# Principal Component Analysis Applied to Continuation Power Flow

A. Fonseca, *Member, IEEE*

***Abstract*—This paper analyzes the applicability of Principal Component Analysis (PCA) to study the high voltage solutions of Continuation Power Flow (CPF). A simulation based process is used for data generation and subsequent PCA application with their results analysis.**

**The simulation considers a single bus – load equivalent and controlled variation of the equivalent system impedance behind the load. Variations of data structure are proposed and analyzed in order to interpret results from power system perspective and exploration about possible hidden correlations obtained from PCA application. Finally, on the basis of these correlations, the analysis of clusters is discussed.**

***Index Terms*—Clustering, Continuation Power Flow (CPF), Data Reduction, Feature Extraction, Principal Components Analysis (PCA).**

## I. Introduction

PRINCIPAL Component Analysis (PCA) is a powerful tool for data processing which includes data reduction and feature extraction. In the context of power systems, CPA presents very interesting perspectives for solution of many kinds of problems, for instance, some interesting approaches are proposed in [1-5].

The authors in [1] and [2] present approaches for real time dynamic vulnerability assessment for power systems and detection of islanding conditions. In [3] and [4], PCA is used to analyze the steady state operational power grid data and expose some correlations. In [5], the researchers describe an algorithm for transformer differential protection based on pattern recognition of the differential current. In summary the potential of PCA for data reduction and feature extraction for power systems is high.

On the other hand, Continuation Power Flow (CPF) has been powerfully used as a tool for planning studies to determine the power transfer capacity for many operating conditions of power systems. PV–Curves or nose-curves are the main result of CPF and they play a major role in understanding and explaining voltage instability.

This article sheds some light to understand the application of PCA to study the high voltage solutions of CPF. It exposes some correlations between principal components of PCA with control and state variables of power system, and shows possible groups formation of these relations.

A. Fonseca is with the Department of Substations Specialized Maintenance from CELEC EP –Transelectric in the Electrical Studies and Protection Area, Quito, Ecuador (e-mail: antonio.fonseca@celec.gob.ec)

## II. Theoretical Analysis

### A. Continuation Power Flow - CPF

Generally, the power transference capacity of the power grid is one of the principal elements to limit stability in a power system. This capacity is linked with the electrical characteristics of the elements of generation, transmission system and associated control devices.

Fig.1 shows the basic model used to analyze the power transference capacity, for simplicity the resistance was neglected.

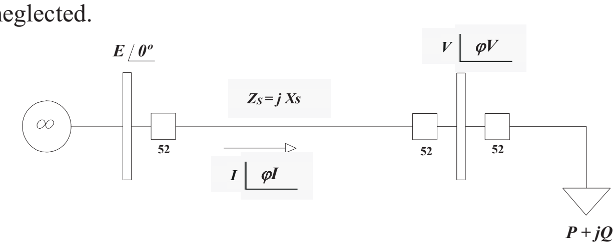



Fig. 1. Two bus transmission system

The basic equations for active and reactive powers allow to define a second order equation, with respect to $V^2$, which has the solution (1)

$$V = \sqrt{\frac{-(2QX_S - E^2) \pm \sqrt{(2QX_S - E^2)^2 - 4X_S{}^2(P^2 + Q^2)}}{2}} \quad (1)$$

There is a real solution if:

$$-P^2 - \frac{E^2}{X_S}Q + \left(\frac{E^2}{2X_S}\right)^2 \geq 0 \quad (2)$$

When (2) is equal to zero, there is a unique voltage solution of power flow; on the other hand, when (2) is greater than zero there are two voltage solutions. Consequently, this equation represents the convergence limit for power flow and correspond to the point of maximum power flow transferred between two points of a transmission grid.

Applying continuous load increase, it is possible to obtain the solutions at each load step. For real power systems, the relations between variables is more complex than (2) and it is necessary to use additional tools to find the solutions at any operating conditions, especially near to the convergence limit. One of the most used tools for this purpose is CPF.

Formally, the CPF uses an iterative process involving predictor and corrector steps, for finding power flow solutions, starting from a known initial solution and the continuous increase of load [6]. This methodology allow to find the entire power flow solution even near the stability limit as shown in Fig. 2.

