9th International Conference.
Electrical Power Quality and Utilisation.
Barcelona, 9-11 October 2007

A Hybrid Method for Sag Source Location in Power Network

Abbas Khosravi, Joaquim Melendez, Joan Colomer
Dept. of Elec., Comp. Science and Automatic Control
University of Girona, Girona, Spain
Email: {khosravi,quimmel,colomer}@eia.udg.es

Jorge Sanchez
ENDESA Distribucion
Barcelona, Spain
Email: jsanchezc@endesa.es

Abstract— **The work presented in this paper belongs to the power quality knowledge area and deals with the voltage sags in power transmission and distribution systems. Propagating throughout the power network, voltage sags can cause plenty of problems for domestic and industrial loads that can financially cost a lot. To impose penalties to responsible party and to improve monitoring and mitigation strategies, sags must be located in the power network. With such a worthwhile objective, this paper comes up with a new method for associating a sag waveform with its origin in transmission and distribution networks. It solves this problem through developing hybrid methods which hire Multiway Principal Component Analysis (MPCA) as a dimension reduction tool. MPCA reexpresses sag waveforms in a new subspace just in a few scores. We train some well-known classifiers with these scores and exploit them for classification of future sags. The capabilities of the proposed method for dimension reduction and classification are examined using the real data gathered from three substations in Catalonia, Spain. The obtained classification rates certify the goodness and powerfulness of the developed hybrid methods as brand-new tools for sag classification.**

I. Introduction

Nowadays, the electricity dependence of industries, commerce and services has provoked the regulation of power quality. The objective is to reduce damages or misbehavior to consumers' devices and processes. From generators to customers, voltage waveforms may suffer alterations that negatively affect the power quality. The most important phenomena that affect voltage quality are short interruptions, voltage dips (or sags), flicker, supply voltage variations and harmonic distortion. Among them, voltage sags are the most significant due to their severity and number of occurrences per year. The definition of a voltage sag according to the IEEE standards is a momentary decrease (10%-90%) in the Root Mean Square (RMS) voltage magnitude where the duration is longer than a half cycle and less than one minute [1]. Common causes of voltage sags are storms, the start-up of large loads at neighboring facilities, and grounding or wiring problems. Either originated in transmission or distribution network, sags easily propagate up and down areas of geographical occurrence point. They can cause interruptions to sensitive end user equipment and industrial devices such as adjustable-speed drives, relays, and robots. Voltage sags cause more than 80% of the industrial customer complaints today, which signifies economic losses averaging $10,000 US per event. More often than not, industrial practitioners and customers put the blame for those financial losses on utilities involved in the distribution network. On the contrary, the investigation into causes and location of voltage sags don't apportion blame for all cases to electric utilities. Many times these are domestic and industrial loads initiating sags in the power network. Through sag source locating, any dispute about the major responsible party can be resolved fairly. Therefore penalties may be imposed to the responsible party for generating a disturbance resulting in customer downtime, as is done for harmonic pollution in some countries [2]. Besides, sag source location is necessary for power quality troubleshooting, diagnosis, and mitigation strategy development. Also, huge amount of data collected by power quality monitors has need of efficient paradigms for quick and reliable analysis. Taking into account all these explanations, automatic analysis and source detection of power events especially voltage sags have become an essential requirement for power quality monitoring.

Apart from the demand of industry for more research on this subject, classification of voltage disturbance is a thought-provoking problem for academicians. Tackling this problem from different points of view, researchers have developed diverse techniques. In [3], authors are interested to the voltage sag origins and categorize them in three classes using some features fed to a fuzzy system. A new method to locate the source of voltage sag in a power distribution system using polarity of the real current component relative to the monitoring point has been introduced in [4]. A comprehensive and profound review of voltage sags as well as their stochastic assessment has been given in [1]. A phasorial analysis and an unsupervised method have been compared in [5]. In [6] decision about location of occurred disturbance is made based on changes in two indices named disturbance power and disturbance energy. Another possible way of detecting voltage sag source is to use the seen impedance and its angle before and after sag occurrence [7]. Extracting some temporal descriptors such as duration and depth of sags as well as their fall and recovery slops have been used in with a Learning Algorithm for Multivariate Data Analysis (LAMDA) in [5] [8]. Other attributes used with that goal involve phasorial analysis to obtain the initial phase angle shift or the phase angle difference between current and voltage or the power factor angle [4] [9]. Implementing different methods for sag source location and obtained results in [10] clearly show that the available methods are not totally reliable and more research is needed to extend the location methods.

