a natural hindrance for automation of this method.
- The auto-covariance function is a measure of the dependency of successive samples in a time-series. The ACF of a periodic signal is also periodic with the same period, while the ACF of a white noise sequence is only non-zero at lag zero. As a consequence, the ACF of a signal corrupted with white-noise is unaffected at all non-zero lags. The estimation of the period of a sampled signal involves the calculation of zero-crossings. However, the applicability of this method is limited due to one or more of the following reason:
 - The noise corrupting the data can be coloured, thereby affecting the ACF of the measurement at lags other than zero as well. Consequently, the zero-crossings do not represent the periodic behaviour of the measurement.
 - Detection of low frequency or large period oscillations may require large amounts of data.
 - Presence of multiple oscillations in a single measurement complicates the computation of zero-crossings.
 - The ACF-based method is univariate in nature.
- ICA-based methods impose the condition that sources are statistically independent, which is a useful mathematical property for the isolation of sources. Although the ICA-based methods and their variants overcome the drawback of the PCA-based method, it is hard and often not possible to verify the statistical assumptions on the sources.
- In general, the required number of principal or independent components has been determined by ignoring the cross-term in the spectral relationships among variables. The commonly used SVD (defined below) gives the correct number of linear relationships only when the frequency

¹A transformation matrix exists that relates the basis vectors from PCA to the desired spectral shapes (non-negative shapes), but a practical method remains far from reach¹²

components in each variable are orthogonal and/or non-overlapping, *i.e.*, the cross-term is zero. The singular-value decomposition of a matrix $X_{N \times M}$ is defined as:

$$X = USV^T$$

where U is an $N \times M$ column-orthogonal matrix, S is an $M \times M$ matrix of singular values and V is an $M \times M$ orthogonal matrix. The squared singular values of X are the singular values (or the eigenvalues) of the covariance matrix $X^T X$.

The cross-term involving the common frequency components arises when one converts a linear relationship among variables in time-domain into the corresponding relationship in the spectral domain. For example, if a variable x_3 is linearly related to two other variables x_1 and x_2 , $x_3 = a_1 x_1 + a_2 x_2$, then

$$\begin{aligned} X_3(\omega) &= a_1 X_1(\omega) + a_2 X_2(\omega) \\ \longrightarrow |X_3(\omega)|^2 &= a_1^2 |X_1(\omega)|^2 + a_2^2 |X_2(\omega)|^2 + 2a_1 a_2 |X_1(\omega)| |X_2(\omega)| \cos(\phi(\omega)) \end{aligned}$$

where $\phi(\omega)$ is the phase difference between $X_1(\omega)$ and $X_2(\omega)$ at the frequency ω .

In the final equation above, the term $2a_1 a_2 |X_1(\omega)| |X_2(\omega)| \cos(\phi(\omega))$ is the cross-term. This term is zero if (i) either $|X_1(\omega)|$ or $|X_2(\omega)|$ (non-overlapping) or both are zero, or (ii) $X_1(\omega)$ and $X_2(\omega)$ are orthogonal, *i.e.*, $\phi(\omega) = \pi/2$.

In general, due to the presence of the cross-term and the non-negativity constraint, it is likely that SVD yields only a lower bound on the number of basis shapes since it relies on orthogonal unconstrained decomposition. The issue of basis size determination has also remained an open-ended in the NMF literature.¹³

- In summary, there is a need for a simple and effective tool for the detection and isolation of plant-wide oscillations. This tool should also have a feature that requires minimal human intervention, that is the methodology can be easily automated.

In view of the above issues, it is proposed to use Non-Negative Matrix Factorization (NMF)¹⁴ for the detection and isolation of plant-wide oscillations. NMF does not make any statistical assumptions on the sources, but only relies on the non-negativity constraint on the basis spectrum, which is a natural property of the power spectrum. Thus, NMF can handle a larger class of data sets that could contain effects of dependent sources. In this respect, NMF is therefore, a more general method. Simple and effective tools based on NMF results are shown to provide valuable insights into the spectral characteristics of the measurements.

Non-negativity of basis shapes in the context of image analysis was addressed in the work by Lee and Seung.¹⁴ In their work, Lee and Seung¹⁴ re-formulated the problem of basis vector decomposition of a matrix X , which contains only non-negative entries, into a product of matrices of non-negative basis vectors and therefore the weights as well. The resulting method attained popularity under the term ‘Non-Negative Matrix Factorization’ (NMF). The ideas in [14] were related to a certain extent to the idea of Positive Matrix Factorization (PMF) introduced by Paatero.¹⁵ NMF is a very important factorization in linear algebra and is widely used in environmental science, spectroscopy and chemometrics (see [16],[17]). NMF has been applied particularly in image analysis, classification and pattern recognition where a physical interpretation of the basis vectors (as well as the weights) was essential to the understanding of the underlying phenomena.^{14,18}

Initialization of NMF algorithms is an open-ended problem in the NMF literature.^{19,13} The algorithm is guaranteed to converge for any feasible random initialization, but yields a different

set of basis vectors with each run. In this work, this issue is addressed by using the absolute basis vectors from the SVD in contrast to the random initialization. Such a choice has two advantages. Firstly, SVD gives the right shape (possibly with sign reversed) when there is one dominant frequency and secondly, it has been observed to invariably capture the most dominant shapes.

NMF only provides an approximate non-negative matrix factorization and does not carry any explicit information on the rank of the matrix being analyzed. As a result, the determination of the number of basis shapes is specific to each application and hence a uniform measure is lacking. A consequence of this fact along with the fact that the NMF involves the solution to a non-convex optimization problem has been that there are no specific rules about the order in which the basis shapes are obtained. In other words, permutation is an issue in NMF-based methods. Both these issues are addressed in this paper by the use of a measure known as the Pseudo-Singular Value (PSV), which is defined in Section 3.3. The PSV is similar to the singular values obtained in SVD in that it gives an idea of the amount of power captured by each basis shape relative to the total power present in the data set. The total power is defined as the sum of the (normalized) power over all variables at each frequency. A plot of the total power vs. frequency presents a compact picture of the frequency-domain behaviour of the plant. It is an effective tool in plant-wide oscillation detection as will be illustrated in Section 5.

NMF essentially breaks down the total power into power components that are localized in a narrow frequency range. The computation of each of these power components is provided in Section 3.2 and illustrated in Sections 4.2 and 5. The basis and weights obtained from NMF offer useful insights into the features of the data set but there is a need to further simplify and condense the information present in these quantities given that the number of basis vectors and weights can increase with the size of the data set. With this motivation, an index defined as the *Strength Factor (SF)* is proposed, which provides a true assessment of the oscillation strength captured by each basis shape in each variable, and the relative strengths of the power components in each variable. The strength factor can then be used in conjunction with process knowledge to diagnose the root-cause of oscillations.

Three industrial case studies are illustrated to highlight the usefulness and effectiveness of the NMF method in achieving the objective of multivariate spectral analysis. A comparison with the traditional spectral PCA method is also illustrated in one of the case studies.

The paper is organized as follows. Section 2 gives a brief introduction to the concept of NMF and the common Euclidean distance-based algorithm. In Section 3, the concept of total power and its decomposition along with the notion of pseudo-singular value is presented. Subsequently in section 4, the strength factor is developed and the proposed ideas are illustrated on a simulation example. Applications to three industrial case studies are presented in Section 5. The paper ends with a few concluding remarks in Section 6.

2 Non-negative Matrix Factorization (NMF)

PCA is a popular projection (or dimensionality reduction) method in which the linear correlation among variables is unmasked by expressing the columns of the data matrix as a linear combination of orthogonal basis vectors typically in a lower dimensional space. The projections are achieved by decomposing X as

$$X \approx TP \quad \text{or} \quad X = TP + E \tag{1}$$

where X is an $N \times M$ data matrix ($M \ll N$), P is an $r \times M$ matrix and T is an $N \times r$ matrix. N represents the number of observations and M represents the number of variables or measurements, while r is the dimensionality of the reduced space. The matrix P provides a reduced set of basis vectors for the row space while the matrix T provides a reduced set of basis vectors in the column (*variable*) space. A special property of these basis vectors is that they form an orthogonal basis set. The dimensionality reduction is achieved by a rank r approximation of the matrix X using singular value decomposition (SVD) of X . Typically an eigenvalue analysis of the covariance matrix is carried out in place of SVD of X for computational ease. Popular measures that are used to choose the order of the basis space r are the 80% variance, the scree Test and the 1-eigenvalue criteria, all of which depend on the eigenvalues of the covariance matrix. The approximation error is contained in E , which is zero if $r = M$ or if the last $(M - r)$ singular values of X are all identically zero. The PCA approximation is optimal in the sense that the 2-norm of the error is minimized.

Despite the simplicity and the ease of computations in PCA, it is a fact that there is no control over the sign of the entries of the quantities T and P . This limitation arises due to the lack of a constraint on the sign in the computation of SVD of X . In addition, the orthogonality requirement in PCA results in a very distributed representation that uses cancellations to capture variability. Therefore, PCA is useful when the objective is merely to obtain a set of orthogonal basis vectors. However, it has serious disadvantages in applications where meaningful interpretation of the basis vectors as well as the weights is vital for further analysis. In the problem of detecting plant-wide oscillations X is the matrix of spectra, with the effect that entries of the matrix X are *all non-negative*. In such situations, it is useful and meaningful to have basis vectors and weights that are also non-negative

Non-negative matrices arise in several other areas such as multivariate image analysis, pattern recognition and multivariate spectral analysis. One of the key objectives in these applications is to obtain a parts-based representation of an object, image or a face; or the objective can be to obtain the basis spectral shapes or parts that constitute the spectra of measurements so that measurements with common spectral features can be clustered together. As argued above, PCA (and also several other related methods) do not provide the means for a parts-based representation.

In order to preserve the special structure (such as non-negativity), Lee and Seung¹⁴ proposed the Non-negative Matrix Factorization (NMF) scheme, where a matrix with non-negative elements is *approximated* as a product of two matrices B and W

$$X_{N \times M} \approx B_{N \times r} W_{r \times M} \quad \text{or} \quad X_{N \times M} = B_{N \times r} W_{r \times M} + E \quad (2)$$

such that both B and W contain non-negative entries. A rank-reduced approximation of X is thus,

$$X_{i\mu} \approx \sum_{k=1}^r B_{ik} W_{k\mu} \quad (3)$$

where $X_{i\mu}$ represents the $i = 1 \dots n$ elements of the μ^{th} column of X , B_{ik} represents the $i = 1 \dots n$ elements of the k^{th} column of B and $W_{k\mu}$ represents the k^{th} element of the μ^{th} column of W . Typically, B and W are rank-reduced matrices of sizes $N \times r$ and $r \times M$ respectively.

Depending on the context of application, B and W take the roles of the basis vectors and the coefficients or weights (co-ordinates) of linear combinations. If X is the matrix of spectra arranged in columns, the columns of B provide the *basic spectral shapes*, while the matrix W give the

coefficients or *weights* of combination².