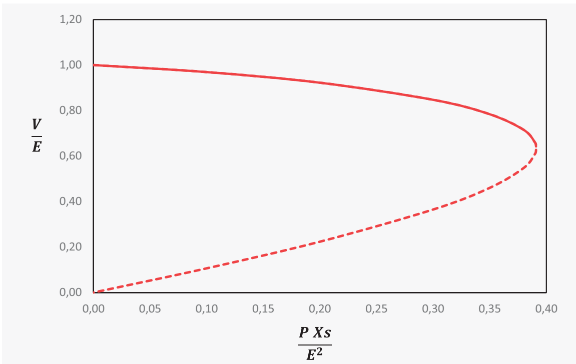



Fig. 2. Entire PV- Curve. Solution domain for 0.97 inductive power factor

The continuation power flow is robust and flexible, however the method is time consuming. A better performance is obtained for computing conventional power flow solutions at up to their normal convergence, then switch to continuous methods. If consider only the first step, the solutions correspond to high voltage solutions of CPF, as shown in the continuous line in Fig. 2.

Additionally, if in this model $Z_s$ changes, it is possible to analyze in a simple way, the effects of transmission contingencies on the power transference capacity. If (2) changes into the impedance ratio ($X_{s\text{-Thevenin}} / Z_{l\text{-Load}}$) domain , it is possible to show the domain of power flow solution in function of the impedance ratio $X_s/Z_l$ as indicated in the continuous line in Fig. 3.

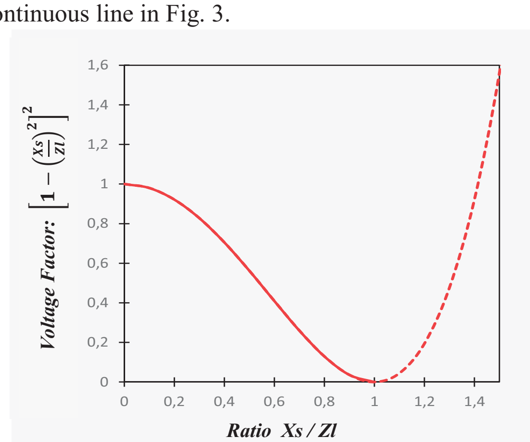



Fig. 3. Power flow solution domain in function of impedance ratio for two bus transmission system

This Voltage Factor has an inflection point as shown in (3).

$$\frac{d^2}{d\left(X_s/X_l\right)^2}\left[1-\left(\frac{X_s}{X_l}\right)^2\right]^2 = 0 \quad \Rightarrow \frac{X_s}{X_l} = \frac{\sqrt{3}}{3} \approx 0{,}58 \qquad (3)$$

In this paper high voltage power flow solutions, for different values of ratio $X_s/X_l$, are simulated. PCA will be applied at these cases.

*B. Principal Component Analysis – PCA [7]*

The Principal Component Analysis is a powerful tool for data processing. It is characterized by a high dimensionality reduction with minimal loss of information.

Let **X** be a normalized data matrix, the covariance matrix **S** is derived from (4)

$$\mathbf{S} = \frac{1}{n-1}.\mathbf{X}^T\mathbf{X}. \qquad (4)$$

Because of the properties of this matrix, it is possible to generate other diagonal matrix D with their corresponding eigenvectors E. This diagonal variance matrix D is given by (5)

$$\mathbf{D} = \mathbf{E}^T.\mathbf{S}.\mathbf{E} \qquad (5)$$

The diagonal of D contains the variances $\sigma^2_k$ of X, and correspond with the eigenvalues $\lambda_k$ of matrix S. This eigenvalues are identical for both matrix: S and D.

Analyzing the descending order for all $\lambda_k$, it is possible to extract the principal orthogonal components of the data matrix X. This is doneto cover an adequate level of variability for the original data. This reduction generates a lower dimensional matrix $E_q$.

Once the number of components have been chosen, the projections of matrix X over the principal components are determined by (6).

$$\mathbf{F} = \mathbf{X}.\mathbf{E}_q \qquad (6)$$

## III. Application of PCA to CFP

Data Matrix X is obtained following the sequence as shown in Fig. 4.

