

# Unusual Voltage Sag Event Detection in Power Systems

V. Barrera Núñez, J. Meléndez Frigola, S. Herraiz Jaramillo, J. Sánchez Losada

**Abstract**— Three multivariate statistical tools (Principal Component Analysis, Factor Analysis, Analysis Discriminant) have been tested to characterize and model the sags registered in distribution substations. Those models use several features to represent the magnitude, duration and unbalanced grade of sags. They have been obtained from voltage and current waveforms. The techniques are tested and compared using 69 registers of sags. The advantages and drawbacks of each technique are listed.

**Index Terms**— discriminant analysis, factor analysis, power quality monitoring, principal component analysis, voltage sag (dip).

## I. INTRODUCTION

THE importance of Power Quality monitoring has increased in the last years [1]. Consequently, it has caused sudden advances in monitoring devices. Electric parameters are permanently monitored and registered, producing huge volume of data. The end-user has a large amount of data that can be used to assess and reduce the power quality problems [2][5]. Hence, statistical techniques have to be exploited to take advantage of information contained in those registers. Principal Component Analysis (PCA), Factor Analysis (FA) and Discriminant Analysis (DA) have been tested in this paper to characterize voltage sags and determine unusual ones.

This paper is organized in twelve sections. In the second section the features extracted from sag waveforms are described. Next, we present the registers of voltage sags. In the fourth to sixth sections, PCA, FA and DA are briefly explained. Later, in the seventh to ninth sections, the results of each technique are shown. Finally, comparison results, conclusion and future works are discussed.

## II. FEATURES OF VOLTAGE SAGS

We have extracted, from the sags waveforms, a set of features useful to describe duration, magnitude and unbalance degree of sags. They are summarized in Table I. Each feature have been selected because the information it provides:

1) *RCV* and *PNF*: Minimum *Remaining Complex Voltage* and *Positive-Negative Factor F* contain information about

magnitude and unbalance grade of sags, respectively. *Six-phase* and *Symmetrical component* are two algorithms to calculate them [4][3]. The first algorithm was used in this work. *RCV* is a voltage signal that represents the three phase voltages.

2) *L<sub>v</sub>*: *Loss of voltage* is defined as the integral of the RMS voltage drop during the event. *L<sub>v</sub>* describes magnitude and duration of sag [3]. In this analysis *L<sub>v</sub>* was computed from *RCV* instead of the three gathered voltage signals, according to the equation 1:

$$L_v = \sum_{sag} (1 - RCV_{p.u.}(t)) \quad (1)$$

TABLE I  
FEATURES TO UNUSUAL VOLTAGE SAG EVENT DETERMINATION

Feature	Duration	Magnitude	Unbalance
RCV		X	
PNF			X
L <sub>v</sub>	X	X	
GV <sub>0</sub>	X		X
I <sub>ratio</sub>		X	
T	X		
N <sub>seg</sub>	X		

3) *GV<sub>0</sub>*: This is a new feature we propose in this paper. *Gain of zero sequence voltage* corresponds as the integral of the RMS zero sequence voltage during the event. It has been computed using the equation 2:

$$GV_0 = \sum_{sag} \frac{V_0(t)_{rms}}{V_{pre-fault}} \quad (2)$$

4) *I<sub>ratio</sub>*: Ratio between the fundamental frequency component of the currents during and before of voltage sag event. It was previously defined in [6].

5) *T*: Voltage sag event duration measured in samples.

6) *N<sub>seg</sub>*: *Nseg/2* corresponds to number of sag transitory states.

Duration, magnitude and unbalanced grade were selected as sag characteristics in order to determine this unusualness.

## III. DESCRIPTION OF VOLTAGE SAGS:

The study has been carried out with 69 registers of sags. All the registers contain 40 cycles, sampled at 128 samples per cycle (50Hz).

This work has been partially supported by the Spanish government under the project DPI2006-09370 and ENDESA DISTRIBUCIÓN.

The authors are with the eXIT Group in the Institute of Informatics and Applications of the University of Girona, Spain, Girona, Campus Montilivi, 17071, e-mail: [vbarrera@eia.udg.es](mailto:vbarrera@eia.udg.es), [quimmel@eia.udg.es](mailto:quimmel@eia.udg.es), [sherraiz@eia.udg.es](mailto:sherraiz@eia.udg.es).

