UNIFYING PCA AND MULTISCALE APPROACHES TO FAULT DETECTION AND ISOLATION

Seongkyu Yoon[1] and John F. MacGregor[2]

Dept. Chemical Engineering, McMaster University, Hamilton Ontario Canada L8S 4L7
[1] yoons@mcmaster.ca
[2] macgreg@mcmaster.ca

Process signals represent the cumulative effects of many underlying process phenomena. Multiresolution analysis is used to decompose the cumulative process effects. The decomposed process measurements are rearranged according to their scales, and PCA is applied to these multiscale data to capture process variable correlations occurring at different scales. Choosing an orthonormal mother wavelet allows each principal component to be a function of the process variables at only one scale level. The proposed method can identify when a multiscale approach is needed. The conventional PCA as well as MSPCA models are shown as the limiting cases of the proposed model. A procedure for both fault detection and isolation is presented. The proposed method is discussed and illustrated in detail using simulated data from a CSTR system. A comparison study is done through Monte Carlo simulation. The proposed method significantly enhances FDI performance by using additional scale information.

Keywords: Fault detection, Fault isolation, Fault, Scales, PCA, Multiresolution analysis

1. INTRODUCTION

Process measurement signals usually represent the cumulative effects of many underlying process phenomena such as process dynamics, measurement noise, external disturbances, process degradation etc. Each effect manifests on a different scale that is a reciprocal of frequency. Faults occurring at different locations, times and frequencies are considered one of the contributing events. In case of a complex fault, its signature is very likely to be confused with multiscale nature of the process signals since the effect of the complex fault is propagated into other measurements through process dynamics and control actions.

One can decompose process signals such that all the contributing events are approximately discriminated according to their scale contents and a fault is distinguished from other events. The decomposition is done with the multiresolution analysis via wavelet transformation which characterizes signal in the domain of time and scale (Mallat, 1989; Daubechies, 1990). Recently, multiscale PCA (MSPCA) has been formulated (Bakshi, 1998). By the MSPCA, one simultaneously extracts process correlations across data and accounts for auto-correlation within sensor data. This involves the decomposition of variables on a selected family of wavelets and the development of separate PCA models at each scale. The models at important scales are then combined in an efficient scale-recursive manner to yield a model for those scales of interest. The MSPCA formulation makes it suitable to work with process signals having a multiscale nature. Fault detection has been demonstrated with the MSPCA concept, but fault isolation has not been dealt with.

This paper presents an indirect usage of the process dynamics for FDI. A unified framework that generalizes MSPCA as well as PCA is proposed. An FDI procedure using the proposed method is presented. The proposed method is illustrated using a simulated data from a CSTR system with feedback control. The FDI performance of the proposed method is assessed through Monte Carlo simulation.

2. DECOMPOSITION OF A FINITE SIGNAL

The decomposition of signals according to their scale contents can be implemented with multi-channel filter banks built by cascading two-channel banks. Since a two-channel filter bank splits a signal into a lowpass version, and a highpass version, this decomposition is recursively applied on the lowpass version (Vetteri and Kovacevic, 1995). This leads a hierachy of multiresolution decomposition. In practice, the only data that can be processed by an algorithm are discrete. In the discrete version of the multiresolution analysis, it is assumed that the initial data already represents an approximation at a certain scale that is related to the sampling interval. By convention, this scale is fixed at j=0. This is the finest resolution, associated with the space V_0. Then, a finite number of decomposition steps J leads to a coarsest resolution associated with V_J. The input is decomposed into a very coarse resolution which exists in V_J and added details which exist in the spaces W_j, j = 1 ,…, J, , where V_J's are called approximation spaces and W_j's detail spaces.

The action of a two channel analysis filter on an infinite signal column vector $\mathbf{x}$ is represented by using the analysis filter matrix, $\mathbf{F}_a^{\mathrm{T}}=[\mathbf{H}_1^{\mathrm{T}}\,\mathbf{H}_0^{\mathrm{T}}]^{\mathrm{T}}$ where $\mathbf{H}_i$ has the effect of filtering the signal by $H_0(z)$ and subsampling by 2 (represented by the shift by 2 in $\mathbf{H}_0$) where i=1 for a high pass and i=1 for a low pass filter. The projections on the low-pass component are recursively decomposed to obtain coarser approximations. The projections on the high-pass component (wavelets at each scale) contain the finer details. Thus,