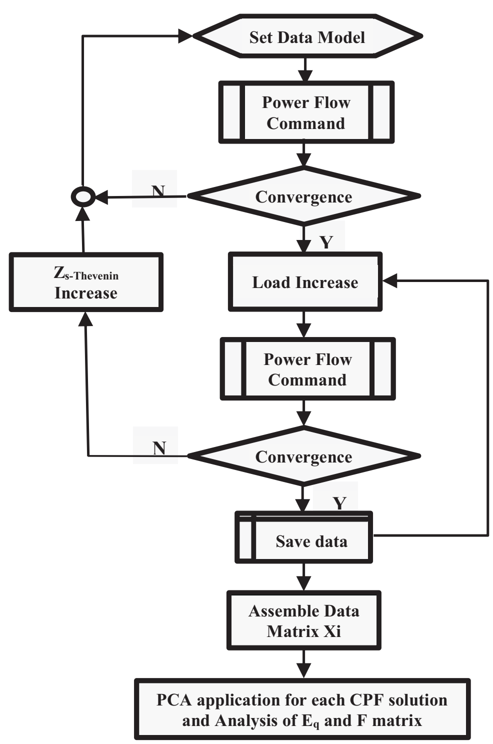



Fig. 4. Chart Flow for application of PCA to high voltage solution of CPF

The simulations consider the system in Fig. 1, and real data for a radial transmission substation at 69 kV [8]. These data correspond to an equivalent impedance of $Z_s = 0.0071 + j0.0985$ pu, short circuit current of 8.5 kA, load power and power factor around 120 MVA and 0.97 lagging, respectively. The power base for analysis is 100 MVA.

To assemble data matrix **X**, several cases are proposed as indicated in Table I, over this data matrix PCA is applied. The variables considered correspond to voltage, current, both in magnitude and angle, active and reactive power.

TABLE I
DATA MATRIX STRUCTURE

| Case | Principal Variables | Subcases |
|---|---|---|
| A | V, φV, I, φI, P, Q | - |
| B | V, I, P, Q | - |
| C | V, I | P, Q |
| D | P, Q | V, I |
| E | V | P, Q |
| F | I | P, Q |

The $Z_{S\text{-Thevenin}}$ considering an increase of 5% of the nominal impedance produces 49 general cases with approximately 150 data samples for each variable.

Fig. 5 shows load impedance $Z_{load}$ and Thevenin impedance $Z_s$ for a continuation power flow considering the increase of $Z_s$ for 3 cases: initial case ($Z_s = 0.0071 + j0.0985$ pu), middle case ($Z_S = 0.0221 + j0.3054$ pu) and final case ($Z_S = 0.0243 + j0.3349$ pu). For middle case the ratio for $X_S/X_l$ at initial point is approximate 0.588.

For permanent determination of $Z_{s\text{-Thevenin}}$ the methodology of three measurement points described in reference [9] is used. For each case, the limit for convergence is the point where two impedances are equal, in accordance with basic theory exposed in Fig. 3. Although the three cases have different $Z_{load}$ at initial point, the power consumed in this point is the same because the load is considered constant power model.

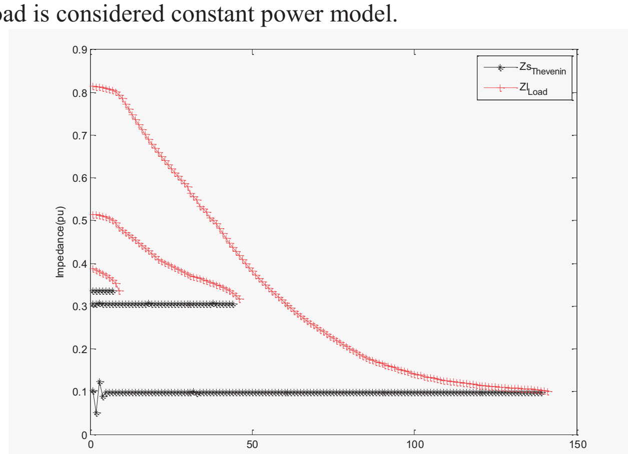


Fig. 5. High voltage solutions of CPF for three cases of $Z_{s\text{-thevenin}}$ variation.

Fig. 6 exposes profiles of voltage, active and reactive power for 49 general cases obtained from appliance of chart flow indicated in Fig. 4.

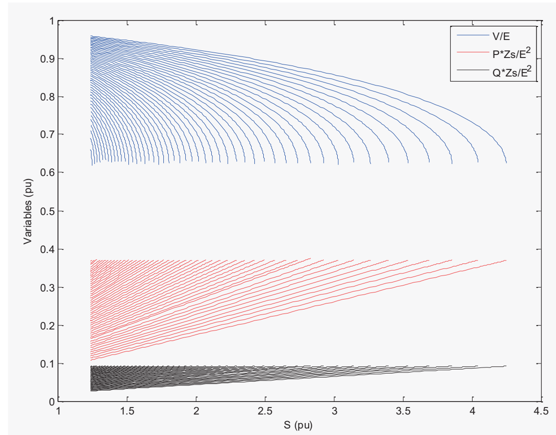


Fig. 6. Voltage, active and reactive profiles for high voltage solutions of CPF

The application of PCA to CPF is analyzed by means of $\mathbf{E}_q$ and **F** matrix which correspond to COEFFS and SCORE matrix, respectively.

