[International Journal of Power System Operation and Energy Management](https://www.interscience.in/ijpsoem)

Volume 1 | Issue 1 Article 14

July 2011

FCM and Statistical Based Approach for Classification and Location of Faults in Electrical Distribution System

PRATUL ARVIND Er.
Indian Institute of Technology - Roorkee,India, pratularvindiitr@gmail.com

Rudra Prakash Maheshwari Dr.
Indian Institute of Technology Roorkee,India, rpmaheshwari@gmail.com

FCM and Statistical Based Approach for Classification and Location of Faults in Electrical Distribution System

Pratul Arvind[1], Rudra prakash Maheswari[2]
[1,2]Department of Electrical Engineering, Indian Institute of Technology, Roorkee, India
e-mail id: pratularvindiitr@gmail.com[1], rpmaheshwari@gmail.com[2]

***Abstract*—Electric Power Distribution System is a complex network of electrical power system. Also, large number of lines on a distribution system experiences regular faults which lead to high value of current. Speedy and precise fault location plays a pivotal role in accelerating system restoration which is a need of modern day. Unlike transmission system which involves a simple connection, distribution system has a very complicated structure thereby making it a herculean task to design the network for computational analysis. In this paper, the authors have simulated IEEE 13- node distribution system using PSCAD which is an unbalanced system and current samples are generated at the substation end. A Fuzzy c-mean (FCM) and statistical based approach has been used. Samples are transformed as clusters by use of FCM and fed to Expectation- Maximization (EM) algorithm for classifying and locating faults in an unbalanced distribution system. Further, it is to be kept in mind that the combination has not been used for the above purpose as per the literature available till date.**

Keywords-PSCAD, IEEE 13-node feeder, FCM, EM.

I. INTRODUCTION

Electric Power Distribution System is a complicated network of electrical power system. Analogous to humans' circulatory system if transmission system can be termed as the arteries of human body then distribution system are the capillaries. Unlike transmission system which involves an easy connection, Distribution system [1] comprises of number of radial feeders which has to be highly reliable and efficient under normal and contingency condition. Transmission system had been a broad area for researchers due to its simplified structure, carries major portion of power over long distances and also considering the impact of the faults that would have on these kinds of lines. But presently due to the increased urbanization and industrialization, the amount of power carried by the distribution grids has also enhanced quite considerably. The large numbers of lines in a distribution system experience regular faults which lead to high value of line current. With the availability of inadequate system information and presence of high impedance faults, identifying and locating faults in a distribution system pose a major challenge to the utility operators. Further, in digital protection schemes, for proper operation of protective relays, correct determination of fault type is a prerequisite. Speedy and precise fault location plays a significant role in accelerating system restoration, reducing outage time and significantly improving system reliability. The methods proposed for fault location in transmission lines [2] are not easily applicable to distribution systems. Ratan Das et-al in [3] presents the design and development of a prototype fault locator which estimates the location of shunt faults on radial sub transmission and distribution lines based on the fundamental frequency component of voltages and currents measured at the line terminal. A review of the classical techniques and knowledge based approaches can be noticed in [4] and [5] which recommend a hybrid approach for locating faults. The methods for locating faults in electrical distribution systems may be broadly classified into three categories. The first deals with the methods that detect components of high frequency in travelling waves, the second includes methods that compute fault impedance from the rms values of current and voltages measured at the fundamental frequency, and the last but not the least is based on methods of visual inspection that consist of patrolling and checking the faulted feeder [6] and [7]. Several methods have been proposed for fault location [8] in power distribution systems. Most of them estimate the equivalent distance to the fault based on the impedance estimation as seen from the substation. The common drawback of the impedance-based methods is the multiple-estimation problem given by the existence of multiple points in the power distribution systems that fulfill the equivalent impedance condition. Consequently, these methods provide precise but uncertain fault locations. With the introduction of digital signal processing tools in power system, wavelet transform came into play for extraction of current features that can be subjected to algorithm meant for appropriate location of faults but yet an errorless fault location could not be achieved. Pratul et al [9] used Gabor transform to collect the features for determining the thresholds for fault location in an unbalanced distribution system. N-ary tree structure has also been proposed in [10] for locating faults due to the highly branched and the non-homogeneity nature of the distribution systems. Also, a combination of artificial neural network and support vector machine can be seen in [11] but training in neural network is itself a cumbersome process. Current features extracted from wavelet multi resolution approach [12] have not been able to fetch accurate result.

