Two Methods for Voltage Sag Source Location

V. Barrera Núñez, X. Berjaga Moliner, J. Melendez Frigola, S. Herraiz Jaramillo, J. Sánchez and M. Castro

Abstract—**This paper presents and compares two approaches to estimate the origin (upstream or downstream) of voltage sag registered in distribution substations. The first approach is based on the application of a single rule dealing with features extracted from the impedances during the fault whereas the second method exploit the variability of waveforms from an statistical point of view. Both approaches have been tested with voltage sags registered in distribution substations and advantages, drawbacks and comparative results are presented. .**

***Index Terms*— Fault location, position measurement, power quality monitoring, voltage sag (dip) source detection.**

I. Nomenclature

PCSC: Phase change sequence current algorithm
RS: Resistance sign algorithm
DR: Distance relay algorithm
MANOVA: Multivariate analysis of variance
AD: Discriminant analysis
PCA: Principal component analysis
MPCA: Multiway Principal Component Analysis
CBR: Case Based Reasoning

II. Introduction

Due to the impact on sensitive industrial loads and costs led by the damages and maintenance costs, voltage sags have focused the attention of power quality studies leaded by the utilities. Unfortunately, those disturbances propagate through the power system affecting loads connected in whole network. Hence, the responsibility for the generation of disturbances on the system must be assessed, being the automatic location of their origin one of the most interesting aspects involved. Given a register and two basic steps have to be accomplished: the first one is to isolate the origin upstream or downstream; next an appropriate algorithm has to be used to accurate locate the origin in the network. In this work we tackle the first step under the basis that the waveforms of voltages and currents are available. Recent studies reinforce the importance of developing robust and simple algorithms for the relative (upstream/downstream) fault location [1].

The first approach we propose is based on a single rule extracted from the analysis of features used in previous source location algorithms proposed in the literature. A multivariate statistic analysis of available registers described by such a set of features revealed the composition of this rule. The second approach takes advantage of information contained on sag waveforms. A model based on PCA using the complete registers of voltage and current signals has been used to classify the waveforms in two classes (downstream/upstream).

The problem of estimating the fault location from the registers of sags is not new. Chouhy [1] has described and compared five source location algorithms using synthetic data [2] to [5]. In this comparison, the DR algorithm obtained the best results [5], whereas the RS algorithm obtained poor results [3]. Author has also tested PCSC algorithm in [6]. DR, RS and PCSC algorithms where selected to perform the statistical analysis of the features involved in the first method proposed in this paper.

On the other hand, Khosravi et al. ([7][8]) presents a methodology based on the MPCA to build a statistical model using the voltage and current waveforms that is used in the classification of sags according to their origin. The classification is done based on the projection of new sags on the projection space defined by the model and analyzing the distance to the center of the model (T^2 statistic) and the projection error (Q statistic). Later, Melendez et al. ([9]) refined the method by adding a CBR step in the decision procedure that allows improving the classification rate based on the similarity of the new sag with those previously diagnosed. This last approach is being compared in this papers with the one rule method described previously.

This paper is organized in six additional sections. In the third section the set of registers of sags is described. Next, a brief description of confusion matrix and ROC curves is performed because the first has been used in the comparison and the second in the selection of the best MPCA model of the second approach. In the fifth and sixth sections, both methods are presented. Later, in the seventh section the results yield by both methods are summarized and compared. Finally, the conclusions are given.

This work has been partially funded by the Spanish government under the project DPI2006-09370, MEC (BES-2007-14942) and FIE grants, and ENDESA DISTRIBUCIÓN.

V. Barrera, X. Berjaga J. Melendez and S. Herraiz are with the eXiT Group in the Institute of Informatics and Applications (IIiA) of the University of Girona, Spain, Girona, Campus Montilivi, 17071, e-mail: [vbarrera, xberjaga, quimmel, sherraiz]@eia.udg.edu. Homepage: http://eia.udg.edu/~[vbarrera, xberjaga, quimmel, sherraiz].

J. Sánchez and M. Castro are with the PQ Department of Endesa Distribución, Barcelona (Spain), e-mail: jslosada@fecsa.es, mcastro@enher.es

The eXiT is part of Automation Engineering and Distributed Systems (AEDS) research group, awarded with a consolidated distinction (SGR-00296) for the 2005-2008 period in the Consolidated Research Group (SGR) project of the Generalitat de Catalunya.