## *A. PCA for Data Structure: V- φV- I - φI – P - Q*

In this case the data matrix Xi correspond in structure as indicated in (7).

$$Xi = \begin{array}{c} \text{CPF Step Solution} \downarrow \\ 1 \\ 2 \\ \vdots \\ n \end{array} \overset{\text{Variables} \rightarrow}{\begin{bmatrix} V & \varphi V & I & \varphi I & P & Q \\ & & & & & \\ & & & & & \end{bmatrix}} \quad (7)$$

There will be 49 matrix **Xi**, the application of PCA to original data and normalized data, generates Fig. 7 and Fig. 8, respectively.

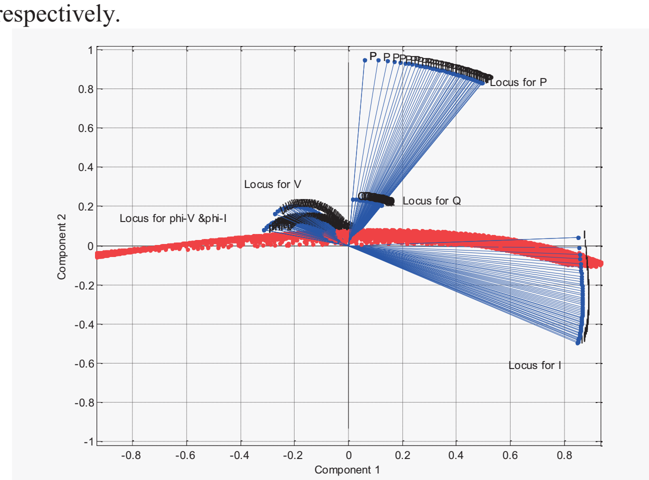


Fig. 7. COEFFS and SCORES for Data Structure: V-φV- I-φI- P-Q
Original Per Unit Data

A revision of the PCA results reveals that all variance of the data can be explained with two principal components.

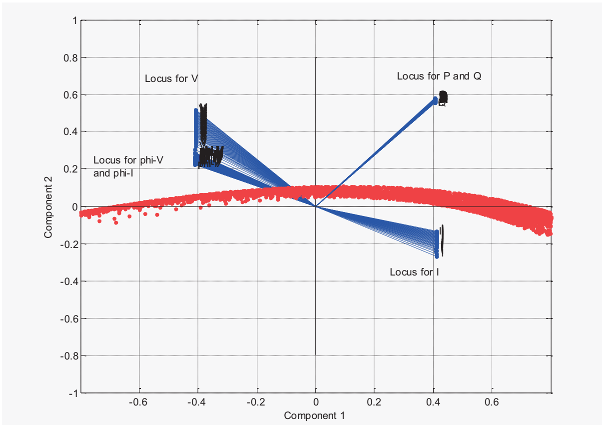



Fig. 8. COEFFS and SCORES for Data Structure: V-φV- I-φI- P-Q Normalized Data

From the previous figures it is possible to reduce the data matrix without considering variables φV and φI, which are in phase, and could be included into variable V.

On the other hand, the variables P and Q are in phase because the load model considered constant power factor. Additionally, the data normalization process hides the specific weight of these variables when they are processed together, in consequence data normalizing will be used only when P and Q are treated separately.

## *B. PCA for Data Structure: VIPQ*

In this case the data matrix has the structure indicated in (8)

$$\text{CPF Step Solution} \downarrow \quad Xi = \begin{array}{c} 1 \\ 2 \\ . \\ . \\ n \end{array} \overset{\text{Variables} \rightarrow}{\begin{bmatrix} V & I & P & Q \\ & & & \\ & & & \end{bmatrix}} \qquad (8)$$

The application of PCA for every case generates Fig. 9. In similar manner as the last case, all variance of the data can be explained with two principal components.

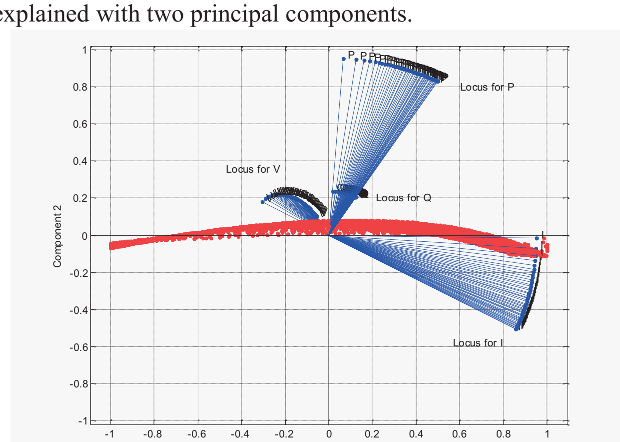



Fig. 9. COEFFS and SCORES for Data Structure: VIPQ

Variables with major positive correlation for first and second principal components are I and P, respectively.