$$\mathbf{x}_a^{(j)} = \mathbf{H}_0\mathbf{x}_a^{(j-1)} \text{ and } \mathbf{x}_d^{(j)} = \mathbf{H}_1\mathbf{x}_a^{(j-1)} \tag{1}$$

where, $\mathbf{x}_a^{(j)}$and $\mathbf{x}_d^{(j)}$ are row vectors whose lengths are n/2^j where n is a signal length. Assuming n is relatively large number, and the transformation matrix, $\mathbf{W}_a$ is expressed as

$$\begin{aligned}\mathbf{W}_a &= \underbrace{Diag\{\overbrace{\mathbf{I},\cdots,\mathbf{I}}^{2^{(J-1)}-1},\mathbf{F}_a\}\cdots Diag\{\mathbf{I},\mathbf{I},\mathbf{I},\mathbf{F}_a\}Diag\{\mathbf{I},\mathbf{F}_a\}}_{J-1}\mathbf{F}_a \\ &= [\mathbf{H}_1 \quad \mathbf{H}_1\mathbf{H}_0 \quad \cdots \quad \mathbf{H}_1\underbrace{\mathbf{H}_0\cdots\mathbf{H}_0}_{J-1} \quad \underbrace{\mathbf{H}_0\cdots\mathbf{H}_0}_{J}]^{\mathrm{T}} \\ &= [\mathbf{H}_1^{(1)} \quad \mathbf{H}_1^{(2)} \quad \cdots \quad \mathbf{H}_1^{(J)} \quad \mathbf{H}_0^{(J)}]^{\mathrm{T}}\end{aligned} \tag{2}$$

where $\mathbf{H}_1^{(j)}$ is the matrix containing wavelet filter coefficients corresponding to scale j and $\mathbf{H}_0^{(J)}$ is the matrix of scaling function filter coefficients at the coarsest scale. After J decompositions, the transformation of $\mathbf{x}$, or $\mathbf{x}_a^{(0)}$ is expressed in terms of $\mathbf{x}_a^{(j)}$ and $\mathbf{x}_a^{(d)}$;

$$\mathbf{W}_a\mathbf{x}_a^{(0)} = [\mathbf{x}_d^{(1)} \quad \mathbf{x}_d^{(2)} \quad \cdots \quad \mathbf{x}_d^{(J)} \quad \mathbf{x}_a^{(J)}]^{\mathrm{T}} \tag{3}$$

Thus, $\mathbf{W}_a\mathbf{x}$ is the same size as the original data sequence, but due to the wavelet decomposition, the deterministic component at lowest frequency is concentrated in a relatively small number of coefficients in $\mathbf{x}_a^{(J)}$, while the stochastic component in each variable is approximately decorrelated and is spread over all components in $\mathbf{x}_a^{(d)}$, j = 1, …, J, according to its power spectrum.

Similarly, the synthesis filter matrix, $\mathbf{F}_S^{\mathrm{T}}=[\mathbf{G}_1^{\mathrm{T}}\,\mathbf{G}_0^{\mathrm{T}}]^{\mathrm{T}}$ can be represented with its high pass and low pass components matrices, $\mathbf{G}_1$ and $\mathbf{G}_0$. Based on it, the transformation matrix for the reconstruction $\mathbf{W}_S$ is obtained. Therefore, the reconstructed signal is

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{W}_s\mathbf{W}_a\mathbf{x} = \mathbf{W}_s[\mathbf{x}_d^{(1)} \quad \mathbf{x}_d^{(2)} \quad \cdots \quad \mathbf{x}_d^{(J)} \quad \mathbf{x}_a^{(J)}]^{\mathrm{T}} \\ &= \mathbf{G}_1^{(1)}\mathbf{x}_d^{(1)} + \mathbf{G}_1^{(2)}\mathbf{x}_d^{(2)} + \cdots + \mathbf{G}_1^{(J)}\mathbf{x}_d^{(J)} + \mathbf{G}_0^{(J)}\mathbf{x}_a^{(J)} \\ &= \mathbf{G}_1\mathbf{x}_d^{(1)} + \mathbf{G}_0\mathbf{G}_1\mathbf{x}_d^{(2)} + \cdots \\ &\quad + \underbrace{\mathbf{G}_0\mathbf{G}_0\cdots\mathbf{G}_0}_{J-1}\mathbf{G}_1\mathbf{x}_d^{(J)} + \underbrace{\mathbf{G}_0\mathbf{G}_0\cdots\mathbf{G}_0}_{J}\mathbf{x}_a^{(J)}\end{aligned} \tag{4}$$

The filter bank necessarily operates over infinite signals (the matrix is infinite along both dimensions) and the filters do not change with time. The application to finite lengths will generally involve either distortion at the boundary or the introduction of some redundancy. If a finite unitary matrix has the same block structure as the infinite unitary matrix $\mathbf{T}$, one can get a square unitary matrix for any size for a given filter set. It was shown how this problem could be overcome in the case of two-channel orthogonal filter banks by using boundary filters (Herley, 1995).