It is useful to observe that the decomposition in equation (2) is similar to the decomposition in PCA as given in equation (1). A major difference lies in the fact that the basis vectors in PCA are orthogonal and distributed, while in NMF the basis shapes are non-orthogonal but well-localized. This fact was demonstrated by¹⁸ in image analysis, where PCA gave rise to basis images that are distorted versions of the original images, while NMF basis images carried localized features that related to the intuitive notions of the parts of the original data. *In this respect, for the plant-wide oscillation detection problem, the basis shapes from NMF can be expected to bear resemblance to the original spectra in its basis vectors whereas PCA can often give rise to distorted basis shapes.*

2.1 Algorithms for NMF

Solutions to non-negative matrix factorizations involve iterative optimization techniques, requiring optimization of a cost function that describes the “closeness” of BW to X . Two primary algorithms based on two different cost functions were first proposed by.¹⁴ These algorithms are based on iterative searches with rules for updates at every iteration, details of which can be found in [19]. The type of update rule is dependent on the nature of cost function that is involved. One of the most popular cost function involves the use of Euclidean distance between X and its approximation BW . Since this algorithm is used in this paper, a brief review is provided below. The reader may refer to^{19,20,16,21} for the second cost function (an entropy-based measure) and the subsequent variants of these two algorithms.

Euclidean Measure

The cost function is based on the Euclidean distance between two matrices A and \hat{A} ,

$$\|A - \hat{A}\|^2 = \sum_{ij} (A_{ij} - \hat{A}_{ij})^2$$

The lower bound of this measure is clearly zero, and occurs if and only if $A = \hat{A}$.

The objective function is to:

Minimize $\|X - BW\|^2$ with respect to B and W , such that each element of B and W respectively, $\{b_{ik}\}, \{w_{kj}\} \geq 0$.

Note that the above cost function is convex in only one of B or W , but not convex in both variables together. Therefore, it is hard to find an algorithm that can reach the global minima. However, many numerical optimization techniques exist that can be applied to find local minima. Two multiplicative update rules exist which can at least guarantee a locally optimal solution.

These multiplicative update rules are stated below. Proofs for the convergence of these algorithms can be found in.¹⁹

Update for Euclidean measure

1. Initialize B and W to two random non-negative matrices.
2. Update B and W until the termination condition is achieved. The multiplicative update

²The spectra is arranged in columns so that it is easy to relate to the NMF literature.^{16,18,13,17} This arrangement is in contrast to the convention used in chemometric literature; however, the choice of this arrangement has no bearing either on the analysis or the results.

rule to be used is:

$$W_{aj} \leftarrow W_{aj} \frac{[B^T X]_{aj}}{[B^T B W]_{aj}} \quad \text{and} \quad B_{ka} \leftarrow B_{ka} \frac{[X W^T]_{ka}}{[B^T W W^T]_{ka}}$$

where W_{aj} refers to the a^{th} row of W and B_{ka} refers to the a^{th} column of B , where $a = 1 \cdots r$; r is the dimension of the reduced space.

Note that when $X = BW$, the multiplicative factors are unity, which implies that perfect reconstruction is necessarily a fixed point of the update rules.

Proofs of the convergence of the above mentioned multiplicative rules make use of an auxiliary function similar to that used in the Expectation-Maximization algorithms and are provided in.¹⁹

The final solution is almost always a local minimum since global optimum is seldom achievable. In the work by Donoho and Stodden,²² the authors establish conditions under which NMF can give unique solutions in the context of image analysis from a geometric standpoint.

Normalization The power spectrum of each variable is normalized such that the sum of the spectrum (over all frequencies) is unity. The zero frequency component is also removed by mean-centering the data. This is identical to the zero-mean, unit-variance standardization of the time-domain data. In NMF, the basis shapes are normalized to possess unit 1-norm so as to enhance their resemblance with the power spectra. Besides, it is a necessary step in the algorithm to ensure that the final solution is a meaningful approximation.

3 Issues and improvements in NMF implementation

As aforementioned in Section 1, there are two key issues associated with the solution to NMF of X . These issues are (i) initialization and (ii) choice of basis size. Each of these issues is discussed below. Although the discussion is in the context of engineering applications, the ideas can be applied to other situations as well.

3.1 Initialization of NMF Algorithms

Initialization of NMF algorithm remains an open problem to date (see [23], [24]). Random initialization of B and W to matrices with positive elements is commonly used in practice and is guaranteed to provide convergence to a local optimum. Since the final solution is generally a local optimum, the basis vectors can differ from one run to the other with random initialization, which means that different initial conditions can lead to different final basis set with the same terminating criterion. Another disadvantage with random initialization is that it does not guarantee the capture of the most dominant shapes in the data set. In the work by Wild *et al.*,²³ a structured initialization is done using the spherical K -means clustering algorithm. The given data is clustered into K spherically disjoint centroids and the centroid vectors are used as an initial estimate of the basis vectors. A limitation of the use of this technique is that it requires a knowledge of the number of centroids, *i.e.*, basis vectors.

In this work, it is proposed to use absolute values of the basis vectors given by SVD to initialize the NMF algorithm, *i.e.*, $B(0) = |T|$. The matrix W is also initialized in a similar way, *i.e.*, $W(0) = |P|$.

The motive behind using SVD for initialization is that when the number of basis shapes equals one, the absolute values of the basis shape and weights given by SVD are the desired basis shape and weights.

In the authors' experience with the application of NMF to industrial case studies, it has been observed that the initialization of NMF algorithms with the proposed method has led to the capture of the most dominant shapes. A theoretical analysis to verify this observation would be useful but beyond the scope of this discussion.

3.2 Total power and NMF decomposition

In order to proceed to the discussion on the determination of basis size, it is useful to understand the concept of total power and its decomposition by NMF.

Denote a set of N regularly sampled observations of the j^{th} process measurement as x_j ($j = 1, \dots, M$) and $X_j(f_l)$, $l = 0, \dots, \frac{N}{2} - 1$ as its *normalized* power spectrum at $N/2$ frequencies $f_l = l/N$. Then the total power at each frequency is defined as

$$X_T(f_l) = \sum_{j=1}^M X_j(f_l) \quad (4)$$

The maximum value of $X_T(f_l)$ is M , the number of process measurements. The total power is a good measure of the overall spectral behaviour in the plant. Since peaks at common frequencies in the measurements add up, large spectral peaks in the plot of total power vs. frequency indicate that there could be several variables oscillating together at that frequency (frequencies). A single peak close to or equal to M implies the presence of a single dominant oscillation in the plant, while a flat spectrum would indicate no dominant oscillations in any of the variables.

The NMF of the spectral matrix $X = [X_1 \ X_2 \ \dots \ X_M] \approx BW$ can be interpreted as a decomposition of the total power X_T by each basis shape. Assuming that an r -dimensional approximation is sought, the power component due to the k^{th} basis shape is given by

$$C_k = \sum_{j=1}^M B_k W_{kj} \quad k = 1, \dots, r$$

where B_k is the k^{th} column of B , representing the k^{th} basis shape and W_{kj} is the weight of the j^{th} variable corresponding to B_k . Observe that each C_k is a spectrum-like column vector with the size equal to the number of frequencies.

Under ideal conditions, it would be expected that $X_T = \sum_{k=1}^r C_k$. The squared Euclidean distance between the X_T and C_k paves way for the definition of the Pseudo-Singular Value for the NMF.

The interpretation of a multivariate spectral analysis tool decomposing the total power into its respective constituents is a novelty of this work. The decomposition is enabled by NMF only because of the non-negativity constraints on the basis and weights. The interpretation lays the foundation for defining a ‘singular value’ like measure for the NMF and assessing the strength of different oscillations in a measurement corrupted with noise.

3.3 Pseudo-Singular Values for NMF

Choosing the size of the basis space remains an unresolved issue to date. Unlike in PCA, there exists no “optimal” rule for NMF-based decompositions. Therefore, the size of the basis space is determined either by trial and error or using some heuristics. A generally used rule-of-thumb for choosing r is such that $r < \frac{NM}{M+M}$, where N represents the no. of observations, while M represents the number of measurements. In the work by,¹⁷ the authors use a heuristic method, wherein the dimensionality is chosen by comparing the root-mean-square error obtained with NMF and SVD methods as the dimensionality is increased. Another commonly used heuristic method is to continue to increase the basis space dimension until one begins to see a repetition of basis vectors, *i.e.*, when the basis shapes appear to look alike. However, there is no guarantee that an over-specification of the basis size necessarily lead to repetition of basis shapes, particularly in the presence of noise.

In this work, a novel measure is introduced to determine the basis size. This measure is termed as the Pseudo-Singular Value (PSV) and is based on the *incremental* variance captured by each power component relative to the total power. *The incremental calculation takes into account the absence of the orthogonality constraint on the basis shapes* and also allows one to rank the contribution of each basis shape.

PSV of the j^{th} basis shape is denoted by ρ_j and is defined as,

$$\begin{aligned} \rho_j &= \frac{\left\| \sum_{k=1}^j C_k \right\|_2^2 - \left\| \sum_{k=1}^{j-1} C_k \right\|_2^2}{\|X_T\|_2^2} \quad j = 2, \dots, M \\ &= \frac{\|C_1\|_2^2}{\|X_T\|_2^2} \quad j = 1 \end{aligned} \tag{5}$$

where C_k is the k^{th} power component defined in equation (5) and X_T is the total power defined in equation (4).

The PSV is similar to the (normalized) singular values obtained from the SVD of the covariance matrix in the sense that its value indicates the amount of power captured by the corresponding basis shape, and hence the name.

PSV always lies between 0 and 1 as can be seen from equation (5). A value of ρ_j close to or equal to unity implies that the corresponding basis power component describes the total power X_T completely (or a major part of it), while values of ρ_j close to zero imply that those basis shapes have negligible power components (the basis shapes by themselves can have significant amplitudes but the weights can be negligible).

Remarks:

- The definition of PSV above gives rise to the perspective of a “basis power component”, which is more appropriate to use in spectral analysis rather than the “basis shape”.
- The PSV is a measure of contribution of a power component to the total power. Using this measure, the power components are ordered in the decreasing order of contribution.
- By definition, *it follows that the fraction of variance (of the spectral matrix) captured by r basis shapes of NMF is identical to the sum of the corresponding r PSVs.*

- The procedure for calculating PSV involves computation of all the power components at once. If the first r ordered power components are sufficient to explain the total power in the data set, then the remaining PSVs of the last $(M - r)$ power components are zero.
- In practice, however, the lowest value can be negligibly small due to the presence of noise in the data and/or the non-linear relationships between the spectra. Besides, overlapping (or repetitive) basis shapes can also force some or all of the $(M - r)$ PSVs to take non-zero values. In such a case, the first r components are referred to as the dominant components.
- Power components with negligible PSVs are not considered for analysis in practice, with the effect that they correspond to either noise or power components of negligible strength. If a priori information indicates a need for analysis of frequencies present in these residual (weak/neglected) power components, then such an analysis can be done without altering any of the other results.
- The weak components/noise/harmonics are captured by the $(M - r)$ components. This alleviates the need for filtering to a large extent (see Section 5).
- Although the definition of PSV in equation (5) is dependent on the order in which the power components are obtained, a two-stage calculation of PSV makes it quite robust to this order. In the first stage, the PSVs are calculated for the components obtained in the order given by the NMF algorithm. In the second stage, the components are re-ordered with respect to the one with the highest value and the PSVs are recalculated. The component(s) with the highest PSV(s) are the most robust ones with respect to the order in which PSV is calculated.