Important features were found: P is in quadrature with variables I and V, and V has opposite correlation with I for all components.

## *C. PCA for Data Structure VI. Subcases: VIP and VIQ*

In this case the data matrix has the structure indicated in (9)

$$\text{CPF Step Solution} \downarrow \quad Xi = \begin{array}{c} 1 \\ 2 \\ . \\ . \\ n \end{array} \overset{\text{Variables} \rightarrow}{\begin{bmatrix} V & I & P \text{ or } Q \\ & & \\ & & \end{bmatrix}} \qquad (9)$$

The application of PCA to structure VIP and VIQ generates Fig. 10 and Fig. 11, respectively.

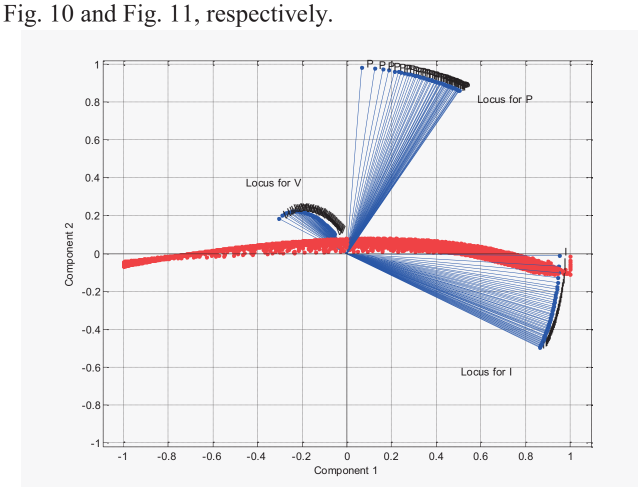



Fig. 10. COEFFS and SCORES for Data Structure: VIP

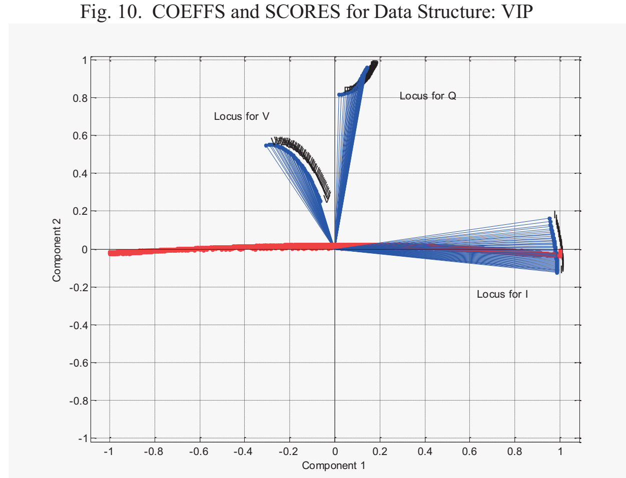



Fig. 11. COEFFS and SCORES for Data Structure: VIQ

Again, a revision of the PCA results reveals that all variance of the data can be explained with two principal components.

### *D. PCA for Data Structure PQ. Subcases: PQV and PQI*

In this case the data matrix has the structure as indicated in (10)

$$\text{Variables} \rightarrow \quad V \text{ or } I \quad P \quad Q$$

$$Xi = \begin{matrix} \text{CPF Step Solution} \\ 1 \\ 2 \\ . \\ . \\ n \end{matrix} \begin{bmatrix} & & \\ & & \\ & & \\ & & \end{bmatrix} \qquad (10)$$

The application of PCA to structure PQV and PQI generates Fig. 12 and Fig. 13, respectively. A revision of the PCA results reveals that all variance of the data can be explained with two principal components.

