Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Process Control 15 (2005) 931–941
JOURNAL OF PROCESS CONTROL
www.elsevier.com/locate/jprocont

Short communication

PSCMAP: A new tool for plant-wide oscillation detection

A.K. Tangirala [a], S.L. Shah [b,*], N.F. Thornhill [c]

[a] Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[b] Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada T6G 2G6
[c] Department of Electronic and Electrical Engineering, University College London, Turrington Place, London WC1E 7JE, UK

Received 20 August 2003; received in revised form 26 November 2004; accepted 11 January 2005

Abstract

A commonly encountered issue in process industry is concerned with the detection of plant-wide oscillations. In this paper, a new visualization tool termed as the power spectral correlation map (PSCMAP) is proposed for this purpose. The proposed colour map is based on a new measure defined as the power spectral correlation index (PSCI). A simple clustering algorithm is developed to group blocks of variables with similar spectral shapes. The combined visualization tool is shown to be simple, effective and powerful in collecting variables with common frequency-domain behaviour in a multivariate process. The potential of the combined technique is illustrated by an application to two industrial processes, (i) a simulated pulp and paper process and (ii) a SE Asia refinery.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Power spectral correlation map (PSCMAP); Power spectral correlation index (PSCI); Plant-wide oscillation detection

1. Motivation

The detection of plant-wide oscillations is concerned with the identification of a set of signals (or measurements) oscillating at similar frequencies, but not necessarily in phase with each other. Such a problem is of great practical significance since it involves the identification of common cause disturbances whose effects propagate to many units and thus may impact overall process performance. The causes of oscillations, their influence on poor performance, and their benefits to the economy and safety of the plant are well outlined in the literatures [1–6].

Time-domain methods, i.e., methods that capture and explore the temporal aspects of the measurements, are limited in their applicability to the oscillation detection problem by the lack of knowledge of process order and time delays (phase lags among measurements). On the other hand, frequency-domain methods such as spectrum-based methods, which extract the frequency-domain information from the raw measurements, are intuitively natural choices for the oscillation detection problem. An important advantage of spectral techniques over time-domain-based methods is that they do not require the knowledge of process order and time delays for dynamic systems. The significance of spectral analysis lies in the fact that oscillations are commonly indicative of the presence of valve non-linearities, poor controller tuning and oscillatory disturbances. Of course, lack of oscillations does not necessarily mean good performance index of that loop, and the converse is also true, i.e., bad performance index need not necessarily be associated with oscillations.

Motivated with these ideas, tools such as high density plots (HDP) and spectral principal component analysis (SPCA) have been proposed by Thornhill et al. [7,8] and successfully applied to various multivariate industrial processes for the purpose of plant-wide oscillation detection.

* Corresponding author. Tel.: +1 780 492 5162; fax: +1 780 492 2881.
E-mail address: sirish.shah@ualberta.ca (S.L. Shah).

0959-1524/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2005.01.005

In this article, we propose a new visualization and analysis tool, namely the power spectral correlation map (PSCMAP), for analyzing variables with similar power spectral shapes. The proposed tool is based on a new measure defined as the power spectral correlation index (PSCI). The PSCI is the correlation between the power spectra of two signals, and is a measure of the similarity of the spectral shapes of the signals. This measure is defined in such a way that it lies between the values 0 and 1. If two measurements have similar dominant frequencies, then their PSCI takes a high value (close to unity).

The primary application of PSCMAP is to gather measurements with similar dominant oscillations in a plant with several oscillatory variables in a simple and effective way. The main advantage of PSCMAP is the fact that it completely eliminates the need for inclusion of lags or phase differences among process measurements as would be required in a temporal analysis. The utility of PSCMAP is further enhanced by a regrouping of the variables with similar correlation indices. For this purpose, we propose a simple clustering algorithm. The shapes associated with each group can be obtained by looking at the shape of what is known as a *pivot* variable for each group (defined in Section 3). This pivot variable is in some sense representative of the basic spectral shapes for each group. In this sense, PSCMAP offers advantage over the classical SPCA method, wherein it is difficult to attach any physical meaning to the basis vectors.

The utility of the proposed method is illustrated on two industrial applications—a simulated pulp and paper process and a SE Asian refinery.

This paper is aimed at process engineers who can deploy these tools to obtain valuable insights into the plant operation with minimal computational efforts while using historic data.

The presentation is organized as follows. Section 2 introduces the notion of PSCI. In Section 3 the concept of PSCMAP is introduced with a presentation of the proposed clustering algorithm. Section 4 contains the application of the proposed technique to three industrial case studies. The paper ends with concluding remarks in Section 5.

2. Power spectral correlation index

The power spectral correlation index (PSCI) is defined as the correlation between the power spectra of two different measurements. It is a measure of the similarity of spectral shapes, i.e., measure of the commonness of frequencies of oscillations. The procedure to calculate the correlation is illustrated in the block diagram shown in Fig. 1.

The DFTs that are used to calculate the spectrum are calculated after removal of means from the time-series

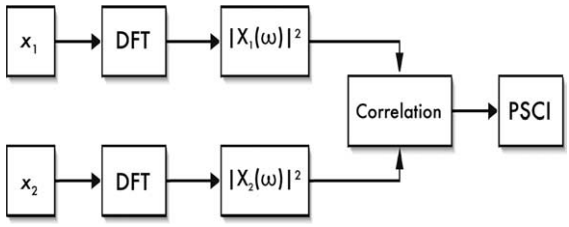

Fig. 1. Illustration of the procedure for calculating power spectral correlation index between two variables x_1 and x_2.

data. However, the correlation used in the calculation of PSCI is calculated without the removal of mean of the spectra (reasons for which are explained in Section 3.3). The PSCI for any two spectra $|X_i(\omega)|^2$ and $|X_j(\omega)|^2$ is calculated as:

$$\text{Correlation}(|X_i(\omega)|^2, |X_j(\omega)|^2) = \frac{\sum_{\omega_k} |X_i(\omega_k)|^2 |X_j(\omega_k)|^2}{\sqrt{\sum_{\omega_k} |X_i(\omega_k)|^4 \sum_{\omega_k} |X_j(\omega_k)|^4}} \quad (1)$$

As a result, PSCI always lies between 0 and 1.