The sequence of RMS values has been calculated using the Short Fourier Transform (SFT) in one cycle (128-samples) sliding window to estimate the magnitude of the nominal frequency (50Hz). Matlab® has been used to compute the features whereas the statistical analysis has been performed on SPSS®<sup>1</sup>. All the features were standardized with zero mean and unit variance ( $\mu=0$ ;  $\sigma=1$ ). Table II summarized the correlation among these features.

TABLE II  
DATASET CORRELATION MATRIX

	RCV	PNF	LV	GV0	I <sub>ratio</sub>	T	N <sub>seg</sub>
RCV	1	0,63	-0,47	0,20	-0,61	0,02	0,07
PNF	0,63	1	-0,24	0,108	-0,39	0,03	-0,18
LV	-0,47	-0,24	1	0,06	0,21	0,51	-0,02
GV0	0,20	0,108	0,06	1	0,06	0,39	0,04
I <sub>ratio</sub>	-0,61	-0,39	0,21	0,06	1	0,00	0,19
T	0,02	0,03	0,51	0,39	0,00	1	0,26
N <sub>seg</sub>	0,07	-0,18	-0,02	0,04	0,19	0,26	1

The features  $GV_0$ ,  $T$  and  $N_{seg}$  have low linear correlation (correlation coefficients are near to zero, especially  $N_{seg}$ ).

TABLE III  
NORMALITY TEST

Signif. level	RCV	PNF	Lv	V <sub>0</sub>	I <sub>ratio</sub>	T	N <sub>seg</sub>
KS	0,02	0	0	0	0	0	0
SW	0,002	0	0	0	0	0	0

In order to determine the normality of each feature (variable) the Kolmogorov-Smirnov (KS) and Shapiro-Wilk (SW) normality tests were performed (Table III). KS and SW test the null hypothesis  $H_0$ : *The distribution of the population is normal*. Hence, both tests rejected it; therefore the distribution of these features is not Gaussian.

#### IV. PRINCIPAL COMPONENT ANALYSIS

It is based on the analysis of the covariance matrix of a set of  $m$  observations represented by  $n$  variables. The singular value decomposition of this matrix is used to obtain a set of eigen-values and eigenvectors. And these are used to project the original data onto a lower dimension space preserving the same information and rejecting the noise. These new variables, or scores, are the result of the linear combination of the original variables [1]. Then, the set of data can be expressed as a linear combination of  $r$  (lower than  $n$ ) new variables,  $t_i$  assuming an error  $E$  [9]:

$$X = \sum_{i=1}^r t_i p_i + E \quad (3)$$

Where  $t_i$  and  $p_i$  are named scores and loading vectors respectively and are computed to reflect relevant relation amongst samples ( $t_i$ ). While  $p_i$  highlights the correlation among variables and they correspond to eigenvectors of covariance matrix (C).

$$Cp_i = \lambda p_i \quad (4)$$

PCA assumes that the eigenvectors with bigger eigen-values

are the best ones for expressing the data upon based on the maximum variance criteria.

$$t_i = Xp_i, \quad i = 1, \dots, r \quad (5)$$

The first  $r$  principal components build up a new space with a lower dimensionality than the original one. Projection of the data to the  $i$ -th axis in this new space can be done using linear transformation presented in equation 5 [9].

##### A. PCA Previous Tests

In order to assess if the data are suitable to perform a PCA these tests are performed.

1) *Barlett's Test*: In PCA is convenient that the variables are correlated, so the correlation matrix differs from identity matrix). This test allows statistically determining if the variables are correlated or not. Barlett's tests the null hypothesis:  $H_0$ : *The correlation matrix is the identity matrix*. With a confidence interval equal to 95% and Barlett significance value lower than 0,05 implies that variables are correlated (hypothesis  $H_0$  is rejected).

2) *Kaiser–Meyer–Olkin coefficient*: KMO measures the proportion between simple correlation and partial correlation. KMO greater than 0,7 is recommendable to perform PCA, view equation 6.

$$KMO = \frac{\sum_{i=1}^n \sum_{j \neq i}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j \neq i}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j \neq i}^n p_{ij}^2} \quad (6)$$

Where  $r_{ij}$  and  $p_{ij}$  are the simple and partial correlation coefficients between variables respectively.

#### V. FACTOR ANALYSIS

Factor analysis is a method for investigating whether a number of variables of interest  $X_1, X_2, \dots, X_n$ , are linearly related to a smaller number of unobservable or subjacent variables known as *factors* [10].

$$X = \mu + \sum_{i=1}^r t_i p_i + E \quad (7)$$

Where,  $t_i$  and  $p_i$  are scores and loadings vectors respectively.  $\mu$  is the mean of scores to  $k$ -th observation and  $E$  are the errors.

