2013 IREP Symposium-Bulk Power System Dynamics and Control –IX (IREP), August 25-30, 2013, Rethymnon, Greece

Detection and Visualization of Power System Disturbances using Principal Component Analysis

E. Barocio, Bikash C. Pal
Department of Electrical and Electronic Engineering
Imperial College London, London, SW7 2AZ, UK

Davide Fabozzi, Nina F. Thornhill
Department of Chemical Engineering
Imperial College London, London, SW7 2AZ, UK

Abstract

In this paper, a multivariate statistical projection method based on Principal Component Analysis (PCA) is proposed for detecting and extracting unusual or anomalous events from wide-area monitoring data. The method combines PCA with statistical test to detect and analyze anomalous dynamic events from measured data.

Simulations based on a transient stability model of the New England Test System are used to demonstrate the ability of the method to detect and extract system events from wide-area data.

1 Introduction

During the last decade, detecting and visualizing power system events has emerged as a new and promising research area. Event detection and analysis techniques are key parts of Wide-area Monitoring and Control Systems (WAMCS). The recently published IEEE task force report on the identification of electromechanical modes stresses the importance of data processing and visualization as means to monitor the operational status of the system [1].

A variety of techniques have been proposed for automatic extraction and characterization of dynamic features from measurements during ambient and transient operation. Parametric and non-parametric mode estimation algorithms have been specifically designed for detecting the impact of system disturbances on the dynamic stability margin of the system [1]. However, on-line estimation and visualization of modal parameters are very challenging, and may require long records of data and pre-filtering to obtain useful information [2].

Data visualization and analysis tools may provide operators with a more complete understanding of evolving system dynamics [3]. In [4], Frequency Disturbance Recorder (FDR) measurements were used to provide a real-time view of the power system frequency behaviour. Using voltage and frequency measurements it is possible to determine the location and extent of system damage by tracking changes in key modal parameters.

More recently, statistical techniques with the ability to collect and process vast amounts of data have been proposed. In [5], a multidimensional characteristic ellipsoidal technique was used to detect system disturbances and stressed operation conditions. Along with these efforts, statistical techniques based on multivariate dimension reduction schemes, have been proposed to detect system faults [6], the temporal variability and spatial distribution of wind power [7], and the identification of coherent generators [8].

In this paper, a multivariate statistical projection method based on Principal Component Analysis (PCA) is proposed for detecting, localizing and characterizing system damage on a global basis. The technique can be used to detect anomalous dynamic events from large and complex data sets, locate the source of damage and provide some estimate of its extent and distribution. Conceptually, the PCA-based approach transforms the data in order to remove correlations among variables and reduce their dimensionality without significant loss of information [9].

When combined with a Wide-Area Monitoring System (WAMS), the proposed framework allows determining security margins with respect to faults or changes in the system operating conditions.

Simulations based on a transient stability model of the New England test system are used to demonstrate the ability of the method to detect and extract system events from wide-area data.

The issues of automation of the process, unsupervised fault detection and the incorporation of operational constraints are discussed. The directions for future research and development are also identified.

2. System Reduction by Principal Component Analysis Modeling

Wide-area measurement systems provide a large amount of information about system dynamic behaviour that needs to be processed and correlated in near real-time for detection and visualization of system disturbances.

Given a set of simultaneous observations at locations $\mathbf{x}_l$, $l=1,\ldots,m$ of a transient process, the observation (data) matrix can be defined as:

$$\boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \cdots \quad \boldsymbol{x}_m] \tag{1}$$

where each column represents the time evolution of a physical variable (the individual sensor signals).

Figure 1 provides a physical interpretation of the proposed framework.

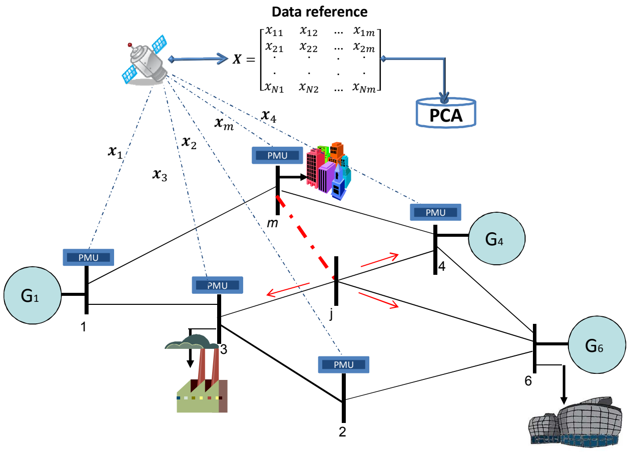

Fig. 1 Overview of the proposed approach.

Power system measurements are multi-scale, noisy and may contain missing data, trends and outliers, which mask the presence of system damage, and make the analysis and visualization of specific system behaviour very challenging. Moreover, monitoring large interconnected systems may be data intensive prompting the need for the use of efficient model reduction techniques.

In what follows, PCA is used to identify, extract and monitor critical dynamic events from multivariate, noisy data on a global basis. Techniques for real-time event detection and the determination of the source location are proposed and tested. The dimensionality reduction is used to facilitate the visualization and classification of large and complex data sets.

2.2 Principal Component Analysis for Event Detection

The event detection and prediction involves comparing, in near real-time, spatio-temporal data against event data patterns. Due to the high dimensional nature of the data, model reduction techniques are needed to capture the relevant dynamic features.

To deal with model reduction of large datasets with minimum loss of information, high-dimensional data is projected onto a lower-dimensional space which contains most of the variance of the original data [9].

More formally, the original data space $\boldsymbol{X}^* \in \mathbb{R}^{N,m}$ is decomposed as the product of matrix $\widehat{\boldsymbol{T}}$ and $\widehat{\boldsymbol{P}}$ plus a residual matrix $\boldsymbol{E}$, as:

$$\boldsymbol{X}^* = \widehat{\boldsymbol{T}}\widehat{\boldsymbol{P}}^T + \boldsymbol{E} = \sum_{i=1}^{K} \boldsymbol{t}_i \boldsymbol{p}_i^T + \sum_{i=K+1}^{p} \boldsymbol{t}_i \boldsymbol{p}_i^T \tag{2}$$

where $\widehat{\boldsymbol{T}} \in \mathbb{R}^{N,K}$ and $\widehat{\boldsymbol{P}} \in \mathbb{R}^{m,K}$ are related with the principal component scores and loadings. $\boldsymbol{E} \in \mathbb{R}^{N,m}$, represent the residual variability which is not captured by the Principal Components (PCs) scores and K is the number of PCs [9]. On the other hand, the $\boldsymbol{t}_i$ score vectors contain information on how the samples relate to each other and the columns of the loading matrix, $\boldsymbol{p}_i$, are eigenvectors of the covariance matrix, $\boldsymbol{S} \in \mathbb{R}^{m,m}$ defined by [9]

$$\boldsymbol{S} = \frac{1}{N-1}\boldsymbol{X}^{*T}\boldsymbol{X}^* \tag{3}$$

The loadings provide information as to which physical variables contribute to the most to individual principal components.