4 Analysis of NMF results

Once the “optimal” B , W and basis power components are obtained, what remains is an appropriate interpretation of these quantities to perform a measurement-level analysis. The desired features of this quantification are that it should:

- Have the ability to capture the “strength” of the power component in presence of noise by suitably neglecting the effects of noise
- Be able to rightly assess the relative strengths of the different power components that may be present in a measurement.
- Possess simplicity, be effective and be able to quantify this information in a compact form.

The weights by themselves are a good reflection of the strength of each power component in a variable, but can fail to meet the first requirement due to the fact that the presence of noise can reduce the relative amplitude of an oscillation in the spectrum of that measurement. Consequently, this results in a “small” weight relative to those which have the same oscillation but possess good SNR leading one to erroneously believe that the particular measurement has a weak oscillation.

With this motivation, a new measure known as the Strength Factor is presented in the following section.

4.1 Strength Factor

The underlying idea is analogous to the idea of total power. *NMF can be interpreted to break up the power spectrum of a measurement into its constituent components.*

After the PSVs are calculated and the components are ordered, the basis shapes and the corresponding weights are re-ordered accordingly. Then, the strength of the k^{th} power component in the j^{th} measurement is defined using the idea of increments,

$$\begin{aligned}
 SF_{kj} &= \frac{\left\| \sum_{p=1}^k B_p W_{pj} \right\|_2^2 - \left\| \sum_{p=1}^{k-1} B_p W_{pj} \right\|_2^2}{\|X_j\|_2^2} \quad k = 2, \dots, M \quad \forall j \\
 &= \frac{\|B_1 W_{1j}\|_2^2}{\|X_j\|_2^2} \quad k = 1 \quad \forall j
 \end{aligned} \tag{6}$$

where X_j is the power spectrum of the j^{th} measurement. The strength factor can be interpreted analogously as PSV.

The strength factor always lies between 0 and 1. Unity value of SF_{kj} implies that the k^{th} power component entirely captures the spectral behaviour of the j^{th} measurement, while a zero value indicates that the k^{th} power component has no presence in that measurement.

The following remarks are due in view of the definition given above.

Remarks

- A useful property of the strength factor is that the $\sum_{k=1}^r SF_{kj} \leq 1 \forall j$. When the equality is satisfied, it means that the r power components have completely explained the spectral behaviour of that measurement.
- SF can distinguish between those measurements possessing small spectral peaks due to the presence of noise and those measurements which primarily contain noise (see Tags 5 and 2 of the Entech case study in Section 5).
- When $\sum_{k=1}^r SF_{kj}$ yields values close to unity, then very little is left to be explained by the remaining $M - r$ power components. The remainder could be noise or very weak oscillatory behaviour.

A condensed plot of the strength factors is a useful visual tool because of its effectiveness and compactness. Once the strength factors are obtained, the variables can be ranked according to their “strengths” in the direction of each power component.

4.2 Simulation Example

The purpose of this section is to illustrate the proposed ideas on a simple example. The superiority of NMF over using the traditional PCA in multivariate oscillation detection is also exemplified.

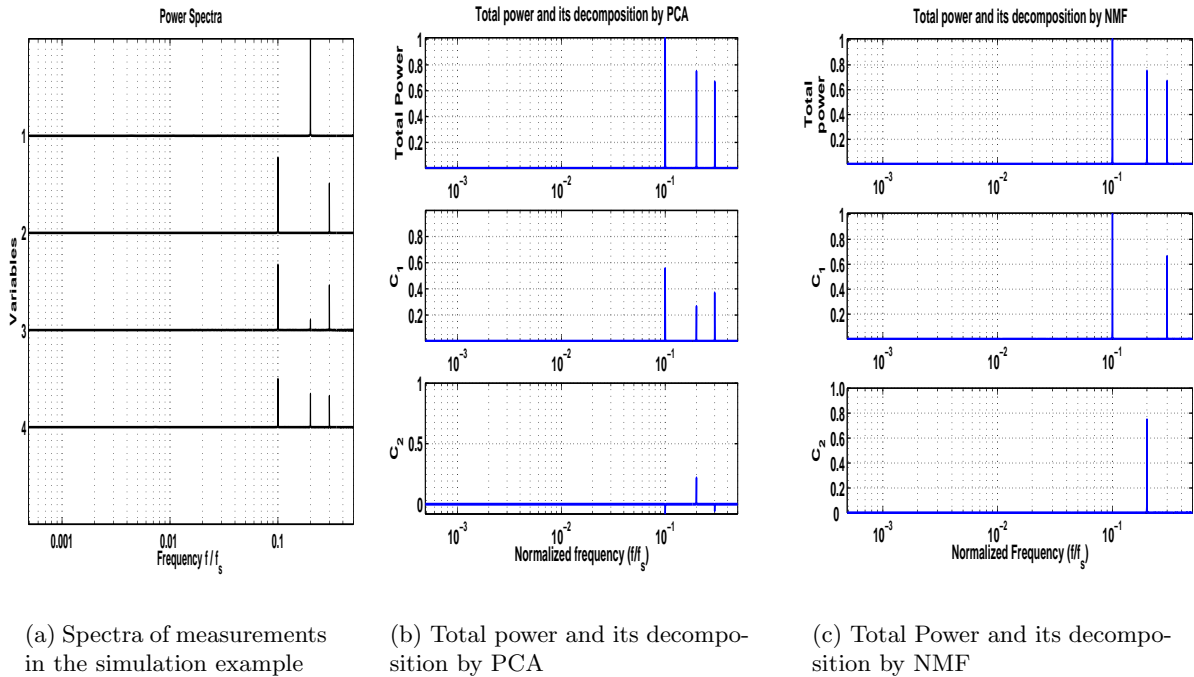


Figure 1: Comparison of the power decomposition obtained from PCA and NMF respectively

The illustrative example consists of four signals x_1 , x_2 , x_3 and x_4 . Each of these signals are generated as

$$\begin{aligned}
 x_1[n] &= \sin(2\pi(0.2)n) \\
 x_2[n] &= \sin(2\pi(0.1)n) + \sin(2\pi(0.3)n) \\
 x_3[n] &= x_1[n] + 2x_2[n] \\
 x_4[n] &= x_1[n] + x_2[n]
 \end{aligned}$$

Measurements of each of these signals is assumed to be corrupted with uncorrelated Gaussian noise such that the SNR value is 10.

The power spectral matrix is formed as $X = [X_1 \ X_2 \ X_3 \ X_4]$ where X_j is the normalized power spectrum of the j^{th} measurement.

The NMF algorithm was initialized with the absolute values of the eigenvectors and weights from SVD. SVD of the spectral matrix yielded $\sigma = [0.57 \ 0.33 \ 0 \ 0]$. Therefore, SVD indicates the presence of two basis vectors.

For the NMF analysis, the pseudo-singular values obtained were $\rho = [0.69 \ 0.31 \ 0 \ 0]$. These values clearly suggest that two spectral components can explain the total power in the data set.

The total power decomposition by NMF is shown in Figure 1(c) with the total power shown at the top, and the *power components* for $k = 1, 2$ extracted by NMF shown below it. It is clear that NMF has effectively separated the two power components without any ambiguity. Observe that the total power (less the insignificant amount of noise) can be obtained by adding up the two power components. It also highlights the physical significance of the basis shapes obtained by NMF.

In contrast, the results from PCA shown in Figure 1(b) cannot be given any such interpretation.

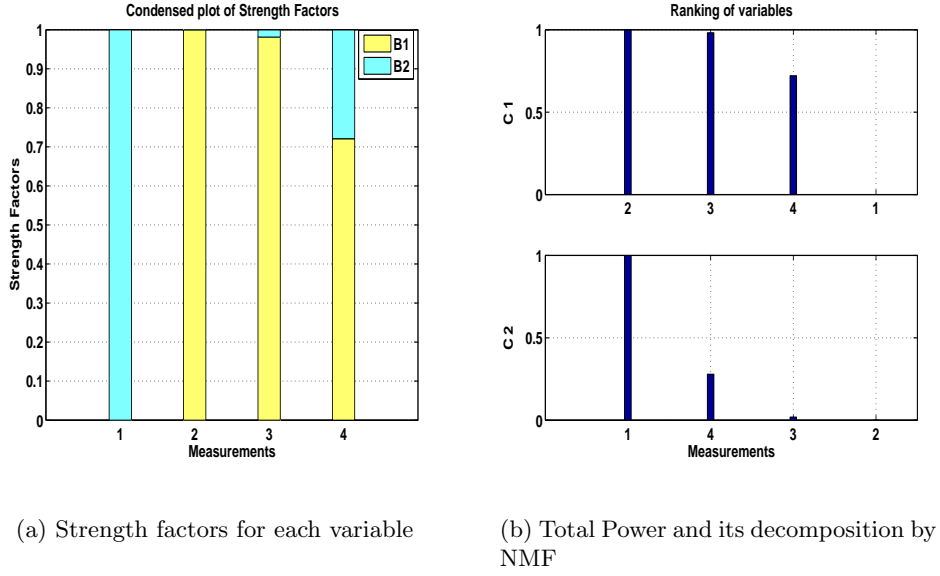


Figure 2: Strength factors and the ranking of measurements for each power component

The power components not only lack any meaning (with the exception of the first one), but also do not add up to produce the total power present in the noise-free part of the signal. *In order to produce the noise-free total power, one would require more than two power components. This is the key limitation of PCA.* Thus, PCA fails to provide a platform for separating the power components in a meaningful and compact way.

The relative strengths of the power components in Figure 1(c) are given by the strength factors, which are plotted in Figure 2(a) using a bar plot³. The interpretation of this plot is useful for subsequent case studies.

Each bar in Figure 2(a) corresponds to a measurement and is a stacked bar of strength factors corresponding to each power component. Each power component is represented by a single colour. *The extent of colour in each bar represents the extent of the corresponding power component in that measurement.* The first component (containing 0.1 and 0.3 cycles/sample) and the second component are denoted by cyan and yellow colours respectively. The bar plots for measurements 1 and 2 confirm their frequency content. Bars 3 and 4 contain colours in proportion to the “amount” of the respective components. Another important observation to make is that $\sum_{k=1}^2 S_{kj} = 1 \forall j$ (each bar touches unity), which means that the spectral behaviour of the data set has been entirely captured by the two power components.

Finally, the measurements are ranked according to their strengths in Figure 2(b). Ranking is done by first dividing each S_{kj} , $k = 1, 2$ by the maximum for each k , that is, after setting the strength of the strongest measurement to unity for each component⁴. NMF has clearly succeeded in doing a correct analysis of the spectral data.

³A gray scale can be also used to illustrate this plot, however a colour plot is chosen for better visualization

⁴Normalizing this way for ranking purposes brings about a uniformity for each component.