In the present paper, the authors have simulated IEEE 13- node distribution system using PSCAD which is an unbalanced system and current samples are generated at the substation end. The current samples are subjected to FCM to obtain clusters and fed to expectation maximization algorithm [13]. The paper presents an alternative solution to the problems associated with interruptions by means of a statistical modeling of current sample database applied to determine the fault location in power distribution systems to reduce the system restoration time.

II. GENERATION OF CURRENT SAMPLES

One of the crucial blocks in locating fault in a distribution system involves its design on computer interface software on which algorithm can be verified. The IEEE 13 node radial feeder [14] shown in Figure. 1 is considered as reference for generation of current samples at the substation end. The purpose behind publishing IEEE 13 node feeder data is to make available a common set of data that can be used by program developers and users so that the appropriateness of their solutions can be verified. Though the feeder is very small yet it displays some very interesting characteristics such as it is short and relatively highly loaded for a 4.16 kV feeder, has one substation voltage regulator consisting of three single-phase units connected in wye, overhead and underground lines are also present with variety of phasing. It is further equipped with shunt capacitor banks, in-line transformer and unbalanced spot and distributed loads. It is considered as very unbalanced system.

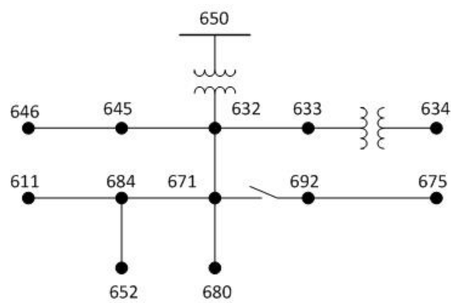

Figure 1: IEEE 13 – node feeder

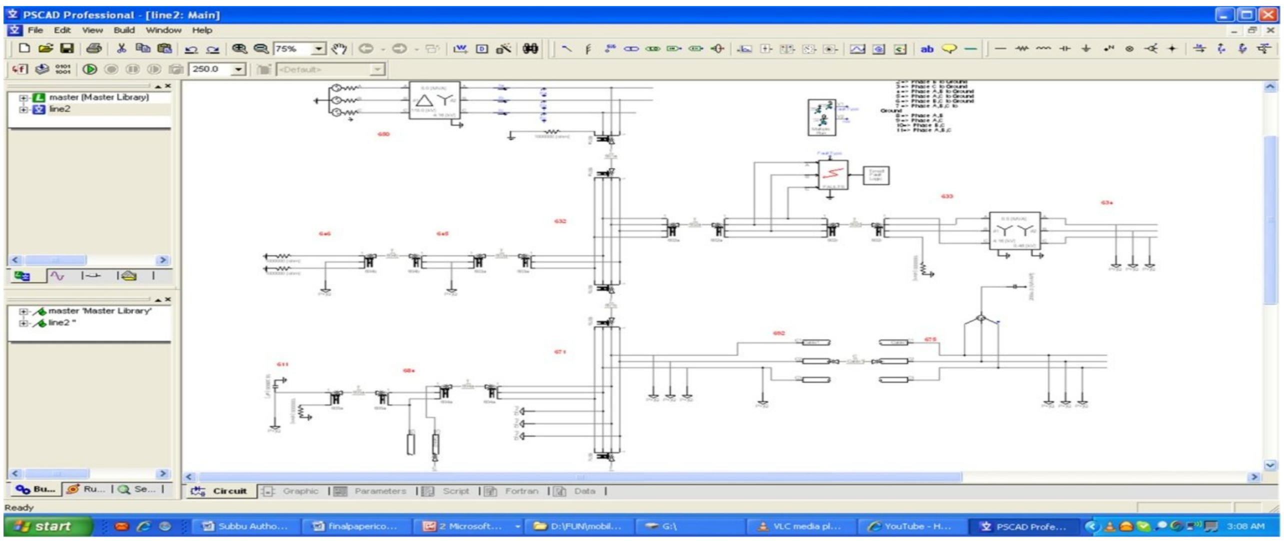

Figure 2: Simulation of IEEE 13- node feeder in PSCAD

The feeder shown in Figure 2 is designed and simulated in PSCAD [15] which is a powerful electromagnetic transient simulation program most suitable for time domain simulations of the systems. The graphical interface of the software makes it very easy to build the circuit and observe the results within a single integrated environment. For the purpose of modeling certain assumptions were considered such as the elimination of voltage regulator, purging of cable, and distributed load being replaced as spot load at the end of the segment. The Frequency Dependent (Phase) Model is considered since it is numerically accurate and robust transmission line model available. Three – Phase current samples is obtained at substation end after extensive simulation [16] by creating all nine types of faults (single – line to ground fault, double – line to ground fault, line –line) fault) respectively at various locations i. e between nodes 632 – 633, 632 – 671 at different fault resistance value ranging from 0 ohm to 30 ohm at a step of 10 ohm and at 180^0 fault inception angle over various locations between the node. The simulation has the duration of run for 1 second; with fault occurring at 0.34165 sec and duration of fault being 0.1 sec. the peak absolute values of the current samples are being considered. The rms values of the current samples are obtained using PSCAD. Line 632 – 633 is considered as zone 1 and 632 – 671 is considered as zone 2 respectively. Various locations are taken between nodes 632 – 633, 632 – 671 in step of 10% of total length of the line in the respective zones because they are the three – phase connections as per the given data. The samples thus obtained are utilized as the input for the algorithm presented later.

