Comparison of Principal Component Analysis Techniques for PMU Data Event Detection

L. Souto, J. Meléndez, S. Herraiz

Intelligent Systems and Control Engineering group

Universitat de Girona

Girona, Girona, Spain

laiz.souto, joaquim.melendez, sergio.herraiz AT udg.edu

Abstract—**Principal component analysis (PCA) is a dimension-ality reduction technique often applied to process and detect events in large amounts of data collected by phasor measurement units (PMU) at transmission and distribution level. This article considers five different approaches to select an appropriate number of principal components, builds the statistical model of the PMU data online over a sliding window of 10 seconds and 1 minute, and evaluates the computation times and the accuracy of correct event detections with use of two statistical tests in a $1-$hour data file from the UT-Austin Independent Texas Synchrophasor Network with phasor quantities collected at different PMU substations.**

Index Terms—**fault detection, phasor measurement units, power system faults, principal component analysis**

I. Introduction

THE increasing digitalisation of electric power systems is generating vast quantities of data at different locations, voltage levels, and time intervals. These data include distinct electrical quantities collected by smart meters and phasor measurement units (PMU) that can be exploited to characterise energy behavioural patterns and identify anomalies of different nature. However, due to the high complexity and massive amounts of data, it is a difficult task to visualise and identify patterns, outliers, and abnormal behaviours at relevant scales.

In this scenario, dimensionality reduction techniques are appealing to reduce a dataset with optimality and minimal loss of information. Among the dimensionality reduction techniques commonly applied to electric power systems, principal component analysis (PCA) is one of the most widely used (see, e.g. [1]–[5]). It builds a data-driven model of observations in which the covariance structure is described with a reduced number of dimensions through a few linear combinations of the original variables that express major trends in the dataset. Much as the strategies to define an appropriate number of principal components are well known in literature (see, e.g. [6], [7]), there is a lack of consensus about how to adjust them to detect specific events of interest in PMU data. Therefore, a systematic evaluation of those procedures is necessary to perform dimensionality reduction and event detection with PCA effectively.

Fitting into this context, this paper provides an in-depth comparative analysis of five different approaches to select an adequate number of principal components that define the statistical model of a PMU network when it comes to the accuracy of correct event detections with usage of distinct statistical tests. To do so, it relies on phasor quantities measured at multiple PMU substations to build a PCA model of the network operating conditions in real time, which increases the situational awareness of the analysis, over a sliding window of distinct lengths of time. The analysis is tested in a 1-hour data file from Texas Synchrophasor Network and applicable to power transmission and distribution networks with multiple PMUs installed at different locations without requesting any information about the network topology and its electrical parameters.

The text is organized as follows. The theoretical background is presented in Section II: II-A explains the building of the PCA model, II-C includes five different calculation methods to select the number of principal components, II-B describes event detection in the projection subspace and residual subspace. Afterwards, a study case is shown in Section III, the results and discussion are described in Section IV, and conclusions are finally presented in Section V.

II. Theoretical background on PCA

The PCA methodology presented in this article, adapted from [8], builds a statistical model of the PMU data collected over a time window of duration τ_s and detects anomalies with use of two complementary indicators: the Hotelling's T^2 (T^2) and the square prediction error (SPE) statistics. The former measures the square distance of the projected data to the centre of the model, whereas the latter measures the square distance of the observation to the projection subspace.

Thereby, the method may be divided into three main steps, as presented in Sections II-A–II-B: building the PCA model with PMU data, selection of principal components to the projection subspace, and event detection in the projection subspace and residual subspace.

A. Building the PCA model with PMU data

Let $\mathbf{X}$ be the $n \times m$ observation matrix displayed in (1) with n observations – referred to the number of samples of phasor quantities – and m variables – referred to PMU locations – supposed to be centered (zero mean) and scaled (unit variance)

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} \quad (1)$$

Then, compute the covariance matrix of $\mathbf{X}$ and apply eigenvalue decomposition to obtain two $m \times m$ matrices $\mathbf{V}$ (whose columns are the eigenvectors and contain the principal components) and $\mathbf{\Lambda}$ (diagonal matrix whose elements express the variability in the direction of each principal component or column of $\mathbf{V}$) with (2)

$$\mathbf{V\Lambda V^T} = \frac{1}{n-1}\mathbf{X^T X} \quad (2)$$

Dimensionality reduction in the number of variables can be performed by retaining r principal components of $\mathbf{V}$ ($r < m$) with the largest eigenvalues. Then, the $m \times m$ matrix $\mathbf{V}$ becomes an $m \times r$ matrix $\mathbf{P}$ which defines a projection space of lower dimension representing the r most significant principal components. As the choice of an appropriate value of r is not straightforward, five different methods to select an appropriate value of r are presented later in Section II-C and compared in Section IV.