To detect variations from typical or expected behavior, the observation data is centered and scaled, by subtracting the column means ($\boldsymbol{X}_{avg} \in \mathbb{R}^{1,N}$) and dividing by the standard deviation ($\boldsymbol{X}_{std} \in \mathbb{R}^{1,N}$). This results in a reference data set of the form

$$\boldsymbol{X}^* = \left(\boldsymbol{X} - \boldsymbol{l}\boldsymbol{X}_{avg}\right)./\boldsymbol{l}\boldsymbol{X}_{std} \tag{4}$$

where $\boldsymbol{l}$ is a vector of ones ($\boldsymbol{l} \in \mathbb{R}^{N,1}$) and ./ indicates element-by-element division. The centering and scaling process removes engineering units and also results in a more stationary data set which is more suitable for data monitoring.

From the PCA model described in the equation (2), we can obtain statistical quantities analysing data projections onto PCs model subspace. For instance, let us assume that vector sample $\boldsymbol{x}_i^* \in \mathbb{R}^{1,m}$ is decomposed in two orthogonal vector as follows:

$$\boldsymbol{x}_i^* = \overline{\boldsymbol{x}_i^*} + \boldsymbol{e}_i$$

where:

$$\overline{\boldsymbol{x}_i^*} = \boldsymbol{x}_i^*\widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^T = \boldsymbol{x}_i^* C \tag{5}$$

where $\boldsymbol{C} \in \mathbb{R}^{m,m}$, is the model projection matrix. In a similar way, the projection to residual subspace of data matrices can be carried out using:

$$\boldsymbol{e}_i = \boldsymbol{x}_i^* - \overline{\boldsymbol{x}_i^*} = \boldsymbol{x}_i^*\left(\boldsymbol{I} - \widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^T\right) = \boldsymbol{x}_i^*(\boldsymbol{I} - \boldsymbol{C}) \tag{6}$$

where $\boldsymbol{e}_i \in \mathbb{R}^{1,m}$ and $(\boldsymbol{I} - \boldsymbol{C})$ is called the residual projection matrix and $\widehat{\boldsymbol{I}} \in \mathbb{R}^{m,m}$ is identity matrix.

The projections on the space model given in (5) and residual space in (6) has direct relationship with multivariable control charts [10].

The reduced model is the first statistical approach to define the secure operation margin and also to provide a framework for real time monitoring changes in the selected features of the system.

3. Estimation of Secure Operated Margin and Detection of System Disturbances.

Based on the above development two statistical tests used in statistical multivariable process control are adopted [10] to estimate the operating security margin and detecting the system disturbances.

3.1 Estimation of Operation Security Margin

Initially, two multivariate statistical indices are used in conjunction with PCA reduced model to define secure operating margins in statistical terms.

As a first step, the Hoteling's statistic [11] for observations $\boldsymbol{x}_i^* \in \mathbb{R}^{1,m}$ is calculated as:

$$\mathrm{T}_i^2 = \widehat{\boldsymbol{t}}_i\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{t}}_i^{\boldsymbol{T}} = \boldsymbol{x}_i^*\widehat{\boldsymbol{P}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{P}}^T\boldsymbol{x}_i^{*T} \quad , \quad i = 1,2,..,N \tag{7}$$

where $\widehat{\boldsymbol{\Lambda}} \in \mathbb{R}^{K,K}$, $\widehat{\boldsymbol{t}}_i \in \mathbb{R}^{i,,K}$ and $\widehat{\boldsymbol{P}}$ are the diagonal matrix of selected eigenvalues of $\boldsymbol{S}$, i^{th} score row vector of $\widehat{\boldsymbol{T}}$ and loading matrix obtained from the reduced model, respectively. The Hoteling's test represents the total amount of inherent variability that is inside the system behaviour. The statistical distribution of its values follows the *F*-distribution [10].

We consider that the data reference system is under normal conditions for a given significance level α, if:

$$\mathrm{T}_{\mathrm{i}}^2 \leq \mathrm{T}_{\boldsymbol{\alpha}}^{\mathbf{2}} = \frac{(N^2-1)K}{N(N^2-K)} \boldsymbol{F}_{\boldsymbol{\alpha}}(K, N-K) \tag{8}$$

where N is the number of samples used to construct the reference data model, K is the number of selected PCs and $\boldsymbol{F}_{\alpha}(K, N-K)$ is the critical value for an $\boldsymbol{F}$ distribution with N and (N-K) degrees of freedom at the (1-α) confidence level. Typically, α takes values between 1% and 5%.

A complementary test on the Q_i statistic, defined as the sum of squared residual of the selected PCs, is used to measure the variation in the residual space not accounted for by the PCA model,

$$Q_i = \boldsymbol{e}_i \boldsymbol{e}_i{}^T = \boldsymbol{x}_i^* \left(\boldsymbol{I} - \widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^T\right) \boldsymbol{x}_i^{*T} \tag{9}$$

which is also interpreted as a measure of the "best fit" of the PCA model, showing the distance between the actual and the predicted data.

A suitable threshold for statistical significance of Q_i is given by:

$$Q_i \leq Q_\alpha = \theta_1 \left[\frac{c_\alpha h_o \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_o (h_o - 1)}{\theta_1^2}\right]^{1/h_o} \tag{10}$$

with

$$\theta_i = \sum_{a=K+1}^{M} \lambda_{a\,;\,i=1,2,3}^{i} \;;\; h_o = 1 - \frac{2\theta_1\theta_3}{3\theta_1^2}$$

where Q_α is the upper confidence limit for the residual model Q_i with significant level and c_α is the normal deviation corresponding to the upper (1- α) percentile.

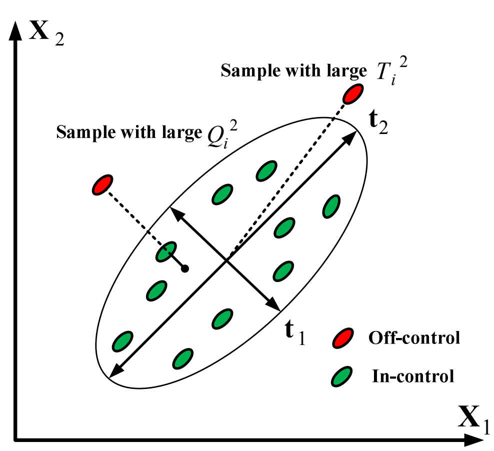

Fig. 2: Data projection on two PCs.