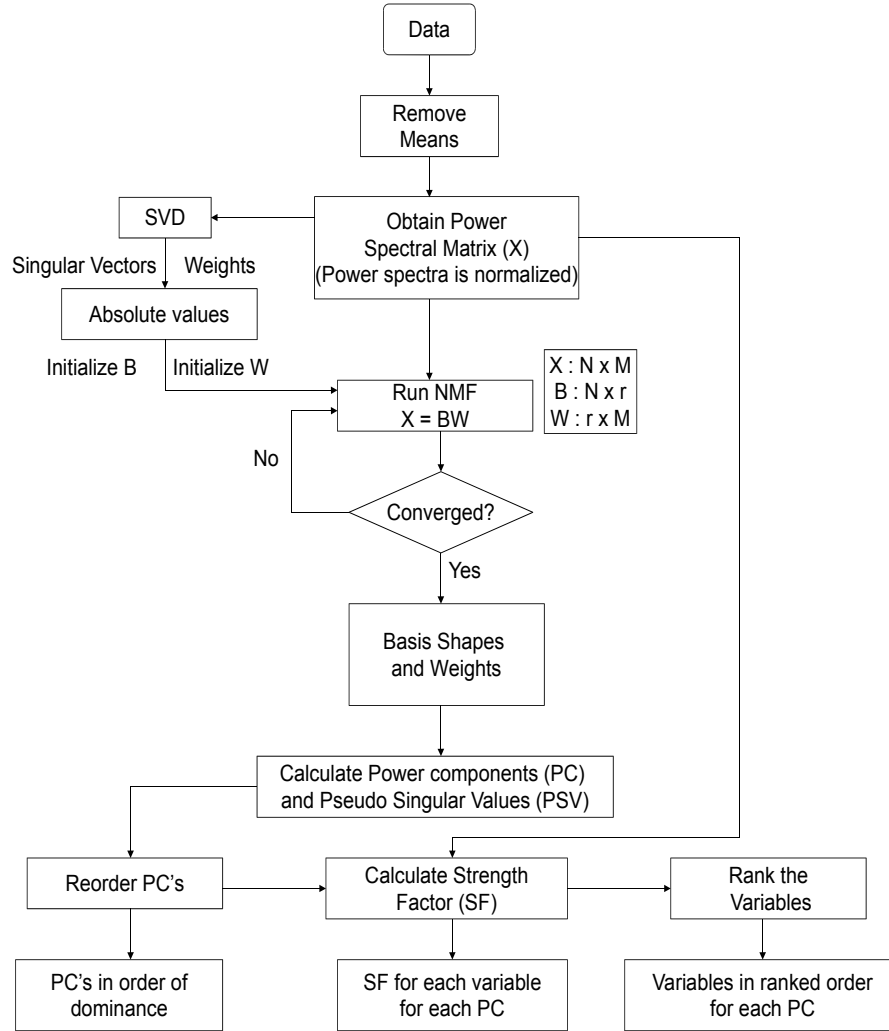


Figure 3: A general procedure for using NMF in multivariate spectral analysis

4.3 A general procedure

A general procedure for applying NMF to routine operating data is provided in Figure 3. The following points are useful in practice while using this algorithm:

- the first r dominant shapes can be chosen depending on the application and the quality of data. This is a user specification. As remarked in Section 3.3, the choice of r does not alter the course of analysis. It should be noted that the purpose of choosing “dominant” components is to first analyze the significant oscillations.
- The residual component analysis (not shown in the figure) is useful to confirm the presence of noisy measurements, weak frequencies and harmonics, if any. This step alleviates the need for filtering to a large extent and also provides insights when the objective is to study the plant for the presence of frequencies.

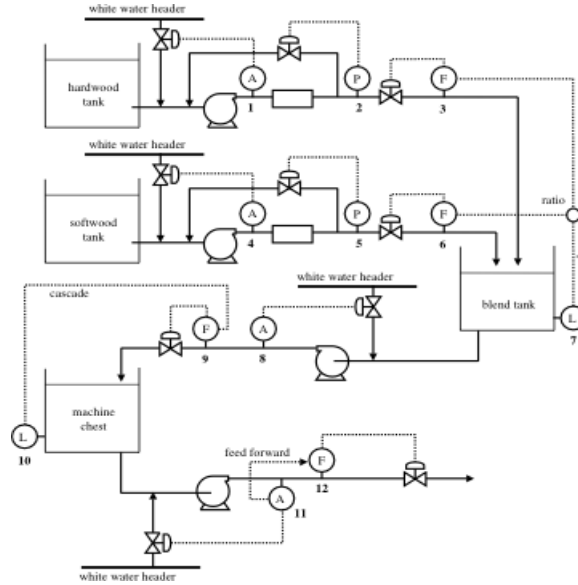


Figure 4: Schematic of the simulated process of Entech Control Inc.

5 Industrial Case Studies

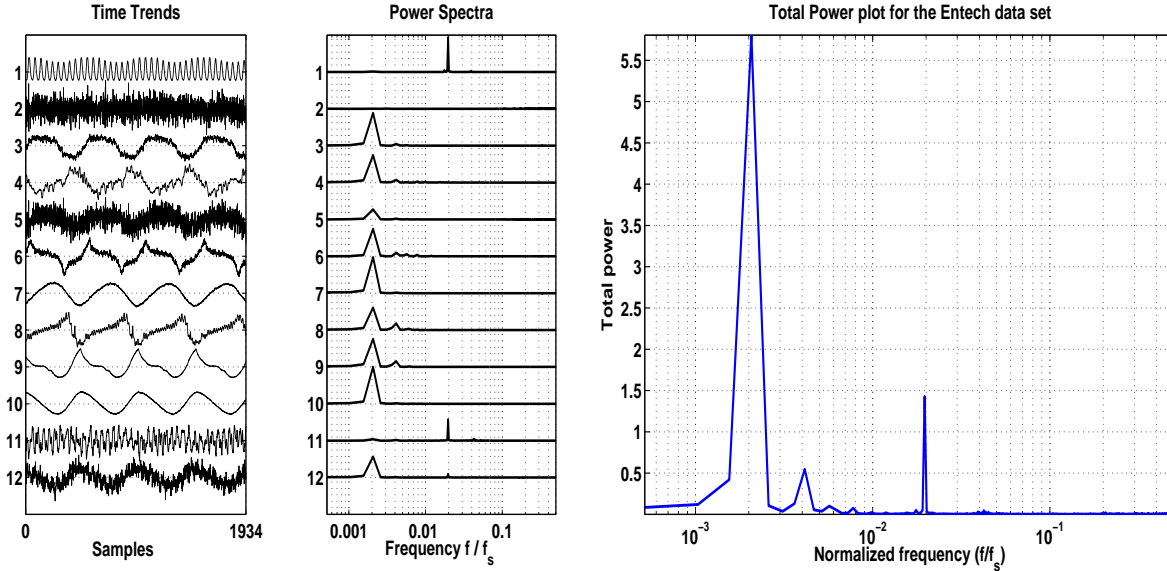
In this section, three case studies are taken up to demonstrate the potential of NMF and the proposed ideas in plant-wide oscillation detection and isolation. Two of these case studies have already been documented in earlier works on the use of PCA.¹ They are taken up here to show the relative ease with which one can extract the same information, thereby showing the advantages of NMF over PCA. The third data set has not been analyzed before and is the largest data set among the three sets with complex dynamics.

5.1 Case Study 1: Entech Data

The data set is from a simulated industrial process, courtesy of Entech Control Inc. The simulated process shown in Figure 4 consists of a pulp manufacturing process, where the hardwood and softwood pulps are mixed to give a stream of desired composition. The data set comprises 1934 samples from 12 process measurements (tags) associated with 12 control loops. The objective of this analysis is to detect oscillations in the loops and isolate those loops with common oscillations. A high density plot (HDP)¹ of the time trends and the corresponding spectra of the 12 *controller errors* (SP - PV) is shown in Figure 5(a).

Total Power Plot vs. High Density Plot

The first exercise is to obtain a plot of the total power X_T , which is shown in Figure 5(b). The plot clearly shows that there are three dominant frequencies whereas an individual scrutiny of the variables from the HDP was observed to contain two major peaks in [1]. This point exemplifies the effectiveness of the total power plot over the HDP when the objective is to obtain a collective picture of the multivariate process. In fact, the total power plot and HDP complement each other well. PCA of the spectral data was reported to capture most of the information in the spectra using two basis vectors.¹ The ambiguity of including the third basis vector could be easily resolved using the total power plot.



(a) High Density Plot of the simulated Entech data

(b) Total power plot for the Entech data set

Figure 5: High Density Plot and the Total Power Plot for the Entech data

5.1.1 NMF Analysis

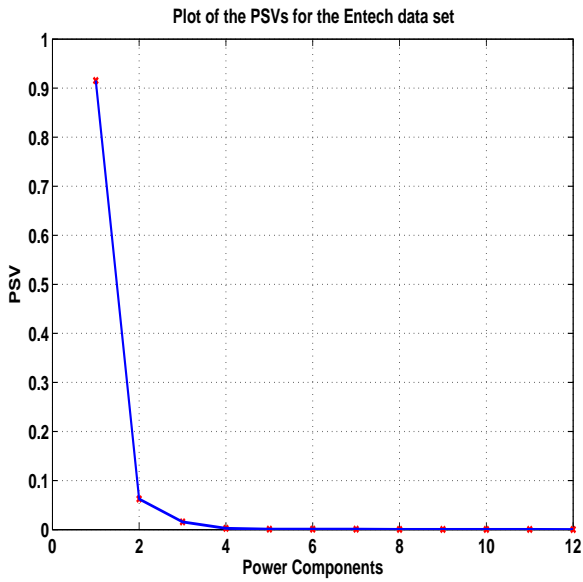
SVD of the spectral matrix was performed to initialize the NMF algorithm. A plot of the PSVs is shown in Figure 6(a). It can be seen that values beyond the third PSV are virtually zero. Therefore three PSVs are chosen, which is also in agreement with the observations from the total power plot.

The resulting power decomposition shown in Figure 6(b) clearly demonstrates the ability of NMF to decompose the total power into its respective components. Besides, it is also confirmed that the PSVs serve as a true indicator of the dominant power components by rightly arranging the power components in their order of dominance.

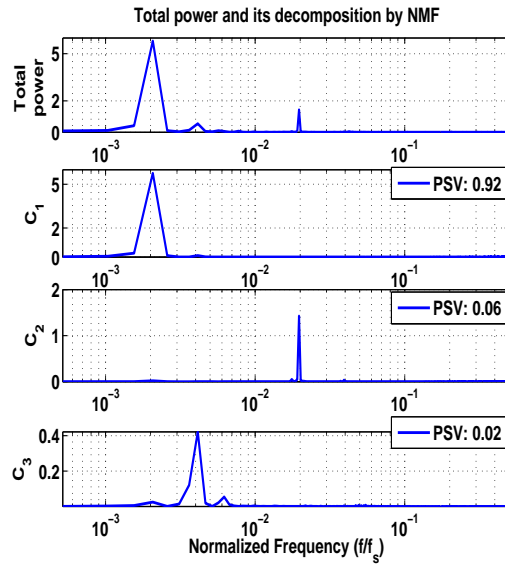
5.1.2 Distribution of total power in measurements

Figures 7(a) and 7(b) show the strength factors and the ranking of each variable with respect to the three power components. They clearly bring out the ability of the proposed measures in identifying, quantifying and ranking the different frequency components in a multivariate data set. The robustness of strength factor over the weight needs a mention here. For tag 5, the weight value corresponding to the first basis shape (containing the 0.002 cycles/sample) was 0.2608, leading us to believe that tag 5 has a weak oscillation of 0.002 cycles/sample. But the fact is that tag 5 primarily contains this oscillations but is also corrupted by relatively high amounts of noise, which makes the oscillation appear weak. However, the strength factor is able to detect this fact and therefore it rightly assigns a value of 0.96 to tag 5.

It should be remarked that NMF has extracted the first harmonic of 0.002 cycles/sample is obtained directly without the need for filtering, as was done in [1] and [6].