The usage of $\mathbf{P}$ instead of $\mathbf{V}$ to transform $\mathbf{X}$ into the principal components representation space results in a projection onto a space of lower dimension in which some information contained in the original data is lost. As a matter of a fact, since $\mathbf{V}$ is a unitary matrix, the inverse operation is carried out with the transpose, i.e. $\mathbf{VV^T} = \mathbf{I}$, but $\mathbf{P}$ is not unitary, therefore $\mathbf{PP^T} \neq \mathbf{I}$. The scores and the transformation of scores into the original data with $\mathbf{P}$ can be calculated with (3)

$$\begin{cases} \mathbf{t} = \mathbf{xP} & \text{and} \quad \hat{\mathbf{x}} = \mathbf{tP^T} \\ \mathbf{T} = \mathbf{XP} & \text{and} \quad \hat{\mathbf{X}} = \mathbf{TP^T} \end{cases} \quad (3)$$

where $\mathbf{t}$ and $\hat{\mathbf{x}}$ denote the score and projection of a single observation $\mathbf{x}$ (top) and $\mathbf{T}$ and $\hat{\mathbf{X}}$ denote the score matrix and projection matrix of the whole dataset $\mathbf{X}$ (bottom). The difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is the residual matrix $\tilde{\mathbf{X}}$ which resumes the information contained in the $m - r$ components from the residual space for each observation and can be calculated with the residual matrix $\tilde{\mathbf{C}}$. Thereby, the complete PCA model can be described as in (4)

$$\tilde{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X}\left(\mathbf{I} - \mathbf{PP^T}\right) = \mathbf{X}\tilde{\mathbf{C}} \quad (4)$$

B. Detection in the projection subspace and residual subspace

Event detection in the projection subspace is evaluated with the T^2 index, which computes a weighted distance of the projected data to the centre of the model using λ_i as a weight using (5)

$$T_x^2 = \sum_{i=1}^{r} \frac{t_i^2}{\lambda_i} \quad (5)$$

For a single observation $\mathbf{x}$ whose score vector is $\mathbf{t}$, the $\mathbf{T_x^2}$ index is given by (6)

$$T_x^2 = \mathbf{t\Lambda_r}^{-1}\mathbf{t^T} \quad (6)$$

with

$$\mathbf{\Lambda_r}^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_r} \end{bmatrix} \quad (7)$$

The statistical limit T_{lim}^2 is calculated analytically with (8)

$$T_\alpha^2 = \frac{r\left(n^2 - 1\right)}{n\left(n - r\right)}F_\alpha\left(r, n - r\right) \quad (8)$$

where α is the confidence level and $F_\alpha\left(r, n - r\right)$ is the critical point of the Fischer-Snedecor distribution for r and $n - r$ degrees of freedom. Any result that surpasses T_{lim}^2 is tagged as faulty for the T^2 statistics.

In turn, event detection in the residual subspace is evaluated with the SPE index, which evaluates the variation out of the projection space defined by the r principal components through the error component $\tilde{\mathbf{x}}$. The SPE of an observation $\mathbf{x}$, Q_x is given by (9)

$$Q_x = \left(\mathbf{x} - \hat{\mathbf{x}}\right)\left(\mathbf{x} - \hat{\mathbf{x}}\right)^T = \|\tilde{\mathbf{x}}\|^2 \quad (9)$$

The statistical limit Q_{lim} is calculated analytically with (10)

$$Q_\alpha = \theta_1\left[1 + \frac{h_0 c_\alpha\sqrt{2\theta_2}}{\theta_1} + \frac{h_0\theta_2\left(h_0 - 1\right)}{\theta_1^2}\right]^{\frac{1}{h_0}} \quad (10)$$

with

$$\theta_k = \sum_{i=r+1}^{m} \lambda_i^k, k = \{1, 2, 3\} \quad \text{and} \quad h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (11)$$

where c_α is the normal deviation for a confidence level α. Any result that surpasses Q_{lim} is tagged as faulty for the SPE statistics.