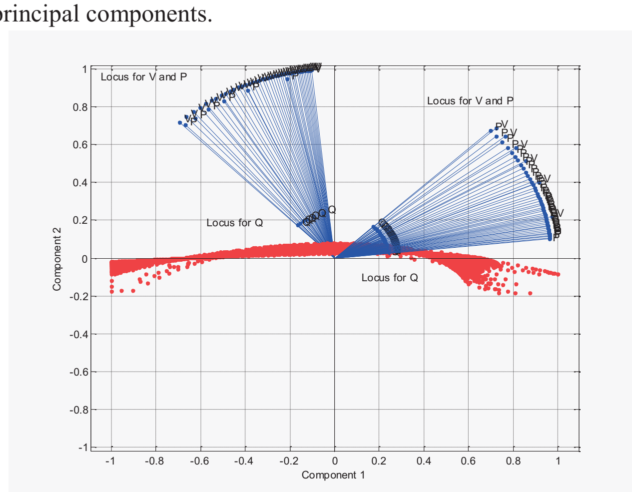



Fig. 12. COEFFS and SCORES for Data Structure: PQV

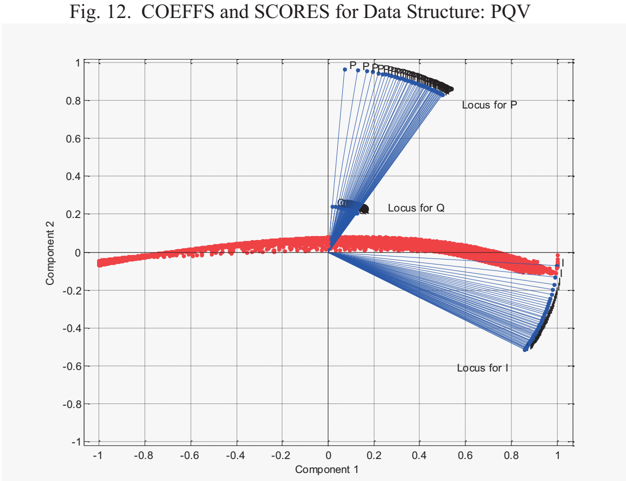



Fig. 13. COEFFS and SCORES for Data Structure: PQI

There is an important feature for PQV structure, in some point of impedance increase, the three variables change the directions of their principal components. This property can be used to develop groups and will be treated in the next section.

### *E. PCA for Data Structure: V. Subcases VP and VQ*

In this case the data matrix has the structure indicated in (11)

$$\text{Variables} \rightarrow \quad V \quad P \quad \text{or } Q$$

$$Xi = \begin{matrix} \text{CPF Step Solution} \\ 1 \\ 2 \\ . \\ . \\ n \end{matrix} \begin{bmatrix} & & \\ & & \\ & & \\ & & \end{bmatrix} \qquad (11)$$

The application of PCA to structure VP and VQ generates Fig. 14 and Fig. 15, respectively.

According to the case D - subcase PQV, this process considered normalized data.

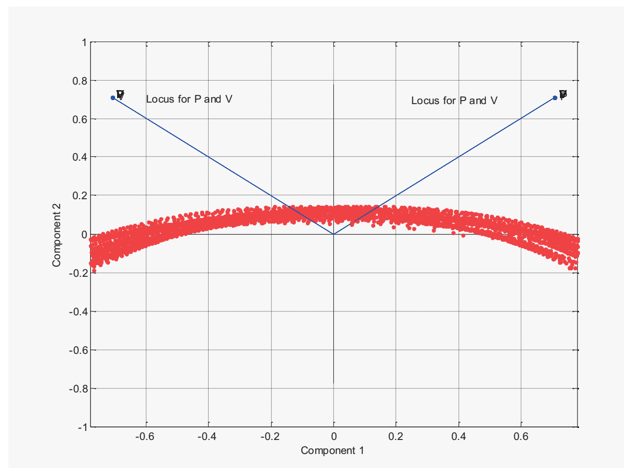



Fig. 14. COEFFS and SCORES for Data Structure: VP

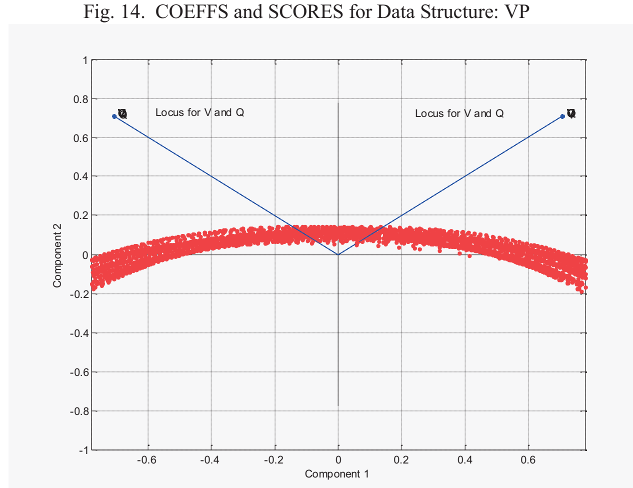
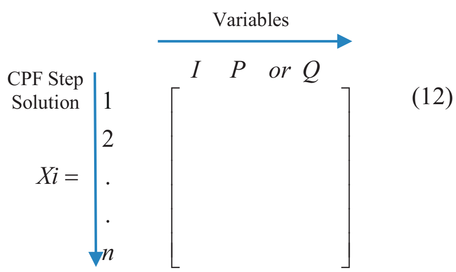



Fig. 15. COEFFS and SCORES for Data Structure: VQ

The form of these graphs are similar as the last case, i.e. the three variables change the directions of their principal components in some point of loading factor.

