Karhunen-Loéve (PCA) based detection of multiple oscillations in multiple measurement signals from large-scale process plants

P. F. Odgaard and M. V. Wickerhauser

***Abstract*— In the perspective of optimizing the control and operation of large scale process plants, it is important to detect and to locate oscillations in the plants. This paper presents a scheme for detecting and localizing multiple oscillations in multiple measurements from such a large-scale power plant. The scheme is based on a Karhunen-Loève analysis of the data from the plant. The proposed scheme is subsequently tested on two sets of data: a set of synthetic data and a set of data from a coal-fired power plant. In both cases the scheme detects the beginning of the oscillation within only a few samples. In addition the oscillation localization has also shown its potential by localizing the oscillations in both data sets.**

I. INTRODUCTION

In optimizing large-scale process plants the focus has mainly been put on optimizing the nominal performance. Often the second step is to detect and accommodate non-intend behavior such as faults in the plant. An example on such a large-scale-plant is a coal-fired power plant. These non-intended behaviors can have many different appearances. One of these behaviors is oscillating elements. Such non-intended oscillations in large-scale process plants can be problematic. These can be a result of a failure in a part of the plant which can spread out to the entire plant. It is as consequence highly important to detect such oscillations. Unfortunately, detection of these oscillations is not as easy as it seems. The measured signals contain numerous other signals parts than just the possible oscillations, some examples on other non-expected part signal components are noise and disturbances.

The research in this area has been focused on time and/or statistic based oscillation detection methods. Most of it is dealing with sequentially analysis of one signal. In [1] zero crossings are used to detect oscillations, by integrating the absolute control error of a number of intervals between zero crossings. A method based on the auto-correlation function was presented in [2]. All these methods analyze the different measurements independently, and are not designed to detect multiple oscillations. However, some examples on schemes for detecting multiple oscillations are: [3], [4], [5], [6] and [7].

P.F. Odgaard is at KK-electronic a/s, Pontoppidanstræde 101, DK-9220 Aalborg East, Denmark, `pfo@kk-electronic.dk`, during the research for this paper he was at Department of Electronic Systems, Aalborg University, Denmark

M. V. Wickerhauser is at Department of Mathematics, Washington University, One Brookings Drive, St.Louis, MO 63130, USA, `victor@math.wustl.edu`

Normally a high number of measurements are generated from a large-scale power/process plant. Meaning that measurements are available from different places in the plant. It might be preferable to base the oscillation detection on the entire set of measurements or a subset of these, instead of analyzing these measurement signals independently, as the previous methods suggest. The approach has a number of advantages. More than one oscillation might occur at a given time in a large-scale process plant. However, the ratio of energy in the different oscillations can be assumed to be non constant in the different measurements, due to different dynamic behavior of the parts of the plant. This means that it is possible to separate the different oscillations. This can also be used to localize the signal in which the given oscillations are occurring and suggest candidates for the cause of the oscillation.

Another way to extract multiple signal trends (such as oscillations) from a set of signals is to compute a Karhunen-Loève basis of the data set. In this terminology signal trends and oscillations are strongly related, in way that an oscillation is a periodic signal trend. Assume that the analysis of the signals are performed on sequences of samples which maximal contain a few periods of the oscillation. Consequently signal trends and oscillations can be assumed to be non-separable.

The general trends in the signal set will be approximated by only a few most approximating vectors in the Karhunen-Loève basis, see [9] and [10]. In [11] and [12] a method is developed for oscillation localization by analyzing the measurements shifted to the frequency basis. This is done based on the assumption that an oscillation has been detected. In this work the principal component analysis is used, which is method strongly related to the Karhunen-Loève basis analysis.

However, using a Karhunen-Loève basis of the time domain signals can be used to detect the oscillations as well as localize the oscillation. The localization is done by finding a set of candidate signals representing the possible roots of the oscillation. This method will succeed if the oscillations are the dominating components in the analyzed signals. I.e. the remaining signal components can either be neglected or are not repeated through the signals. This idea is similar to dynamic PCA, e.g. see [8].

In this paper the plant is first defined and described. The subsequent section describes the Karhunen-Loève basis and the computation of it. This leads to an oscillation detection scheme presented in this paper, as well as a scheme for localizing oscillating elements. The oscillation detection and localization schemes are subsequently applied to a couple of examples. In the end a conclusion is drawn.