Taken together, both the statistical tests let us identify whether the event is associated with unusual variability within the normal subspace defined by the selected PCs or by variations not explained by the retained PCs. Figure 1 illustrates two types of unusual variation captured by the statistical test [10].

3.2 Detection of System Disturbances

In this paper, we used two ways to detect system disturbances which are used in a complementary manner:

- Control charts
- Score plots

Control charts are obtained from the indices (7) and (8), and are especially useful when the numbers of PCs are higher than 3 and cannot be represented by a tridimensional ellipsoid.

Once that secured operating margin has been estimated, the tracking process to detect system disturbances is carried out. The new observations of each PMU are mean-centred and scaled with respect to the mean and standard deviation obtained in (4). These observations are denoted by $\widehat{\boldsymbol{x}}_j{}^* \in \mathbb{R}^{1,m}$; the statistical indices are then updated as follows:

$$\mathrm{T}_j^2 = \widehat{\boldsymbol{x}}_j{}^* \widehat{\boldsymbol{P}} \widehat{\boldsymbol{\Lambda}} \widehat{\boldsymbol{P}}^T \widehat{\boldsymbol{x}}_j{}^{*T} \tag{11}$$

$$Q_j = \widehat{\boldsymbol{x}}_j{}^* \left(\boldsymbol{I} - \widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^T\right) \widehat{\boldsymbol{x}}_j{}^{*T} \tag{12}$$

Both indices are compared again with their respective thresholds as computed in (8) and (10). If a sequence of the new values of T_j^2 or Q_j is above a given threshold, the fault process is occurring.

In the second approach, the scores associated with the selected PCs have been used. To illustrate this index, consider that a suitable number (e.g. 2 or 3) of PCs represent the secure operated margin; the selected scores denoted by $\widehat{\boldsymbol{T}}$, the eigenvalues of $\boldsymbol{S}$ and their corresponding eigenvectors, provide the orientation and the semi-axis length of the ellipse or 3D ellipsoid representation [11]. Figure 3 depicts the result of modeling the data system $\boldsymbol{X}^*$ with the 2 or 3 largest PCs.

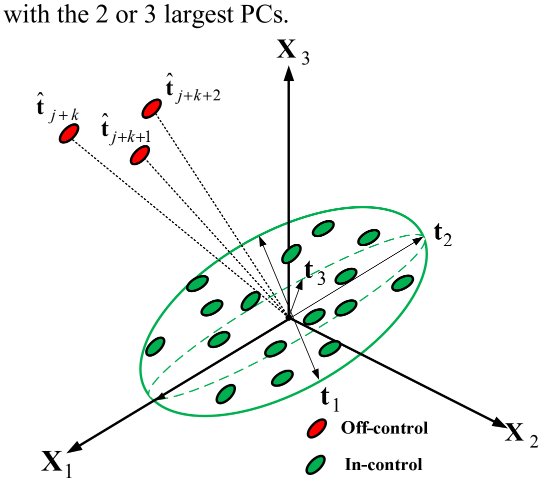

Fig. 3: Data projection on score plot.

The ellipsoid defines a region where the system is operating under normal conditions. The new observations denoted by $\widehat{\boldsymbol{x}}_j{}^* \in \mathbb{R}^{1,m}$, are used to estimate the new row score:

$$\widehat{\boldsymbol{t}}_j = \widehat{\boldsymbol{x}}_j{}^* \widehat{\boldsymbol{P}} \quad , \quad j = N+1, N+2, \ldots, L \tag{13}$$

within the control region (2D ellipse or 3D ellipsoid), one can map the data $\widehat{\boldsymbol{t}}_j \in \mathbb{R}^{1,K}$ for a segment onto the same axes. If the new observation lies outside the ellipsoid for a given period of the time is identified as a fault -- see Fig. 3. Otherwise they would be considered as a false alarm. Both cases indicate changes on the data covariance, which are also reflected in the indices (7) and (9).

3.3 Estimation of Fault Localization Based on Contributions Plots.

Additional information may be extracted from the contribution plots to identify the source of the fault. For instance, consider that a successive number of new observations fall outside the ellipsoidal surface or that the new values for T_j^2 and Q_j are above the critical values given in (8) and (10). This indicates that one event has occurred, but it does not provide information about the nature or source of the event. In this regard, contribution plots may be used to determine which PMU detects the event first, providing quick estimation about where the fault occurs as well as how the fault is spreading across the system.

According to [12], the contribution plots may reveal the group of variables making the highest contribution to the model PCA or to the residuals. In the context of our problem, the individual contribution of the m^{th} variable (PMU) to the K^{th} score is denoted as:

$$c_m{}^{\left(\mathrm{T}_j^2\right)} = p_{K,m} \widehat{\boldsymbol{x}}_m{}^* \left(\frac{\hat{t}_K}{\lambda_K}\right) \tag{14}$$

where $\hat{t}_K$ and λ_K , represent the value and the variance of the *K*-score respectively and $\hat{p}_{K,m}$ is the element of the matrix $\widehat{\boldsymbol{P}}$**,** and $\hat{x}_m$ is m^{th} value of the current data vector $\widehat{\boldsymbol{x}}_{\boldsymbol{i}}{}^*$**.** In this paper, we examined the contribution from each PMU at each time period rather than individual contributions. Using the total contribution defined in [12], we can evaluate the contribution of the m^{th} time period to the T_j^2, which is taken as the sum of all contributions to the m^{th} variable:

$$tc_m{}^{\left(\mathrm{T}_j^2\right)} = \sum_{i=1}^{r} c_m{}^{\left(\mathrm{T}_j^2\right)} \tag{15}$$

where *r* represents the number of abnormally large score variables. Variable contribution can also be computed to Q_j statistic, more details are given in [12].

4. On-line Monitoring and Fault Analysis

The basic procedure for monitoring the system based on the PCA method can be derived according to the section 3.1 that covers PCA model building and sections 3.2 and 3.3 that deal with fault detection and fault contribution. The proposed algorithm for a fault detection and identification consists of the following steps:

1. Data selection and standardization:
   - Selection of the reference data: the size of initial matrix $\boldsymbol{X}$ with *N* observations and *m* variables must be long enough to characterize the normal system variability.
   - Scaling and centering of the input data matrix that is used for process model building.
2. PCA model building:
   - Covariance matrix calculation based on standardized input data matrix.
   - Eigenvalues and eigenvectors calculation.
   - Number of PCs selection that will be retained in the model.
   - Determination of confidence limits.
3. Test of the new data:
   - Scaling and centering of the new data.
   - Projection of new data to the existing space.
   - T_i^2 and Q_i values calculation for each sample.
   - If T_i^2 or Q_i sample values exceed the confidence limits, abnormal event has occurred.
   - If the event persists for 2 second or 20 samples, using a sampling frequency of 10 Hz, a fault is occurring in the system.
4. Fault localization
   - Draw the total contribution plot to estimate the fault localization and evaluate its propagation across the system.