C. Selection of principal components to the projection subspace

Five different methods to select an appropriate value of r are taken into consideration, as described in the following subsections.

1) Kaiser criterion: In this method, r is selected such that all principal components whose eigenvalues are below the average variance are dropped from the matrix $\mathbf{P}$, in agreement with (12). In other words, this criterion consists in retaining all r principal components whose variance is larger than one, as $\mathbf{X}$ is a scaled matrix. This ensures that every principal component selected contains at least as many information as a single original variable in terms of variance.

$$r := \{\max\ j \in \{1, \cdots, m\} \mid \lambda_j \geq \frac{1}{m}\sum_{i=1}^{m}\lambda_i\} \quad (12)$$

2) Automatic scree plot: In this method, the eigenvalues λ_i are plotted decreasingly as a function of their element number i in the matrix $\boldsymbol{\Lambda}$ and the chosen value of r corresponds to the eigenvalue whose distance to the origin of the coordinate system is the shortest, in agreement with (13). The idea is to search for an elbow in the plot, which always displays a downward curve, from which the eigenvalues are approximately equal.

$$r := \{i \in \{1, \cdots, m\} \mid \min_i \sqrt{\lambda_i^2 + i^2}\} \qquad (13)$$

3) Explained variance: In this method, a minimum percentage of the total variance $Var\,(\%)$ is previously defined and r is taken as the smallest integer satisfying (14)

$$r := \frac{1}{m} \sum_{i=1}^{r} \lambda_i \times 100 \geq Var\,(\%) \qquad (14)$$

4) Variance reconstruction error: In this method, further explained in [6], the optimal value of r is determined by the minimum variance reconstruction error (VRE), in agreement with (15), considering a faulty observation $\mathbf{x_f}$ represented by an m-dimensional unitary vector $\xi_{\mathbf{i}}$ multiplied by a fault magnitude f and the correlation matrix of reconstruction error $\mathbf{R}$. This procedure results in the best reconstruction of the variables, as the VRE decreases monotonically in the residual subspace and increases in the projection subspace with the number of principal components, and the selection of r can be adjusted to detect specific events of interest defined by $\mathbf{x_f}$.

$$r := \{j \in \{1, \cdots, m\} \mid \min_j \xi_{\mathbf{i}}^T \mathbf{R} \xi_{\mathbf{i}}\} \qquad (15)$$

5) Statistical detectability: In this criterion, based on [8], r is chosen such that the smallest detectable events can be detected statistically in the projection subspace and residual subspace, according to Section II-B, considering that $T_x^2 \geq T_{lim}^2$ must hold so that an event can be detected with T^2 statistics and $Q_x \geq Q_{lim}$ must hold so that an event can be detected with SPE statistics. For a faulty observation $\mathbf{x_f}$, r is selected such that (16) holds with a single PCA model or (17) holds with two PCA models built separately for T^2 and SPE statistics, which enables adjustments to detect specific events of interest defined by $\mathbf{x_f}$.

$$r := \{\min\ i \in \{1, \cdots, m\} \mid T_{x_f}^2 \geq T_{lim}^2\ and\ Q_{x_f} \geq Q_{lim}\} \qquad (16)$$

$$r := \left\{ \begin{array}{l} r_{T^2} := \{\min\ i \in \{1, \cdots, m\} \mid T_{x_f}^2 \geq T_{lim}^2\} \\ r_Q := \{\min\ i \in \{1, \cdots, m\} \mid Q_{x_f} \geq Q_{lim}\} \end{array} \right. \qquad (17)$$