Despite all the research done so far, automatic classification of sags and other events in power systems has not been widely treated and correct classification rates for the actual events are not as high as classification results used in areas such as pattern recognition, speech recognition, and so on.

This paper deals with the problem of locating voltage sags in the power networks in a completely different fashion. Being prompted by the promising applications of MPCA in other areas of engineering, this paper hires it for developing a new algorithm for sag source location in the power network. Capitalizing on the dimension reduction capability of MPCA, here we come up with some hybrid approaches for locating sags in transmission and distribution networks. The capabilities of the proposed method for dimension reduction and detecting sag sources in the power network are examined and verified using the real data gathered from some substations in Catalonia, Spain. The results presented in some rates certify the goodness and powerfulness of the developed hybrid approaches as a brand new tool in power quality domain for locating sags.

The remaining part of this paper is structured as follows. First in section 2, a brief introduction of MPCA is given. Section 3 comes up with the details of the proposed method. Numerical results then are presented in section 4 accompanied by some discussions. Finally, we finish paper with some conclusions and statements for future work.

II. MPCA

Principal Component Analysis (PCA) is a multivariate statistical technique that projects data onto linear subspaces that are the most descriptive of variance in a data set. Linear combinations of the measured variables make up new latent variables describing the direction of subspace. In PCA, these latent variables are referred to as principal components. After the major events or states have been accounted for by the dominant principal components, the remaining components are interpreted as being the result of random errors or noise in the data. These components are grouped together in an error term. As such, PCA acts as a filtering technique [11].

The general form of PCA relates the subspaces to the original scaled data set:

$$X = \sum_{i=1}^{k} t_i p_i^T \tag{1}$$

Here t_i are known as score vectors and p_i are the loading vectors/principal components (i.e. eigenvectors of covariance matrix of X). The scores describe where the data points are projected in the PCA subspace, while the loading vectors describe the linear combination of the original variables that constitute each principal component. Put in other words, they determine that how much each variable contributes to construction of each principal component. Once the eigenvalues, λ_i, and the eigenvectors, p_i, have been determined, the scores, t_i, can be easily calculated from:

$$t_i = X p_i. \tag{2}$$

PCA can be used for statistical control application by applying two different statistics, Hotelling's T^2 statistic and Q-statistic ($Q_{residual}$). These statistics can be used to construct multivariate control charts with confidence limits for the entire PCA model.

One stipulation of PCA is that it can only be used for modeling two-dimensional data. MPCA is an extension of PCA that has been developed for the purpose of being able to apply PCA to data sets consisting of three dimensional arrays, as is typically the case in applications such as image processing and batch process monitoring.

Algorithmically, MPCA is consistent with PCA and has the same goals and benefits. The key difference is a pre-processing step known as unfolding required to convert a three dimensional data matrix to a large two dimensional matrix. Unfolding simply involves a rearrangement of the data. Rearrangement can be done in sex possible ways that, only three of which are mathematically unique: time-wise, batch-wise, and variable wise. Let's consider $X_{I \times J \times K}$ as the original three dimensional data set where I, J, and K stand for number of batches/experiments, variables, and sample times. One type of batch-wise unfolding has been presented in Fig. 1. It puts all observations of each variable for all batches in one block.

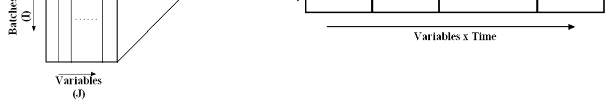

Fig. 1. Batch-wise unfolding of the three dimensional matrix leading to a two dimensional one ($I \times (K \times J)$)

The decision on how to unfold the data depends on the particular objectives of the model or the analysis. In practice, the time-wise and the batch wise options are most common.