II. System Description

The system in mind in this paper, is a large-scale power plant, where oscillations are required to be detected directly from the sensors signals measured at different locations at the plant. However, the performance of the oscillation detection is highly decreased if the signals contain transients like step response etc. This means that in most cases it is preferable to feed the oscillation detection algorithm with residuals of the measurements and estimated sensor values. In order to reach the potential of this algorithm, the detection algorithm requires a number of different measurements. In which it is assumed that the repeated signal components from the signal is the trends (oscillations) in the signal, which the algorithm shall detect. Phase delays between these measured signals are not considered in this work. It is also assumed that the remaining signal components are disturbances and noise. The time sequences of these are not repeated in the different sensor signals. This means that the system can be described by (1).

$$\mathbf{y}[n] = \sum_{0<i<m} \mathbf{y}_{\mathrm{o},i}[n] + \mathbf{n}_{\mathrm{m}}[n], \tag{1}$$

where $\mathbf{y}[n] \in \mathcal{R}^k$ is a vector of the analyzed signal at the discrete time n, k is the number of sampled and analyzed measurements, $\mathbf{y}_{\mathrm{o},i}[n]$ is the i'th vector of oscillating signal component, $\sum\limits_{0<i<m} \mathbf{y}_{\mathrm{o},i}[n]$ is the sum of vectors of the oscillation signal components, and $\mathbf{n}_{\mathrm{m}}[n]$ is a vector of the remaining signal components, which are uncorrelated and are assumed to be normal distributed with zero mean and the variance d.

In the following the measured data will be analyzed as a block of data, with data sampled at time n as the newest elements. If the length of the data block is denoted M, the data block $\mathbf{Y}[n]$ can be defined as in (2).

$$\mathbf{Y}[n] = \begin{bmatrix} y_1[n-M+1] & y_1[n-M+2] & \cdots & y_1[n] \\ y_2[n-M+1] & y_2[n-M+2] & \cdots & y_2[n] \\ & \vdots & & \\ y_k[n-M+1] & y_k[n-M+2] & \cdots & y_k[n] \end{bmatrix}, \tag{2}$$

in which all the k measurement are stored for the last M samples. I.e. the first row contains the M most recent samples of measurement 1, the second row contains the M most recent samples of measurement 2, etc.

III. Karhunen-Loève analysis

The core of this algorithm is the Karhunen-Loève analysis which consists of computing a Karhunen-Loève basis and its related eigenvalues. Given a matrix, $\mathbf{Y}$, of k row vectors in $\mathcal{R}^m$, where $k > m$, the Karhunen-Loève basis minimizes the average linear approximation error of the vectors in the set, [9]. Another advantage of the Karhunen-Loève basis is that it approximates the general structures of all the signals in $\mathbf{Y}$ with just a few basis vectors, see [9] and [10] and [13].

The Karhunen-Loève basis is computed based on $\mathbf{Y}$. First of all it is assumed that the row vectors in $\mathbf{Y}$ has zero mean, if not a preliminary step is introduced in order to fulfill that assumption. The Karhunen-Loève basis, $\mathcal{K}$, can be defined as

$$\mathcal{K} = \{v_1, \cdots, v_m\}, \tag{3}$$

$\mathcal{K}$ is an orthonormal basis of eigenvectors of the matrix $\mathbf{YY}^T$, ordered in such a way that v_n is associated with the eigenvalue λ_n, and $\lambda_i \geq \lambda_j$ for $i > j$. A matrix of the basis vectors can following be defined as

$$\mathbf{K}_{\mathrm{L}} = \left[v_1, \cdots, v_m\right]. \tag{4}$$

In other words the Karhunen-Loève basis is the eigenvectors of the autocorrelation of $\mathbf{Y}$. The eigenvalues of the autocorrelation takes the values of the variances of the related Karhunen-Loève basis vectors. The approximating properties of the Karhunen-Loève basis vectors are sorted in increasing order, that means that if the basis consists of p vectors the basis vector p is the most approximating basis vector, and the corresponding eigenvector has the largest value. In addition the general structures in all the vectors in $\mathbf{Y}$ are represented by only a few basis vectors. The remaining basis vectors represent the signal parts which are not general for $\mathbf{Y}$, i.e. noise in the signal, etc.