As shown in the schematic, the phase information is excluded due to taking the magnitude of the DFT of the measurements. In the detection of plant-wide oscillations, the objective is to collect variables with similar oscillatory behaviour. A time-domain correlation of the measurements will reveal similarities such as common similarities among different measurements only if they are aligned properly. A proper alignment of data requires prior knowledge of process delays, orders and measurement lags. Obtaining and implementing this information is not a trivial task in most applications. Besides, calculation of phase differences is sensitive to the presence of noise and the frequency resolution [9]. On the other hand, the power spectral correlation method meets the same objective without any need for computing the phase differences or measurement lags. In other words, the PSCMAP is a tool that is blind to phase differences. This is illustrated by an example in Appendix A.

Remark. The power spectral correlation index is different from the well-known *coherency* (see [10]). Coherency is a function of frequency and is defined as the normalized cross-covariance function between the Fourier transforms of two signals. As a result it is complex-valued. In contrast, the PSCI is real-valued and is the normalized covariance between the power spectra of two signals as defined in Eq. (1). Therefore, the PSCI is not a function of the frequency. One of the main differences is that coherency is dependent on the phase difference between two signals, whereas PSCI is invariant with respect to phase difference. This property makes PSCI the preferred metric for data analysis as it does not

require the phase or time-lag information between different signals.

3. Power spectral correlation map

For multivariate processes, the PSCI is a matrix of size $m \times m$, where m is the number of measured variables. In order to provide an effective interpretation of the PSCI, the matrix is plotted as a colour map, which is termed as the power spectral correlation map. The intensity as well as the type of colour in the map is assigned in proportion to the value of the correlation index. This mapping is performed according to the choice of colour and the number of shades in that colour.

An important aspect of this colour map is to be able to automatically re-arrange and group variables together with similar shapes, i.e., variables, which oscillate at a common frequency and have therefore similar values of PSCI. For this purpose, a simple clustering algorithm is proposed, that allows the user to cluster variables with similar shapes based on a threshold value, γ. A brief discussion on the choice of threshold is presented below in Section 3.2. The PSCMAP is essentially characterized by the (i) colour code and (ii) the clustering algorithm to group variables with similar spectral shapes. These features are explained below.

3.1. Colour code

The colour code is an important factor in the visualization ability of the PSCMAP. It also determines the sensitivity of the colour map to variations in the correlation index. As the number of colour shades increases, the PSCMAP is able to display more variations in the correlation index, but at the cost of losing the ability to interpret the colour map. Thus, there is a trade-off between the sensitivity and the interpretability of the map. The choice of colour code depends on the user as well as the application in context. A general recommendation is to choose colour code belonging to one colour rather than choosing from a range of different colours. Then, one can vary the number of shades in that colour to suit the needs of the application.

In the published version of this work four colour shades of grey are chosen.[1] It is observed that this choice was sufficient for the applications in context. Since the PSCI values lie between 0 and 1, each colour shade represented the band of values (0–0.25), (0.25–0.5), (0.5–0.75) and (0.75–1) respectively. More divisions could be obtained by increasing the number of colour shades. In a colour map with these shades, we would be generally interested in spotting the cells with the darkest colour since they represent high values of correlation.

The next step is to clearly group those cells with common colour so that variables with similar oscillations can be clustered together. When the dimension of the data set is small and/or when the number of oscillations are regular and limited to two or three frequencies, it is possible to perform a visual re-arrangement of the colour map. However, when the data set is large and/or when there are multiple oscillations, it is important to automate this process. In the next section, a simple clustering algorithm is presented for this purpose.

3.2. Clustering algorithm

The proposed clustering algorithm sequentially extracts blocks of maximum number of correlated variables from the PSCI matrix until a single variable is left. The algorithm is based on two steps: (i) identifying a *pivot* variable at each stage and (ii) extracting the block of variables that are correlated with this *pivot* variable. The *pivot* variable (at each stage) is defined as that variable which is strongly correlated with the maximum number of variables (remaining at each stage). The maximum number of iterations in this algorithm is at worst equal to the number of variables, which occurs when all the loops have no commonality in their spectral shapes. In a general case, when there are common oscillations in loops, this number can be much less than the number of variables.

A key aspect of this algorithm is that the user need not specify the number of clusters. The only specification is on the *threshold* γ, which is the value the user attaches to denote the minimum significant value of the correlation index. The value of this threshold can also be linked with the colour code. For instance, if four colour shades are chosen to display the colour map, then the threshold value can be set to 0.75, which means all those pairs of variables with PSCI values in the fourth colour shade would be considered significant.

The steps for implementing this algorithm are given as follows (denote the PSCI matrix by Σ_ω):

(1) Determine the number of entries in each column of Σ_ω that are greater than the threshold γ. The column with the maximum number is the *pivot* variable.
(2) Extract the subset of Σ_ω corresponding to variables which are strongly correlated with the *pivot* variables based on the threshold γ.
(3) Update the PSCI matrix with the correlation indices of the remaining variables.
(4) Repeat the above steps until a single variable is left.

[1] The electronic version of this article uses shades of red—white, yellow, orange and red.

Thus, at every stage the algorithm extracts that block with the largest number of strongly correlated variables. In addition, the spectra of the pivot variables are representative of the basis spectral shapes present in the data set. This is similar to the spectral PCA methodology where the basis shapes and the clusters are extracted using an orthogonal decomposition of the spectra [7].

3.3. Remarks on the properties of PSCMAP

- The PSCMAP essentially relies on PSCI, which is a measure of similarity of spectral shapes. From the definition of PSCI in Eq. (1), it is observed that for two variables to have a high PSCI, it is necessary that they possess spectral peaks in identical frequency bins (since we are dealing with a discrete set of frequencies). This property makes PSCI a strong candidate for plant-wide oscillation detection. On the other hand, it also means that if the spectra of two signals are such that their frequencies are apart by one bin, then PSCI will render them uncorrelated. Theoretically, this is an agreeable result, since these two signals indeed differ in their frequency behaviour and therefore they cannot be considered similar. However, if this disparity in the spectral shapes is due to computation issues (such as resolution of frequency channels) or if for practical reasons this difference in frequencies is negligible, then the analyst can choose to compute both spectra at a lower resolution (for example, by reducing the number of samples) so that the frequency bins either overlap or match exactly. In some cases, the user can even choose to perform smoothing of the spectra using a window function of appropriate length. In essence, the PSCMAP works at the level of frequency resolution that is chosen to compute the spectra.
- The power spectral correlation definition in Eq. (1) differs from the usual definition of correlation in statistics, which is computed by subtracting the mean from the spectra. Therefore, this is a non-centered (or uncentered) correlation. This metric is used so as to ensure that two non-overlapping spectra have zero correlation. If the mean was subtracted, then the correlation between non-overlapping spectra would be a negative quantity, making it difficult to interpret such values. Non-centered correlation is equivalent to the cosine of the angle between two vectors.