To improve the visualization of system disturbances, we use the score plot. Refer to section 3.2 for details about how to obtain this plot.

5. Numerical Results

As an illustrative example, the proposed algorithm is applied to the New England test system [13]. The measurements of the voltage phasor outputs of 20 PMUs were recorded at a sampling frequency of 10 Hz. 16 measurements are related to the generation nodes and the last four are related to strategic network nodes, selected by an extensive study based on centrality measures [14]. The marks in green colour indicate the locations of PMU nodes in Fig. 4. Two cases of interest will be considered in this research.

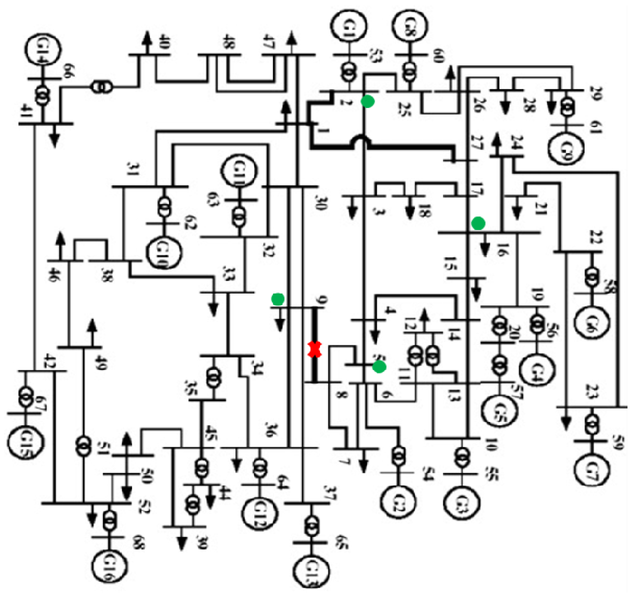

Fig. 4: Test system.

5.1 Building PCA model

Using the proposed data analysis framework, a statistical model of dimension $\boldsymbol{X} \in \mathbb{R}^{4000,20}$ is obtained to construct the PCA model. The reference data considers 20 measurements of voltage phasor collected by the PMU for 200 sec (4000 observations) at the normal operating conditions.

Considering that the best performance of PCA is under Gaussian distribution, an off-line analysis is first conducted; the Q-Q plot and its histogram are computed to check the statistical distribution of the reference data. As an example, in Fig. 5 shows the Q-Q plot and its histogram for the voltage at node 2. Both tests confirm that reference data follows a normal distribution, providing statistical support to the construction of PCA model. Similar results were obtained for the other nodes of the system. Results shown in Fig. 5 also suggest that the PCs may have a similar statistical distribution.

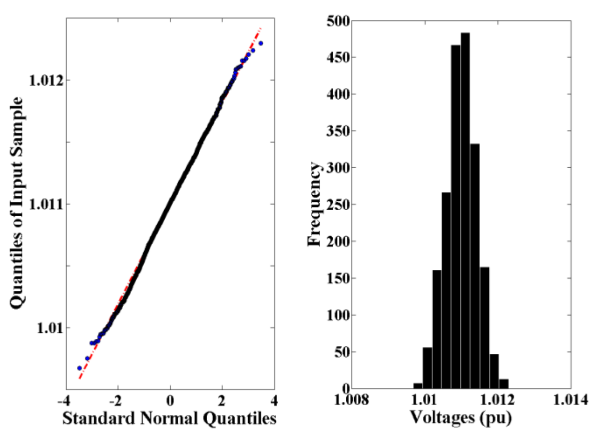

Fig.5: Q-Q plot and Histogram of voltage data at node 2.

A critical step in PCA modeling is the determination of the number of PCs to be retained in the model. This choice has a relevant impact in the accuracy of reduced model. There are several proposed approaches in the literature to determine K such as the Parallel Analysis (PA), Cumulative Percent Variance (CPV) and cross validation [15]. Here, the PA and CPV are used to determine the number of PCs.

The PA method involves a Monte Carlo approach to generate a large number of eigenvalues based on simulated data sets. These eigenvalues are used to build confidence intervals used to construct a threshold; furthermore, all components with eigenvalues below this threshold value are considered spurious. On the other hand, the CPV is a visual and simple method that computes the cumulative variance percent of all eigenvalues, in the following way:

$$CVP(K) = \sum_{i=1=}^{M} \frac{\lambda_i}{trace(\sum \lambda_i)} 100 \qquad (17)$$

The results of (18) may be plotted to determine the percentage of variability. Now, based on the eigenvalue decomposition of $\boldsymbol{X}^{*}$ given in (4), the eigenvalues were calculated. The PA plots display the percentile values against the original eigenvalues. Fig. 6 shows such plots. It can be appreciated that, for the first three factors (92.5%), the eigenvalues from the random data are clearly lower than those from the original data. In Fig. 6, the plot of CPV is depicted, according to which only four out of 20 eigenvalues are significant. Considering both analysis, four eigenvalues represent 95.63% of the total data variance and are selected as the optimal number of PCs from which a PCA model can be built.

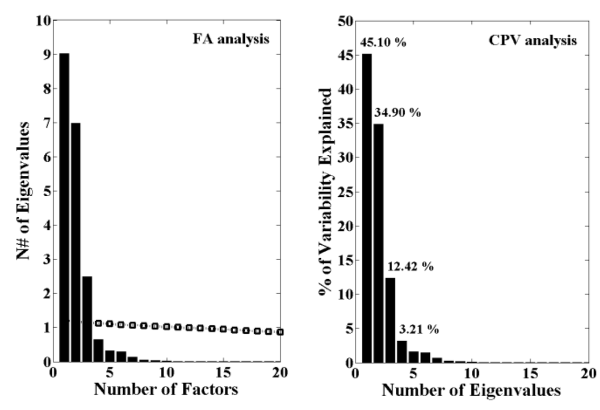

Fig. 6: PA results and screen percent plot of CPV technique.

Fig. 7 shows the score plot of reference data. Voltages values are symmetrically spread around one equilibrium point. This suggests that the system is operating at normal conditions. Figure 7 shows different voltage levels of each record collected by the PMU for the reference data. This plot provides a general overview of the system behavior at normal operating conditions.