Once a data set has been rearranged, a PCA model is formulated in the usual fashion. Loading plots, score plots, Hotelling T^2, and Q-statistic can be calculated for MPCA models in the same manner as for regular PCA models.

In a nutshell, PCA and its extension MPCA:

- handle large number of correlated variables well,
- take advantage of structure in data,
- provide data compression, and
- provide explanation of which variables (in the original space) are responsible of an out-of control- situation (bad fitting the statistical model) [12].

Taking advantage of these properties, in the next section we develop a procedure for sag classification based on MPCA.

In literature there are countless sources for and applications of PCA and MPCA in different areas of science and engineering. We just refer the interested readers for further information to [13] [14] [15] [16] and references therein.

III. Hybrid Sag Source Location

As we explained in the previous section, PCA and MPCA are powerful tools for reducing dimensionality of huge data sets through projecting and reexpressing samples upon some principal components. In this research we use these new representations (scores) as features which have been extracted from data by MPCA and then put them into some classifiers to solve the problem of sag source location. Put in other words, sag source location in the power network will be equivalent with classifying sags in High Voltage (HV) and Medium Voltage (MV) classes.

We first compute RMS values of currents and voltages for each voltage sag and then arrange them in a three dimensional matrix like that one shown in Fig. 1. In this figure, we replace batches with sags and put sex variables (three voltages and three currents) along the horizontal axis. The third dimension is unchanged. Therefore, dimensionality of the final HV and MV matrices will be $numberofsags \times 6 \times time$.

Since RMS magnitudes of voltages and currents are completely different, scaling is a preprocessing stage needed to avoid overweighing of voltage variables towards currents. Without scaling, voltages appear dominant in the analysis due solely to their relative magnitude with respect to currents. Among different scaling methods introduced in [14], we apply autoscaling which results in zero-mean centered data with a unit variance.

Fig. 2 shows different stages of the proposed method after database construction. First we create a HV/MV MPCA model using the HV/MV prepared databases. The same procedure shown in Fig. 1 is followed for unfolding the three dimensional matrices when creating models. In the models created using these unfolded matrices, we can easily discuss how each variable contributes to the created models based on loading plots.

Through model creation, we transform the RMS waveforms of voltages and currents to a point in an k dimensional space, where k is the number of principal component retained in the models. These k scores are the features/attributes that we feed them to the well-known nonlinear classifiers. In this manner, a classifier can be trained easily and then exploited for solving the problem of sag source location in the power network. After projecting HV and MV sags to the built MPCA models, we pass computed scores to some considered classifiers and train them using these data. Here we can employ a wide variety of different nonlinear classifiers such as Neural Network (NN), Support Vector Machines (SVMs), Decision Trees (DT), etc.

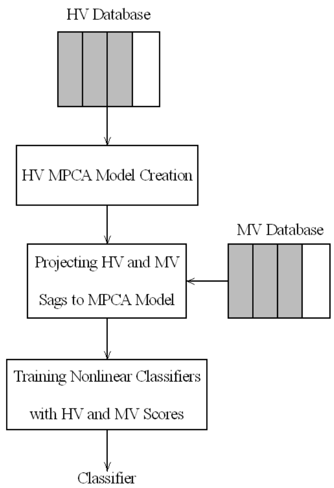

Fig. 2. Different stages in the proposed hybrid method

After training stage which is completed off-line, the exploitation of the developed method for classification of sags in online applications is pretty straightforward and simple. It's done just in two stage as shown in Fig. 3. Projecting sags to MPCA models and then putting score into the trained classifier is the whole procedure that we should follow to solve the problem of sag source location in the power network.

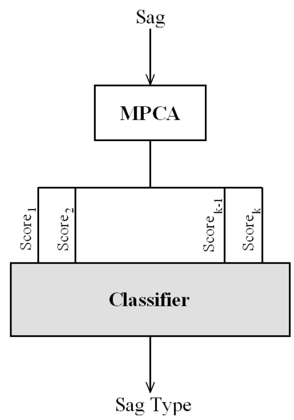

Fig. 3. Online Exploitation

Benefits of MPCA with respect to other classification strategies reside in its generality and the fact that no additional parameters are needed to be adjusted. The previous stages could be accomplished very quickly and the computation mass is roughly low which is to some extent important in online applications. Another noticeable aspect of the proposed method is its robustness. No matter how sever noise corrupts some data, it captures and considers its effects as meaningless variation.