This criterion considers the worst-case scenario to build one-size-fits-all PCA models for the overall grid. This choice shall be able to detect the smallest theoretical values of $T_{x_f}^2$ and Q_{x_f} computed with each individual variable possibly involved in the event for 16 or 17 and consequently presents the highest theoretical statistical detectability without requiring different PCA models for each set of variables.
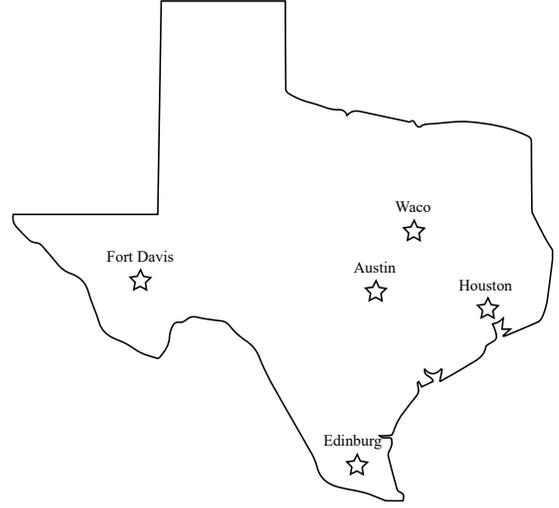
Fig. 1. Map of PMU substations within the Texas Synchrophasor Network

III. CASE STUDY

The procedures described in Section II-C to select an appropriate number of principal components and perform event detection are tested with PMU data from the UT-Austin Independent Texas Synchrophasor Network available in [9] (see website). A map of the locations of the PMUs, installed at distinct transmission and distribution voltage levels within the Electric Reliability Council of Texas (ERCOT), is shown in Fig. 1.

Only low-frequency oscillations below $15\ Hz$ can be found in the dataset, as the phasor quantities are gathered at $f_s = 30\ Hz$. Thus, this article is concerned about low-frequency voltage transients which last no more than a few hundred milliseconds, as highlighted in Fig. 2; the detection of faster dynamic events is out of the scope of this article, as it requires a higher sampling frequency of PMUs. The events found in the dataset, whose types and inner causes are unknown, are listed in Table I. Overall, they are expected to be isolated events at a single location (which ensures detection of the smallest events in the worst-case scenario) and occur once or less every 10 seconds. Their magnitudes are greater than $1\ \%$ of the nominal voltage, varying from a few Volts at low voltage level (substation at Fort Davis) to a few thousand Volts at high voltage level (substations at Edinburg and Waco). Duration of disturbances is also heterogeneous, lasting from a few milliseconds (impulses) to a few hundred milliseconds (transients).

Relying on voltage magnitudes, the PCA model is built online over a sliding window. This approach captures the dynamic time-varying nature of power systems and adapts the PCA model to the most recent operating conditions. The five different methods to select the number of principal components described in Section II-C are considered to build the PCA model and further tested and compared in terms of performance of T^2 and SPE statistics over a $10-$second and $1-$minute sliding window, supposedly associated with
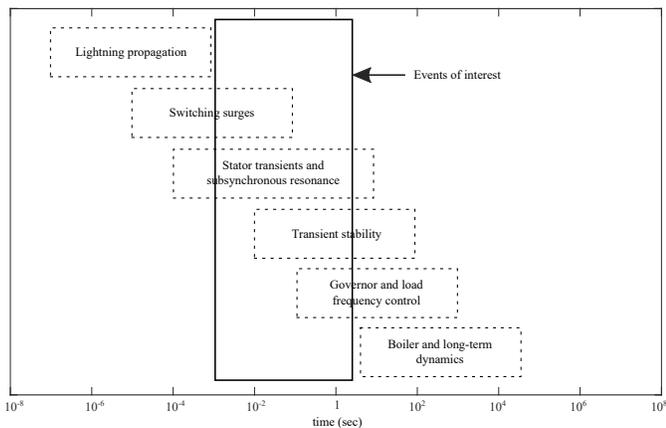
Fig. 2. Time ranges of distinct dynamic phenomena in power systems (adapted from [10])