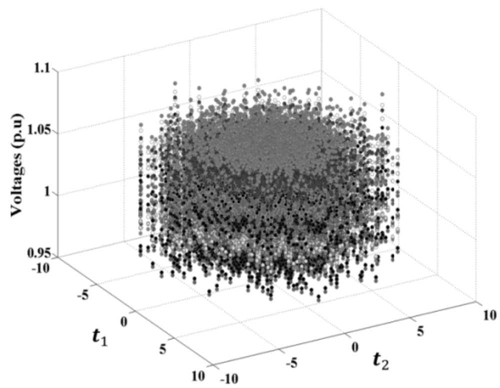

Fig. 7: Score plot of data reference.

Now, using the reference model, we project the reference data information into T_i^2 and Q_i limits, using equations (7), (8), (9), and (10). This procedure allows evaluating the quality of secure operated margin. The thresholds for T_i^2 and Q_i are calibrated considering: 4000 observations, four PCs and a significant level of confidence of 99%.

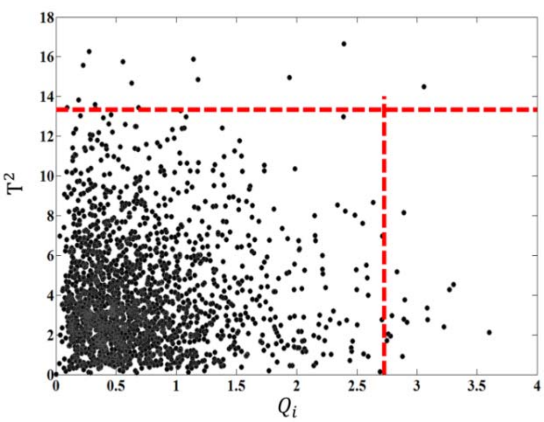

Fig. 8: Scatter Plot T_i^2 vs Q_i for normal operating condition.

Figure 8 presents the scatter plot of T_i^2 vs Q_i, for normal operating condition. The red dashed lines define the 99% of confidence limits for T_i^2 and Q_i respectively. In spite of this accuracy, some observations are out of the limits. Particularly, some values of the T_i^2 are associated to the inherent system variability that lies within the PMU. This inherent variability is maintained in the PCA model, indicating the amount of variability that may be considered as "normal" or false alarms. Out of the total 4000 data points of the PMUs, 13 data points (0.25%) were outside the T_i^2 control limit and 16 data points (0.4%) were outside the Q_i control limit.

Once the secured operating margin is estimated, the on-line monitoring process will be presented in the following sections.

5.2 Case study I: Detection and Visualization of Strong Perturbations.

A three-phase fault is applied at bus 8 cleared in 100 ms by opening the line between nodes 8 and 9. Records of 300 sec were generated, assuming a random process on the loads. At 200 sec a three phase fault was applied. The simulation ended at 300 sec.

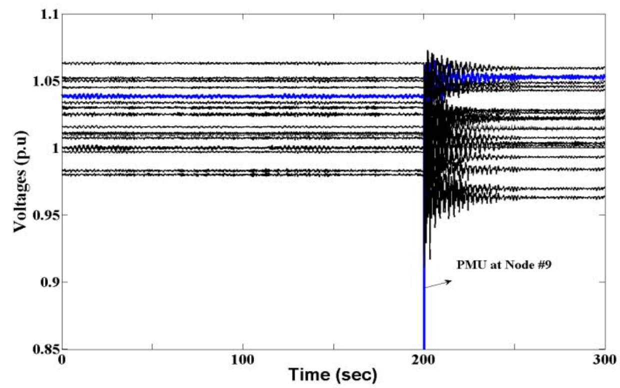

Fig. 9: Records with a three phase fault at line 8-9.

From Fig. 9, it can be noted that the PMU located at node 9, capture the impact of three phase fault. After the fault is cleared by opening the line 8-9, the system found another equilibrium point at approximately 50 seconds from the starting transient.

Following the procedure for on line monitoring described in Section 4, the control charts for T_i^2 and Q_i is computed considering the thresholds at 99% of confidence. Figure 10 shows the T_i^2 and Q_i control charts plots based on the PCA model. From these figures, we can clearly see that there are two operating conditions; the first 200 sec constitute the normal operating region and the last 100 sec of the simulation are the fault. In both control charts appear a few false alarms at the beginning of the simulation.

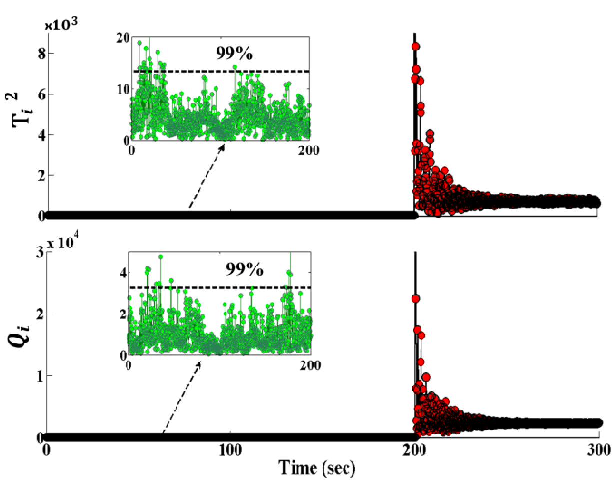

Fig. 10: Monitoring results using T_i^2 and Q_i control chart.

Figure 11 shows the score-domain trajectories projected into ellipse (2D) or ellipsoid (3D) to present a different perspective of the fault evolution. Considering the three largest PCs and using (12), the information is mapped inside the ellipse or ellipsoid. The points in green colour indicate the tracking of the first 200 seconds of the simulation. The points in red colour are generated following the fault inception (blue point) indicating the onset of the transient behaviour. Two operating conditions are clearly identified.

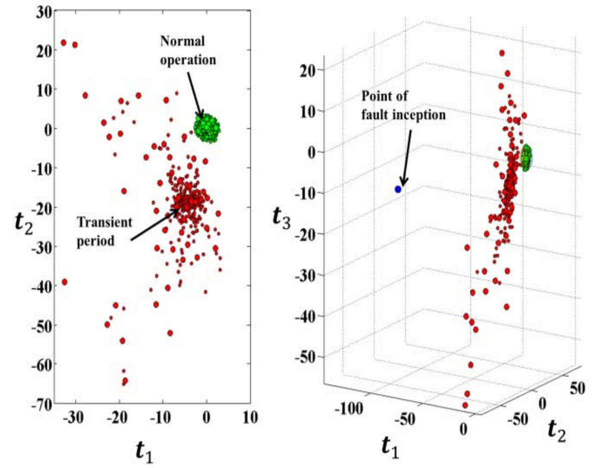

Fig. 11: Score plots for a three phase fault at line 8-9.