It is interesting to consider this point that in the developed method we deal with current and voltage waveforms as just some signals without taking care of from which domain they come. In contrast with developed methods in literature that often try to locate sags in the network based on extracting and computing some electrical concepts and measures from waveforms such as apparent impedance, we do this more simply and easily without resorting to those things.

It's noteworthy mentioning that applying MPCA as a dimension reduction tool in the developed hybrid approaches will lead to great results if the data are well separable based on scores. And as we will see later in the next chapter when working with real data, this assumption is almost completely true.

IV. Experiment with Real Data

The proposed method has been tested using sets of real data registered in a couple of 110/25 KV substations of the Spanish distribution network in the course of one year. The databases consist of time histories of three phase voltages and three currents. According to the registered information, HV and MV sags have been occurred at different hours, days and seasons during the year. Admittedly, they properly represent various types of sags (balanced and unbalanced) and completely reflect the network behavior after their occurrence as well. Why variety of sags is so much important stems from the fact that the proposed methods all are data-driven. Therefore, the more diverse sags are, the more powerful method is for sag classification. Fortunately, the used databases all have this characteristic.

Data from power quality monitors are downloaded to computers. In some MS-Excel files all waveforms have been recorded for some periods before and after sags occurrence giving us possibility to analyze them with different tools. Monitoring softwares in substations measure voltage sag magnitude and duration as well as other temporal features for each voltage phase and save all important instants in those MS-Excel files.

Waveforms have been sampled at 6400 Hz resulting 128 samples per period (20 ms). Each register contains 39 periods with 4993 samples for each variable. RMS values for three voltages and three currents are calculated for HV and MV data sets using a one period sliding window Fast Fourier Transform (FFT) to estimate the magnitude of nominal frequency (50 Hz) of the network during sag occurrence. Applying FFT to voltage and current waveforms before RMS computation is necessary for filtering those harmonics which have been added to measures.

TABLE I
Overall classification rates using hybrid approaches fed by scores

	Classifier		
Rates (%) for	MLP	RBF	DT
1st Fold	83.6	96.4	87.3
2nd Fold	87.3	90.9	85.4
3rd Fold	94.5	92.7	87.3
4th Fold	85.4	87.3	89.1
Average Rates	87.7	91.8	87.3

There is also another tiny point about working with real databases. As we mentioned earlier, MPCA models are data-driven and validity of those is strongly dependent on the selection of data used. From the other side, recording devices in the measuring points record any occurred event in the network in the databases. So, when working with real databases, a precise inspection of data for removing outliers is vital for success of the proposed method. Here outlier terminology refers to as any event else sags. e.g. interruptions, not recovered sags, swells, etc. This could be simply done through visual inspection or through developing some simple algorithms for detecting and removing them. Doing this stage is equivalent to getting sure that in the MPCA analysis of a chemical process, all available batches belong to the same process.

To investigate the performance of the method, four fold cross validation technique is introduced through splitting HV and MV data sets into training and test subsets. In each fold, we consider 75% of all sags for MPCA model creation and the rest, 25%, as not-yet-seen sags for model validation (each sag is used three times for model creation and once for projection). To make sure that sags in both training and test subsets are completely unrelated to each other, we first change the order of them in the data sets and organize them by chance.

We employ three well-known classifiers for equipping MPCA with nonlinear data mining tools. They are:

- Multilayer Perceptron (MLP),
- Radial Basic Function (RBF) network,
- Decision tree (DT).

These classifiers have been well investigated in literature and their implementation could be found in [17]. The type of DT used here in this research is J48 which is a specific version of well-known C4.5 decision tree [18].