### *F. PCA for Data Structure: I. subcases IP and IQ*

In this case the data matrix has the structure indicated in (12)




The application of PCA to structure IP and IQ generates Fig. 16 and Fig. 17, respectively. These figures are in accordance to the case D - subcase PQI.

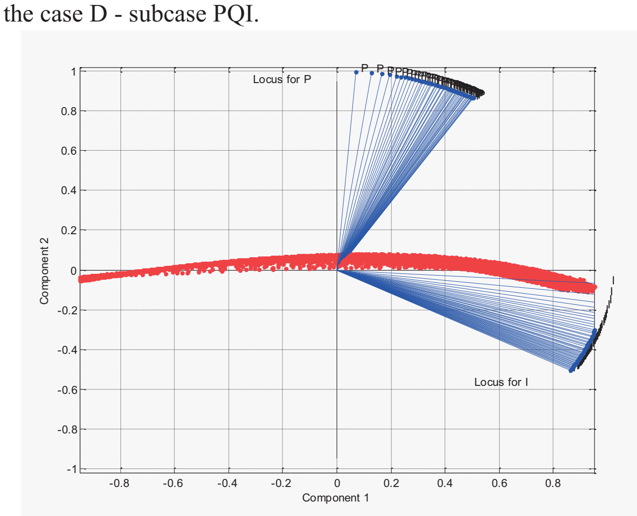



Fig. 16. COEFFS and SCORES for Data Structure: IP

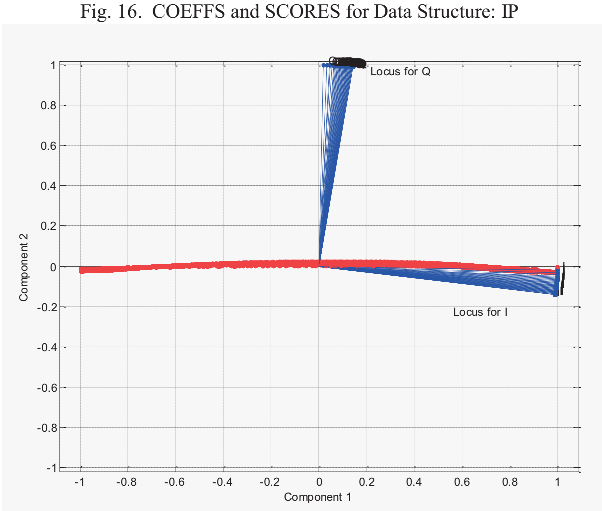



Fig. 17. COEFFS and SCORES for Data Structure: IQ

Variables with major positive correlation for first principal components are I. Additionally, P and Q are in quadrature with variable I.

## IV. Discussion and Possible Clusters

PCA application to high voltage solutions of CPF has been made for some data structure. The cases D and E have an interesting feature related with changes of directions for principal components, in some point of impedance ratio $X_S/X_l$.

A detailed review of the case where this situation occur, determine that the ratio for $X_S/X_l$ at initial point is approximate 0.588, this value is in accordance to (3).

The results interpretation implies that for $Z_S/X_l$ ratio lower than 0.58 the variable of greatest variance is P. On the other hand, when $X_S/X_l$ ratio is greater than 0.58, the variable of greatest variance is V. The last condition occurs due to the power system becomes weak by an increase of $X_S$.

This feature is proposed in order to group PV curves by use of k-means clustering. For this objective, the analysis of the first principal component for variables V and P, will be grouped to determine when the system becomes weak by an increase of $X_S/X_l$ ratio greater than 0.58.

Fig. 18 indicates the result of cluster for data matrix **X** considering PQV structure. Four groups are proposed, groups 2 and 3 correspond with the $X_S/X_l$ ratio greater than 0.58 for variables P and V, respectively. Groups 1 and 4 correspond when the $X_S/X_l$ ratio is lower than 0.58 for variables V and P, respectively.