Once the fault is detected, the total contribution of each monitored node is computed to estimate the fault localisation. As a first approach the total contribution at fault inception is shown in Fig. 12.

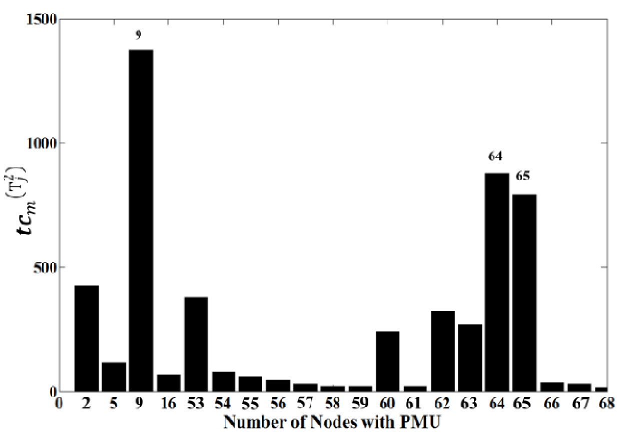

Fig. 12 Total contribution map for T_i^2 at fault inception point.

From Fig. 12, we can see that there are several PMUs that detect the abrupt change in the system. The PMUs located at nodes 9, 64, 65, 2 and 62 let us obtain a first approach about the fault's effects; suggesting the fault localization is around those nodes.

According to Fig. 10, the fault persists for a long period, almost 30 seconds after finding another equilibrium point. Furthermore, the total contribution map shows in Fig. 12 may be useful to visualize the fault evolution across the system.

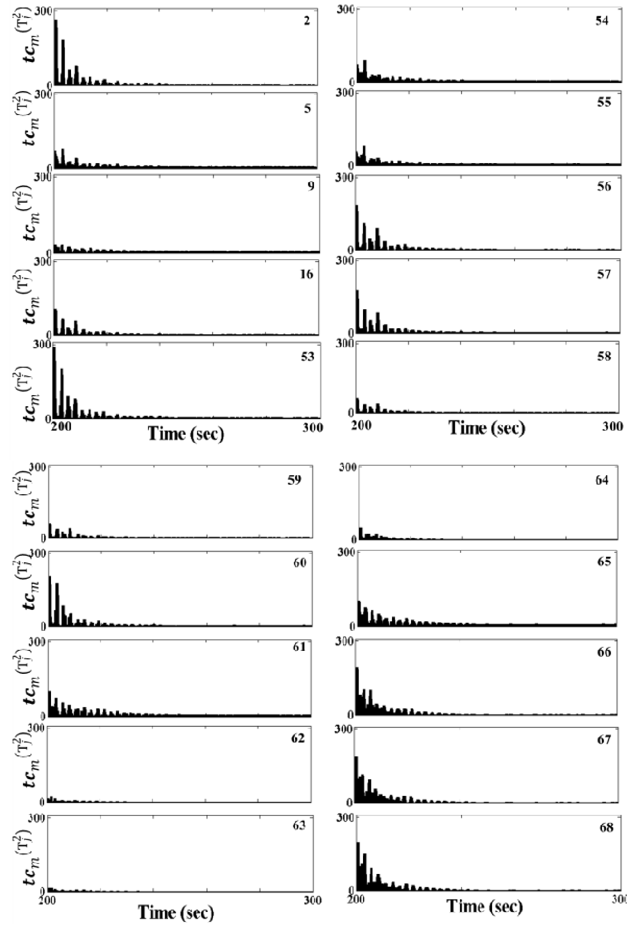

Fig. 13: Total contribution map visualization for T_i^2 .

From Fig. 13, we can identify that during the first 50 sec after the fault is cleared, the PMUs located at nodes 2, 53, 56, 57, 60, 66, 67 and 68, present the highest contributions, followed by PMUs located at nodes 5,16,54,55,64 and 65. The results suggest intense dynamic participation of 11 generators from the differents areas of the system and confirm that the fault is spreading across the system.

5.3 Case study II: Detection and Visualization of Anomaly Events.

System disturbances can be triggered by different causes. One of them could be related with the sensor failure due to communication system failure, bad calibration or more recently by cyber-physical attacks [16]. In any case the control actions based on data that come from faulty sensor (PMU) is at best inefficient and at worst dangerous, due to the potential source of propagation of a cascading fault across the system.

In this case, we assume that one PMU malfunctions and delivers wrong information. To carry out this test we assume that PMU located at node 2 malfunctions after 20 sec of simulation. Figure 14 shows the records collected by the PMUs for 300 sec. The arrow indicates the time of the fault inception.

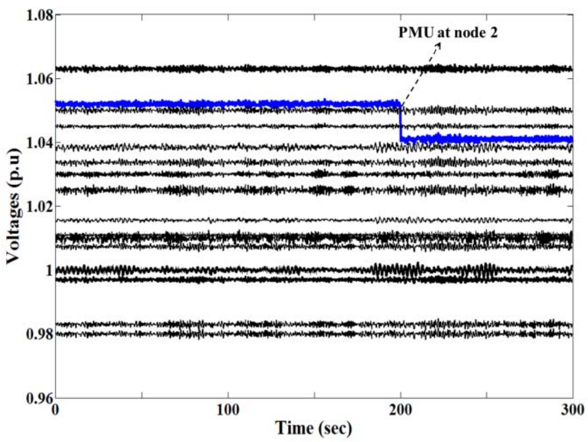

Fig. 14: Transient stability simulation with a malfunction at PMU located at node 2.

Then, following the same procedure for online detection; the control charts T_i^2 and Q_i are computed using the previous threshold. Figure 15 shows the T_i^2 and Q_i statistics plots. Both statistical tests identify the event. There we can see that the failure condition is maintained because the PMU causing the faulty condition is not repaired.

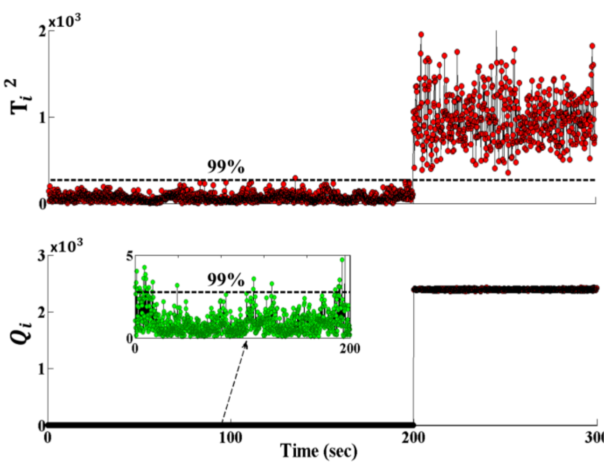

Fig. 15: Monitoring result using T_i^2 and Q_i control chart.