Table I presents overall (HV and MV) results for different classifiers when fed by 9 scores. Average rates for different classifiers are more or less the same, although RBF network performs a bit better than other classifiers. We see the obtained rates regardless of the classifier type are satisfactory high. This means that features extracted by MPCA (scores) are informatively rich. This reexpression of sags just in a few coordinates which is quite useable for sag source location in the power network highlights powerfulness of MPCA for dimension reduction with minimum loss of information.

In another experiment we normalize scores by corresponding eigenvalues. So, the i-th input of classifiers, x_i, will be as follows:

TABLE II

OVERALL CLASSIFICATION RATES USING HYBRID APPROACHES FED BY NORMALIZED SCORES BY CORRESPONDING EIGENVALUES

	Classifier		
Rates (%) for	MLP	RBF	DT
1st Fold	81.8	94.5	92.7
2nd Fold	87.3	90.9	80.0
3rd Fold	96.4	98.2	94.5
4th Fold	85.4	87.3	90.9
Average Rates	87.7	92.7	89.5

$$x_i = \frac{t_i}{\lambda_i} \quad i = 1, \ldots, 9 \tag{3}$$

Our motive for doing this is to avoid overweighing some big scores towards small ones. This simple mathematical computation is done very quickly, hence it is not an disadvantage for this method when compared with the previous one. Besides, since eigenvalues are all positive, score signs which can carry much useful information for classification are not killed. Classification rates after applying this to scores and feeding them to three considered classifiers have been presented in Table II. In comparison with Table I, the overall rates for RBF and DT classifier have been improved which means that normalizing scores by corresponding eigenvalues is a proper movement for enhancing results.

Besides the presented results here, we have conducted lots of other experiments with these data. Taking into account the results presented here and also the results of other experiments, we come to this conclusion that RBF network when fed by scores normalized by eigenvalues is the best classifier among examined ones. Normalizing scores by eigenvalues makes sags more separable and positively contributes to the rate enhancement.

Furthermore, these are only HV MPCA models which are suitable for and useable in hybrid approaches. MV MPCA models are not much suitable in this specific application at all, since obtained scores using these models are not well separable. The maximum average classification rate using those classifiers comes around 75% which is far below the rates presented in Table I and Table II. This is the reason why all MPCA models in Fig. 2 and Fig. 3 have been labeled as HV models.

V. CONCLUSION AND FUTURE WORK

In this paper, we dealt with the problem of sag source location in the power network in an innovative manner. First, we interpret sag source location in transmission and distribution systems as a classification task. Then, we hired well-known and widely used MPCA technique for developing a hybrid method for classifying sags in two HV and MV classes. In this approach, MPCA is used as a powerful dimension reduction algorithm guaranteeing minimum loss of information. Scores of projected sags to MPCA models were used for training different nonlinear classifiers (MLP, RBF network, and DT). Created MPCA model and trained classifier constitute a highly powerful hybrid tool for properly solving sag source location problem in the power quality area. Experiments done with sags recorded in some substations in Catalonia, Spain proved the powerfulness of the proposed method. Obtained classification rates for different classifiers are satisfactorily high, although rates obtained from MPCA-RBF based one is a bit better than two other ones.

Despite obtained results so far, there are lots of rooms for improvement. The first direction for further research is to properly determine how many scores we should put into classifiers for getting the highest possible results. A method for selecting the best feature/scores should be done off-line when training the nonlinear classifiers.

The presented hybrid approach here is lineat in the first part (dimension reduction)and nonlinear in the second part (classifier). More attempts for making it completely nonlinear through equipping the first part with nonlinear dimension reduction algorithms can enhance overall rates. Using Kernel PCA (KPCA), we will be able to capture and model nonlinear variation of data much more properly than linear PCA.

Finally, both PCA and MPCA are unsupervised learning algorithms for dimension reduction. When capturing and modeling variation, they don't consider labels of the data. As a consequence, they may project and reexpress samples from different classes similarly. Looking for and applying linear/nonlinear supervised dimension reduction algorithm is worth a try. Since we have already got good results with MPCA, those methods also should be roughly similar to it. For instance, Partial Least Square (PLS) is a method that has both properties.

In future work and researches, we will move in the mentioned directions for improving reliability of the developed methods for sag source location in the power network.