978-1-4244-1770-4/08/$25.00 ©2008 IEEE

III. Gathered Voltage Sag

ENDESA[1] provided the set of sag events registered in three substations (25kV) classified as upstream or downstream. Each register was sampled at 128 samples per cycle (50Hz) and contains 40 cycles (**Fig. 1**).

A. *Preprocessing*

Short Fourier Transform (SFT) in one cycle with a 128-sample sliding window has been used to estimate the RMS voltage and current sequences at the frequency (50 Hz) of nominal power.

B. *Voltage Sag Selection*

TABLE I
VOLTAGE SAG EVENTS USED IN THE STUDY

	Initial amount		Excluded		Total		
Subst.	Down	Up	Down	Up	Down	Up	Down+Up
A	43	49	0	15	43	34	77
B	26	35	8	8	18	27	45
C	12	56	0	17	12	39	51
	81	140	8	40	73	100	173

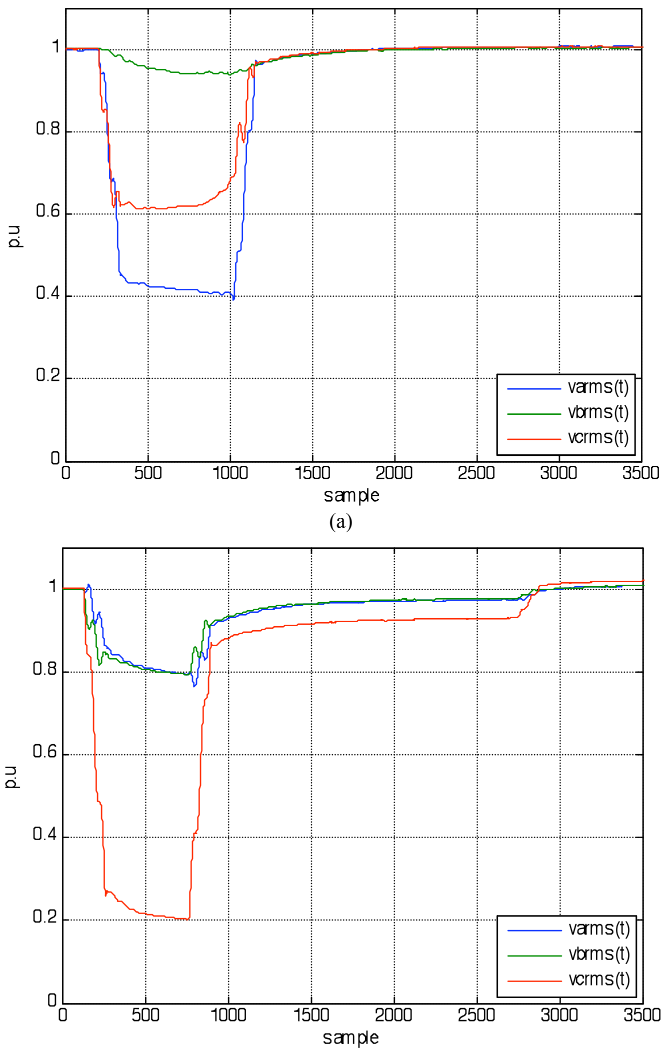

Fig. 1. Example of Voltage sag events: (a) Downstream, (b) Upstream. RMS sequence calculated with a 128-sample sliding window.

[1] Spanish acronym of Energy distribution company of Barcelona-Spain

The initial set of available sags contained 81-downstream and 140-upstream sag events. This original set has been cleansed and those registers, without complete waveforms (without pre or post fault state. See for example **Fig. 2**.), have been eliminated [10]. Finally a database of 173 sags (73-downstream, 100-upstream) has been used, Table I.

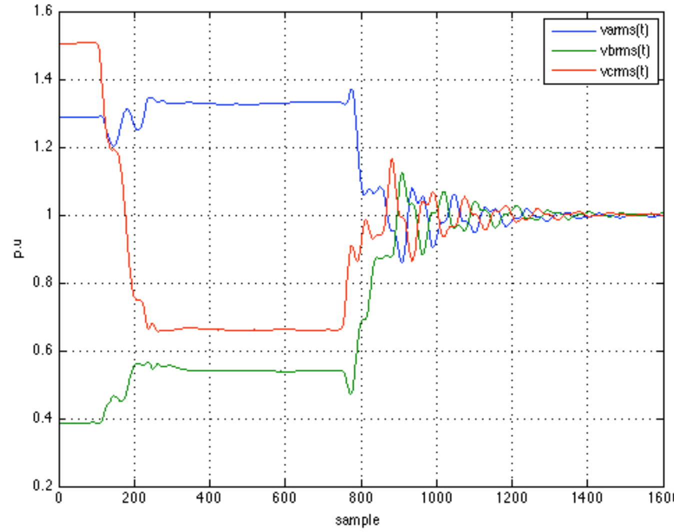

Fig. 2. Excluded sag event (25 kV). It event does not steady state. The source locations algorithms require prefault state.