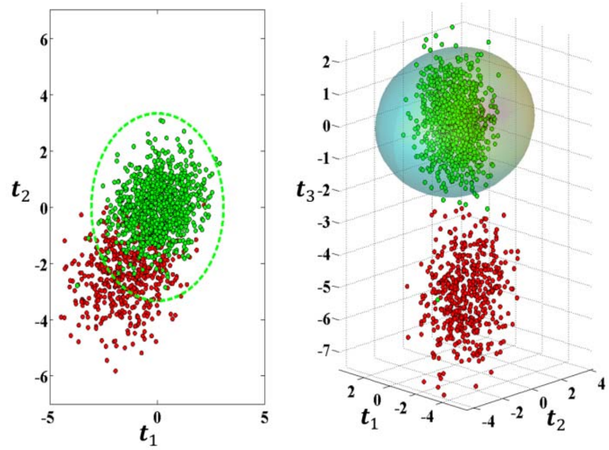

Fig. 16: Score plots for malfunction of PMU located at node 2.

Similar results are obtained by the score plot shown in the Fig. 16. Two closely operating conditions are identified, this suggests a change in the equilibrium point or operating condition. Both visualization perspectives are complementary to avoid a wrong intepretation of result. Further, the contribution plot is computed to obtain a visualization of the system behavior. Figure 17 shows the total contribution for the fault inception. The $tc_m^{\left(\mathrm{T}_j^2\right)}$ value for the PMU located at node 2 is dominant over the values of the other PMUs. This provides the first insight about the nature of the event.

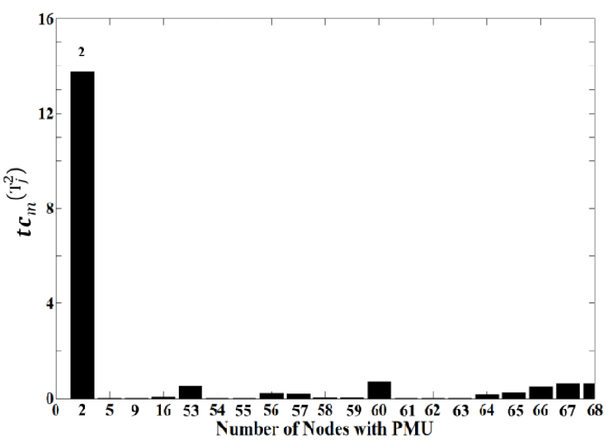

Fig. 16: Total contribution map for T_i^2 when the fault system is detected.

As the event persists for a long period the total contribution map is built to have a better visualization of the fault. From Fig. 17, we can confirm that the detected fault is associated with the malfunctioning of the PMU located at node 2.

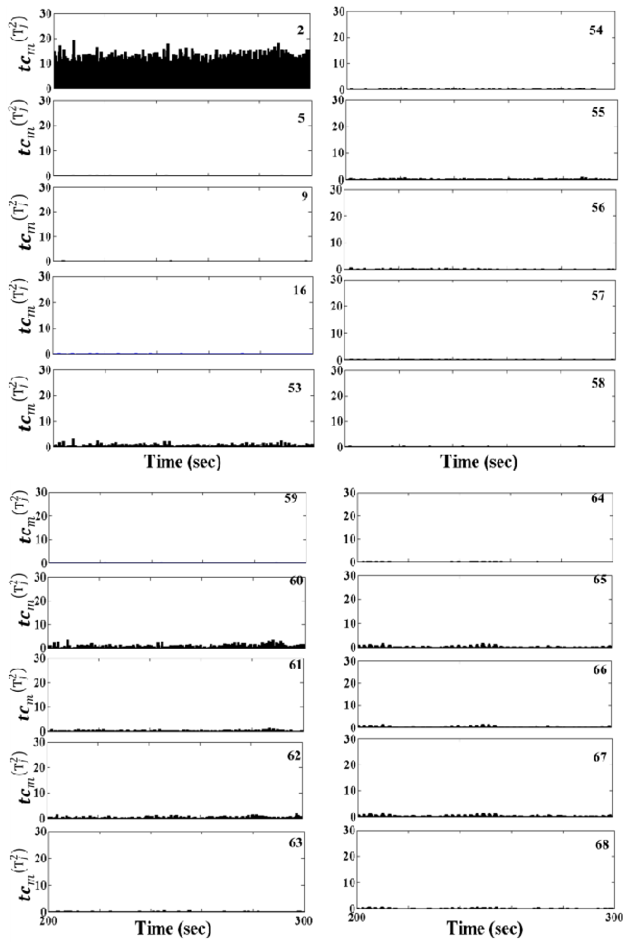

Fig. 17: Total contribution map results for T_i^2, considering an unexpected malfunction of PMU located at node 2.

6. Discussion

Both examples have shown that the PCA methodology detects and estimates correctly system disturbances and their locations. The total contribution map also provides a good estimation about the fault propagation and localization. Our proposed methodology is simple and the algorithm is low in computational cost, allowing building visualization tools that may be useful for the power system operator.

However, several challenges remain unaddressed. For instance, as the PCA model and its threshold in the beginning not adaptive, new stable operating conditions may be classified as faults or they may not be detected. Both circumstances are illustrated in the examples presented in the section 5.2 and 5.3. The first example, as shown in Fig. 10, the system reach two stable operation condition, the second one is classified as a fault. In the second example, (see Fig. 14), if the PMU intermittence malfunctions and its values do not change significatively, the PCA model may not be able to detect these changes.

The preliminary results inspire the development of more advances in PCA models and methods in the context of the power system monitoring. This would constitute an important step forward in grounding recent PCA work. Furthermore, different extensions of PCA method can be tested to enhance the monitoring capability. For instance, recursive PCA or kernel PCA could be applied to track the non-stationary system behavior; both approaches can handle time-varying operating conditions [17], [18]. Another technique that may be used to identify events and disturbances that appear in many different time scales is the multi resolution analysis based PCA model [19]. A different analysis could be applied to handle historic massive data, first using high dimensional PCA to expanding the visualization to more dimensions [20] or use batch process monitoring concept which is very well established in multivariate statistical methods [21], (see appendix A).

7. Conclusions

The preliminary results suggest that PCA allows determination of the secure operated margin, as well as detection of disturbances in the system using a visual representation. The total contribution maps can assist the operator with this centralized information, providing a quick map of the system behavior. The extension of the PCA method to detect multiple faults and small perturbations and to deal with non-stationary signals will be considered in forthcoming works.