III. APPROACH FOR FAULT LOCATION

An approach to resolve the problem of fault location in distribution system subjected to different kinds of fault by examining system behavior is presented. After simulating a distribution system with different types fault over a range of fault resistance and at various locations, current waveforms are recorded at the substation end. These samples are pre-processed in PSCAD to obtain the rms value. Each recorded event has relevant information that enables data classification with certain types of classes established in the model. Detectable groups were taken into account in a preliminary data analysis. The goal is set to achieve by associating groups to zones in order to establish correspondence between fault location and data classification within the groups. Fuzzy c- mean is then applied on these current samples thus obtained and are subjected to expectation – maximization algorithm for fault classification and fault location of zones respectively. A detailed algorithm for FCM and EM-algorithm applied for the above purpose is presented below:

A. Fuzzy c-Means (FCM) Clustering

Fuzzy *c*-means (FCM) clustering was developed by Dunn [17] in 1974. This was further generalized by Bezdek [18] in 1981 and has become popular. It is considered as a derivative of *k*-means clustering. Clustering data allows the conformation of meaningful groups in an analytical way, which helps to classify data according to similarities or affinities. The clustering algorithms are based on the use of metric differences for the distance estimation. Various types of clustering methods have been developed. Out of them fuzzy clustering finds its vital representation in the field of data mining, artificial intelligence, numerical taxonomy, pattern recognition, image analysis, image processing, and medicine,. It is widely used because of fuzzy membership, since fuzzy sets could allow membership functions to all clusters in a data set so that it is very suitable for cluster analysis. Fuzzy C-means algorithm is based on the minimization of a criterion function. FCM clustering algorithm [19] is applied because of good performance and less execution time to obtain clustered data. In the proposed work, authors have used Haojun Sun *et al.* [20] algorithm to fix the number of clusters.

Suppose a matrix of n data elements (fault signal), each of size $s(s=3)$ is represented as $X=(x_1,x_2,.....,x_n)$. FCM establishes the clustering by iteratively minimizing the objective function as given in Eq. (1)

Objective function: $$O_m(U,C)=\sum_{i=1}^{c}\sum_{j=1}^{n}U_{ij}^{m}D^2(x_j,C_i) \quad (1)$$

Constraint: $$\sum_{i=1}^{c}U_{ij}=1; \quad \forall j \quad (2)$$

Where, U_{ij} is membership of the j^{th} data in the i^{th} cluster C_i, m stands for the fuzziness of the system $(m=2)$ and D represents the distance between the cluster center and data point.