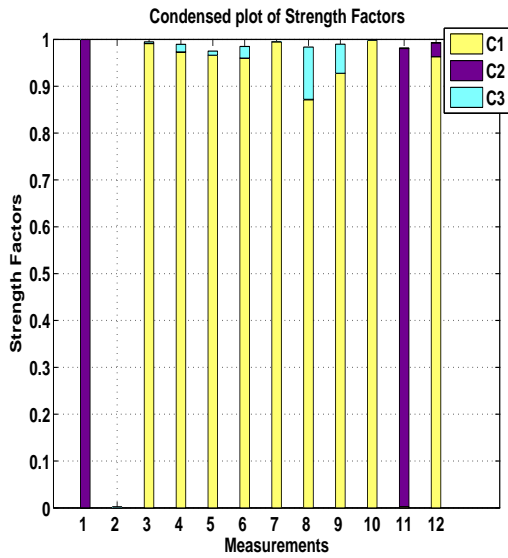


(a) Pseudo-Singular Values of the spectral data set

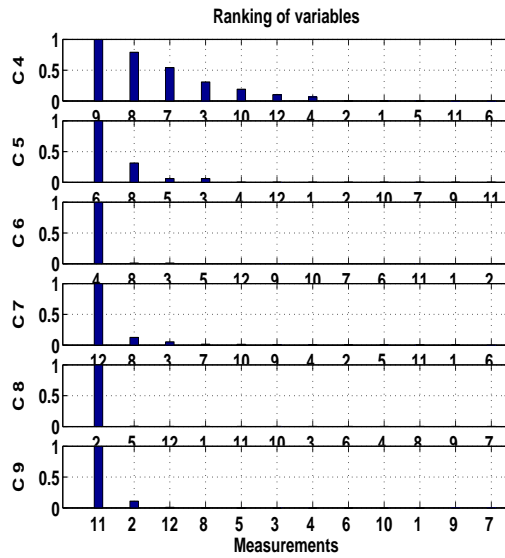


(b) Decomposition of the total power by NMF

Figure 6: PSV plot and the power decomposition provided by NMF



(a) Strength Factors



(b) Ranking of measurements

Figure 7: Strength Factors and ranking of measurements from the Entech process

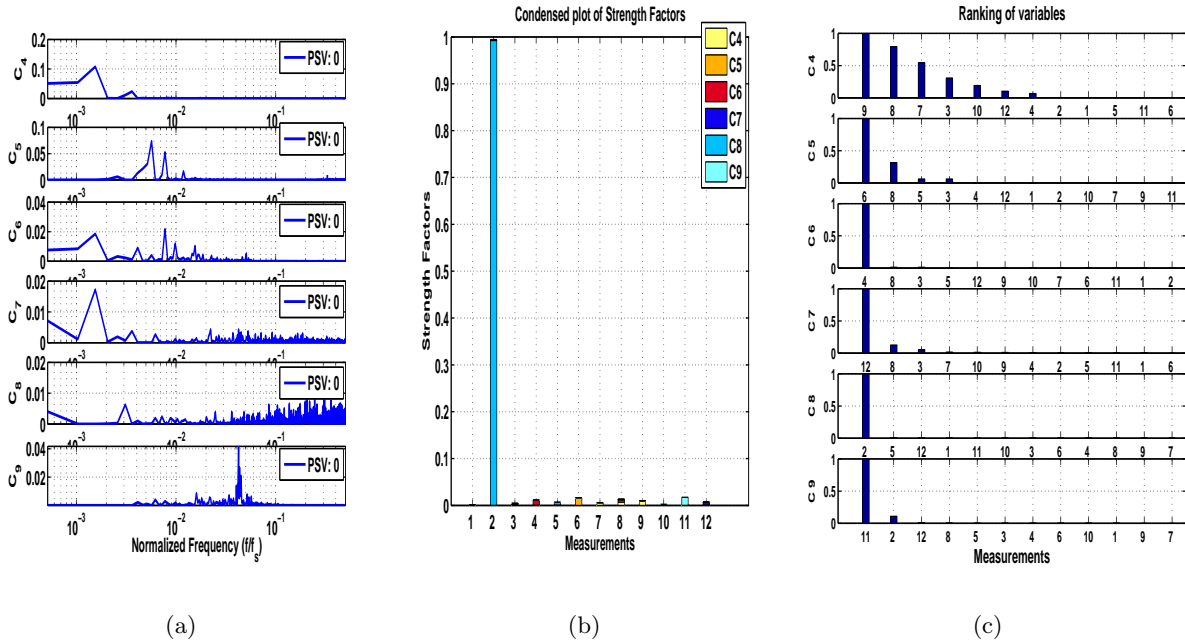


Figure 8: Residual component analysis for the Entech data set

5.1.3 Residual component analysis vs. Filtering

As remarked in Section 3.3, the residual components can be explored for weak/noisy components and any harmonics. For this purpose, 6 out of the 8 residual components were analyzed. Since most of the spectral information has already been captured as is evident from Figures 6(b) and 7(a), the residuals can be expected to primarily contain noise.

The results are shown in Figure 8(a). The following observations are due to Figures 8(b) and 8(c)

- Component 6 contains the remaining harmonics of 0.002 cycles/sample, which is stronger in tag 6 when compared to tag 8. This is an indication that the former has more harmonics than the latter and that the associated loop contains a non-linearity. In fact, loop 6 was confirmed to be a sticky valve.¹ Observe that this measurement does not have the maximum power at 0.002 cycles/sample, but the power is rather distributed over the harmonics. *Therefore, the measurement with the maximum strength is not necessarily the source carrier, particularly when the source is a non-linearity.*
- Figure 8(b) confirms that measurement 2 primarily is made up of noise.
- Finally, the residual analysis detects an additional frequency in tag 11 at about 0.0424 samples/cycle. This is a weak component and does not correspond to any harmonic⁵. In this case, this measurement has a higher strength at 0.02 cycles/sample than 11.

The ability of NMF to detect and localize harmonics in the variables is well demonstrated here. Filtering measurements to detect harmonics can be a fairly cumbersome task, particularly when

⁵In fact, Tag 1 was found to contain a tightly tuned controller operating at its limit cycle¹

Table 1: Summary of the Entech data analysis

Component	Frequencies	Measurements (Ranked)
1	0.002	10, 7, 3, 5, 4, 12, 6, 9, 8
2	0.02	1, 11 12 (weak)
3	0.004	{8, 9, 6, 4} {5, 3, 11} (weak)
5	Harmonics of 0.004	6, 8 (very weak)
9	0.0424	11 (very weak)

these frequencies are closely spaced. Furthermore, it is emphasized here that the purpose of filtering is only to verify the presence of additional frequencies, but the strengths of these frequencies have to be assessed only in relation to the original data set. Filtering can be avoided to a large extent by dividing the frequency axis into three regimes of low-, mid- and high-frequency regimes as is recently done by.¹¹ However, in their work, the results from the analysis on each segmented data set has to be put together in order to obtain the overall picture. The residuals here obviate the need for filtering and any related methods, and also provide the relative strengths in the measurements.

Table 1 summarizes the important observations.

NMF vs. PCA

The information extracted by NMF on the data set is more methodical, natural and elegant when compared to the PCA method used in [1], where an elaborate analysis was necessary. No filter has been applied here, and yet the same information has been obtained with relative ease. In addition, a new frequency has been detected in Tag 11 which was hitherto undetected. The residual analysis involved in PCA using the squared prediction error (SPE) requires relatively substantial manual intervention and interpretation than the simple residual component analysis illustrated in this work.

5.2 Case Study 2: South East Asian Refinery Data

This second case study involves the detection and isolation of oscillatory loops belonging to a SE Asia refinery.^{25,1}

The refinery process, whose simplified schematic is shown in Figure 9, consists of a hydrogen plant.²⁵ Measurements from 37 process tags were sampled at 1 min. intervals for a period of 24 hours. A smaller subset of 512 samples were chosen for the spectral analysis. The objective of this analysis is to detect the presence of oscillations and isolate those loops with common oscillations. Spectral analysis of this data using High Density Plots and Spectral PCA had appeared in the work by.¹

The total plot shown in Figure 10(a) gives an overall picture of the spectral behaviour in the plant. It is evident that there are 4 regular oscillations and a series of oscillations with broad peaks between the frequencies 0.02 and 0.06 cycles/sample. The most dominant oscillation has a frequency of 0.06 samples/cycle. The PSV plot shown in Figure 10(b) suggests that 4 dominant components are present in the plant, which is also in agreement with the earlier observations. Only 10 components were extracted since the maximum size is 37 and that the number of dominant components is well below ten.

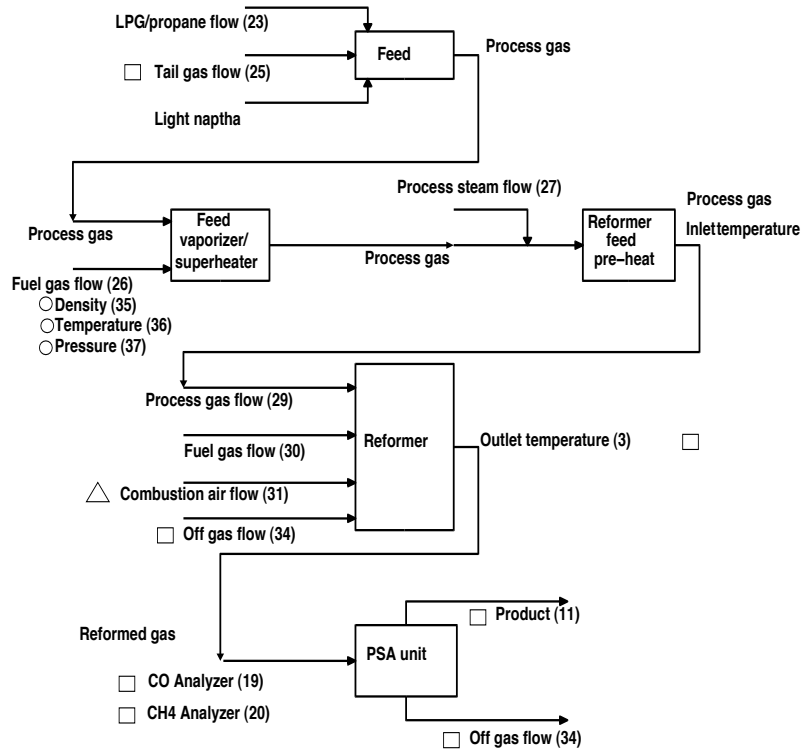
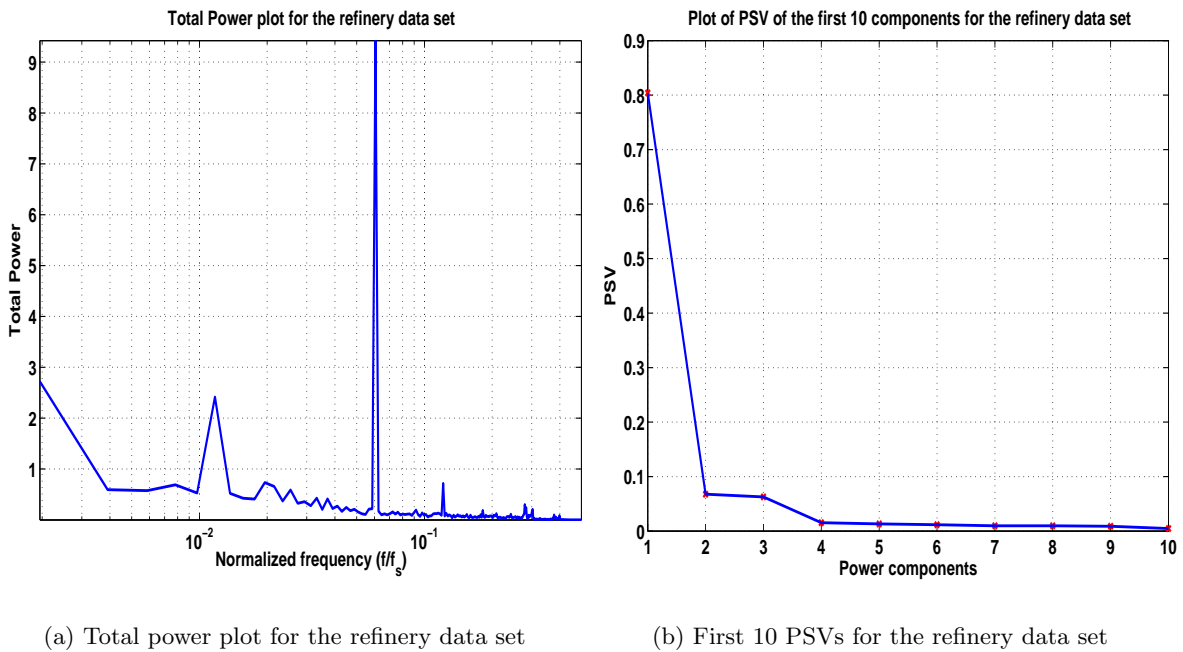