8. Acknowledgments

The financial support from Marie Curie FP7-IAPP Project "Using real-time measurements for monitoring and management of power transmission dynamics for the smart grid- REAL-SMART", Contract No. PIAP-GA 2009-251304.

9. Appendix A: Multi-Way PCA Model

To understand the nature of batch data for power system monitoring problem, consider that batch process data includes time-varying trajectories of all the measured process variables $\boldsymbol{X}_i$, $i=1,\ldots N$ collected under normal operation for a number of batches. Data are then arranged in a three-way matrix $\boldsymbol{X} \in \mathbb{R}^{L,m,N}$, where L is the number of batches, m is the number of variables such as voltage, angle current or frequency and N is the number of times each batch is sampled. In the context of WAMS data, a batch is a short segment of PMU data. The three-way matrix is unfolded into a two-way array ($\boldsymbol{X} \in \mathbb{R}^{L,mN}$ or $\boldsymbol{X} \in \mathbb{R}^{LN,m}$) so that the direction of the batches is maintained as shown in the Fig. 18. In essence the monitoring methodology proposed in this paper can be extended to incorporate Multi-way PCA model. More details can be found in the reference [21].

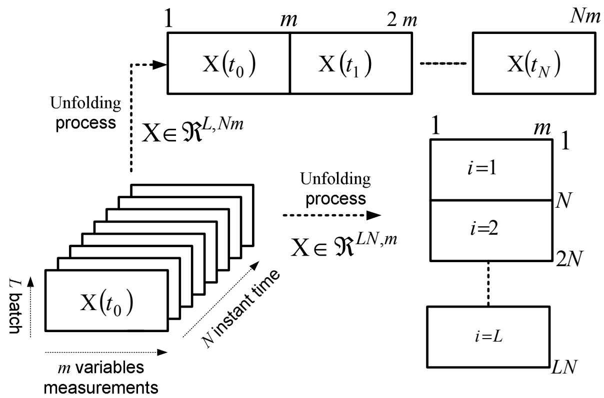

Fig. 18: Multiway unfolding process of a three-way batch data set.

10. References

[1] IEEE Task Force, "Identification of electromechanical modes in power systems", IEEE/PES, Special publication TP462, June 2012.

[2] Arturo R. Messina, I. Moreno N, J. J. Nuño, "Monitoring the health of large interconnected power systems: A near real-time perspective", 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS), August 29-31, Mexico , 2012.

[3] S-J.S. Tsai, J. Zou, Y. Zhang, Y. Liu, "Frequency visualization in large electric power systems*", IEEE/PES General Meeting*, San Francisco, Cal, 2005.

[4] R. Matthew Gardner, B. Jordan, Yilu Liu, "Wide-area visualization strategy based in FNET measurements", *IEEE/PES General Meeting*, Calgary, CAN, 2009.

[5] J. Ma, Y.V. Makarov, R. Diao, P.V. Etingov, J.E. Dagle, E. De Tuglie, "The characteristic ellipsoid methodology and its application in power systems*", IEEE Transc. On Power Systems*, Vol. 27, No. 4, pp. 2206-2214, Nov. 2012.

[6] Ya-Gang Zhang, Zeng-Ping Wang, Jin-Fang Zhang, Jing Ma, "Fault localization in electrical power systems: A pattern recognition approach" Electrical Power and Energy Systems, Vol. 33, pp. 791–798, 2011.

[7] Daniel J. Burke, Mark J. O'Malley, "A study of principal component analysis applied to spatially distributed wind power" IEEE Transac. On Power Systems, Vol. 26, No. 4, pp. 2084- 2092, Nov. 2011.

[8] Krishna K. Anaparthi, Balarko Chaudhuri, Nina F. Thornhill, Bikash C. Pal, "Coherency identification in Power Systems through Principal Component Analysis", IEEE Transac. On Power Systems, Vol. 20, No. 3, pp. 1658-1659, August 2005.

[9] Jolliffe**,** I.T, "Principal Component Analysis", Series: Springer Series in Statistics, 2nd ed. 2002, XXIX, 489 p, ISBN 978-0-387-95442-4.

[10] Barry M. Wise and Neal B. Gallagher, "The process chemometrics approach to process monitoring and fault detection", Journal Proceeding Cont., Vol. 6, No. 6, pp. 329-348, 1996.

[11] Richard A. Johnson, Dean W. Wichern,"Applied multivariate statistical analysis", sixth edition, Pearson, Prentice Hall, ISBN-13: 978-0131877153, 2007.

[12] Johan A. Westerhuis, Stephen P. Gurden, Age K. Smilde, "Generalized contribution plots in multivariate statistical process monitoring", Chemometrics and Intelligent Laboratory Systems, vol. 51, pp. 95-114,2000.

[13] B. Pal and B. Chaudhuri, "Robust Control in Power Systems", Springer, USA, ISBN-13: 9781441938534 2005.

[14] Francisco Gutierrez, E. Barocio, F. Uribe, and P. Zuniga, "Vulnerability analysis of power grids using modified centrality measures", Discrete Dynamics in Nature and Society, http://dx.doi.org/10.1155/2013/135731, Volume 2013.

[15] Pedro R. Peres-Neto, Donald A. Jackson, Keith M. Somers, "How many principal components? Stopping rules for determining the number of non-trivial axes revisited", Elsevier, Computational Statistics and Data Analysis, vol. 49, pp. 974-997, 2005.

[16] Daniel P. Shepard and Todd E. Humphreys, "Evaluation of the vulnerability of Phasor Measurement Units to GPS Spoofing Attacks", International Journal of Critical Infrastructure Protection, Elsevier, Vol. 5, Issues 3-4, pp. 146-153, 2012.

[17] Weihua Li, H. Henry Yue, Sergio Valle-Cervantes, S. Joe Qin, "Recursive PCA for adaptive process monitoring" Journal of Process Control, Vol.10, pp. 471-486, 2000.

[18] Xueqin Liu, Uwe Kruger, Tim Littler, Lei Xie , Shuqing Wang, "Moving window kernel PCA for adaptive monitoring of nonlinear processes", Chemometrics and Intelligent Laboratory Systems, Vol. 96, pp. 132-143, 2009.

[19] Choi, S.; Morris, J.; Lee, I. "Nonlinear multi-scale modelling for fault detection and identification", Chemical Engineering Science, Vol. 63, No. 8, pp. 2252-2266, 2008.

[20] Nina F. Thornhill, Hallgeir Melbo and Jan Wiik, "Multidimensional visualization and clustering of historical process data", Ind. Eng. Chem. Res. Vol. 45, pp. 5971-5985, 2006.

[21] P. Nomikos, J.F. MacGregor "Monitoring batch processes using Multiway Principal Component Analysis", AIChE Journal Vol.40 N. 8, p-p, 1361-1375, 1994.