3. PCA OF MULTISCALE DATA

As a data sequence is transformed and expressed in terms of its contributions at different scales, the same transformation can be applied to all the measurement data. Let $\mathbf{X}$ be the observation matrix of dimension, $n\times m$. All the columns of $\mathbf{X}$ can be decomposed into the details at all levels and the approximation at the coarsest level.

$$\begin{aligned}\mathbf{W}_s\mathbf{W}_a\mathbf{X} &= \mathbf{G}_1^{(1)}\mathbf{H}_1^{(1)}\mathbf{X} + \mathbf{G}_1^{(2)}\mathbf{H}_1^{(2)}\mathbf{X} + \cdots \\ &\quad + \mathbf{G}_1^{(j)}\mathbf{H}_1^{(j)}\mathbf{X} + \cdots \\ &\quad + \mathbf{G}_1^{(J)}\mathbf{H}_1^{(J)}\mathbf{X} + \mathbf{G}_0^{(J)}\mathbf{H}_0^{(J)}\mathbf{X} \\ &= \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_j + \cdots + \mathbf{X}_J + \mathbf{X}_{J+1}\end{aligned} \tag{5}$$

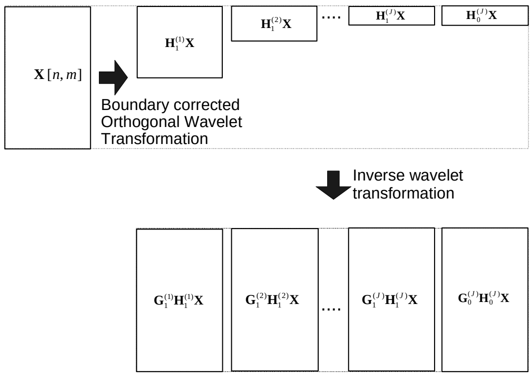

Fig. 1. Generation of multiscale data

where each term represents scale contribution of the original data matrix. All the terms have the same number of rows and columns. One can characterize the original data by three parameters of variable(m), sample time(n), and scale(J). It is splitting all the scale contributions and laying the terms side by side to produce a two-dimensional matrix $\mathbf{X}_G$ of size $[n \times (J+1)m]$ (Fig. 1). The $\mathbf{X}_G$ is defined

$$\mathbf{X}_G = \mathbf{V}Diag(\mathbf{X})_{J+1} \quad (6)$$

where $\mathbf{V}$ is the rearranged transformation matrix and $Diag(\mathbf{X})_{J+1}$ is a diagonal matrix whose diagonal component is $\mathbf{X}$ and off-diagonal terms are zero matrices with the same dimension as $\mathbf{X}$. Its dimension becomes $[(J+1)n,\ (J+1)m]$. Once we choose an orthogonal mother wavelet, $\mathbf{G}_0 = \mathbf{H}_0^*$, $\mathbf{G}_1 = \mathbf{H}_1^*$, then $\mathbf{G}_0^{(J)} = (\mathbf{H}_0^{(J)})^{\mathrm{T}}$, $\mathbf{G}_1^{(j)} = (\mathbf{H}_1^{(j)})^{\mathrm{T}}$. $\mathbf{V}$ is simplified in term of the analysis filter matrices. PCA model of multiscale data is obtained by applying PCA on $\mathbf{X}_G$. Alternatively, one can obtain the GPCA model by ordering the principal components of all the scale blocks as follows;

- Perform regular PCA on $\mathbf{X}_j$ using PCA algorithm to obtain $\mathbf{P}_j$ and eigenvalues, λ_j which contain all the eigenvalues of each sale block.
- According to the eigenvalues magnitudes, arrange all the corresponding loadings.
- Obtain the loadings of GPCA model with the arranged block loadings. For example, when the largest eigenvalue is the largest one from j scale block, then the first GPCA loadings is;

$$\mathbf{P}_{G,1}^T = \left[\underbrace{\mathbf{0}(1,m) \ \cdots \ \mathbf{0}(1,m)}_{(j-1)m} \quad \mathbf{P}_{b,j}^T \quad \underbrace{\mathbf{0}(1,m) \ \cdots \ \mathbf{0}(1,m)}_{(J+1-j)m} \right]$$

Then, the second loading of GPCA model is obtained by selecting the loading related with the largest eigenvalue among the remaining eigenvalues except for the one used in the previous step.