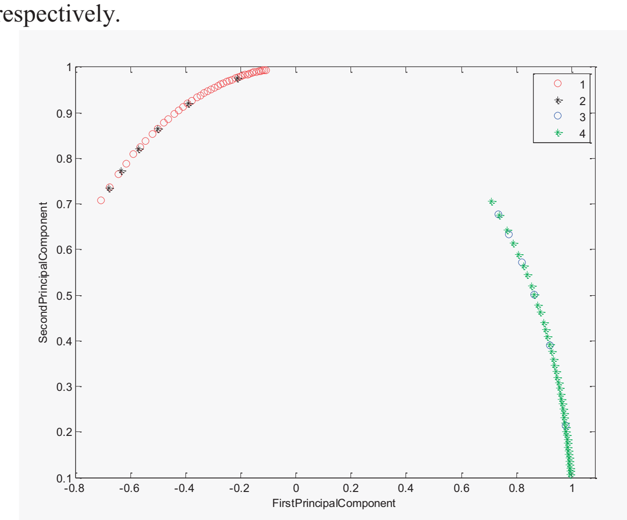



Fig. 18. Cluster for PQV Structure

Also the relationship found, in complement with data structure, could be used as a classifier criteria to determine the weakness degree of the power system for new PV curves. This task can be executed using discriminant analysis for example.

## V. Conclusions

For the basic power system model analyzed in this document, PCA applied to CPF gives some interesting relations between variables. The principal relation observed correspond to PCA application to PQV structure, in this case a feature related with directions changes of principal components in some point of impedances ratio $X_S/X_l$ was found.

For the basic model analyzed, the direction of principal components changes when $Z_S/X_l$ ratio is lower than 0.58 which correspond with inflection point for Voltage Factor. Before this point, the variable of greatest variance is Active Power, but when the system becomes weak by an increase of $X_S$ ($X_S/X_l$ ratio greater than 0.58), the variable of greatest variance is the Voltage.

The analysis of the first principal components for variables Voltage and Active Power allows to group PV-curves profile by use of k-means clustering. For this objective, the components will be grouped for determine when the system becomes weak by an increase of $X_S/X_l$ ratio greater than 0.58.

Although the variables: Voltage and Active Power, have no linear relation, PCA gives interesting and valid results for the basic model analyzed. This analysis can be expanded considering a higher power system model, where the relations between variables is more complex, especially when the system is near the stability limit.

## References


[1] J. C. Cepeda, G. S. Member, D. G. Colomé, and N. J. Castrillón, "Dynamic Vulnerability Assessment due to Transient Instability based on Data Mining Analysis for Smart Grid Applications," *Assessment*, pp. 1–7, 2011.

[2] X. Liu, D. M. Laverty, R. J. Best, K. Li, S. Member, D. J. Morrow, S. Mcloone, and S. Member, "Principal Component Analysis of Wide-Area Phasor Measurements for Islanding Detection — A Geometric View," vol. 30, no. 2, pp. 976–985, 2015.

[3] S. Das and P. S. N. Rao, "Principal component analysis based compression scheme for power system steady state operational data," *ISGT2011-India*, pp. 95–100, 2011.

[4] R. B. R. Bo and F. L. F. Li, "Power flow studies using principal component analysis," *2008 40th North Am. Power Symp.*, no. 1, pp. 1–6, 2008.

[5] E. Vazquez, I. I. Mijares, O. L. Chacon, and a. Conde, "Transformer differential protection using principal component analysis," *2006 IEEE Power Eng. Soc. Gen. Meet.*, 2006.

[6] P. Kundur, *Power System Stability and Control*, McGraw-Hil. 1994.

[7] I. Erlich, "Computational Intelligence for Power Systems," 2006.

[8] CELEC-EP-Transelectric, "Post Operative Report - Post_Abr-27-2015 - Perfiles- Curvas Entregas SE Policentro," 2015.

[9] S. a. Arefifar and W. Xu, "Online tracking of power system impedance parameters and field experiences," *IEEE Trans. Power Deliv.*, vol. 24, no. 4, pp. 1781–1788, 2009.


## Biography

**Antonio Fonseca** (M'10), was born in Quito, Ecuador in 1979. In March 2003, he received the Electrical Engineering diploma and in July 2009 the degree of MSc. in Electrical Engineering, from "Escuela Politécnica Nacional" (EPN) Quito–Ecuador. At the EPN he has been an invited professor, for under and graduate regular courses in the Electrical Engineering Faculty. At this time, he works at the Substations Maintenance Department in CELEC EP – Transelectric. His main research interests are electrical protections and automation systems, analysis of faulted power systems and voltage stability analysis.