The meaning of the factors must be induced from seeing which variables are most heavily loaded on which factors.

FA and PCA methods are related but differ. PCA is generally used when purpose is data reduction. FA is generally used when the research purpose is to identify latent variables which contribute to the common variance of the set of measured variables. FA distinguishes the specific variability while PCA the global variability [10].

##### A. FA Previous Tests

Barlett's test and KMO coefficient revision have to be performed to analyze the suitability of this method. But also additional two test are recommended:

<sup>1</sup> Statistical software to data analysis

1) *Determinant of correlation matrix*: If determinant of the correlation matrix is close to zero the variables are highly correlated.

2) *Measure of Sampling Adequacy - MSA*: It is a coefficient similar to KMO but MSA is for each variable. MSA greater than 0,5 to perform FA is suggested.

$$MSA_i = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2} \quad (8)$$

## VI. DISCRIMINANT ANALYSIS

Discriminant analysis is a technique for classifying a set of observations into predefined classes. The purpose is to determine the class of an observation based on a set of variables known as predictors (*discriminating variables*). The model is built based on a set of observations for which the classes are known (*criterion variable*). The *criterion variable* contains the class of each observation. Based on the set of observations, the technique constructs a set of linear functions of the predictors, known as *discriminant functions*, such that [8]:

$$L = +b_1x_1 + b_2x_2 + \dots + b_nx_n + c \quad (9)$$

Where,

$b_i$ , discriminant coefficient.

$x_i$ , discriminating variables.

$c$ , constant term.

$L$  is analogous to multiple regression, but the  $b$ 's are discriminant coefficients which maximize the distance between the means of the class.

There is one discriminant function for 2-class discriminant analysis, but for higher order DA, the number of functions (each with its own cut-off value) is the least of  $(k - 1)$ , where  $k$  is the number of classes in the criterion variable, or  $p$ , the number of discriminating variables. Each discriminant function is orthogonal to the others [8].

The selected features are *discriminating variables* in this analysis. The *criterion variable* is a vector that indicates if the sag is unusual or not.

## VII. PCA RESULTS

PCA has been applied using the whole set of features except  $N_{seg}$ . It has been excluded due to its poor correlation level. A significance value in Bartlett's test (0,01) has been obtained (lower than the original 0,05). Therefore, the null hypothesis ( $H_0$ ) can be rejected, which means that the features are significantly correlated and suitable to apply the PCA method.

The KMO coefficient equal to 0,55 was obtained. Although it is lower than the recommended 0,7, it is close enough to continue with the method taking it in consideration.

Three principal components were selected because explain 81,1% of the global variability according to Table IV. Moreover, observing the loads vectors ( $p_i$ ) in Table V, the last

three components explain low variability of data.

TABLE IV  
EIGEN-VALUES AND ACCUMULATED EXPLAINED VARIANCE

	1	2	3	4	5	6
$\lambda$	2,34	1,63	0,90	0,58	0,36	0,20
% $\lambda$	39,07	27,13	14,94	9,62	5,94	3,31
$\sum\% \lambda$	39,07	66,19	81,14	90,76	96,70	100

TABLE V  
LOADINGS VECTORS

Feature	Principal Component					
	1	2	3	4	5	6
<i>RCV</i>	0,91	0,20	0,04	-0,07	0,14	0,33
<i>PNF</i>	0,75	0,23	-0,08	0,60	-0,10	-0,13
<i>LV</i>	-0,63	0,51	-0,46	0,13	-0,28	0,19
<i>GV0</i>	0,08	0,70	0,65	-0,16	-0,26	-0,04
<i>Iratio</i>	-0,73	-0,10	0,48	0,42	0,19	0,15
<i>T</i>	-0,18	0,89	-0,18	-0,05	0,38	-0,11

First and second loading vectors are depicted in Fig. 1. It shows that sags with low depth are located on the right hand side (positive first component axis), while the sags with high depth (high sag magnitude) will appear on the left hand side. The same occurs with the second principal component, the longest sags are located in the positive axis. In third component we can observe that the unbalanced grade is described because the greatest coefficient is  $GV_0$ . Hence, these three components explain depth, duration and unbalance grade respectively.