- The scores can be obtained in two ways: One can use the block score according to the loading selected, or use $\mathbf{X}_G$ and the GPCA loadings $\mathbf{P}_{G,1}$; $\mathbf{T}_{G,1} = \mathbf{X}_G\mathbf{P}_{G,1}$
- Deflate residuals; $\mathbf{E}_G = \mathbf{X}_G - \mathbf{T}_{G,1}\mathbf{P}_{G,1}^{\mathrm{T}}$
- Return to step 2 and repeat the procedure until one obtains all the GPCA loadings with all block loadings.

One may capture the correlations among process variables as well as among scales. Due to the usage of an orthogonal mother wavelet, the PCA model inherits a few useful properties from the wavelet orthonormality and the related filter features.

Discrimination of process variations: Due to the orthogonality between scale blocks, the off-diagonal terms of $\mathbf{X}_G^{\mathrm{T}}\mathbf{X}_G$ have all zero values. The covariance matrix of $\mathbf{X}_G^{\mathrm{T}}\mathbf{X}_G$ becomes a block diagonal matrix. Therefore, there will be no interactions between blocks, and the principal component will be a function of variables within only a scale block.

Generalization of PCA and MSPCA: The diagonal terms of $\mathbf{X}_G^{\mathrm{T}}\mathbf{X}_G$ have non-zero values. Their block diagonal terms of $\mathbf{X}_G^{\mathrm{T}}\mathbf{X}_G$ are the same as those of individual blocks in MSPCA, and the sum of the block diagonal terms becomes an identity matrix. When one considers the zero level of decomposition, GPCA becomes PCA. MSPCA is a special case of GPCA without the information on the percents of process variations explained by the principal components. Thus the GPCA model unifies PCA as well as MSPCA.

Clustering of multiscale data: When all the scale characteristics of dominant events in a process may not be shown in the orders of $2^j T_S$, but unevenly spaced from lower to higher scale, one may not need to investigate all the decomposed scales. One then can cluster the several scales at which the similar correlations are shown. This will significantly reduce the number of scales to be analyzed and simplify the multiscale analysis.

Detrending: Once a multiscale data based PCA model has been built, one can identify a scale or a principal component which is not needed for process monitoring. One may further identify a useless process variation at the identified scale. One can then remove one or all the principal components within the suspected scale. Removing all the principal components within the scale is equal to a bandwidth filter. When a scale includes both the useful and useless process variations, the GPCA based modeling method provides enhanced flexibility in handling the useless process variation for the process monitoring and obtaining a sensitive monitoring model.

4. FAULT DETECTION AND ISOLATION

With the notion of GPCA, one has additional information that is a scale contribution for fault analysis. One can estimate how much each scale contributes to a fault in terms of Hotelling's T^2. The overall T^2 can be partitioned into a contribution from each scale as follows:

$$\mathbf{T}_j^2 = \sum_{i=1}^{m} x_{G,(j-1)m+i}^2 w_{(j-1)m+i} \quad (7)$$

where $j = 1,\cdots,J+1$, $w_k = \sum_{i=1}^{A}\left(\frac{p_{ki}}{s_{t_i}}\right)^2$, p_{ki} is k-th variable weight in i-th latent variable, m is number of variables, and J is the number of scales. Once the contributing scale for a fault is identified, one can look at the individual variable contributions to the contributing scale. It can be expressed as follows:

$$\mathbf{T}_{jk}^2 = x_{G,(j-1)m+k}^2 w_{(j-1)m+k} \,. \quad (8)$$

The above equation is the scale variable contribution to T^2 relative to zero. The expression for contribution to a change in T^2 over some time interval is:

$$\Delta\mathbf{T}_{jk}^2 = \mathbf{T}_{jk,t}^2 - \mathbf{T}_{jk,t-1}^2 = \Delta x_{G,(j-1)m+k}^2 w_{(j-1)m+k}$$