A high eigenvalue means that the corresponding eigenvector supports a large part of the energy of the row vectors in the matrix $\mathbf{Y}$. This means that the basis vectors (eigenvectors) which approximate general trends in the data set have eigenvalues of high numerical value, and the remaining basis vectors which support the noise or non-repeated signal components have eigenvalues of low numerical values. This can be illustrated by Example III.1.

Example III.1 *Given a matrix of data with 20 measurements, each with 10 samples. 15 of the measurements contain only noise, the remaining 5 measurements contain as well an oscillation of a given frequency. Applying a Karhunen-Loève analysis on this matrix of data will result in: one basis vector supporting the general trend, which is the oscillation, and the corresponding eigenvalue would take a high numerical value. All the remaining eigenvectors would support the noise in the data matrix and their related eigenvalues would take low numerical values.*

The opposite situation is illustrated by Example III.2.

Example III.2 *Given matrix of data with 20 measurements, each with 10 samples. All the measurement vectors* *contain only noise. An applied Karhunen Loève analysis of this matrix would result in basis vectors supporting the noise and eigenvalues would be close in numerical values to each other.*

IV. Method

By the definition of the Karhunen-Loève basis, it would approximate oscillations and other signal trends with only a few bases vectors, given the basis supporting the trends in the sampled data set.

A. Oscillation Detection

A Karhunen-Loève analysis is subsequently performed on the data matrix $\mathbf{Y}[n]$, meaning that the analysis is performed on the M most recent samples of each of the measurements. The outcome of this is a basis contained in the matrix $\mathbf{K}_{\text{L}}[n]$, and $\lambda[n]$ which is a vector of the eigenvalues corresponding to basis vectors in $\mathbf{K}_{\text{L}}[n]$. $\lambda[n]$ contains information of the occurrence of oscillations or other general signal trends in $\mathbf{Y}[n]$. A way to transform this information into a scheme for detecting oscillations is to compare the variance of $\lambda[n]$ with a threshold ϵ. The variance of $\lambda[n]$ will depend on the number of high numerical valued elements in $\lambda[n]$. I.e. the occurrence of one or more oscillations or other general trends in $\mathbf{Y}[n]$ will result in a variance of $\lambda[n]$ which is higher than ϵ, whereas the variance of $\lambda[n]$ will be lower than ϵ in the case of none oscillations in the data matrix, since the energy in the signals will be supported by numerous basis vectors resulting in eigenvalues close to each other in numerical values. As in Example III.2, the variance will consequently be close to zero. ϵ depends on the number of oscillations in the signals, meaning it is needed to adapt the threshold to the given application. This means that a detection can be formulated like in (5). $O_{\text{d}}[n]$ is a detection signal, which takes the value 1 if an oscillation is detected and the value 0 if not.

$$O_{\text{d}}[n] = \begin{cases} 1 & \text{if } \operatorname{var}(\lambda[n]) \geq \epsilon, \\ 0 & \text{if } \operatorname{var}(\lambda[n]) < \epsilon \end{cases}. \tag{5}$$

B. Extraction the oscillations as well as the number of oscillations

In terms of finding the number of oscillations in the data matrix $\mathbf{Y}[n]$ the attention is again addressed to $\lambda[n]$. The elements in $\lambda[n]$ would as stated previously have high numerical values if their corresponding eigenvectors contain a general trend from the data matrix. The task can be done counting the number of elements in $\lambda[n]$ which takes high numerical value. In the vector $\lambda[n]$ the elements are sorted in increasing order. Meaning that the most approximating vector is the last one and hereby corresponds to the last element in $\lambda[n]$, and the second most approximating basis vector is the second last one, and hereby corresponds to the second last element in $\lambda[n]$, etc.

This means that the number of oscillations in the data can be found by stepping back through $\lambda[n]$ as long as the elements in $\lambda[n]$ are higher than a threshold κ. The threshold κ also need to be determined based on the application. This can be formulated as an algorithm, in which j is a counting variable, N_{o} is the number of oscillations in the data, and $\lambda_m[n]$ denotes the m'th element in $\lambda[n]$. The algorithm for counting the number of oscillations in the data can be written as

$j = 0$
$N_{\text{o}} = 0$
WHILE $\lambda_{m-j}[n] \geq \kappa$,
  $j = j + 1$
  $N_{\text{o}} = N_{\text{o}} + 1$
END.