IV. Confusion Matrix And Roc Curves

In order to compare the results obtained by both approaches a confusion matrix was used. A confusion matrix is a representation of classification results as this presented in Table II. It shows the differences between the true and predicted classes for a set of labeled examples [11].

TABLE II
CONFUSION MATRIX

		Real class	
		Reference class	No reference Class
Predicted Class	Reference class	TP	FP
	No reference class	FN	TN

Where,

- TP stands for true positive (cases correctly predicted as the reference class).
- TN stands for true negative (cases correctly classified as non reference class).
- FP for false positive (cases classified as the reference class with its real class is the non reference class).
- FN for false negative (cases classified as a non reference class with its real class is the reference class).

The evaluation of these indices allows computing several performance parameters of the classifier. Special attention is put on *sensitivity* and *specificity* to compute de *Receiver Operation Characteristic* - ROC curve:

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

A ROC curve is a two-dimensional graph where the y-axis represents *sensitivity* and x-axis represents *1–specificity* of the classifier obtained for different values of the decision threshold used in the classification. To compare the performance of several models through a ROC curve the *Area Under the ROC curve* – AUC can be used. ROC curve and AUC are used to select the best PCA model parameters in the approach based on waveform.

V. One Rule Based Approach

A. Introduction

The proposed rule uses two features in the antecedent to determine the sag origin. The proposed rule is the following:

Proposed rule:

"IF Rey<0 AND Zsag<Zss THEN downstream ELSE upstream END"

Where *Zsag* is the impedance during the event and *Zss* is the steady state pre fault impedance. *Rey* is the estimated resistance from the imaginary part of the sequence components.

This rule was obtained from a multivariable statistical analysis. MANOVA[2] and DA[3] statistical techniques were used in the analysis.

B. Voltage Sag Source Location Algorithms

A brief description of the selected algorithms is included in this section in order to show the features analyzed in the MANOVA and DA techniques. The first algorithm only uses the current signal, while the others use both voltage and current signals.

1) Phase Change in Sequence Current - PCSC: It estimates the origin using the change of the phase angle $\Delta\phi$ of the positive-sequence component of the current between the fault and prefault conditions. The PCSC rule is:

PCSC Rule:

"IF $\Delta\phi$>0 THEN upstream ELSE downstream END"

Where $\Delta\phi$ is set between $-\pi$ to π.

One cycle before the fault is used for computing prefault phasor and another cycle after the fault inception is used for estimation of fault current phasor [6].

2) Resistance Sign - RS: It obtains the source location from the sign of the real part of the estimated impedance. RS algorithm uses the positive sequence voltage and current component. The impedance is estimated taking *n* cycles of the voltage and current signals including prefault and fault cycles. The number of cycle is determined by the reversion of the power flow. Two impedances are estimated, therefore two resistances too (*Rex* and *Rey*), *Rex* is due to the estimated impedance from the real part of the sequence components; and *Rey* is due to the estimated impedance from imaginary part of the sequence components. An extended explanation about this is performed in [2]. The RS rule is:

RS Rule:

"IF Rex>0 AND Rey>0 THEN upstream ELSE IF Rex<0 and Rey<0 THEN downstream ELSE not conclusive test END"

Hence, if both resistances have different sign the test is not conclusive. Positive sign means that source location is upstream and negative sign is associated to downstream sag [2].

3) Distance Relay - DR: Its principle is based on the change of magnitude and angle of the impedance seen by the relay before and after the fault. For a downstream fault the seen impedance during the fault will decrease with respect the seen impedance in steady state; its phase angle will change too. The DR rule is:

DR Rule:

"IF |Zsag|<|Zss| AND $\angle$Zsag>0 THEN downstream ELSE upstream END"

According to the different type of faults, *Zss* and *Zsag* the proper voltage-current pair has to be taken in estimating them [5].