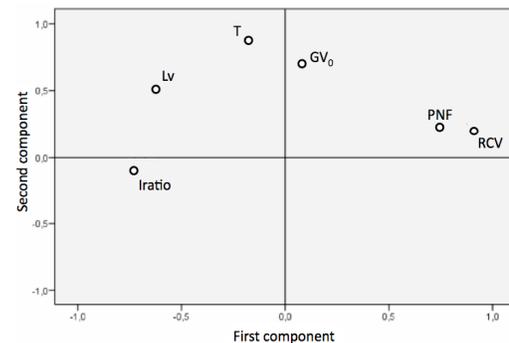


Fig. 1. Loadings vectors ( $p_i$ ). First and second component. The first component determines depth and second component the duration of sag.

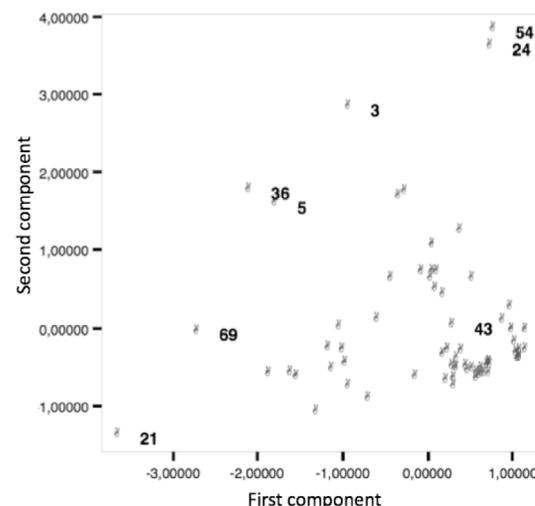


Fig. 2. Scores vectors. Voltage sag events projected on first and second principal component. Depth and duration of sag respectively.

When the original data is projected on the PCA space, we observe clearly that the sags number 54 is the longest sag of the dataset, while sag number 21 is the deepest and shortest; view Fig. 2. In third component the sag 54 appears as the most unbalanced.

### VIII. FA RESULTS

Determinant of correlation matrix, significance value obtained from Barlett's test, KMO and MSA coefficients in Table VI are shown.

TABLE VI  
FA SUITABLE TEST RESULTS

Excluded	D	KMO	Barlett	MSA			
				Nseg	V0	T	Lv
Nothing	0,1	0,48	0	0,02	0,36	0,43	
$N_{seg}$	0,14	0,55	0		0,47	0,46	
$N_{seg}, T$	0,25	0,56	0		0,21		
$N_{seg}, V_0$	0,19	0,54	0			0,36	0,14

In order to obtain a suitable set of features  $N_{seg}$  was excluded. Thus, MSA coefficients are close to 0,5 when  $N_{seg}$  is not included. A similar behavior was found in PCA too.

The number of factors through the *Chi-square goodness-of-fit Test* and some analysis was determined. It tests the hypothesis:  $H_0$ : *The adequate number of factors is lower or equal than extracted number of factors* and  $H_1$ : *The adequate number of factors is greater than extracted number of factors*. Then, Chi-Test suggests using two factors with a confidence interval of 95%, resulting in this case that  $H_0$  is accepted ( $0,173 > 0,05$ ) according to Table VII.

TABLE VII  
GOODNESS OF FIT TEST

Number of factors	Significance (Chi-square)
1	0
2	0,173

TABLE VIII  
FACTORIAL MATRIX

Feature	Factor		
	1	2	3
RCV	0,90	0,00	-0,00
PNF	0,63	0,06	-0,05
$L_v$	-0,47	0,81	-0,22
$GV_0$	0,20	0,41	0,81
$I_{ratio}$	-0,61	-0,03	0,24
T	0,02	0,69	0,14

However, with a number of factors equal to three, the meaning of each factor has a better interpretability than using only two. Observing the factorial coefficients shown in Table VIII, it is possible to determine the meaning of each factor.

Hence, first factor describes the depth of sags; second factor describes the loss of voltage and duration; and the third factor describes the degree of unbalance.

Sag number 36 is the sag event with a major loss of voltage and duration, while sag number 21 is the deepest; view Fig. 3. According to PCA the sag number 54 is the longest and most unbalanced, while in FA is only the most unbalanced (the highest score on third factor).