The ordering of $\lambda[n]$ automatically gives the basis vectors supporting the oscillations. The oscillations in the data are supported by the N_{o} last column vectors in $\mathbf{K}_{\text{L}}[n]$. The set of basis vectors supporting the oscillation can be extracted to those shown in (6).

$$\{\mathbf{v}_{m-N_{\text{o}}+1}[n], \cdots, \mathbf{v}_m[n]\}, \tag{6}$$

These vectors supporting the oscillations in the plant, might be very useful for the operators and engineers, who have to handle the oscillations in the system. Since these vectors show the oscillations, the operator might get a better understanding of the problem causing the oscillations. Additional signal analyzes can help localizing some candidates as origin of the oscillation, but that is not always enough for finding the cause of oscillation.

E.g. the problem might be localized to be in a given pump, but does not say anything about the cause of the oscillation in the pump. However, seeing the basis vector supporting the oscillation might help the operators and engineers understanding the cause of the problem in this given pump.

C. Localization of oscillations

The oscillations can be located in the measurements by computing the inner product of each vector of measurements with the basis vectors supporting the oscillations, see (7).

$$\mathbf{C}_{\text{o}} = \mathbf{Y} \cdot \begin{bmatrix} \mathbf{v}_{m-N_{\text{o}}+1}[n] & \cdots & \mathbf{v}_m[n] \end{bmatrix}, \tag{7}$$

where $\mathbf{C}_{\text{o}}$ is a matrix of the oscillation coefficients. The row vectors in $\mathbf{C}_{\text{o}}$ contain the different coefficients for each measurement.

High values of these inner products mean that oscillations are present in the given signal. However, due to gains and feedback loops in the plant, the highest coefficient of a given oscillation does not mean that origin of the oscillation is at the location of the corresponding measurement.

A way to use these coefficients of oscillations in the given measurement signal, is to take out a number of signals with the largest coefficients and view these as candidates causing the oscillation. The next step is to use knowledge of the plant to rule out some candidates for causing the oscillation. It could be to determine if any of the candidates depend directly one of the other candidates etc.

This means that localization scheme is given as follows

1) Compute the oscillation coefficients: $\mathbf{C}_o = \mathbf{Y} \cdot \left[\mathbf{K}_L[n]\{:, m - N_o + 1\} \quad \cdots \quad \mathbf{K}_L[n]\{:, m\}\right]$.
2) For each oscillation find the p measurements with the highest coefficients by sorting the column vectors of $\mathbf{C}_o$.
3) Eliminate elements in the candidate groups based on knowledge of the plant. E.g. if oscillations are detected in both part A and B where B directly depends on A, then eliminate B from the candidate list.

This algorithm leads to a group of candidates of the origin of each oscillation in $\mathbf{Y}[n]$.

D. Data preprocessing

Before the detection and localization algorithm is applied to the data matrix. Each measurement (in each row vectors), shall be preprocessed in order to achieve zero mean of the row vectors, and normalized by the maximum possible value of each measurement. The first requirement is due to the Karhunen-Loève analysis, the second is in order to localize the root of the oscillation.

V. EXAMPLES

In this section the oscillation detection method is applied to two data sets: a set of synthetic data and a set of data from a coal-fired power plant.

A. Detecting harmonics in random noise

The data set consists of 40 vectors each of 600 samples. One sine signal (2.2 Hz sampled at 16Hz) signal is occurring in the first 10 vectors and another sine signal (3.1 Hz sampled again at 16 Hz) is occurring in the next 10 vectors. All these oscillations have different amplitudes. Random noise added to all the elements in the data matrix. Both oscillations start at sample 101 and they reach their maximal amplitude at sample 150, see Fig. 1.

The detection algorithm is subsequently applied to this set of synthetic data. In this detection the window length has been set to 14. (Notice it is the window length, which is required to be smaller than the number of measurements and not the length of the entire data block). The variance of $\lambda[n]$ for this given data set can be seen in Fig. 2. It can be seen that the oscillations are detected just a few samples after the beginning of the oscillations. The oscillations are detected at sample 106.