B. FCM Algorithm

Flow chart of FCM algorithm is shown in Figure 3. The implementation steps are given below:
Input: fault signal data; Output: Clustered data;

- Initialize the cluster centers C_i.
- Calculate the distance D between the cluster center and data point by using Eq. (3)

$$D^2(x_j,C_i)=\left\|x_j-C_i\right\|^2 \quad (3)$$

- Calculate the membership values by using Eq. (4)

$$U_{ij}=\left(\sum_{k=1}^{c}\left(\frac{D(x_j,C_i)}{D(x_j,C_k)}\right)^{2/(m-1)}\right)^{-1} \quad (4)$$

- Update the cluster centers using Eq. (5)

$$C_i=\frac{\sum_{j=1}^{n}U_{ij}^{m}x_j}{\sum_{j=1}^{n}U_{ij}^{m}} \quad (5)$$

- The iterative process starts:

1. Update the membership values U_{ij} by using Eq.(4)
2. Update the cluster centers C_i by using Eq. (5).
3. Update the distance D using Eq. (3).
4. If $\left|C_{new}-C_{old}\right|>\varepsilon; \quad (\varepsilon=0.001)$ then go to step1.
5. Else stop.

- Assign each fault signal to a specific cluster for which the membership is maximal.

IV. Expectation – Maximization Algorithm

From information of the groups obtained using the FCM algorithm (discussed in section – III), initial values for the centers are estimated. The initial value of covariance matrix is taken as the identity matrix and the mixture coefficients are then calculated with the proportion of data in each group, in relation to the sample. Once initial parameters are obtained, the estimation of the mixture model parameters is initiated by the Expectation - Maximization algorithm [21], which is an iterative procedure until the desired convergence is achieved. EM is an iterative approach to maximum likelihood estimation. Each Iteration of an EM algorithm consists of two steps: an Estimation (E) step and a Maximization (M) step. The M step involves the maximization of a likelihood function that is redefined in each iteration by the E step. The results are the final values of parameters μ (mean vector), V (covariance matrix) and p (weight/ coefficient of mixture) of each group. The steps of the Expectation – Maximization algorithm are as follows:

1. Determine the number of components of the mixture by using the fuzzy cluster-mean algorithm.
2. Determine initial values of parameters of each component $(\hat{\mu}^{(0)},\hat{V}^{(0)},\hat{p}^{(0)})$.
3. Calculate the posterior probability for each observation (Expectation-steep) as shown in the following equations:

$$\hat{\tau}_{ij}=\frac{\hat{p}_i\phi(x_j;\hat{\mu}_i,\hat{V}_i)}{\hat{f}(x_j)} \quad (6)$$

$$\hat{f}(x_j)=\sum_{g=1}^{G}\hat{p}_g\phi(x_j;\hat{\mu}_g,\hat{V}_g) \quad (7)$$

Where $\hat{\tau}_{ij}$ represents the posterior probability of x_j corresponding to the i term, $\phi(x_j;\hat{\mu}_i,\hat{V}_i)$ is the normal multivariate density and $\hat{f}(x_j)$ corresponds to the estimated mixture of distributions for the i terms evaluated in x_j and j is an index which indicates the total amount of data.

4. Update, $\hat{\mu}$, $\hat{V}$, $\hat{p}$ of each component (maximization-step) by using equations (8) – (10). $\hat{p}_i,\hat{\mu}_i,\hat{V}_i$ are the updated estimations.

$$\hat{p}_i=\frac{1}{n}\sum_{j=1}^{n}\hat{\tau}_{ij} \quad (8)$$

$$\hat{\mu}_i=\frac{1}{n}\sum_{j=1}^{n}\frac{\hat{\tau}_{ij}x_j}{\hat{p}_i} \quad (9)$$

$$\hat{V}_i=\frac{1}{n}\sum_{j=1}^{n}\frac{\hat{\tau}_{ij}(x_j-\hat{\mu}_i)(x_j-\hat{\mu}_i)^T}{\hat{p}_i} \quad (10)$$

5. Repeat steps 3 and 4, until desired convergence is obtained.

Subsequently, the organization of groups in classes associated to faults is based in the probability of appearance in each group as given by the mixture model in the following equation:

$$\hat{f}_{FM}(x)=\sum_{g=1}^{G}\hat{p}_g\phi_g(x;\hat{\mu}_g,\hat{V}_g) \quad (11)$$

Where $f_{FM}(x)$ corresponds to mixture model of sample (x) which corresponds to random sample of n observations of dimension d .