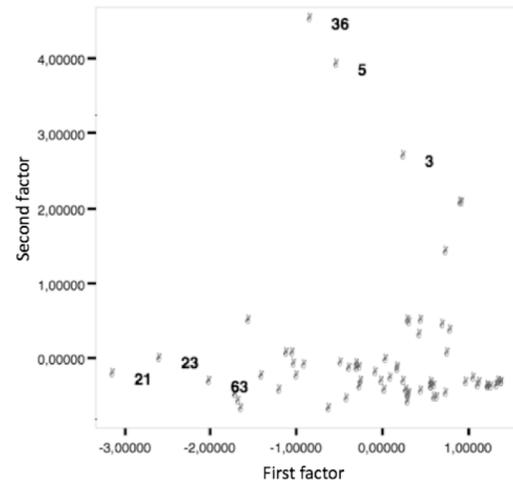


Fig. 3. Voltage sag events projected on first and second factor. Depth and  $L_v + T_{sag}$  of sag respectively.

### IX. DA RESULTS

In order to perform a DA, sags have been labeled with zeros or ones. Zeros is assigned to represent normal or usual sags and ones to unusual ones. A distance criteria defined on the projection space given by the PCA has been used. The sag with a Hostelling distance (it is referred as the  $T^2$  statistic) to the origin (center of the model) greater than 2 units in the PCA space were considered as unusual sags. So, 12 sag events were labeled as unusual.

TABLE IX  
DISCRIMINANT FUNCTION COEFFICIENTS

	RCV	PNF	$L_v$	$GV_0$	$I_{ratio}$	C
Coefficient	4,743	-6,171	0,005	0,001	0,116	0,502

The coefficients of the discriminant functions in Table IX are shown. The features  $N_{seg}$  and  $T_{sag}$  by DA were automatically excluded, due to their low discrimination. However, the features  $L_v$ ,  $V_0$  and  $I_{ratio}$  should be excluded too, because their coefficients are near to zero, as a result, their contribution is very low. Hence, the DA model only discriminates according to depth of sags.

In order to determine if the covariance matrices of groups are significantly equals, Box's test was performed. It tests the null hypothesis  $H_0$ : *The covariance matrices are equals*. Box's test rejected  $H_0$ , which means that the sag classes have different distribution in their original space. Therefore, Box's test suggests a quadratic discriminant function. But this is not completely true because of the reduced amount of data used. Consequently finally we used a linear discriminant function. Moreover, the quadratic discriminant analysis computational cost is  $O(n^2)$  for n-dimensional vectors [7]. The linear function is:

$$L = 4,7RCV - 6,2PNF + 0,005L_v + 0,001GV_0 + 0,1I_{ratio} + 0,5 \quad (6)$$

The sags will near to  $L=3.1$  are considered as unusual whereas those close to  $L=-0.6$  are the most common.

The F test of Wilks's showed that the means of these groups differ and consequently the null hypothesis  $H_0$  was rejected ( $H_0$ : The group means are equals).

The Box's test showed that the covariance matrices differ between groups. Lack of homogeneity of variances will mean significance tests are unreliable. The lack of homogeneity is due to the dataset is small and split of the dependent variable is very uneven.

In spite of that DA model with 10-fold cross validation was tested. The classification rate was 94.2%. Two usual and two unusual sags were wrongly classified.

## X. COMPARISON OF RESULTS

TABLE X  
FEATURES OF SOME VOLTAGE SAG EVENTS

id	RCV	PNF	LV	GV <sub>0</sub>	I <sub>ratio</sub>	T <sub>sag</sub>
0	0,74	0,92	206,8	181,9	4,17	1656
21	<i>0,18</i>	0,20	409,3	86,3	14,30	961
36	0,39	0,87	<i>1653,7</i>	3,0	2,45	<i>3137</i>
54	0,95	1,00	209,9	<i>3109,6</i>	2,03	<i>4737</i>

The value taken for features corresponding to sags number 21, 36 and 54 are shown in Table X. Those have been compared with the mean value of each feature (id=0). According to PCA an FA results, the relevant features have been italicized. For instance, the deepest sag is the number 21, its *RCV* magnitude is 0,18; four times smaller than *RCV* mean (0,74). Similarly, Table X confirms that sag number 54 is the most unbalanced and longest. The sag number 36 has the highest *L<sub>v</sub>* and high T with respect to average sag. The sags 21 and 54 are plotted in Fig. 4.