Figure 9: Schematic of the SE Asia refinery process



(a) Total power plot for the refinery data set

(b) First 10 PSVs for the refinery data set

Figure 10: Total Power and the first 10 PSVs for the refinery data

5.2.1 Isolation of loops with common oscillations

The dominant components and the weak components extracted by the NMF along with their strength factors are shown in Figures 11 and 12 respectively. Note that in Figure 11(c) only those variables whose strength factors are greater than 0.5 are shown for the sake of clarity in the illustration.

Here, again as in the earlier case study, NMF has successfully extracted the different components in the order of dominance with the help of PSV. It has also demonstrated its superiority in terms of ease and implementation over the combination of PCA and the filtering technique that was employed in.¹ In the paper by,¹ three principal components explained nearly 90% of the variance. In their work, it was necessary to perform an analysis in a 3-D space since PCA requires to look at its principal components simultaneously to find clusters of common behaviour. Clusters had to be located in a 3-D space visually, limiting its applicability, making it a cumbersome exercise and providing very little scope for automation. In fact, these limitations become even more stringent as the required number of principal components becomes large. Related works on enhancing the visualization of multidimensional principal components has been reported by²⁶ and.²⁷ In contrast, the need for multidimensional visualization is eliminated by NMF since it obtains a set of close-to non-overlapping decomposition. Consequently, it is sufficient to look at each component individually.

5.3 Case Study 3: MCC data set

An industrial data set was provided by the courtesy of Mitsubishi Chemical Corporation (MCC), Mizushima, Japan. The provided data set consisted of 59 variables: 28 process variables (PV's), 15 controller outputs (OP's) and 16 indicator variables. Each variable has 7200 observations with a sampling interval of 1 minute corresponding to data for a period of 5 days of operation.

Figure 13 shows the process schematic of the plant. In this section, the proposed procedure is applied to this large data set to detect the oscillations and their propagation in the plant. At first a total power plot is obtained in order to examine the different kinds of oscillations present in this plant. The same is shown in Figure 14(a). Based on the plot, it is clear that the plant contains at least eight dominant oscillations besides the low power ones. Fifteen components were extracted whose PSVs are shown in Figure 14(b).

5.3.1 Isolation of loops with common oscillations

The dominant power components, their strengths in the measurements and the rankings are displayed in Figure 15. From the fact that there are several peaks in the spectral components, it is clear that the plant is experiencing various oscillations under the influence of various sources.

The oscillation that was of concern to the plant personnel was that one with a period of about 2 to 3 hours in the condenser level of a distillation column. The corresponding time-period (frequency) was approximately 144 samples/cycle (≈ 0.0069 cycles/sample). This oscillation actually propagated throughout the plant, causing sub-optimal operation and large economic losses. The source of this oscillation was not clearly known, but the oscillation was visibly observed in the condenser section.

None of the dominant oscillations contain the frequency of interest. Therefore, a residual component analysis is in place here. Exploring the next five components, that is, components 5 to