Similar expressions can be obtained for the scale contribution to the overall SPE given by,

$$\mathrm{SPE}_j = \sum_{i=1}^{m}\left(x_{G,(j-1)m+i} - \hat{x}_{G,(j-1)m+i}\right)^2 \quad (9)$$

where $\hat{x}_{G,k} = (\mathbf{P}_A\mathbf{P}_A^T)_{k\cdot}\, x_{G,k}$ and the j-th variable contribution to the SPE at that scale SPE_{jk} is

$$\mathrm{SPE}_{jk} = \left(x_{G,(j-1)m+k} - \hat{x}_{G,(j-1)m+k}\right)^2 \quad (10)$$

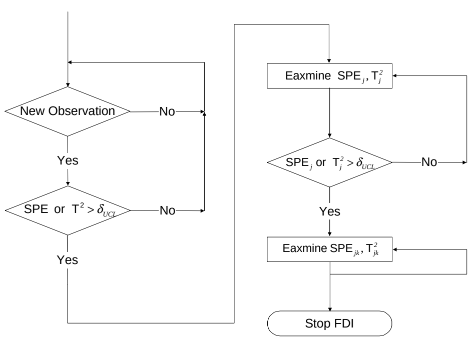

Fig. 2. GPCA model based FDI procedure

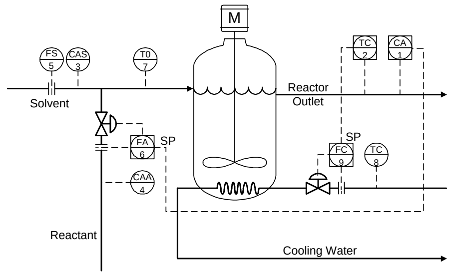

Fig. 3. Process flow diagram of CSTR system

Therefore, after a fault is detected by the SPE or T^2, one can inspect the scale contributions to the SPE or T^2, and then the variable contributions to the contributing scales. Having identified the important scales and the variable contributions to these scales, one may isolate a fault with high accuracy. Figure 2 outlines the FDI procedure based on the GPCA model.

5. SIMULATION STUDY

A nonisothermal continuous stirred-tank reactor (CSTR) model (Marlin, 1995, page 90-92) is used for a case study (Fig. 3). The reaction is 1st order ($A \rightarrow B$) and the reactor system involves heat transfer through cooling coils to remove the heat of reaction. The process has two feed streams (the solvent and the reactant *A*), one product stream, and a cooling water flow to the coils. The reactant (F_A) and the cooling water (F_C) flows control the reactor outlet concentration (C_A) and temperature (T), respectively. However, in this application only the temperature controller is active. Measured process disturbances are the inlet concentrations (C_{AS}, C_{AA}), the inlet temperature (T_O), the solvent flow (F_S), and the cooling water temperature (T_C). All of these disturbances are simulated to have first order autoregressive variations under both normal and fault conditions. In addition, unmeasured stochastic disturbances are also simulated as first order autoregressive behaviors in the reaction rate (k) and heat transfer (UA) constants respectively. A negative correlation between F_S and C_{AS} is added. In all studies, the training data was collected for 200 minutes from the process under routine operation when no faults were present. The data collection interval T_S was 10 seconds.

5.1 Comparisons to PCA and MSPCA

The collected data was decomposed into 4 detail and one approximation blocks corresponding to $2T_S$, $4T_S$, $8T_S$, $16T_S$, and the frequencies lower than $16T_S$. Deubechies-5 wavelet was used as a mother wavelet to generate the multiscale data. Then, PCA and MSPCA/GPCA models were calculated with the single-scale and multiscale data, respectively. Figure 4 shows principal component loadings of the three models. Each cell represents the effects of all 9 variables- C_A, T, C_{AS}, C_{AA}, F_S, F_A, T_O, T_C, and F_C, respectively.

The PCA model consists of 3 principal components. The first principal component explains the control action. It shows that the cooling water flow (F_C), the reactor outlet concentration (C_A) and the reactor outlet temperature (T) are positively correlated. The second shows a negative correlation between the solvent flow (F_S) and the solvent concentration (C_{AS}). The third one seems to be a combination of several effects. The PCA model reasonably describes the relevant process correlations of the CSTR system.