TABLE I
TYPES OF EVENTS

#	Location	Category
1	Edinburg	Impulse, multiple
2	Edinburg	Transient
3	Fort Davis	Impulse, single
4	Edinburg	Impulse, multiple
5	Edinburg	Transient
6	Fort Davis	Impulse, single
7	Fort Davis	Impulse, single
8	Waco	Transient
9	Edinburg	Transient
10	Edinburg	Impulse, multiple
11	Edinburg	Impulse, multiple
12	Edinburg	Impulse, multiple
13	Edinburg	Transient
14	Edinburg	Transient
15	Fort Davis	Impulse, single
16	Edinburg	Impulse, multiple

TABLE II
EVENT DETECTION RESULTS (YES/NO) OVER A 10−S WINDOW

#	Kaiser			Scree plot			Variance			VRE			Detectability			
	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	r	Q
1	3	Y	Y	4	Y	Y	3	Y	Y	3	Y	Y	2	Y	2	Y
2	2	N	Y	3	N	Y	3	N	Y	1	N	Y	4	Y	1	Y
3	2	Y	Y	2	Y	Y	3	N	Y	2	Y	Y	4	Y	1	Y
4	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
5	3	Y	N	3	Y	N	3	Y	N	2	Y	Y	3	Y	1	Y
6	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	N	2	Y
7	2	N	Y	3	N	Y	3	N	Y	2	N	Y	2	N	2	Y
8	2	N	Y	3	Y	N	2	N	Y	2	N	Y	3	Y	3	N
9	3	Y	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	1	Y
10	2	Y	Y	2	Y	Y	2	Y	Y	1	N	Y	2	Y	2	Y
11	2	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
12	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
13	2	Y	Y	3	Y	N	3	N	Y	2	Y	Y	3	Y	1	Y
14	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
15	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
16	2	Y	Y	2	Y	Y	2	Y	Y	1	N	Y	2	Y	2	Y

TABLE III
EVENT DETECTION RESULTS (YES/NO) OVER A 1−MIN WINDOW

#	Kaiser			Scree plot			Variance			VRE			Detectability			
	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	r	Q
1	3	Y	Y	4	Y	N	3	Y	Y	2	Y	Y	2	Y	2	Y
2	2	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
3	2	Y	Y	5	Y	N	3	Y	Y	1	Y	Y	2	Y	2	Y
4	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
5	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
6	2	Y	N	2	Y	N	3	Y	N	1	N	Y	2	Y	1	Y
7	2	Y	Y	3	Y	Y	3	Y	Y	2	Y	Y	4	Y	1	Y
8	2	N	N	3	N	N	3	N	N	1	N	N	2	N	2	N
9	3	Y	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	3	Y
10	2	N	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	3	Y
11	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
12	3	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	3	Y	3	Y
13	2	N	Y	2	N	Y	3	Y	Y	2	N	Y	3	Y	1	Y
14	2	Y	Y	3	Y	Y	2	Y	Y	2	Y	Y	2	Y	2	Y
15	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
16	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y

different types of events and intrinsic characteristics. As a result, 20 different scenarios were produced per detected event (i.e. 2 statistical tests times 5 selection criteria of r times 2 window lengths).

IV. RESULTS AND DISCUSSION

This section presents the results of event detection for the scenarios described in Section III, comparing the accuracy of the methods described in Section II-C to select the number of principal components r for the events shown in Table I. Tables II and III display the event detection results and the corresponding r obtained for each selection method and time window, considering a 10−second and 1−minute sliding window to build the PCA models, respectively. A confidence level $\alpha = 0.95$ is chosen to calculate T^2 and SPE for all selection methods because it does not result in missed detections in this dataset with the statistical detectability criterion over a 10−second sliding window. The locations and event magnitudes listed in Table I are considered to compute the statistical detectability and the VRE. In addition, the explained variance criterion is computed with an explained variance of

75 % because of the average residual subspace defined by the other criteria.

On the whole, considering the correct detections with both T^2 and SPE statistics, the best method to select the number of principal components is the statistical detectability criterion, which detects more events than the other criteria for all lengths of time evaluated. Nevertheless, it is noteworthy that all events of interest are detected with at least one of the T^2 or SPE statistics when a 10−second window is used to build the PCA model, and that the event number 8 is always missed with a 1−minute window, regardless of the method chosen to define an appropriate number of principal components. Additionally, it can be noticed that the combined use of T^2 and SPE statistics leads to a higher number of correct event detections, reduces the number of missed detections, and consequently increases the detection capability of the PCA model.