C. Results Of Source Location Algorithms

The algorithms were implemented in Matlab®. The results obtained for each one are reported in Table III. The upstream class has used as the reference class. The classification rates obtained with the dataset we presented are very poor: 51,4%, 27,7% and 65,9% respectively.

TABLE III
CONFUSION MATRIX OF THE PROPOSED RULE RESULTS

Real Class		Real Class		Real Class		
PCSC (51,4%)		RS (27,7%)		DR (65,9%)		Total (173 sags)
TP	FN	TP	FN	TP	FN	Upstream
56	44	48	0	87	13	100
FP	TN	FP	TN	FP	TN	Downstream
40	33	0	0	46	27	73

The poorest results correspond to RS algorithm. It is because *Rex* and *Rey* have different signs in many cases. The *Rex* does not take negative values. As a result FP and FN are zero. In **Fig. 3** this behavior is depicted and commented. PCSC algorithm obtained the second better classification rate. With a classification rate near to 50% is difficult to know the reasons to obtain good or wrong classification. As a result, it was impossible analyze the PCSC feature ($\Delta\phi$) behavior. DR algorithm obtained the best results (65,9%), however this classification rate is very low for a practical exploitation. This low rate was obtained because results with $\angle$Zsag>0 have been considered falses in many cases to downstream sag source location when they should be classified as true. As a result, 46 downstream source locations were classified as upstream.

D. Statistical Tools: Justification of the classification rule

The multivariate statistical tools used to propose the source location rule are here described.

1) MANOVA: The more important purpose of MANOVA is to explore how independent variables influence on some

[2] MANOVA – Multivariate Analysis of Variance
[3] DA- Discriminant Analysis

patterning of response on the dependent variables.

The sag source location was used as independent variable whereas $\Delta\phi$, *Rex*, *Rey*, *Zsag/Zss*, $\angle Zsag$ features were used as dependent variables. MANOVA allows answering this question: Which is the importance of each feature in the source location? Hence, we will know the influence grade of the source location over each feature. Finally, the features with most influence grade (quality) were selected.

TABLE IV
QUALITY OF THE SOURCE LOCATION EFFECT OVER THE FEATURE

Feature	Definition	Algorithm	Quality
$\Delta\phi$	Difference in phase angle between the positive-sequence component of current during fault and prefault conditions.	PCSC	1,2%
Rex	Real part of the estimated impedance from real part of sequence components.	RS	91,2%
Rey	Real part of the estimated impedance from imaginary part of sequence components.	RS	36,1%
Zsag/Zss	Ratio between fault impedance and steady state impedance.	DR	44,1%
$\angle Zsag$	Phase angle of the impedance during the voltage sag.	DR	0,4%

Table IV shows the quality of the sag source location effect over each feature. Values near to 100% indicate that the most of the variability in this feature is associated with sag source location. The tests of between subjects effect table indicates that the significance value for $\Delta\phi$, $\angle Zsag$ are greater than p=0,05; which means that sag source location effect is not significance over them (Table V). Thus, *Rex, Rey* and *Zsag/Zss* are features that are more associated with the sag source location.

TABLE V
TEST OF BETWEEN-SUBJECT EFFECTS

Feature	Type III Sum of squares	df	Mean square	F	Sig.
$\Delta\phi$	11,92	2	5,96	2,02	0,146
Rex	94,05	2	47,02	891,95	0,00
Rey	3,25	2	1,63	49,91	0,00
Zsag/Zss	292,03	2	146,01	69,12	0,00
$\angle Zsag$	3,84	2	1,92	0,647	0,53

2) DA [12]: It is a technique for classifying a set of observations into predefined classes. The purpose of using it is to determine the class of an observation based on a set of variables known as predictors or input variables. The model is built based on a set of observations for which the classes are known. Based on this training set, the technique constructs a set of linear functions of the predictors, known as *discriminant functions*.

DA has a method so-called *addition by step* implemented in SPSS Inc.[4], which adds step-by-step variables in order to determine the most discriminator variables. In each step the variable that minimize the Wilk's Lambda[5] is added. As a result, with *addition by step method* is possible to obtain the set of features that best describe the sag source location.

Addition by step method was applied to the set of sags. *Rex* and *Rey* were obtained as the most discriminator variables.