  Besides, if the power spectra are normalized to unity power (which is the case in this work), then the *mean of the spectra* are identical and equal to $1/(N/2)$ where N is the number of time-series data. As N becomes large, the mean value of each spectrum goes to zero. Therefore, the effect of not subtracting the mean in this case is negligible.

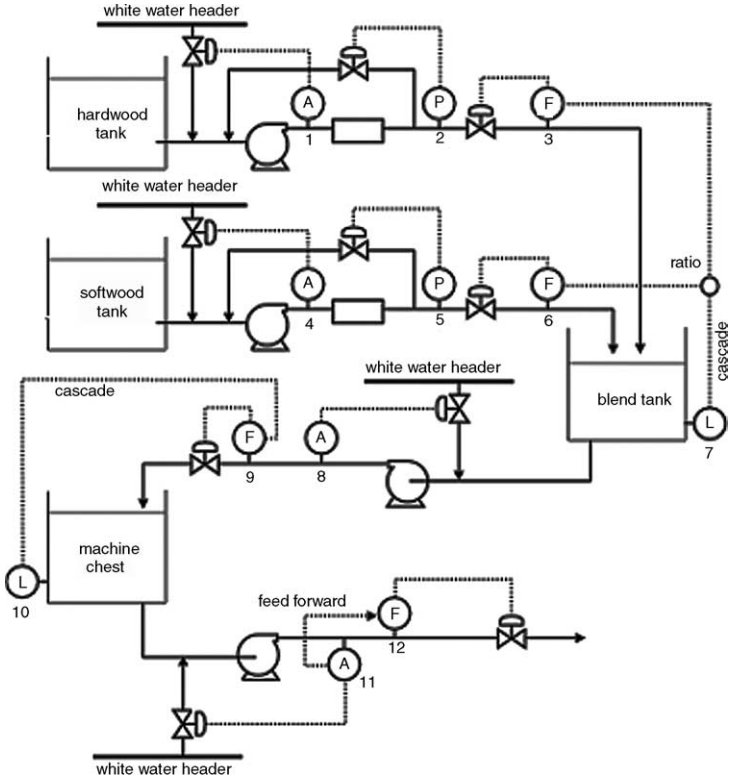

Fig. 2. Schematic of the simulated process of Entech Control Inc.

- The power spectrum calculation should be done after subtracting the *mean of the time-series data*. This is important to remove the effect of the DC component or any steady-state offsets. Thus, the power spectral correlation between two signals $x(t)$ and $y(t) = ax(t) + b$ is always unity irrespective of the values of a and b.
- The dominant oscillation in each member of a group that is identified by the PSCMAP is identical to the dominant oscillation of the pivot variable. This is the basis for identifying the members of each group. Therefore, the PSCMAP is very useful when there is a dominant frequency that influences the loops that are being examined.

If the members of each group have multiple oscillations, the correlation among the members of a group depends on whether the variables have these multiple oscillations in common and whether there is significant power in those multiple oscillations. A discussion of this point is illustrated in the refinery case study in Section 4.2.1.

4. Industrial case studies

4.1. Case study 1: Entech data

We first consider the data set from a simulated industrial process, courtesy of Entech Control Inc.

The simulated process shown in Fig. 2 consists of a pulp manufacturing process, where the hardwood and softwood pulps are mixed to give a stream of desired composition. The data set comprised 1934 samples from

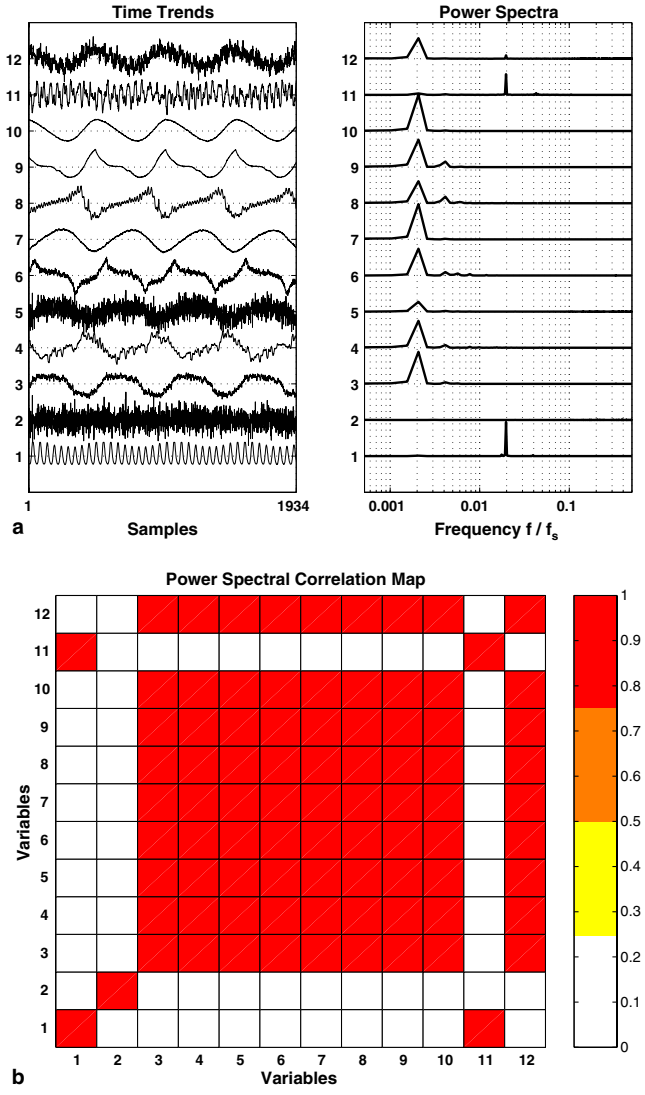

Fig. 3. High density plot (a) and the PSCMAP of the Entech data (b).