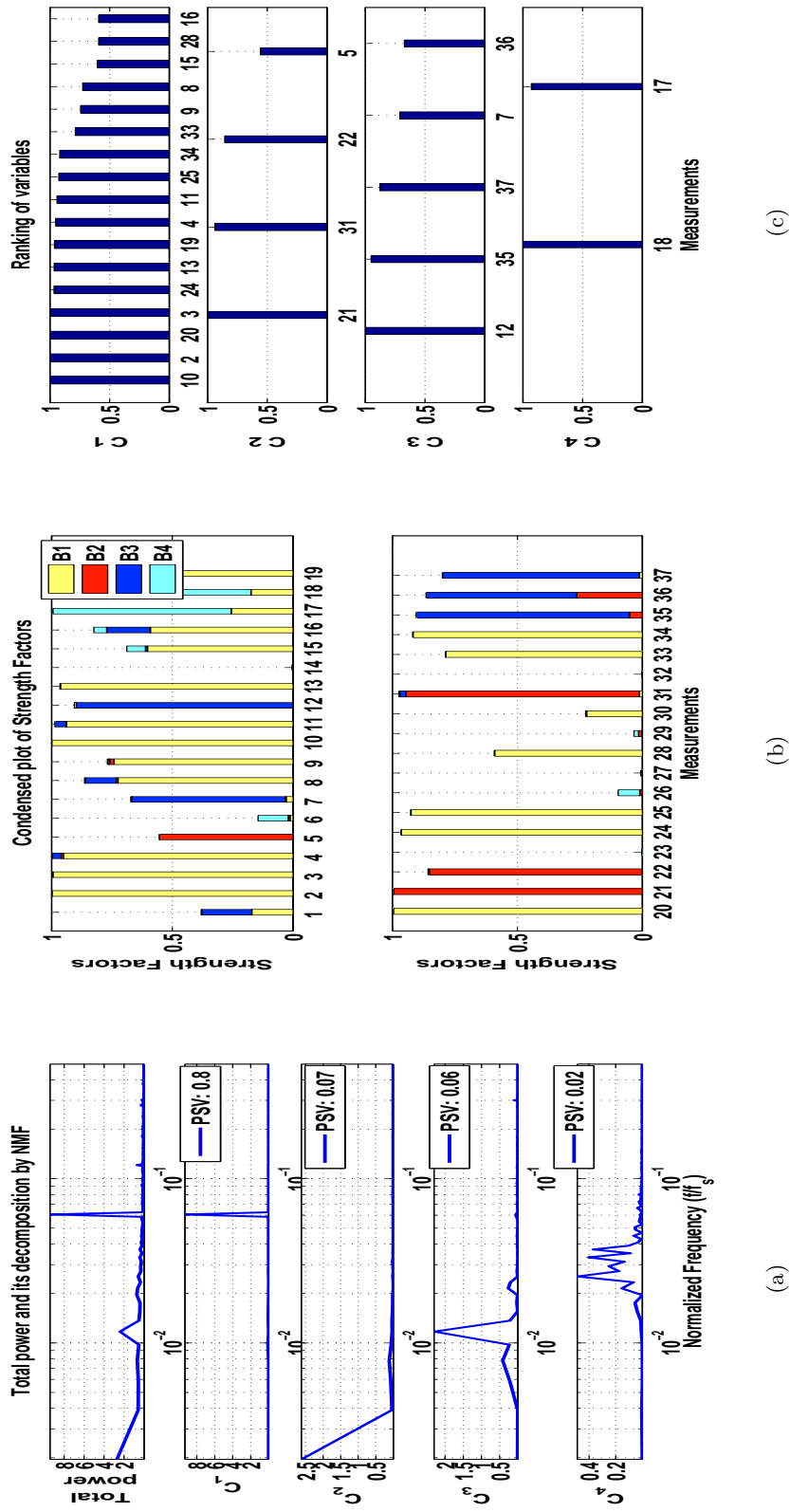


Figure 11: Dominant components of the refinery data set

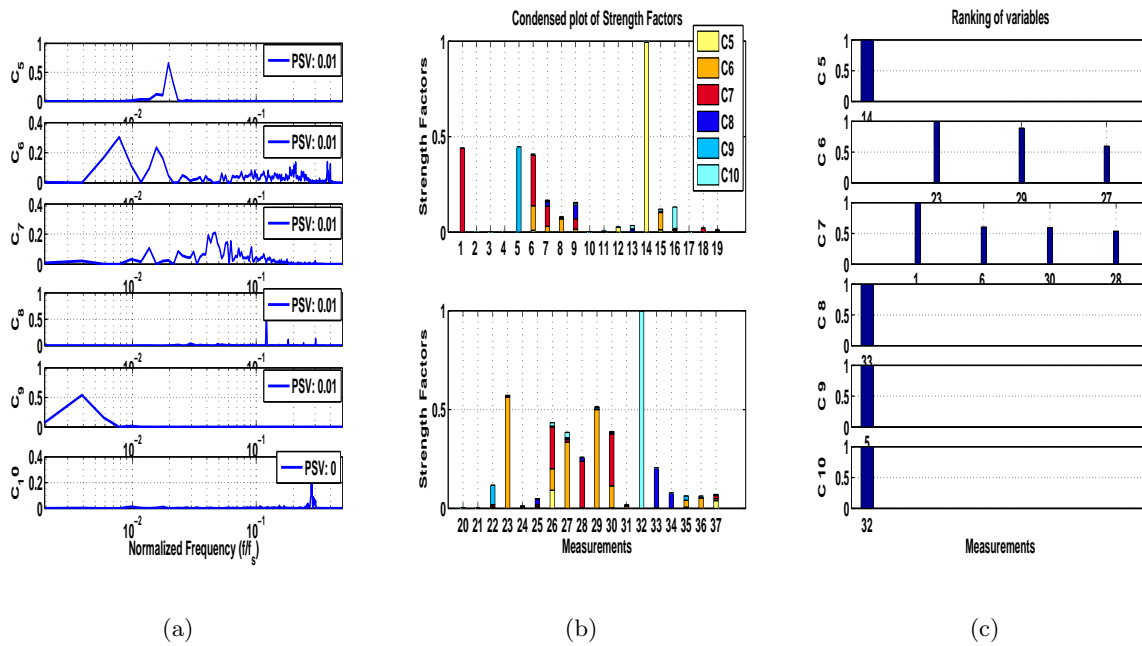


Figure 12: Residual components of the refinery data set

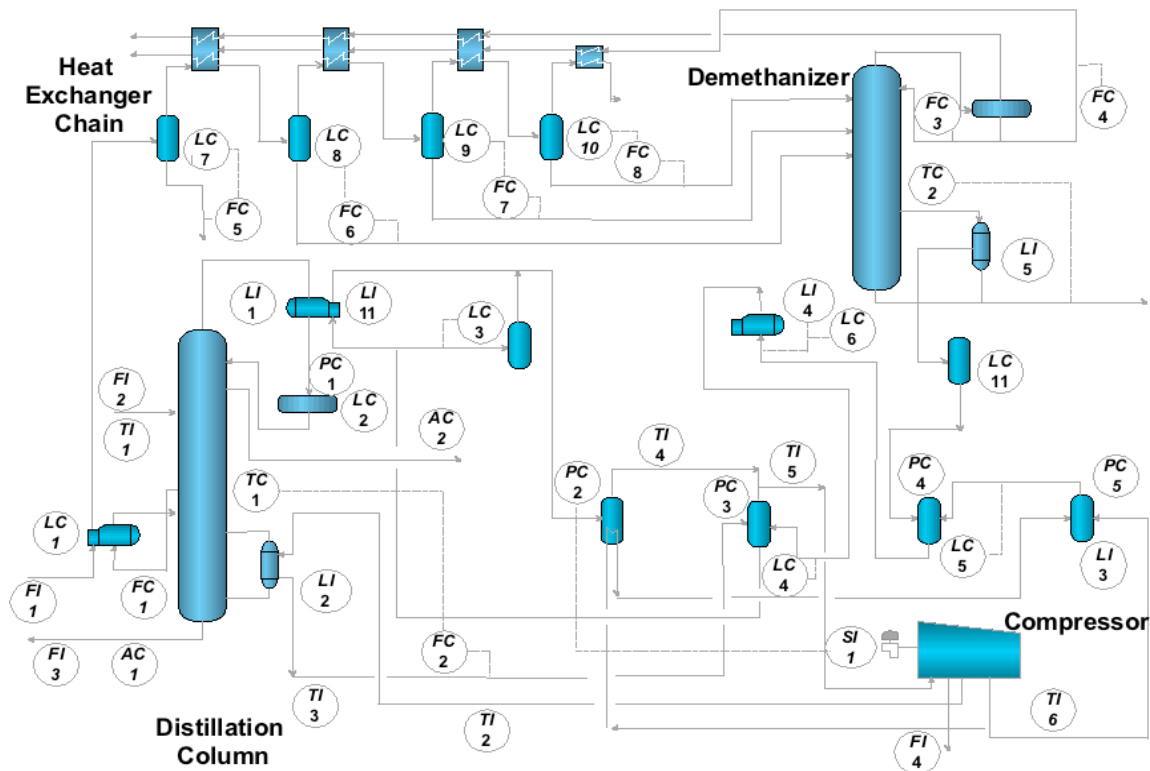
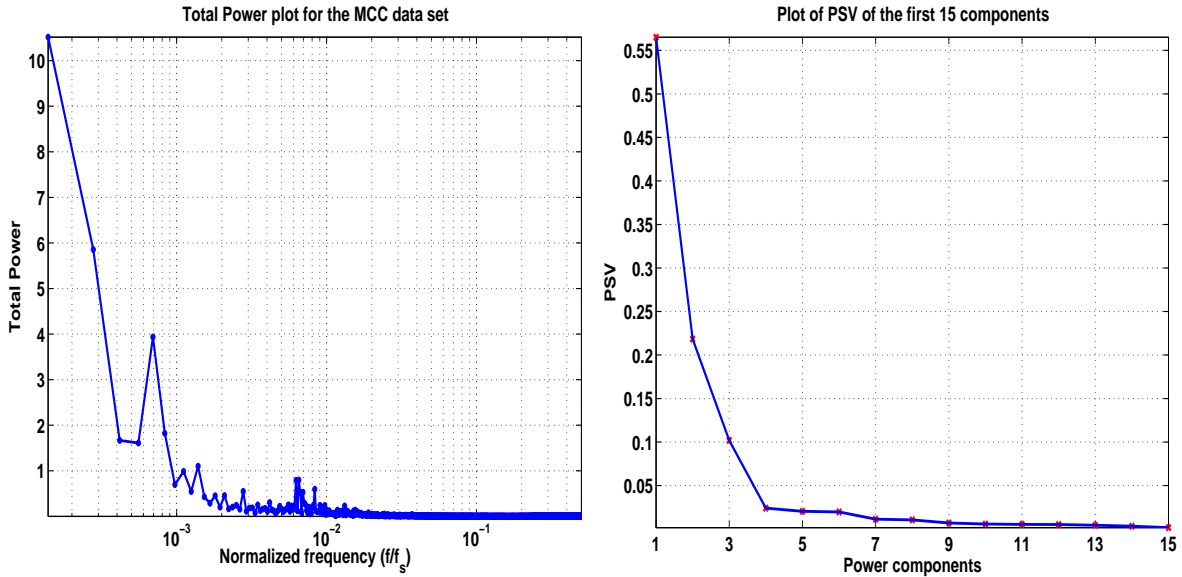


Figure 13: Schematic of the section of the Mizushima plant, Mitsubishi .



(a) Total power plot for the refinery data set

(b) First 10 PSVs for the refinery data set

Figure 14: Total Power and the first 10 PSVs for the MCC data set

10, yields us Figure 16(a). The frequency of interest is spotted in the eighth component, which has frequencies between 0.006 and 0.007 cycles/sample. The corresponding strength factors and rankings for this component are given in Figures 16(b) and 16(c) respectively. Measurements 9, 12 and 25 contain this oscillation to a large extent in that order of strength. The propagation of this oscillation in the plant is shown in Figure 17.

The measurements that have been significantly affected measurements are 9, 12, 25, 10, 32, 29, 36, 51, 40, 28 and 31. The associated tag names are given in Table 2. Most of the loops that are affected are level control loops. Some of the other measurements that are under the weak influence of this oscillation are 2, 11, 41, 22, and 33.

Based on the analysis so far, it can be concluded that one of the tags 9, 12 or 25 are the possible

Table 2: Measurements significantly affected by the 144 samples/cycle oscillation

Measurement	Tag name
9	LI1.PV
12	PC1.PV
25	TI4.PV
10	LI11.PV
32	LC5.PV
29	LC4.PV
36	FI4.PV
51	LC8.PV
40	LC6.PV
28	LC4.CO
31	LC5.CO

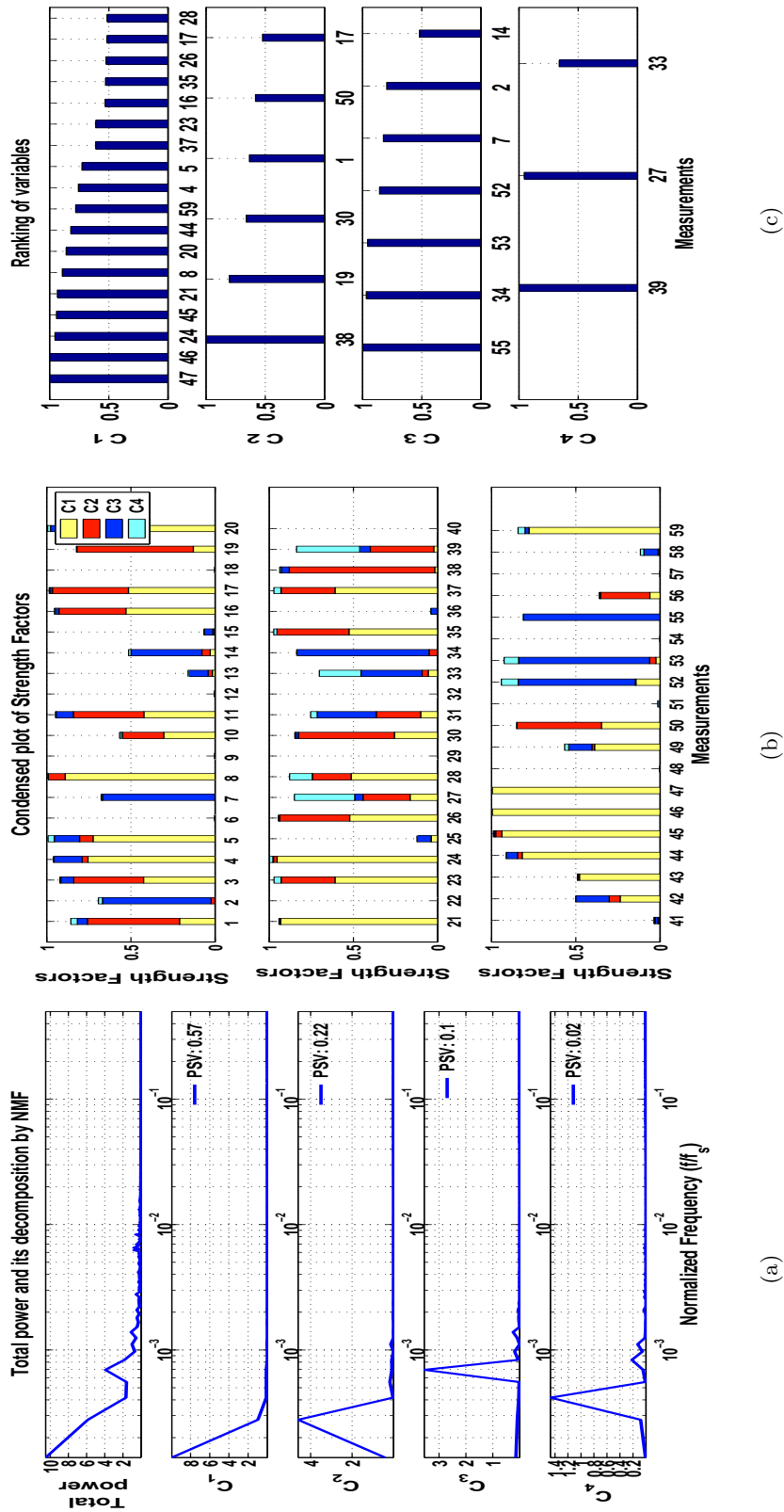
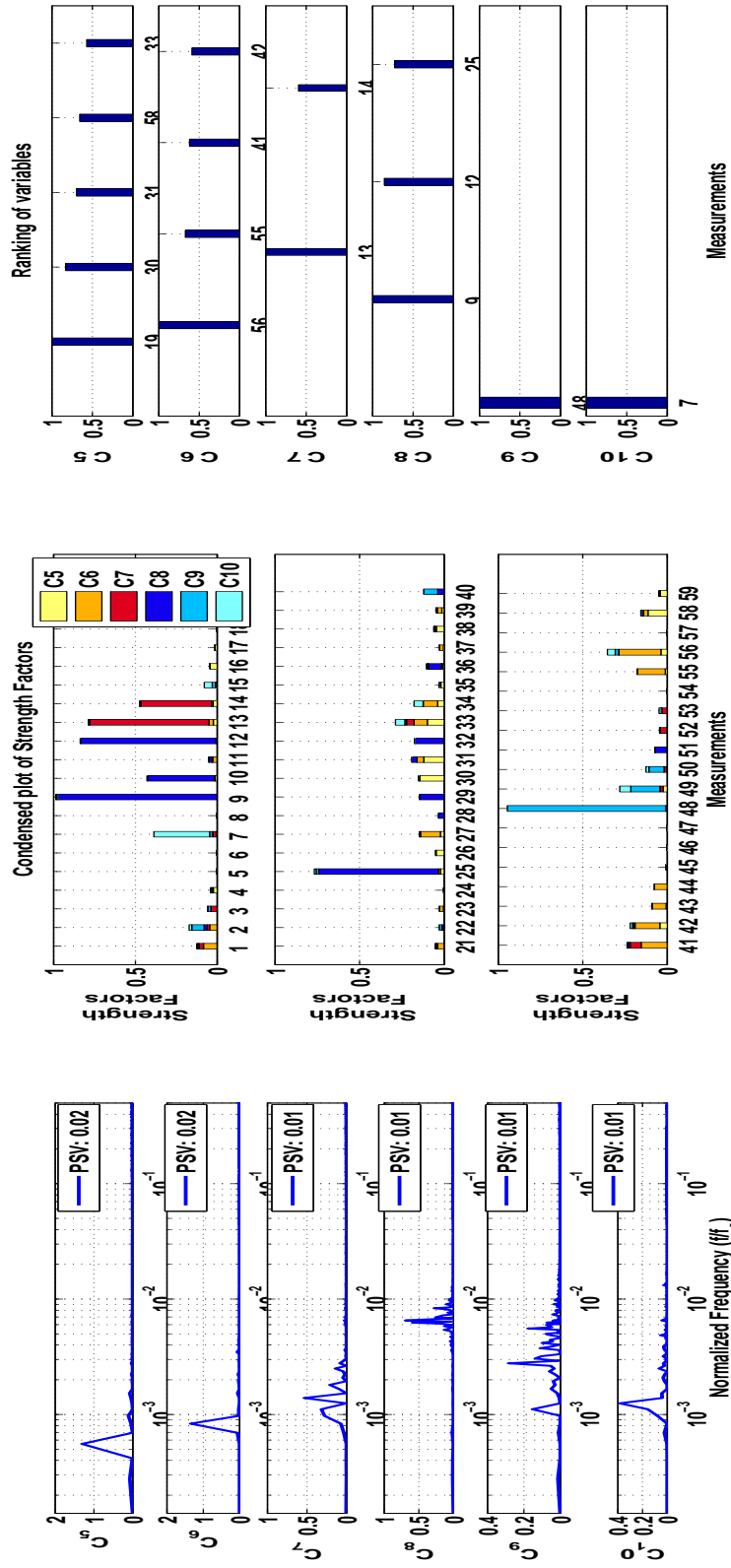