Since the single-scale data was diadyically decomposed into 5 blocks of 4 details and 1 approximation, MSPCA model consists of PCA models of the 5 scale blocks. From left to right in the bottom of figure 4, each column corresponds to the PCA model of the decomposed block of $2T_S$, $4T_S$, $8T_S$, $16T_S$, and the lower frequencies than $16T_S$. The PCA models at all 5 scales show almost the same behavior as that of the single-scale PCA model. This means that the same process correlations existed over all scales.

GPCA model consists of 15 principal components since there are 5 decomposed blocks and each block has 3 principal components. Each principal component of the GPCA model is expressed with 45 variables. It results from the fact that there are 5 decomposed blocks and each block is described with 9 process variables. However, the principal components of GPCA model are actually described with 9 variables at only one scale due to orthogonality among scale blocks. In the GPCA model of figure 4, it is shown the 1st principal component is the function of variables at 4th approximation block and describes a positive correlation among the cooling water flow, the reactor outlet concentration, and the reactor outlet temperature. This correlation is based on the control action at the bandwidth beyond 16 T_S. The effect of the feedback control is most strongly shown at the frequency band lower than $16T_S$. Similarly, the second principal component indicates the similar correlation among the process variables at the frequency band of 8 T_S . Information on the scale contents of the process correlations over the examined frequency bands is available only with the GPCA model.

The process correlations shown at different scales with GPCA model are almost the same as those shown by the regular PCA model. The first 5 principal components of the GPCA model are very similar to the first principal component of the regular PCA model. The 6th ~10th principal components of

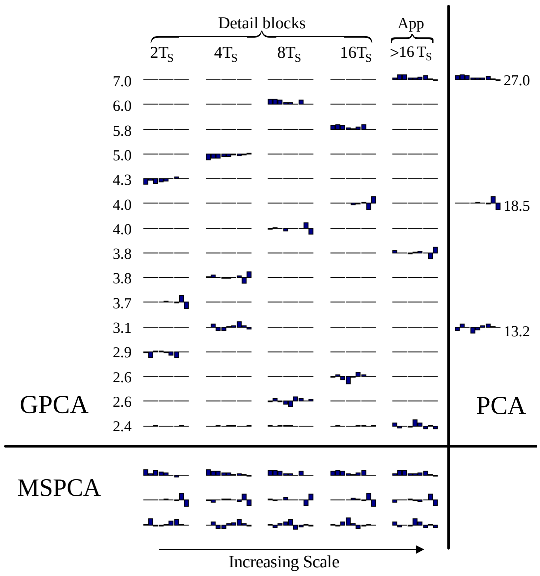

Fig. 4. Comparison of loadings for PCA, MSPCA and GPCA with db-5 and 4 scale levels

the GPCA model in figure 4 appear correspondingly to the second principal component of the regular PCA model. Thus, one may not need to apply multiscale analysis. In this case, PCA becomes a very effective data analysis tool. With the GPCA model, one can judge if a multiscale analysis is needed.

5.2 Comparisons on FDI performance

Unstable contact or an old instrument can cause an excessive oscillatory action on the measurement while its real value remains unchanged. The inlet temperature (T_O) in the CSTR model is assumed to show this behavior (Fig. 5(a)). The fault effect is not propagated into the other variables and is restricted to one frequency band. It indicates that the precision degradation of T_O sensor has a single-scale effect. To examine the FDI performance on this fault, 500 sets of training and testing data were generated with random seeds and at each value of various FSRs (ratio of fault magnitude to square root of a signal variance). With each data set, the type I and type II errors, and the average run lengths (ARL) for fault detection and isolation were calculated. The Type I error of MSPCA is based on the threshold value adjusted with the Bonferroni rule. Figure 6(a) and (b) show the fault detection and the isolation ARLs for PCA, MSPCA and GPCA at various magnitudes of the fault. The isolation ARL was calculated with the variable contributions assuming that the fault is detected and identified. The GPCA/MSPCA outperforms the regular PCA. This is due to the characteristic of the precision degradation whose effect is not spread over wide frequency ranges.

As a second case, a sensor bias of the inlet temperature measurement (T_O) is simulated (Fig. 5(b)). All the other conditions are the same as the above case. Figure 7(a) and 7(b) show the fault detection and the isolation ARLs for the sensor bias. The detection and isolation ARLs of GPCA/MSPCA show almost the same results as that of the regular PCA. This is due to the characteristic of the sensor bias whose effect is spread over wide frequency ranges. The simulation result confirms that the multiscale model does not significantly outperform the regular PCA in such a case.