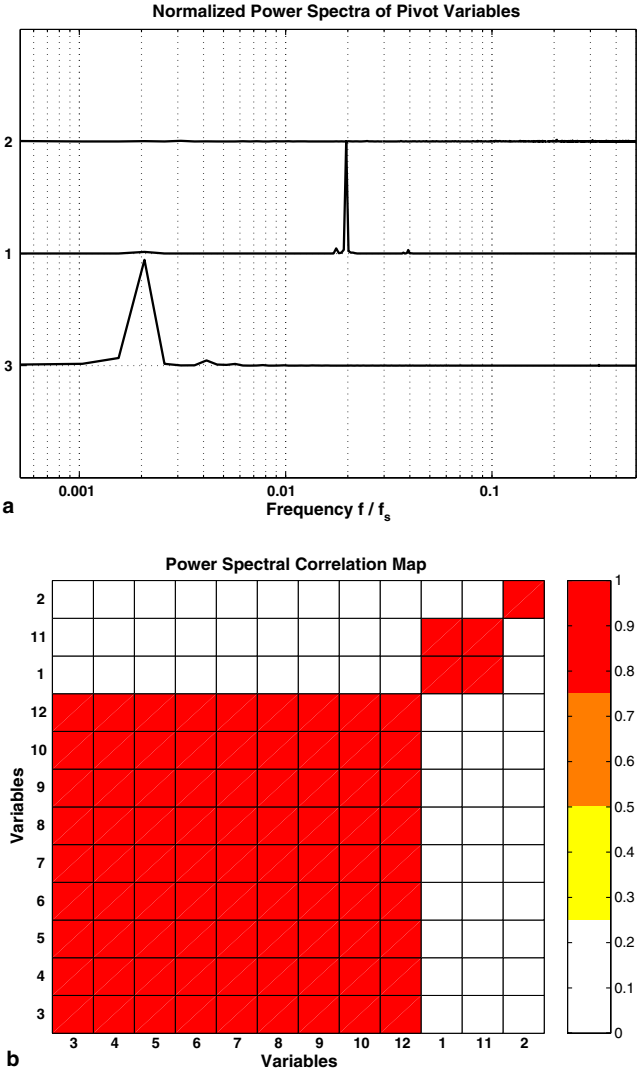

Fig. 4. Re-arranged PSCMAP and the corresponding spectral shapes of the pivot variables. (a) Spectral shapes of the pivot variables from the Entech data analysis and (b) PSCMAP (clustered) of the Entech data.

12 process measurements (tags), each of which was 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 simultaneous display of the time trends and the corresponding spectra of the 12 *controller errors* (SP–PV) is shown in Fig. 3(a), which is also known as the high density plot [7,8]. The spectra on the HDP are normalized such that the total power in each variable is unity.

It can be noticed that tags 3–10 and 12 contain a major peak at a common frequency, while tags 1 and 11 appear to share another common frequency of oscillation. The spectrum of tag 2 resembles that of a "white noise" spectrum, indicative that the associated controller may be doing well. Thus, we are seeking two basic spectral shapes that contain peaks associated with the two groups of measurements.

The role of PSCMAP is to identify loops with common oscillations, and extract the basic shapes for each of these groups.

The raw PSCMAP, i.e., without the re-arrangement, is shown in Fig. 3(b). From a visual observation one can notice a strong correlation between tags 3–10 and 12, and a strong correlation between tags 1 and 11. Tag 2 has an independent spectral shape. These observations are much clearer in Fig. 4(b), which is a re-arrangement of the raw colour map based on the clustering algorithm described in Section 3.2. The resulting PSCMAP contains group of variables that are strongly correlated with each other arranged in the sequence of decreasing member strength. For this case study, t can be seen that the variables of each group do not have any correlation with members of other groups, which implies that they have independent shapes. The plot of the spectra of the pivot variables in Fig. 4(a) confirms this observation. The pivot variables for each group have been identified as tags 3, 1 and 2 respectively. In other words, these tags are representative of the spectral behaviour of the other members of the respective groups. For variables with independent spectral shapes, the pivot variables are the variables themselves.

In order to further illustrate the effectiveness of PSCMAP, the data is filtered so that groups of variables with frequencies higher than 0.002 cycles/sample can be identified. In the original data, this low frequency appears to overshadow the power at higher frequencies for example in tags 6, 8 and 9. The filtered time trends and spectra are shown in Fig. 5(a). An approximated Weiner filter employed in [7] on the same data set has been used for this purpose. The implementation details and the algorithm are provided in [8].

The colour map for the filtered data is shown in Fig. 6(b). The second group of variables has remained unchanged. The first group of variables exhibits a common oscillation now at 0.004 cycles/sample, which happens to be a second harmonic of the 0.002 cycles/sample

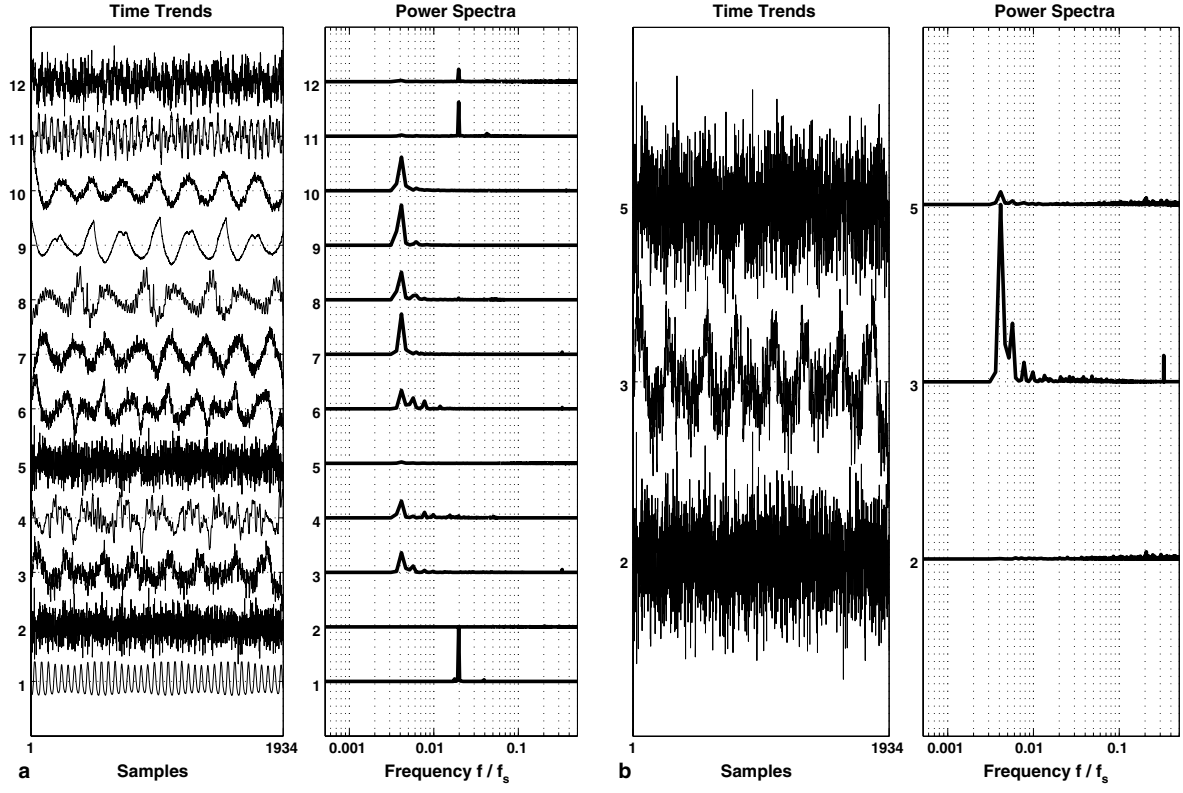

Fig. 5. HDP of the entire and selected filtered Entech data. (a) HDP of filtered Entech data and (b) spectral shapes of tags 2, 3 and 5.