ACKNOWLEDGMENT

The authors gratefully acknowledge the helpful contribution of Mr Esteban Gelso for database construction and Ms R. Naghavizadeh for preparing this paper. This work has been partially supported by the projects Diagnostico de Redes de Distribucin Elctrica Basado en Casos y Modelos -DPI2006-09370- and PTR1995-1020-0P from the research and transfer programs, respectively, of the Spanish Government. Also it has been supported by the project of Development of an intelligent control system applied to a Sequencing Batch Reactor (SBR) for the elimination of organic matter, nitrogen and phosphorus DPI2005-08922-C02-02 supported by the Spanish Government.

REFERENCES

[1] M. Bollen, *Understanding Power Quality Problems: Voltage Sags and Interruptions*. Series on Power Engineering. Piscataway, NJ: IEEE Press, 2000.

[2] E. Belenguer, C. Reineri, N. Aparicio, and R. Mampel, "Voltage sag source location in distribution networks," in *ICREPQ'03*, 2003.

[3] M. Kezunovic and Y. Liao, "A new method for classification and characterization of voltage sags," *J. of Electric Power Systems Research*, vol. 58(1), pp. 27–35, 2001.

[4] N. Hamzah, A. Mohamed, and A.Hussain, “A new approach to locate the voltage sag source using real current component,” *J. of Electric Power Systems Research*, vol. 72, pp. 113–123, 2004.
[5] J. Melndez, J. Mora, D. Llanos, M. Ruiz, J. Colomer, J. Snchez, and X. Corbella, “Classification of short duration faults (voltage sags) in transmission and distribution power systems,” in *ECC03*, 2003.
[6] A. Parsons, W. Grady, E. Powers, and J. Soward, “A direction finder for power quality disturbances based upon disturbance power and energy,” *IEEE Transactions on Power Delivery*, vol. 15(3), pp. 1081–1086, 2000.
[7] A. Pradhan and A. Routray, “Applying distance relay for voltage sag-source detection,” *IEEE Transactions on Power Delivery*, vol. 20(1), pp. 529–531, 2005.
[8] J. Mora, D. Llanos, J. Melendez, J. Colomer, J. Sanchez, and X. Corbella, “Classification of sags measured in a distribution substation based on qualitative and temporal descriptors,” in *17th Int. Conf. On Electricity Distribution, Barcelona, Spain*, 2003.
[9] G. Zheng, X.-M. Yan, and H.-W. L. D. Liu, “Classification of voltage sag based on wavelet transform and wavelet network,” in *Proc. of the Third International Conference on Machine Learning and Cybernetics, August 26-29, Shanghai, China*, 2004.
[10] R. Leborgne, D. Karlsson, and J. Daalder, “Voltage sag source location methods performance under symmetrical and asymmetrical fault conditions,” in *IEEE PES Transmission and Distribution Conference and Exposition Latin America, Venezuela*, 2006.
[11] G. Fery, “Multi resolutional partial least squares and principal component analysis of fluidized bed drying,” Master’s thesis, Department of Chemical Engineering, University of Saskatchewan, Canada, 2005.
[12] B. Wise and N. Gallagher, “The process chemometrics approach to process monitoring and fault detection,” *J. Process Control*, vol. 6(6), pp. 329–348, 1996.
[13] B. Wise, N. Gallgher, S. Butler, D. J. white, and G. Barna, “A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process,” *Chemometrics*, vol. 13, pp. 379–396, 1999.
[14] E. Russell, L. Chiang, and R. Braatz, *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*. Springer-Verlog, London, 2000.
[15] P. Nomikos and J. MacGregor, “Monitoring batch processes using multiway principal component analysis,” *AIChE*, vol. 40(8), pp. 1361–1375, 1994.
[16] T. Kourti, “Process analysis and abnormal situation detection: From theory to practice,” *IEEE Control Systems Magazine*, vol. 22(5), pp. 10–25, 2002.
[17] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Second Edition, Morgan Kaufmann Publishers, 2005.
[18] J. Quinlan, *C4.5: Programs for machine learning*. Morgan Kaufmann, 1993.