Observe that DA and MANOVA results are similar, except that the last one revels the importance of *Zsag/Zss* feature. Consequently, the three features have been selected and the influence in the determination of origin of sags analyzed.

E. Feature selection

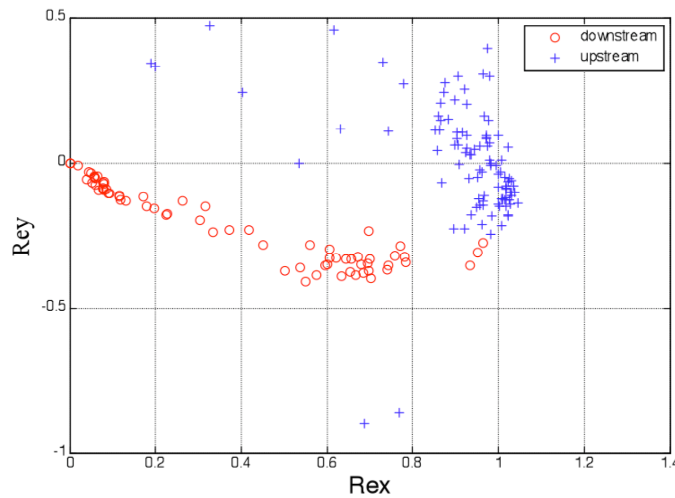

Fig. 3. Rex Vs Rey. Features extracted from RS algorithm.

The *Rex* vs. *Rey* is depicted in **Fig. 3**. *Rex* only takes positives values, whereas *Rey* takes both positive and negatives values. The range of these resistances is important because RS algorithm is based on the sign of them. Therefore, observing the RS rule, RS algorithm will not obtain downstream results, on the contrary many not conclusive result are obtained. Similar results in a test with synthetic sag were found [13]. For these reasons, *Rex* is excluded of the feature selection previously performed.

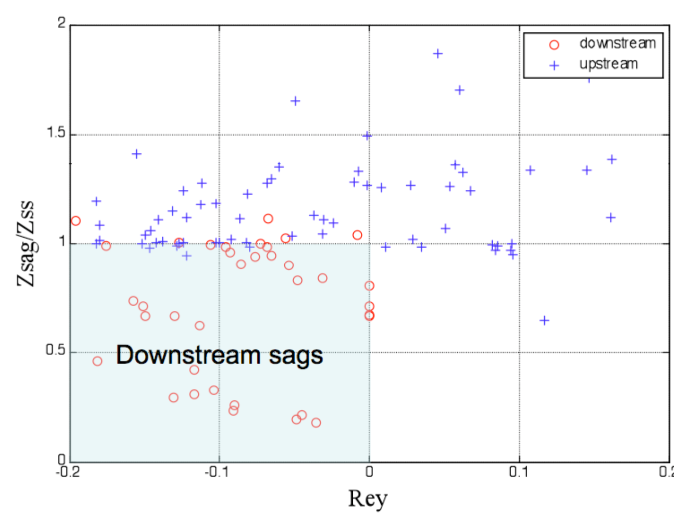

Fig. 4. Rey Vs Zsag/Zss. Features extracted from RS and DR algorithms. The downstream sags are inside of area Rey<0 and Zsag/Zss<1.

A linear classifier (with negative slope) gives a good

[4] SPSS Inc. - Statistical software (www.spss.com)
[5] Probability distribution used to contrast multivariate hypothesis.

solution using *Rex* and *Rey*, but the goal of this analysis is to propose a rule. With rules a training process is not needed.

A plot representing *Rey* vs. *Zsag/Zss* determines clearly the voltage sag origin. From **Fig. 4** it is possible to extract the proposed rule. So, voltage sag origin will be downstream whether *Rey<0* and *Zsag<Zss*. Otherwise, the voltage sag source location will be upstream.

F. Results Of The Approach Based On One Rule

The proposed rule was applied to the set of sags. The results are shown in Table VI.

TABLE VI
CONFUSION MATRIX OF THE PROPOSED RULE RESULTS

	Real Class		Total
	TP	FP	Upstream
(Predicted)	94	6	100
	FN	TN	Downstream
(Predicted)	8	65	73

Observe that the results obtained with the proposed rule are better than the results obtained by the previous algorithms with the same database. The classification rate is 91,9%, which are 26% units higher than the DR algorithm results (65,9%).

