Electric Power Systems Research 136 (2016) 31–42

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

A PCA-based approach for substation clustering for voltage sag studies in the Brazilian new energy context

Jacques Miranda Filho[a,b,*], José Maria de Carvalho Filho[a], Anderson Paulo Paiva[a], Paulo Vitor Grillo de Souza[a], Samuel Tomasin[a]

[a] Federal University of Itajubá UNIFEI – Av. BPS, 1303, bairro Pinheirinho, Caixa Postal 50 Itajubá, MG CEP 37500 903, Brazil
[b] IFES Federal Institute of Espírito Santo – Estrada da Tartaruga, s/n – Muquiçaba, Guarapari, ES CEP 29215-090, Brazil

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form 5 February 2016
Accepted 6 February 2016

Keywords:
Voltage sag
Principal component analysis
Cluster
Regulator

ABSTRACT

Voltage Sags are the most disturbing power quality deviation to the sensitive industrial loads causing production losses and other impacts to the equipment end users and distributors of electricity. A new methodology for estimation of voltage sag patterns and clustering of distribution substations with similar features for voltage sags is the main contribution of this paper. The focus is on the regulation of this phenomenon in Brazil. Network modeling and faults simulations in transmission and distribution levels were performed in a Brazilian actual distribution system with 17 substations in order to get information regarding the number of voltage sags caused in the substations bus bars. Principal Component Analysis was applied on a significant number of variables containing relevant information of power quality and design features of the substations and storing the major Principal Components Scores. Furthermore, Clusters of substations with similar characteristics to voltage sags were formed and assigned a specific Membership to each substation. The expected annual number of voltage sags is the solution. A 95% Confidence Interval for the total number of voltage sags was estimated, leading to the classification of the main variables related to the voltage sags by clusters formed. Clusters of Substations with similarities for regulatory purpose of voltage sags were the principal results, followed by classification of the main variables associated with voltage sags by the clusters formed. A critical analysis of the results and a comparison among different Clustering Methods reaffirmed the assertiveness of the proposed methodology for the management of power quality associated with voltage sags, by the Regulator and the distribution utilities in Brazil.

1. Introduction

The performance of electricity distribution substations is of paramount importance to the distribution utilities in the regulated market of electric power systems. There are several ways to evaluate this performance with respect to voltage sags, or sags, which is the objective of this research. The compatibility between industrial processes requirements and the quality of the electrical supply was studied in [1] taking into account the influence of voltage sags. By applying this approach, new indices for both the supply system and the load buses can be drawn. A comparative analysis of sags among substations in Colombia was studied in [2]. The data acquired from a power quality measurement system is considered, the sags disturbances are classified and quantified according to a proposed level of severity, and an index called Sags Activity index-SAI is proposed allowing to weigh and rank the severity of sags among the affected substations. A new voltage sag classification procedure based on multi-way principal component analysis (MPCA) was studied in [3]. Sampled voltage and current waveforms of previously registered sags are used together with the MPCA technique to obtain a lower dimensional model. This model is then used to project new Sag events and classify them according to their origin in the power system, adding great interest for utilities. The performance of power lines was studied in [4]. The voltage sag indices as non delivered energy, the load variation, and the total cost due to voltage sags were suggested to measure power lines performances. The proposed power line performance can change the traditional performance based on power line outages. The principal component analysis and K-means methods have an extensive range of applications. The usefulness is in reducing the vectors of

* Corresponding author at: Federal University of Itajubá UNIFEI – Av. BPS, 1303, bairro Pinheirinho, Caixa Postal 50 Itajubá, MG CEP 37500 903, Brazil. Tel.: +55 27 999397702.
E-mail address: Jacques.filho@uol.com.br (J.M. Filho).