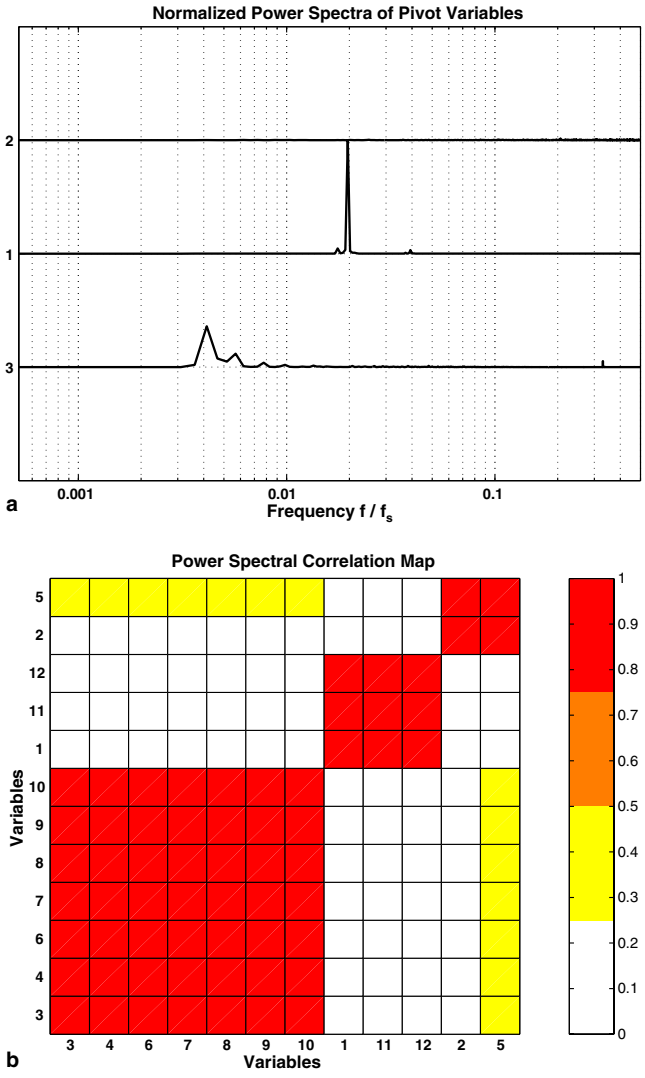

Fig. 6. Re-arranged PSCMAP and the spectral shapes of the pivot variables for the filtered Entech data set. (a) Spectral shapes of the pivot variables from the filtered Entech data analysis and (b) PSCMAP (clustered) of the filtered Entech data.

oscillation. The difference now in the post-filtering scenario is that tag 5 is no longer a member of the first group. Instead, tag 5 exhibits a strong spectral correlation with tag 2, which has a "flat" spectrum as shown in Fig. 6(a), tag 5 should also be expected to contain a similar shape, implying that it does not contain any other significant frequency besides the 0.002 cycles/sample oscillation. Observe that tag 5 is weakly correlated with the members of the first group, which is explained by the spectral shapes of tags 2, 3 and 5 shown separately in Fig. 5(b). A careful look at the spectral shape of tag 5 shows the presence of a very mild peak at the peak frequency of tag 3 (the pivot variable for the first group), but otherwise mostly flat.

From the foregoing discussion it is clear that what requires a deeper and careful visual observation of the multivariate spectra is easily highlighted by the PSCMAP. The results obtained in the paper are comparable with those obtained using spectral PCA [7], however, with simpler computation and improved interpretability. The interpretation with PSCMAP is simpler because the basis vectors obtained in SPCA do not generally carry any resemblance with the original spectra.

One could perform a band-pass filtering on these measurements so that the tag 5 is free of noise at higher frequencies and the 0.004 cycles/sample oscillation is highlighted. This is not addressed in this paper as the focus is on the efficiency of the PSCMAP.

In the second case study, a process with increased complexity and a higher dimensionality is considered.

4.2. Case study II: Refinery process

A simplified schematic of the refinery process is shown in Fig. 7. The data set consisting of 512 samples of 37 measurements sampled at 1 min interval, comprises measurements of temperature, flow, pressure and level loop measurements along with some composition measurements from the gas analyzers. The process contains a recycle loop from the PSA unit to the reformer unit. As in the earlier case study, controller errors (SP–PV) are analyzed for control loop measurements.

The spectral correlation colour maps for this process before and after re-arrangement are shown in Figs. 8 and 9. The effectiveness of the clustering algorithm is clearly demonstrated in these figures.

4.2.1. Analysis

The spectra of the pivot variables obtained for this process are displayed in Fig. 10. Among the pivot variables, only tags 2, 35, 14 and 32 have distinct peaks at different frequencies. The spectral power of tag 5 can be seen to be spread mainly in the lower frequencies. Additionally, tag 17 has among other peaks, some mild peaks common to those of tag 2, which is clearly indicated in the re-arranged PSCMAP (Fig. 9) as mild correlation between the group of tags 17, 18 and the members of the first group. The PSCMAP also indicates weak/mild correlation between the spectrum of tag 30 (one of the pivot variables) and the members of the first group, indicating that the tag 30 has a weak/mild oscillation at the frequency corresponding to the first group. This can be noticed in the tiny peak in the spectrum of tag 30, which could have gone otherwise unnoticed. A similar phenomenon can be noticed in the case of tag 1. The results of this analysis are summarized in Table 1.