VI. APPROACH BASED ON WAVEFORM

This approach consists in building a PCA model from upstream or downstream sags. In this analysis the set of upstream sags were selected to build the model, because the amount of them is greater than downstream sags. So, the goal is to extract the relevant information of upstream sags useful to discriminate them from the downstream sags. The methodology has two general steps [9]:

1) Case Base Preparing and Model Construction in the Principal Component Space: RMS sequence of signals is computed, after that, RMS sequences are standardized, later the PCA model is created and downstream and upstream sags are projected in the PCA space.

2) Model Exploitation: It is based on the similarity criteria between new sags and those in the case base previously diagnosed. New sag is projected in Upstream-PCA model. Later, k_1 nearest neighbors is identified based on Q statistic in PCA space. After, the best k_2 neighbors from k_1 nearest are selected. The k_2 neighbors and a decision threshold (T_h) are used to determine if the new sag is close enough to the upstream sag model.

A. Results Of The Approach Based On Waveform MPCA Models

Validation of the approach has been done using 4-fold cross validation and computing the *sensitivity* and *specificity* for each experiment.

The Upstream-MPCA model was built. The model has been adjusted to capture the 95% of the variability contained in the original data resulting a model with 10 first principal components.

The methodology has been tested using different pairs of values for the k_1 and k_2 retrieved cases and using different thresholds (T_h) to compute the ROC curve depicted in Fig. 5. It can be observed that all the tests had a very good performance: AUC near 1 view Table VII.

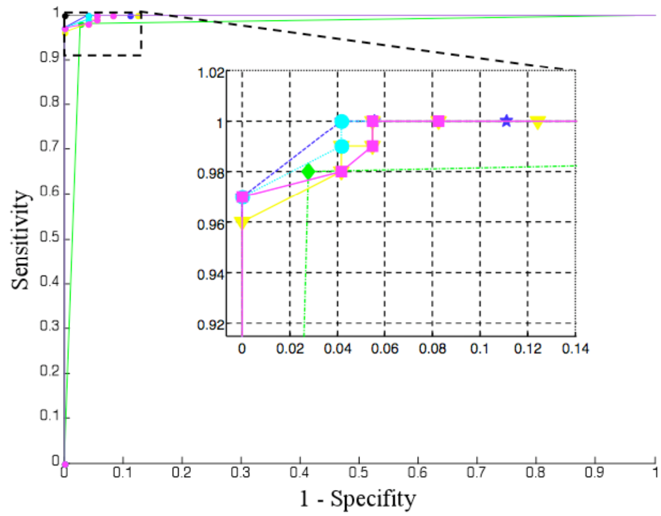

Fig. 5. ROC curves.

TABLE VII
AUC VALUES OF THE CLASSIFIERS

Parmaters	TP	FP	FN	TN	SEN	SPE	AUC
k1=15,k2=1	100	1	0	72	1	0.986	0.993
k1=15,k2=3	100	3	0	70	1	0.958	0.999
k1=15,k2=5	100	5	0	68	1	0.931	0.998
k1=10,k2=3	100	**2**	0	71	1	0.973	**0.999**
k1=10,k2=5	100	4	0	69	1	0.945	0.990

According to Table VII, the best classifier has k_1 = 10 and k_2 = 3 because it presents less FP and high AUC value. In order to select the T_h value for this classifier, a test with several T_h values between 0 and 1 was performed. The best sensitivity and specificity values were obtained with T_h =0,2.

VII. COMPARISON OF THE TWO APPROACHES

The waveform approach obtained the best result (98,8% against 91,9%). The confusion matrix in Table VIII is shown. It presents only 2 errors, while the feature approach has 14 errors in the determination of sag origin.

TABLE VIII
CONFUSION MATRIX OF THE TWO APPROACHES

Approach	TP	FP	FN	TN	SEN	SPE	Hit
Feature	94	6	8	61	0,922	0,910	91,9%±4,1%
Waveform	100	2	0	71	1,000	0,973	98,8%±1,6%

According to table, the error is 8,1% and 1,2%, respectively. In the waveform approach, with an interval confidence of 95% (Normal distribution)[6] the error will fluctuate 1,2%±1,6% and the classification rate 98,8%±1,6%, which means that other similar tests are very probably (0,95) to give a classification rate inside in the interval 97,2% to 100%.

In spite of the difference between the classification rates,

[6] Whether the number of instances is greater than 30, the Normal distribution is possible use it [16].

each approach has additional advantages and drawbacks that have to be taken into account. Table IX summarizes some them.