6. CONCLUSION

In this study, GPCA that is PCA of the multiscale data has been proposed as a modeling method for data showing multiscale features. When process signals represent the cumulative effects of many underlying process phenomena and each of them manifests on a different scale, one can explicitly decompose those process effects over scale levels, and capture the process correlation among variables and scales. GPCA shows what scale ranges should be monitored since it gives scale information of significant process correlation. It also enables one to remove unnecessary effects of process events from further analysis, or to cluster several scale blocks showing similar correlation to efficiently summarize the multiscale process correlations. One can also confirm if the multiscale analysis is needed.

The proposed scheme unifies the existing methods since PCA and MSPCA are limiting cases of GPCA. GPCA thus not only conveys all the information contents shown in both MSPCA and PCA, but also reveals the relative significance of all the principal components over scales, which is unavailable with MSPCA. Based on the proposed method, a procedure for both fault detection and isolation using the GPCA model has been presented. Contributions of the various scales to the overall T^2 and SPE, and contributions of the variables to each scale have been proposed as additional tools for fault isolation. Due to the usage of scale information, FDI performance can be significantly enhanced.

FDI performance has been assessed through Monte Carlo simulation with the CSTR system. When a fault occurs in only one frequency band, multiscale methods will be more effective than the regular methods in detecting and isolating faults. However, it is questionable if the multiscale methods would result in better performance when the fault effect is spread over more than one frequency band. Thus, a monitoring method that gives the best detection and isolation of faults will depend upon the fault characteristics.

REFERENCES

Bakshi B., (1998). Multiscale PCA with Application to Multivariate Statistical Process Monitoring, *AIChE J.*, **44**(7), 1596-1610.

Daubechies, I., (1990). The wavelet transform, time-frequency localization and signal analysis, *IEEE transactions on information theory*, **36**(5), 961-1005.

Herley, C. (1995), Boundary filters for finite-length signals and time-varying filter banks, *IEEE transactions on circuit and systems II: Analog and Digital signal processing*, **42**(2), p.102-114.

Mallat, S. G., (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans on pattern analysis and machine intelligence*, **11**(7), 674-693.

Marlin, T.E., (1995), *Process Control: Designing Processes and Control systems for Dynamic Performances*, McGraw-Hill, New York.

Vetteri, M. and J. Kovacevic, (1995), Wavelets and subband coding, Prentice Hall, Englewood Cliffs, NJ.

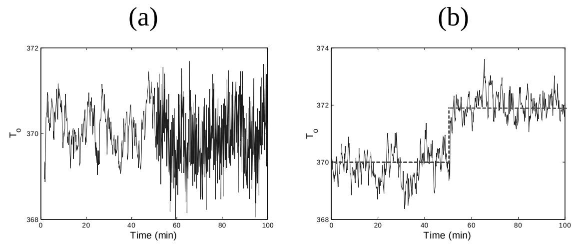

Fig. 5. Simulated faults on T_O sensor; (a) Precision degradation; (b) Sensor bias (Fault at 51 min during [0 100])

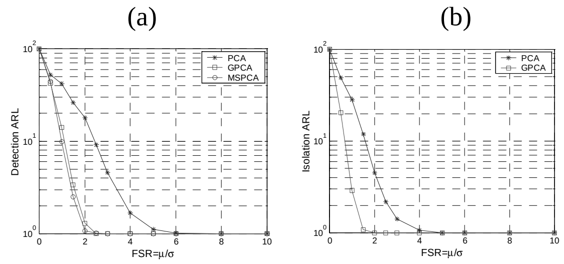

Fig. 6. FDI w.r.t. the precision degradation of T_O sensor; (a) detection ARL; (b) isolation ARL (99% CL)

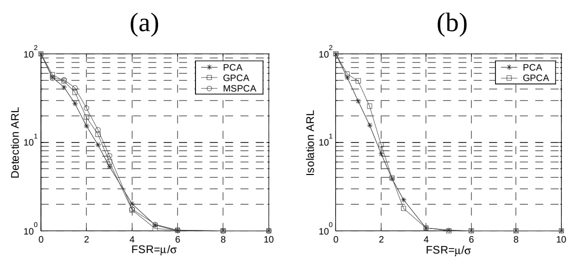

Fig. 7. FDI w.r.t. T_O sensor bias; (a) detection ARL; (b) isolation ARL (99% CL)