An additional point to note is that the first group contains some variables that are strongly correlated with the pivot variable, but mildly correlated with each other. These variables are the sets of tags 8, 15, 16 and 28, 33 respectively. This phenomenon is related to the last point noted in Section 3.3. Fig. 11 shows the spectra

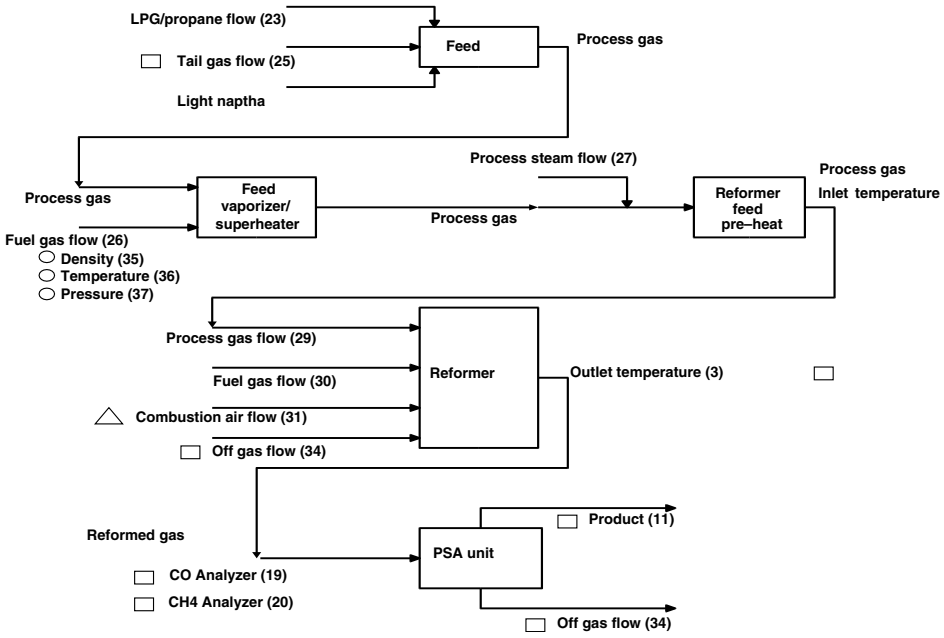

Fig. 7. Schematic of the refinery process.

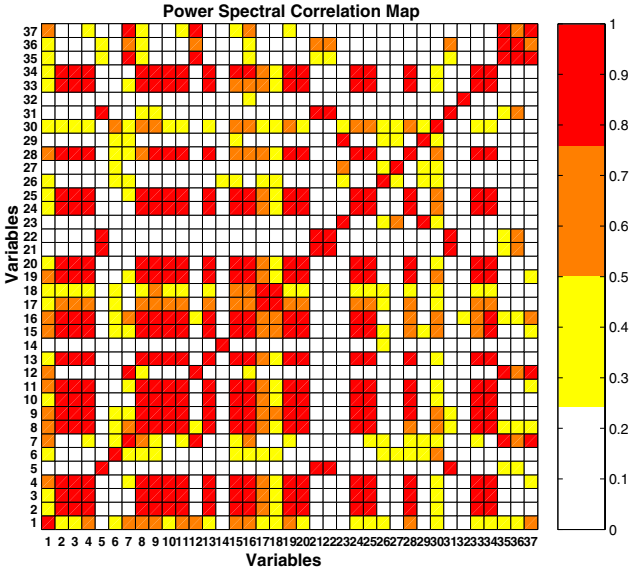

Fig. 8. PSCMAP of the refinery process shown in Fig. 7.

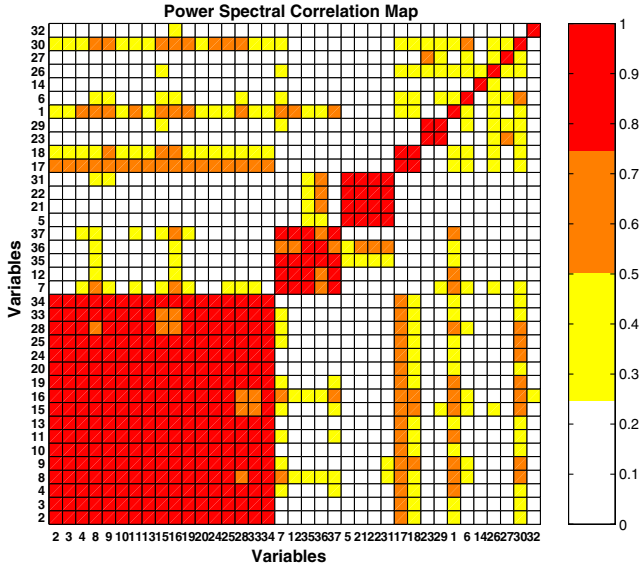

Fig. 9. Re-arranged PSCMAP of the refinery process.

of these set of tags along with the spectral shape of the pivot variable. The dominant oscillations in each of these variables are similar to the dominant oscillation in the pivot variable, i.e., tag 2 in this case. The correlation indices of tags 8, 15 and 16 with tags 28 and 33 are [0.73, 0.6921, 0.6713] and [0.7809 0.7221 0.7270] respectively. The stronger correlation of tag 8 with tag 33 than with tag 28 is because tag 28 contains some additional frequencies around 0.04 cycles/sample which are not present in tags 8 and 33. Similarly, the correlation between tags 15, 16 and tags 28, 33 is mild because of the presence of other significantly differing peaks in both these pairs.

Therefore, the proposed method is able to identify those set of variables which share a common dominant frequency. Using the filtering techniques employed in [7], further analysis can be performed to extract the correlation structure in other frequency regions. However, the purpose of this analysis is to only demonstrate the potential of PSCMAP. For a detailed analysis of this data, the reader is referred to [7], where the authors use SPCA and HDP as the key tools.

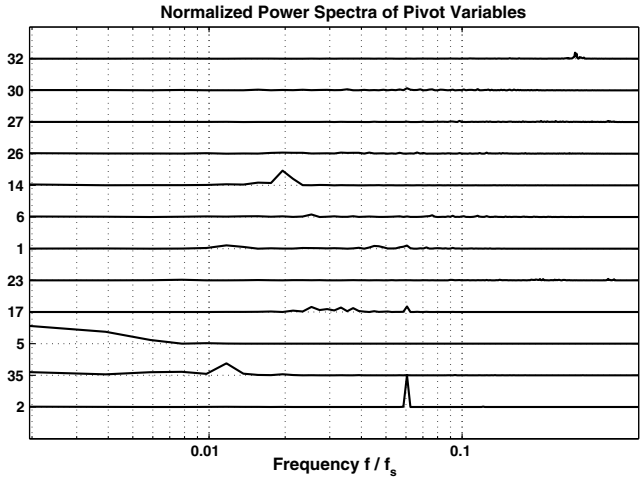

Fig. 10. Spectra of the pivot variables for the refinery process.