TABLE IX
ADVANTAGES AND DRAWBACKS OF THE TWO APPROACHES

First Approach (Features)	Second Approach (Waveforms)
• The classification rate is good. • The computational cost is low. • It does not require training and validation processes. • It does not require raw data. The needed information from relays can be obtained. • It is not based on model. • It is possible online implement it.	• The classification rate is excellent. • The computational cost is high. • It requires training and validation processes. • It requires raw data to obtain the training and validation dataset. • It is based on model. • It is possible to online implement it.

VIII. CONCLUSION

Two voltage sag source location approaches have been tested. One approach is uses a statistical model of the waveforms based on MPCA and the other one is simpler and is based on simple features extracted from the RMS waveforms. Both present good results in the estimation of the sag origin. Their advantages and drawbacks have been related.

Now, the wrong classified sags are being analyzed carefully to improve the single rule method. At the same time the methods will be tested with a larger number of sags registered in other substations.

REFERENCES

[1] R. Chouhy L., “Voltage Sags: Single event characterisation, system performance and source location”, PhD thesis, Chalmers University of Technology, 2007.

[2] C. Li, T. Tayjasanant, W. Xu, and X. Liu, "Method for voltage-sag source detection by investigating slope of the system trajectory," IEE Proceedings-Generation, Trans. and Distr., vol. 150, pp. 367-72, 2003.

[3] T. Tayjasanant, C. Li, and W. Xu, "A resistance sign-based method for voltage sag source detection," IEEE TPWD, vol. 20, pp. 2544-51, 2005.

[4] N. Hamzah, A. Mohamed, and A. Hussain, "A new approach to locate the voltage sag source using real current component," Electric Power Systems Research, vol. 72, pp. 113-23, 2004.

[5] A.K. Pradhan and A. Routray, "Applying distance relay for voltage sag source detection," IEEE Transactions on Power Delivery, vol. 20, pp. 529-31, 2005.

[6] A.K Pradhan, A. Routray, S. Madhan, “Fault direction estimation in radial distribution system using phase change in sequence current”, IEEE TPWDelivery, vol. 22, No. 4, pp. 2065-2071, october 2007.

[7] A. Khosravi, J. Meléndez, J. Colomer, “Classification of Voltage Sags Based on MPCA Models”, Lecture Notes in Computer Science, Springer Berlin, ISSN 0302-9743 (Print), 1611-3349 (Online), vol. 4477/2007, pp. 362-369, July 2007.

[8] A. Khosravi, J. Melendez, J. Colomer, “A Hybrid Method for Sag Source Location in Power Network”, 9th International Conference. Electrical Power Quality and Utilisation. Barcelona, 9-11, October 2007.

[9] J. Meléndez, X. Berjaga, S. Herraiz, J. Sánchez and M. Castro, “Classification of Voltage Sags based on k-NN in the Principal Component Space”, International Conference on Renewable Energies and Power Quality – ICREPQ’08, Santander 12-14 March, 2008.

[10] Barrera, V; Meléndez, J; herraiz, Sergio, “Strategies For Dealing With Voltage Sags In Power Quality Monitoring“, Electric Power Systems Research, ELSEVIER, EPSR-D-08-00410, July, 2008, submitted.

[11] Bradley A. P., “The use of the area under the ROC curve in the evaluation of machine learning algorithms”, in Pattern Recognition, Vol. 30, No. 7, pp. 1145-1159.

[12] Russell, E.L., Chiang, L.H., Braatz, R.D.: Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes, London. Springer-Verlog, Berlin Heidelberg New York (2000).

[13] C. Leborgne, D. Karlsson, J. Daalder, “Voltage Sag Source Location Methods Performance Under Symmetrical and Asymmetrical Fault Conditions”, T&D Conference and Exposition: Latin America, IEEE/PES, pp: 1-6, August 2006.

[14] De Mantaras R. L. and Plaza E., “Case-based reasoning: An overview”, in AI Communications, Vol. 10, No 1, pp 21-29.

[15] Aamodt A., Plaze E., “Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches”, in AI Communications 1994, Vol. 7, No. 1, pp. 39-59.

[16] J. Hernández O., M. Ramírez., C. Ferri R., “Introducción a la minería de datos”, ISBN 84-205-4091-9, Prentice Hall, España, 2004.