http://dx.doi.org/10.1016/j.epsr.2016.02.012
0378-7796/

dimensionality of inputs or outputs in certain applications. The following are some examples of applications in electrical engineering. Barrera et al. [5] presented the results of a power quality monitoring work in Bogotá, Colombia, where a total of 249 substations have been observed for two years from 2007 to 2009. Quantitative analysis of measurement results use the Probability Density Function (PDF) and Principal Components Analysis (PCA), which identified the substations voltage dips on the upper and lower limits of the compatibility viewpoint. Jazebi and Vahidi [6] presented aspects of the reconfiguration of distribution networks to solve power quality problems such as voltage harmonics, sags, and mitigation of technical losses. Biswas et al. [7] presented a new formulation for the optimal location of distributed generation, in a combination of technical factors such as minimizing losses and reduction of voltage sags, and economic issues, such as installation costs and maintenance of distributed generation. Romero et al. [8] presented a new methodology for estimation of voltage sags patterns and clustering of fault zones in high and medium voltage. Network modeling and faults simulation were performed in order to get information about voltage sags caused by faults in the transmission system. Voltage sags patterns were identified by means of a K-means clustering algorithm, allowing the determination of fault zones. Using the power quality measurements data base of the major electricity utility of Bogotá, voltage sags were classified according to the previously determined patterns. Hariyanto et al. [9] presented a new approach to identify bus voltage severity profile due to short circuit fault at a certain point in a distribution power system. Data containing 2 variables, depth and duration of voltage sag due to short circuit faults on every buses, are generated and subsequently clustered using K-means Clustering. Clustering data will produce center clusters and cluster membership. To be able to perceive voltage sag severity, center clusters will be converted to Event Severity Index which explains the severity of a voltage sag event based on CBEMA-ITI Curve. Balouji et al. [10] developed a method in which huge amount of Power Quality event data, voltage sags, swells and interruptions, were classified into finite number of classes and spatial classification of those clusters provides characterization of specific parts of the grid. The method is based on k-means clustering of the feature space, voltage rms values. The PCA method has been applied on feature space before K-means clustering. PCA + K-means algorithm has given the best clustering in terms of PQ event characterization. Guder et al. [11] presented a power quality (PQ) knowledge discovery and modeling framework that has been developed for both temporal and spatial PQ event data collected from transformer substations supplying iron and steel (I&S) plants. PQ event characteristics of various I&S plants have been obtained based on clustering and rule discovery techniques. The data are collected by the PQ analyzers, which detect the voltage sags, swells, and interruptions according to the IEC Standard 61000-4-30. The constructed clustering strategy ensures the feasible system monitoring by reducing unmanageable number of PQ events collected by the distributed PQ measurement systems into event clusters count. Duan et al. [12] presented an approach of K-means clustering analysis algorithm to classify and recognize the voltage sag from the measured historical data of large-scale grid in Shenzhen, China. The distances among different sag incidents in distribution diagram are calculated, and when some of them are nearer each other, a cluster center which is called centroid can be set to represent these incidents. The sag amplitude and duration time reflected by these centroids can be regarded as the voltage sag characteristics of similar substations, which will represent the operation condition and find out the weak link of whole power systems. Morales et al. [13] presented an assessment between Principal Component Analysis (PCA) and Wavelet Transform (WT) signal processing techniques applied for Transmission Lines lightning stroke classification. The patterns extraction was developed through Principal Component Analysis and the Wavelet Transform. The pattern classification is developed using Artificial Neural Network (ANN), k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM). The work presents an assessment of lightning stroke classification, providing useful information, especially in extraction and selection of mother functions and the use of PCA. Results show that by using PCA, optimal mother functions can be extracted, presenting a new alternative for relaying protection. In Brazil, Power Quality is regulated by ANEEL [14], which defines quality indicators and standards in order to provide mechanisms to control and monitor the performance of distributors regarding the quality of electricity. Power quality is regulated in the Electricity Distribution Procedures which state that, "The indicator to be used for performance of the knowledge of a particular bus distribution system with respect to Short Duration Voltage Variation – SDVVs is the number of grouped events per track, amplitude and duration, discretized as criteria established from survey measurements." As for the values, "They are not assigned performance standards to these phenomena; distributors must monitor and make available, on an annual basis, the performance of the bus bars monitored" [14]. In recent years, the quality of electricity in Brazil became an even more important subject, due to the growth of sensitive loads to voltage sags and distributed generation, prompting frequent complaints of the industrial sector to the regulatory agency, and the consequent proposed revision of the regulatory standards [14], with new indicators, limits and penalties for its violation by distribution utilities. The electricity distribution system in Brazil is operated by 67 power distribution companies with 3500 distribution substations. This amount of companies and substations require the development and application of a methodology that forms clusters of similar substations for the regulatory purposes of voltage sags. Therefore, the key contribution of this paper is the description of a new Methodology for substations aggregation into clusters for performance analysis in terms of voltage sags for regulatory purposes with the use of Ward's Hierarchical Clustering Method and Scores of Principal Components. The solution is the expected number of voltage sag events per year. The methodology is applied in a Case Study based on a set of variables, and carried out during a Research and Development project in Brazil. The literature survey, briefly described here, shows that an extensive amount of research has been carried out using Principal Component Analysis (PCA) as dimensionality reduction method and feature extraction, and that the methodology proposed is an innovative application.

2. Theory

2.1. Hierarchical clustering methods

Hierarchical clustering techniques have origin on a series of successive mergers or divisions. *Agglomerative hierarchical methods* start with the individual objects, initially building as many clusters as objects. The objects with greater similarity are first grouped, and merged in accordance with their similarities. With the decreasing of the similarity, the subgroups form a single cluster. Linkage methods are good for clustering items and variables: *Single linkage* considers the minimum distance or nearest neighbor; *average linkage* considers the average distance, *complete linkage* considers the maximum distance or farthest neighbor [15]. Multivariate analysis solves problems with many variables, so computational tool is required. Some available clustering software: SAS, STATISTICA, SPSS, and Minitab 17 Statistical Software, used in this project.

2.2. Ward's hierarchical clustering method

Several hierarchical methods are used to form clusters, among them: simple connection, average of the distances, full connection