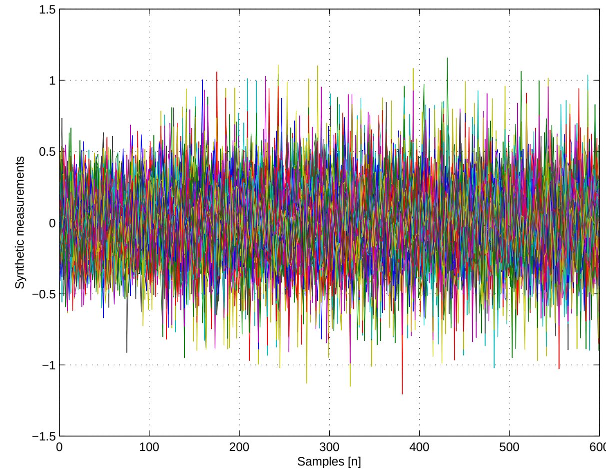

Fig. 1. Plot of the synthetic measurement data. The start of the oscillations at sample 101 is slightly visible.

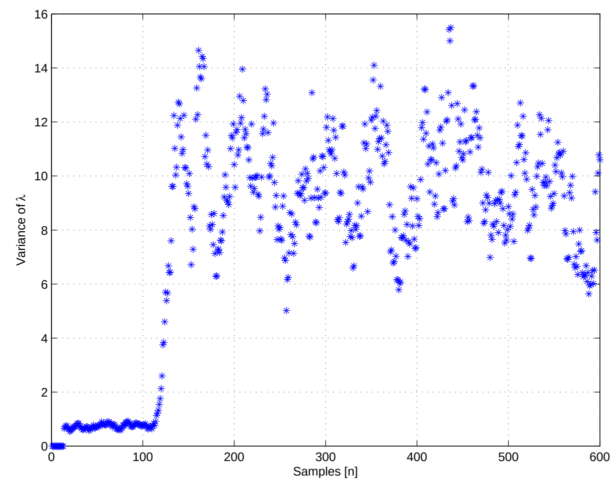

Fig. 2. Plot of the variance of $\lambda[n]$, for the synthetic data. It can be seen that the oscillations are detected just a few samples after the oscillations start, since the oscillations are detected at sample 106.

The two basis vectors supporting the two oscillations in the data set can be seen in Fig. 3. From this plot the sine signals are easily identified. I.e. the algorithm can detect these multiple oscillations.

The localization algorithm was tested as well on this data set, and the result was clear. The algorithm sorted the measurement correct such that the 10 measurement vectors containing the first sine signal were grouped in one group, the 10 measurement vectors containing the second sine signal was as well grouped in another group.

B. Detecting oscillations in data from a coal-fired power plant

In this example a data set from a coal-fired power plant is used. An oscillation starts approximately at sample 90, somewhere in a coal mill in the plant, 85 measurements

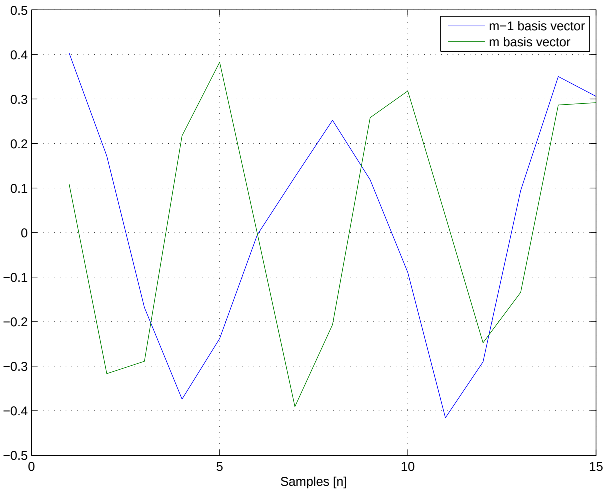

Fig. 3. Plot of the two basis vectors supporting the oscillation, when the oscillations are occurring at their maximum strength.