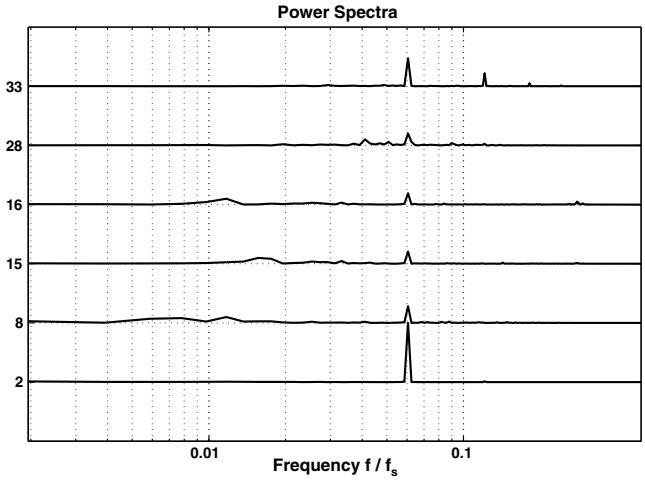

Fig. 11. Spectra of selected variables from the refinery process.

Table 1
Groups of variables with common dominant oscillations for the refinery process

Pivot variable	Group members	Comments
2	2, 3, 4, 8, 9, 10, 11, 13, 15, 16, 17, 19, 20, 24, 25, 28, 33, 34	Oscillating at 0.06 cycles/sample
35	7, 12, 35, 36, 37	Oscillating with frequencies around 0.01 (not pure sinusoids)
5	5, 21, 22, 31	Low frequencies
17	17, 18	Mixed frequencies with dominant peaks around 0.025 and 0.06 cycles/sample
23	23, 29	Mixed frequencies; no isolated peak(s)
1	Mild correlation with the first group	Mixed frequencies
6	Independent shape	Uniform power at all frequencies
14	Independent shape	Oscillation close to 0.02 cycles/sample
26	Independent shape	Uniform power at all frequencies
27	Independent shape	Uniform power at all frequencies
30	Mild correlation with the first group	Tiny peak at 0.06 cycles/sample
32	Independent shape	High frequency peak close to 0.3 cycles/sample

As in the earlier case study, the PSCMAP is able to effectively cluster the variables with common spectral shapes, i.e., common frequencies.

5. Conclusions

A new visualization tool, namely, the PSCMAP has been proposed to identify and extract the measurements from a large plant with several variables oscillating at similar frequencies. The visualization tool is based on a new measure, PSCI, which is a non-centered correlation between the spectra of measurements. The proposed method requires minimal computational complexity and effort. A simple clustering algorithm has been introduced to automate the re-arrangement of variables with similar correlation indices. The combined method has been shown to be able to extract a set of basis spectral shapes from a given data set and is most effective when there is one dominant frequency that influences one set of loops. The effectiveness of the method has been demonstrated on two industrial data sets. Given the large interaction and complexity in a multivariate process, added with the presence of noise, PSCMAP holds a lot of potential in its applications to plant-wide oscillation detection. Furthermore, PSCMAP is an analytical tool and therefore is better suited to assess the similarities in the spectra when compared to a visual-based assessment of the same. The industrial case studies have demonstrated this ability where the colour map is able to highlight the similarities that would have been easily missed by a visual inspection. When compared to the existing SPCA methodology, the proposed method gives basis shapes that are non-orthogonal, but provides easy interpretability of the similarities in the frequency-domain behaviour of measurements.

Finally it is re-iterated that spectral techniques are well-suited to the plant-wide oscillation detection problem, and do not require the knowledge of the process time delays or lags between different measurements.

Their simplicity and ease-of-use make them amenable to a practicing engineer who is concerned with the analysis of routine operating data in detecting the presence of oscillations and obtaining valuable insights into control loop performance.

Acknowledgement

The authors are grateful for the financial support of the Natural Science and Engineering Research Council (NSERC, Canada), Matrikon (Edmonton, Alberta) and the Alberta Science and Research Authority (ASRA) through the NSERC-Matrikon-ASRA Industrial Research Chair in Process Control.

Appendix A. Proof that PSC is independent of phase difference

Consider two sinusoidal signals x_1 and x_2 oscillating with identical frequency ω_0 but with a phase difference ϕ. Now, consider a third signal x_3, which is a linear combination of x_1 and x_2 such that $x_3 = a_1x_1 + a_2x_2$, where $a_1, a_2 \in \mathscr{R}$. From a plant-wide oscillation detection point of view, the objective is to find the basic spectral shape(s) that can explain the spectral behaviour of these three measurements. Since x_1 and x_2 have the same frequency, x_3 has the same frequency, which implies that all three measurements have similar spectral shapes (but with different power). Therefore, there is only one basic spectral shape that can explain the spectral behaviour of all three measurements. On the other hand, consider the relationship between these variables in the frequency domain. It can be shown that these variables are related as:

$$|X_2(\omega)|^2 = |X_1(\omega)|^2$$
$$|X_3(\omega)|^2 = a_1^2|X_1(\omega)|^2 + a_2^2|X_2(\omega)|^2 + 2a_1a_2|X_1||X_2|\cos\phi = \alpha|X_1(\omega)|^2$$

where α is a constant, i.e., independent of frequency. Therefore, it follows that the correlations between the pairs of $|X_1(\omega)|^2$, $|X_2(\omega)|^2$ and $|X_3(\omega)|^2$ is unity, which implies that PSC gives the correct number of basis shapes. More importantly, the phase difference ϕ has no effect on the number of basis shapes in the spectra of the above three variables.

Appendix B. Remarks on use of DFT for process measurements

Practical signals are a mixture of both the deterministic and stochastic components. The deterministic component is due to contributions from process phenomena, while the stochastic component contains contributions from measurement noise, process noise, etc. Oscillation detection problem is concerned with (i) estimation of spectrum and (ii) estimating the frequencies at which the peaks occur in the spectra. A few important questions that arise in this context are: (i) *Is the squared magnitude of DFT still a good estimate of the power spectrum*? (ii) *if yes, what is a good choice of window*? and (iii) *how good are the estimates of the frequencies corresponding to the spectral peaks*?