Figure 15: Dominant components of the MCC unit



(a)

(b)

(c)

Figure 16: Fifth to tenth components of the MCC spectra

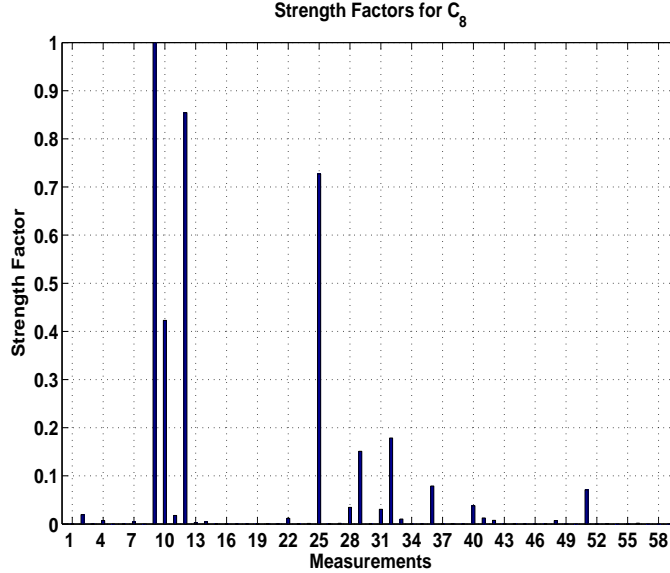


Figure 17: Propagation of the 144 samples/cycle oscillation

sources of oscillation with high probability. Since tag 12 corresponds to a control loop measurement, it is very likely that the source lies in this loop and that source is either a non-linear element or a tightly tuned controller. A diagnostic study conducted by Choudhury *et al*²⁸ showed the presence of a sticky valve in the corresponding loop. Thus, NMF has successfully detected the presence of various oscillations and has rightly guided towards diagnosing the root cause of the oscillation of interest.

6 Conclusions

In this article, a recently developed technique known as the Non-negative Matrix Factorization in the context of image analysis and document clustering has been tailored to address the plant-wide oscillation detection problem. Two of the key issues in NMF, namely the basis size determination and the initialization have been addressed. A new robust measure for basis size determination known as the Pseudo-Singular Value (PSV) has been proposed. In this context, the notions of total power and power component have been introduced for multivariate spectral analysis. A plot of the total power has been shown to be a useful tool in visualizing the behaviour of a process as well as in obtaining a first-hand information on the dominant oscillations in the process. The interpretation that NMF breaks down this total power into its constituents is novel.

The PSV allows one to (i) methodically determine the importance of a component with respect to the total power and (ii) determine the size of the basis space to describe the desired features of a data set. Although these concepts have been introduced in the context of process data analysis, it can be easily applied to non-engineering applications as well. If NMF yields non-overlapping components, PSV reduces to the normalized singular values given by SVD. Initialization of the NMF algorithm has been fixed by kick-starting the algorithm with the absolute values of the basis vectors and weights given by SVD. The use of this strategy when combined with the PSV-based ranking fixes the permutation problem of NMF.

In order to assess the strength of each power component given by NMF in each measurement,

a new measure known as the Strength Factor has been introduced. The SF is a better measure of the presence of a component than the weights of NMF since the latter is sensitive to the presence of noise. Besides, the SF has been shown to carry useful properties which allow one to (i) obtain a robust estimate of the distribution of frequencies in a measurement and (ii) to rank the measurements based on their contributions to the features of each component.

A remarkable feature of the proposed method is that it alleviates/eliminates the need for filtering the data (as is done with existing methods) or dividing the data into different frequency regimes. This is easily facilitated by examining all or a few of the residual components of the NMF.

A general procedure for applying NMF to multivariate spectral analysis has been outlined.

NMF has several advantages over PCA in the analysis of multivariate spectral data, mainly because the former method yields basis shapes with non-negative entries. In addition, NMF has the ability to extract localized features whereas PCA captures the global variation in the data set. The key advantages of using NMF is that it yields components that can be directly related to the individual spectra of the measurements, and that the plot of SF can be interpreted in a 1-dimensional space, whereas the use of PCA method involves the search for clusters in a higher dimensional space (greater than three). In comparison to other techniques such as ICA, NMF is a more general method since it does not make any statistical assumption on the sources, but only makes use of the fundamental non-negativity property of the spectra.

By means of applications to simulated and industrial processes, the potential of the proposed measures and the superiority of NMF over conventional PCA has been demonstrated. The proposed measure with NMF successfully extracted the dominant spectral features and identified distinct clusters of variables with common oscillations in all the cases.

Acknowledgements

The authors gratefully acknowledge the NSERC-Matrikon-Suncore-iCORE Industrial Research Chair program for financial support. The authors are thankful to Mitsubishi Chemical Corporation, Japan for kindly allowing us to provide their data in this study

References

1. Thornhill, N.; Shah, S.; Huang, B.; Vishnubhotla, A. *Control Engineering Practice* **2002**, *10*, 833-846.
2. Harris, T.; Seppala, C.; Jofreit, P.; Surgenor, B. *Control Engineering Practice* **1996**, *9*, 1297-1303.
3. Rengaswamy, R.; Venkatasubramanian, V. *Engineering and Artificial Intelligence* **1995**, *8*, 35-51.
4. Thornhill, N.; Hagglund, T. *Control Engineering Practice* **1997**, *5*, 1343-1354.
5. Bissessur, Y.; Martin, E.; Morris, A. *Control Engineering Practice* **1995**, *7*, 1357-1368.
6. Tangirala, A.; Shah, S.; Thornhill, N. *Journal of Process Control* **2005**, *15*, 931-941.
7. Jiang, H.; Choudhury, M.; Shah, S. Detection And Diagnosis Of Plant-Wide Oscillations Via The Spectral Envelope Method. In *The Proceedings of ADCHEM 2006*; Gramado, Brazil, 2006 Accepted for publication in *Journal of Process Control*.
8. Martin, G.; Turpin, L.; Cline, R. *Hydrocarbon processing* **1991**, *70*, 68-73.
9. Thornhill, N.; Shah, S.; Huang, B. "Detection of distributed oscillations and root-cause diagnosis", in: CHEMFAS4, Cheju Island, Korea, 2001.
10. Thornhill, N.; Huang, B.; Zhang, H. *Journal of Process Control* **2003**, *13*, 91-100.
11. Xia, C.; Howell, J.; Thornhill, N. *Automatica* **2005**, *41*, 2067-2075.
12. Elliott, M.; Walter, G.; Swift, A.; Vandendorpe, K.; Schotland, J.; Leigh, J. *Magnetic Resonance in Medicine* **1999**, *41*, 450-4555.
13. Wild, S. "Seeding Non-negative Matrix Factorization with the Spherical K-Means Clustering", Master's thesis, University of Colorado, 2002.
14. Lee, D. D.; Seung, H. S. *Nature* **1999**, *401*, 788-791.
15. Paatero, P.; Tapper, U. *Environmentrics* **1994**, *5*, 111-126.
16. Sajda, P.; Du, S.; Parra, L. Recovery of constituent spectra using non-negative matrix factorization. In *SPIE 5207*, Vol. 321; 2003.
17. Kim, P.; Tidor, B. *Genome Research* **2003**, *13*, 1706-1718.
18. Guillaumet, D. Color Histogram Classification using NMF. In ; <http://www.cvc.uab.es/~davidg/Papers/CVC57/cvc57.pdf>, 2001.
19. Lee, D.; Seung, H. *Advances in Neural Information Processing* **2001**, *13*, 556-562.
20. Li, Z.; Hou, X.; Zhang, H.; Cheng, Q. "Learning Spatially Localized, Parts-Based Representation", Presented at IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI., 2001.
21. Guillaumet, D.; Vitria, J.; Schiele, B. *Pattern Recognition Letters* **2003**, *24*, 2447-2454.

22. Donoho, D.; Stodden, V. *Advances in Neural Information Processing Systems* **2004**, *17*,.
23. Wild, S.; Curry, J.; Dougherty), A. *Journal of Pattern Recognition* **2004**, *37*, 2217-2232.
24. Wild, S.; Curry, J.; Dougherty), A. Motivating Non-Negative Matrix Factorizations.. In *Eighth SIAM Conference on Applied Linear Algebra*; 2003.
25. Terrible, J.; Shahani, G.; Gagliardi, C.; Baade, W.; Bredehoft, R.; Ralston, M. *Hydrocarbon processing* **1999**, *78*, 43-53.
26. Wang, X.; Medasani, S.; Marhoon, F.; Albazzaz, H. *Industrial Engineering and Chemistry Research* **2004**, *43*, 7036-7048.
27. Thornhill, N.; Melbo, H. Detection of plant-wide disturbances using a spectral classification tree. In *The Proceedings of ADCHEM 2006*; Gramado, Brazil, 2006.
28. Choudhury, M. A. A. S.; Kariwala, V.; Shah, S. L.; Douke, H.; Takada, H.; Thornhill, N. F. A simple test to confirm control valve stiction. In *The 16th IFAC World Congress*; Prague, Czech Republic, 2005.