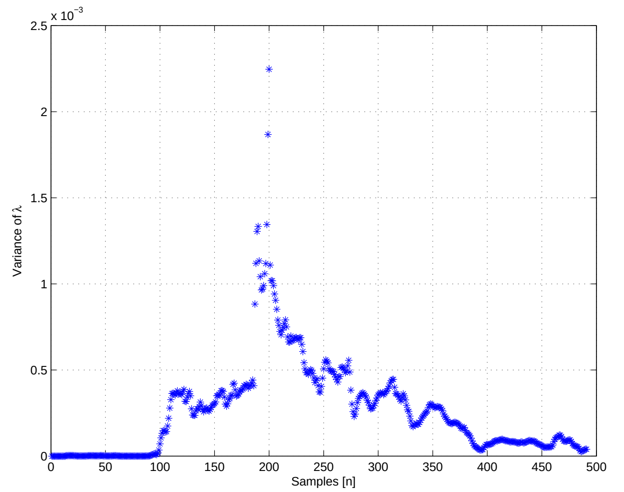

Fig. 5. Plot of the variance of λ for the power plant data. It can be seen that the oscillation is detected at sample 94.

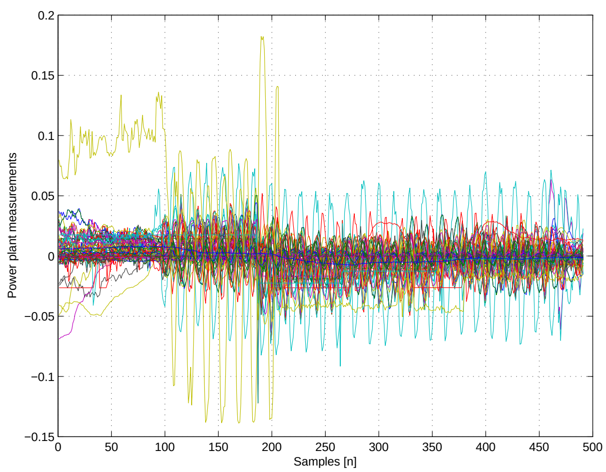

Fig. 4. Plot of the measurement data from a coal fired power plant.

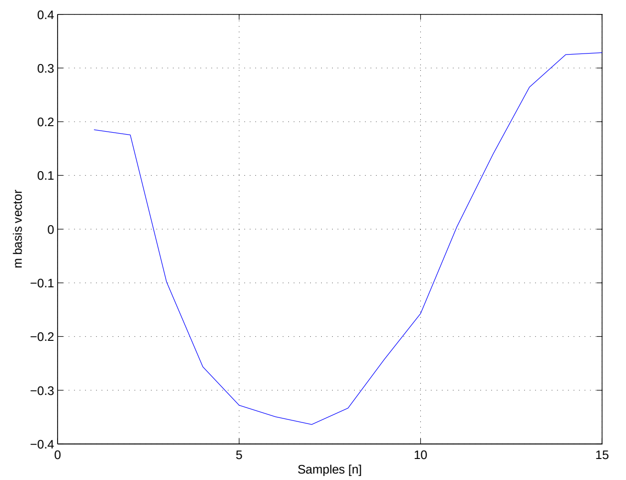

Fig. 6. Plot of the basis vector supporting the oscillation in the measurements.

are provided from different parts of the plant. The data set is prepared for the oscillation detection by removing mean values of the data vectors and scale each data signal by the maximal value of the corresponding sensor. The data set can be seen in Fig. 4.

The oscillation detection scheme is subsequently applied to the data set by using data window lengths of 14 samples I.e. $M = 14$. The achieved variance of λ can be seen in Fig. 5. From this plot it can be seen that the oscillation is detected at sample 94 this is only a few samples later than where the oscillations start to raise in the signals. It is also worth mentioning that the oscillation detection algorithm does not give any false detection. In the meaning that it does not give indications of an oscillation at time locations where no oscillations are occurring. The basis vector supporting the oscillation at sample 200 can be seen in Fig. 6. This figure shows a harmonic signal, which indicates that a part of the plant is oscillating.

The next step is to locate the cause of the oscillation. The oscillating coefficients are shown in increasing order in Fig. 7, together with measurement numbers. From the figure it can be seen that there is one clear candidate for the cause of the oscillation. The measurement with the largest oscillation coefficient, is a power measurement of the motor in one of the coal mills which pulverize the coal before it is blown into the furnace in the power plant. This is the same conclusion given by a more detailed analysis and physical inspection performed on the power plant in order to find the root of the oscillation.