Estimation of spectrum as well as frequencies of oscillatory signals corrupted with noise have been widely studied in literature (see [9,10] for a good treatment). The answer to whether squared magnitude of DFT is a reliable estimate of the spectrum greatly depends on the signal-to-noise ratio (SNR) of the measurement. Process measurements such as those dealt in this paper usually contain a strong oscillation and therefore the SNR is usually high enough to warrant the use the squared magnitude of DFT to estimate the power spectrum.

A condition that is used to determine the safe usage of squared magnitude of DFT to estimate the peak frequencies of sinusoids corrupted with noise [9] is given as,

$$\frac{NA^2\mathrm{PG}}{\sigma_\epsilon^2} \geqslant 100 \tag{B.1}$$

where N is the number of data samples, A is the amplitude of the sinusoid and therefore, with variance $A^2/2$, PG is known as the processing gain of the window and σ_ϵ^2 is the variance of the discrete-time noise corrupting the signal. The processing gain of the window is the ratio of the SNR of the measurement after windowing to the SNR of the actual measurement. For rectangular windows (regular Fourier transform of finite signals), the PG corresponding to DFT with frequency resolution $2\pi/N$ is 0.4. If $N = 2048$, then the above equation implies that $A^2/\sigma_\epsilon^2 \geqslant 100/(2048 \times 0.4) = 0.12 \Rightarrow \mathrm{SNR} = A^2/2\sigma_\epsilon^2 = 0.06$, which is a small threshold value that holds for most process measurements.

A different approach with a similar perspective that warrants the issue of the use of DFT for process data analysis is given below.

Deterministic signal: A deterministic oscillating waveform $x(t)$ with a period of oscillation T_p, amplitude of $2m$ and N samples has a unique value for its DFT. The DFT will be arranged over N frequency channels of which $N/2$ are meaningful, up to the Nyquist frequency. The value of $|X(f)|$ will be mN for the channel whose frequency is $f_\mathrm{p} = 1/T_\mathrm{p}$ and also for its alias, and zero in the other channels. The power in that channel is $\frac{1}{N}|X(f)|^2 = m^2N$ assuming one uses the following definition of power (there are others):

Total signal energy =

$$\sum_{\text{frequency channels}} \text{spectral power in each channel}$$

The point about the deterministic waveform is that the DFT is also deterministic, there are no statistical confidence limits, the answer is correct and exact.

Stochastic signal: A random signal such as uncorrelated random noise has spectral power in all frequency channels and should in theory have a flat spectrum. Suppose its r.m.s. value is m, then $|X(f)| = mN$ for every frequency channel. However, its DFT usually shows a lot of variability because of the stochastic nature of the signal. $|X(f)|$ has an expected value of $m\sqrt{N}$ and standard deviation of $m\sqrt{N}$ [11].

Mixed signal: If a signal has both deterministic and random components then the extent to which the DFT is uncertain depends on the balance of the two components. Consider a sine wave of amplitude $2m$ with added random noise of r.m.s. value m. That oscillating signal would be hard to see by eye from the time trend and any analyst would say the signal was stochastic. The $|X(f)|$ for the mixed signal has magnitude $m\sqrt{N}$ in all channels except for the $f_p = 1/T_p$ channel where the magnitude is $mN + m\sqrt{N}$. The deterministic signal therefore stands out really well above the noise. The DFT of the noise has stochastic variability but the standard deviation in the frequency channel where the signal resides is negligible for any practical value of N (typically $\pm m\sqrt{N}$ on a quantity whose magnitude is $mN + m\sqrt{N}$).

Use of the Welch transform: The Welch transform smooths the spectrum at the expense of worse frequency resolution. The effect of a Welch transform on a pure sine wave is to spread the spectral power across a broader frequency channel. For instance, if the spectrum has been smoothed from 2048 channels to 64, then the resolution ratio is $L = 2048/64 = 32$ and the frequency channels are all 32 times as broad. Therefore it is more difficult to be sure of the true frequency of the deterministic signal. Also, for the deterministic signal (sine wave of amplitude $2m$), the magnitude of $|X(f)|$ reduces from mN to mN/L because it is now spread over a wider frequency channel. The Welch transform reduces the uncertainty in the $|X(f)|$ of a stochastic signal but does not help reduce uncertainty in a mixed signal. For a random signal of r.m.s. value m the value of $|X(f)|$ remains as $m\sqrt{N}$ but the standard deviation in $|X(f)|$ is reduced from $\pm m\sqrt{N}$ to $\pm m\sqrt{N}/L$. Therefore the magnitude of $|X(f)|$ for the mixed signal in the $f_p = 1/T_p$ channel is $mN/L \pm m\sqrt{N}$ so the stochastic contribution is bigger than with the DFT, and the standard deviation is $\pm m\sqrt{N}/L$ which is a worse percentage result than for the DFT. Therefore for a mixed signal with a deterministic component, the DFT is more reliable than the Welch transform.

References

[1] N.F. Thornhill, T. Hagglund, Detection and diagnosis of oscillation in control loops, Contr. Eng. Pract. 5 (1997) 1343–1354.

[2] S.J. Qin, Control performance monitoring—a review and assessment, Comput. Chem. Eng. 23 (1998) 173–186.

[3] L. Desborough, R. Miller, Increasing customer value of industrial performance monitoring—Honeywell's experience, Preprints of CPC6, 2001, pp. 153–186.

[4] N.F. Thornhill, S.L. Shah, B. Huang, Detection of distributed oscillations and root-cause diagnosis. in: CHEMFAS4, Cheju Island, Korea, July 2001.

[5] A. Lankford, Online torque monitoring system reveals unexpected causes of machine problems, Pulp Paper 71 (12) (1997).

[6] M.A.A.S. Choudhury, S.L. Shah, N.F. Thornhill, Diagnosis of poor control loop performance using higher order statistics, Automatica 40 (10) (2004) 1719–1728.

[7] N.F. Thornhill, S.L. Shah, B. Huang, A. Vishnubhotla, Spectral principal component analysis of dynamic process data, Contr. Eng. Pract. 10 (2002) 833–846.

[8] N.F. Thornhill, B. Huang, H. Zhang, Detection of multiple oscillations in control loops, J. Process Contr. 13 (2003) 91–100.

[9] Boaz Porat, A Course in Digital Signal Processing, John Wiley & Sons, New York, USA, 1997.

[10] M.S. Priestley, Spectral analysis and time seriesProbability and Mathematical Statistics, vol. I and II, Academic Press, New York, USA, 1981.

[11] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Receipes, Cambridge University Press, Cambridge, 1998.