In general, the SPE is expected to detect variations associated with changes in the correlation structure and presents small values, whereas the T^2 is expected to detect deviations

from the average normal operating conditions and presents larger values. As a consequence, the SPE is more sensitive than the T^2 and tends to be a better indicator of abnormalities, since changes in the correlation structure of faulty observations are expected to be observed in the residual matrix. In particular, the SPE is expected to present the best performance when the VRE criterion is applied to select r, as it is aimed at minimizing the variance reconstruction error associated with the residual subspace to ensure detection of specific events of interest.

In most cases, the highest number of correct event detections occurs when a 10-second window is applied to build the PCA model with both T^2 and SPE statistics. It happens because the PCA models built over this length of time are more sensitive to the dynamics of the system and consequently are more suitable to detect the events of interest, which last no more than a few hundred milliseconds. Therefore, it is expected that a higher number of missed detections will occur with longer window sizes, which are associated with different phenomena and/or a static representation of the operating conditions of the grid. This explains why the results obtained with a $1-$minute window are slightly worse than those obtained with a $10-$second window in most cases.

Furthermore, a comparative analysis of the five approaches presented in Section II-C to select an adequate value of r over time evinces that the Kaiser criterion, the automatic scree plot, and the explained variance are quantitative and arbitrary, whereas the statistical detectability and the VRE are qualitative. As a matter of a fact, the former criteria rely on the variance of the data and do not allow for adjustments that ensure the detection of specific events of interest, whereas the latter criteria can be adjusted to detect events defined by specific magnitudes f and direction vectors ξ. This explains why the results obtained with the statistical detectability criterion with both T^2 and SPE statistics are the most accurate, which implies that the detectability in the projection subspace and the residual subspace is the best indicator to select an adequate number of principal components and detect distinct events of interest.

V. CONCLUSION

This paper presents a PCA-based strategy for PMU data event detection with usage of T^2 and SPE statistics, considering five different methods to select an appropriate number of principal components which are compared in terms of correct event detections. The results indicate that the PCA methodology is able to identify different types of events, regardless of the approach used to select the number of principal components, and that the results calculated with T^2 and SPE statistics are complementary in some cases, which enhances the detection capability of the PCA model. Nonetheless, the results obtained when the statistical detectability criterion is used to define the number of principal components are the most accurate with both T^2 and SPE statistics, which implies that the detectability in the projection subspace and the residual subspace is the best indicator to select an adequate

number of principal components. Moreover, the window size applied to build the PCA model also contributes to the task and shall be adjusted according to the events of interest.

REFERENCES

[1] Y. Chen, L. Xie, and P. R. Kumar, "Dimensionality reduction and early event detection using online synchrophasor data," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–5.

[2] Y. Ge, A. J. Flueck, D. Kim, J. Ahn, J. Lee, and D. Kwon, "Power system real-time event detection and associated data archival reduction based on synchrophasors," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 2088–2097, July 2015.

[3] T. Xu and T. Overbye, "Real-time event detection and feature extraction using pmu measurement data," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2015, pp. 265–270.

[4] X. Liu, D. M. Laverty, R. J. Best, K. Li, D. J. Morrow, and S. McLoone, "Principal component analysis of wide-area phasor measurements for islanding detection—a geometric view," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 976–985, April 2015.

[5] M. Rafferty, X. Liu, D. M. Laverty, and S. McLoone, "Real-time multiple event detection and classification using moving window pca," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2537–2548, Sep. 2016.

[6] S. J. Qin and R. Dunia, "Determining the number of principal components for best reconstruction," *Journal of Process Control*, vol. 10, no. 2, pp. 245 – 250, 2000.

[7] D. Peña, *Análisis de datos multivariantes*. McGraw-Hill, 2002.

[8] E. Russell, L. Chiang, and R. Braatz, *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*, ser. Advances in Industrial Control. Springer London, 2000.

[9] A. Allen, M. Singh, E. Muljadi, and S. Santoso, "Pmu data event detection: A user guide for power engineers," 2014, link to accompanying zipped data file: http://www.nrel.gov/docs/fy15osti/61664-1.zip.

[10] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Prentice-Hall, 1998.