This example shows the potential of this algorithm to detect oscillation in a data set containing measured data from a power plant, and it can locate a number of candidates for being the cause of the oscillation. Even though

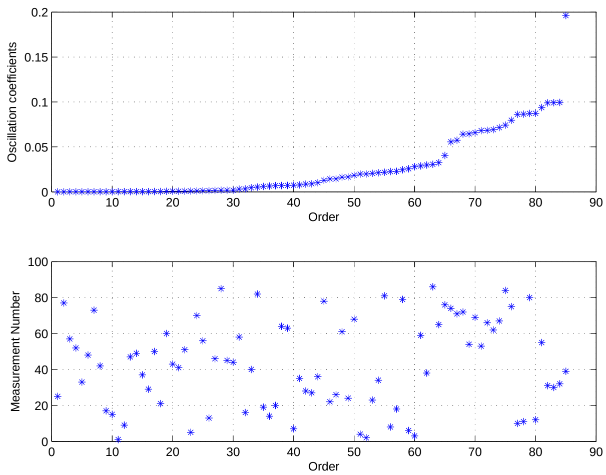

Fig. 7. Plot showing the grouping of signals based on the oscillation. The upper plot shows the oscillation coefficients in sorted order. The lower plot shows the measurement numbers in the same order, meaning that signal with the largest oscillation coefficient is the one to the right.

these examples have shown that the localization algorithm directly localizes the root of the oscillations, this might not be the case for all data sets. Consequently the algorithm should be used to generate a set of candidates, which can be inspected according to their localization order.

VI. Conclusion

In this paper a method for detecting and locating multiple oscillations in data sets from large-scale process plants with multiple measurements are presented. The method uses a Karhunen-Loève analysis to detect and localize the multiple oscillations in the data. The method is subsequently tested on two sets of data. A synthetic set containing two different oscillations, and a data set from a coal-fired power plant. In both cases the oscillation detection scheme detects the oscillation within a few samples after their beginnings. The localization scheme has as well localized the oscillations in both data sets.

VII. Acknowledgment

The authors acknowledge the Danish Ministry of Science Technology and Innovation, for support to the research program CMBC (Center for Model Based Control), grant no 2002-603/4001-93.

VIII. References

[1] T. Hagglund, "A control-loop performance monitor," *Control Engineering Practice*, vol. 3, pp. 1543–1551, November 1995.

[2] T. Miao and D. Seborg, "Automatic detection of excessively oscillatory feedback control loops," in *IEEE International Conference on Control Applications*, Hawaii, USA, August 1999, pp. 359–364.

[3] N. Thornhill, S. Shah, B. Huang, and A. Vishnubhotla, "Spectral principal component analysis of dynamic process data," *Control Engineering Practice*, vol. 10, pp. 833–846, August 2002.

[4] M. Paulnis and J. Cox, "A practical approach for large-scale controller performance assessment, diagnosis and improvement," *Journal of Process Control*, vol. 13, pp. 155–168, 2003.

[5] N. Thornhill, J. Cox, and M. Paulonis, "Diagnosis of plant-wide oscillation through data-driven analysis and process understanding," *Control Engineering Practise*, vol. 11, pp. 1481–1490, December 2003.

[6] C. Xia and J. Howell, "Isolating multiple sources of plant-wide oscillations via spectral idependent component analysis," *Control Engineering Practice*, vol. 13, pp. 1027–1035, August 2005.

[7] X. Zang and J. Howell, "Isolating the source of whole-plant oscillations through biamplitude ratio analysis," *Control Engineering Practice*, vol. 15, pp. 69–76, January 2007.

[8] B. Wise and N. Gallagher, "The process chemometrics approach to process monitoring and fault detection," *Journal of Process Control*, vol. 6, pp. 329–348, 2003.

[9] S. Mallat, *A wavelet tour of signal processing*, 2nd ed. Academic Press, 1999.

[10] M. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, 1st ed. A K Peters, Ltd., 1994.

[11] N. Thornhill, S. Shah, and B. Huang, "Detection and diagnosis of unit-wide oscillations," in *Process Control Implementation 2000*, Glasgow, UK, July 2000.

[12] ——, "Detection of distributed oscillations and root-cause diagnosis," in *IFAC CHEMFES4*, Korea, June 2001, pp. 167–172.

[13] N. Saito, "The least statistically-dependent basis and its applications," in *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, November 1998, pp. 732–736.