TABLE XI  
COMPARATIVE TABLE

PCA	FA	DA
-No probabilistic.	-No probabilistic.	-Probabilistic.
-Expert knowledge is not required.	-Expert knowledge is not required.	-An expert is required.
-It allows knowing the cause of unusualness.	-It allows knowing the cause of unusualness.	Sags have to be previously classified.
		-It does not allow knowing the cause of unusualness.

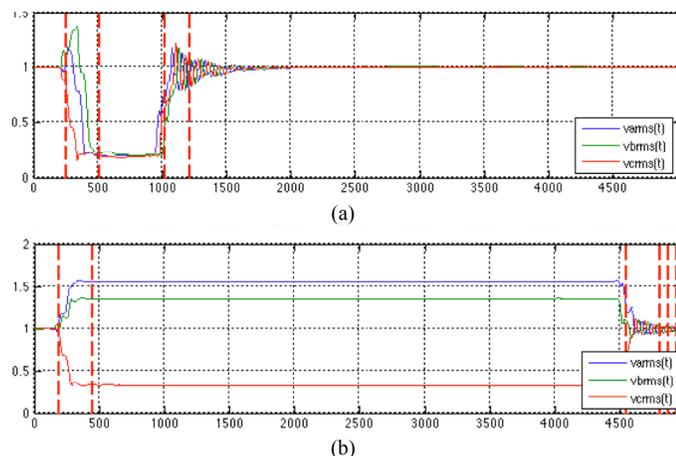


Fig. 4. Voltage sags number 21 and 54 (Vrms Vs Sample). RMS sequence phase voltages. (a) 21, (b) 54.

Some differences between the three approaches are summarized in Table XI are related. However, they have the next common qualities: training process is not tedious, are based on model, can be online implemented.

## XI. CONCLUSIONS

PCA, FA and DA have been assessed to determine the statistically usual and unusual voltage sags. PCA and FA allow knowing the causes of the unusualness, whereas the DA does not, also, an expert person is required to classify the sags as unusual or usual, whereas for using PCA and FA an expert is not required.

The selected features have shown a low correlation level in the registered sags used in this analysis. New features have to be proposed and to register more sags as well.

PCA, FA and DA can be used to determine unusual voltage sags. Advantages and drawbacks of each one have been presented in this paper.

## XII. FUTURE WORKS

A set of features highly correlated has to be proposed. Features that describe the unbalanced grade and duration to proposed feature set have to be added.

To determine unusual sags through one or several statistically criterions, such as: Euclidean distance, Mahalanobis distance, Hotelling's T<sup>2</sup>.

Proposed a methodology based on features and statistically tools to determine unusual voltage sag events.

## XIII. REFERENCES

- [1] J. Mora, V. Barrera, G. Carrillo, "Fault Location in Power Distribution Systems Using a Learning Algorithm for Multivariable Data Analysis". IEEE Transactions on power delivery, vol. 22, no. 3, July 2007, pp. 1715-1721.
- [2] S. Santoso, J. Lamoree, "Power quality data analysis: from raw data to knowledge using knowledge discovery approach", Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference, vol. 1, pp. 172-177, Seattle, Wash, USA, July 2000.
- [3] M. Bollen, D. Sabin, "International Coordination for Voltage Sag Indices", PES TD 2005/2006, May 21-24, 2006, pp.:229 - 234
- [4] M. Bollen, "Algorithms for Characterizing Measured three-Phase Unbalanced Voltage Dips", IEEE Transactions on Power Delivery, vol. 18, no. 3, July 2003.
- [5] McGranahan M, "Trends in Power Quality Monitoring", IEEE Power Engineering Review, October 2001.
- [6] Kyoung K., Jin P., Jong L., Seon A., Seung M., "A Method to determine the Relative Location of Voltage Sag Source for PQ Diagnosis", ICEMS 2005, Proceedings of the Eighth International Conference on Volume 3, 27-29 Sept. 2005 pp.:2192 - 2197.
- [7] S. Omachi, F. Sun, H. Aso1, "A New Approximation Method of the Quadratic Discriminant Function", Aso Lab, Tohoku University, [http://www.aso.ecei.tohoku.ac.jp/publication\\_data/31.pdf](http://www.aso.ecei.tohoku.ac.jp/publication_data/31.pdf).
- [8] McLachlan, Geoffrey J. "Discriminant analysis and statistical pattern recognition". Wiley Series in Probability and Statistics, 2004.
- [9] 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.
- [10] Gorsuch, R. L. (1983). "Factor Analysis". Hillsdale, NJ: Lawrence Erlbaum, 1974.