The fault location has been done using this expectation – maximization algorithm on the clusters obtained after using FCM. It is to be mentioned that the result obtained are shown for classification, and zone identification

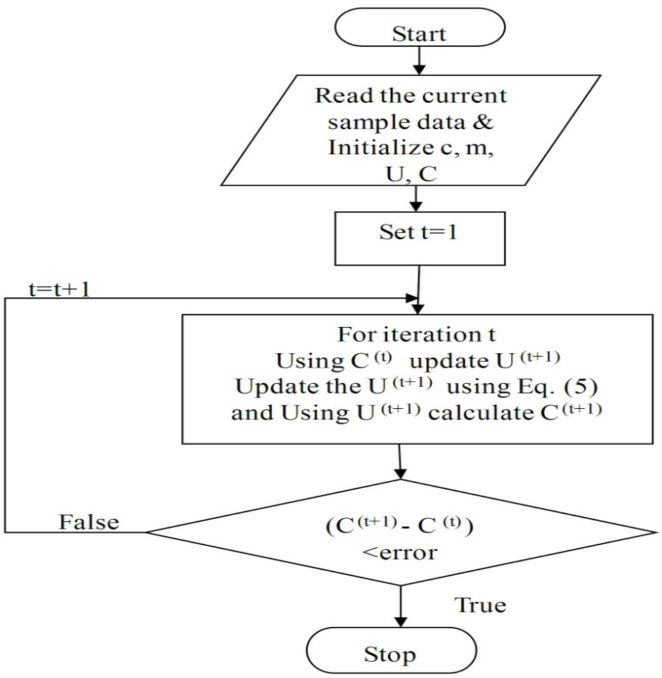

Figure 3: Flowchart for FCM algorithm

V. Results And Discussion

With the statistical model presented in the previous section, fault location is done as per the response available. Current waveforms recorded at the substation after extensive simulation in PSCAD acts as the input. The root mean square (rms) values of the current samples have been used. The approach is aimed to obtain a low economical cost due to constraints in most of the distribution utilities. For obtaining the result, visual discrimination of the zones have been done in preliminary data analysis. A total of 720 current samples were collected from an IEEE 13-node feeder, which as per the IEEE is a very unbalanced system. The current samples were taken for two zones over a resistance ranging from (0 – 10) Ω at ten different locations. It is to be kept in mind that owing to its nature this feeder is neglected by the researchers. The authors are successful in getting the results. Tables represent the results shown for one category for all types of faults such as single line to ground, line to line and double line to ground fault respectively which are very promising.

The result of the proposed algorithm is presented here. Table 1 furnishes information about fault classification after being identified with 0 resistance value. 10 sample each from phase a, b and c from zone 1 and zone 2 respectively are mixed together. This mixture of samples is fed given as input to FCM to obtain cluster centers. Using these cluster centre EM algorithm is applied to yield 100% classification result.

TABLE I. Fault Classification

Fault Type	Phase (A)	%	Phase (B)	%	Phase (C)	%	Sub -total %
Single -line to ground	20/20	100	20/20	100	20/20	100	100
	Phase (AB)	%	Phase (BC)	%	Phase (AC)	%	
Line to line	20/20	100	20/20	100	20/20	100	100
Double line to ground	20/20	100	20/20	100	20/20	100	100
Total						**180/180**	**100**

TABLE II. Samples Identified Under Different Fault Resistance

Fault Type	Resistances						
	R1	%	R2	%	R3	%	Sub - total %
Single -line to ground	25/30	83.33	30/30	100	26/30	86.67	90.00
Line to line	27/30	90.00	20/30	66.67	26/30	86.67	81.11
Double line to ground	25/30	83.33	26/30	86.67	15/30	50.00	73.33
Total	**220/270**						**81.48**

TABLE III. Zone Identification

Fault Type	Zones				
	Z1	%	Z2	%	Sub -total %
Single -line to ground	10/10	100	10/10	100	100
Line to line	10/10	100	10/10	100	100
Double line to ground	08/10	80	07/10	70	75
Total				**55/60**	**91.67**

Table 2 give results of the samples obtained from one-phase over different resistances which are mixed and subjected to above procedure. Output yields 81.48% of the samples are recognized over the correct resistances. Table 3 gives information about different zones. These zones are the different nodes where the current samples have been obtained. Results are presented for one type of fault in all the different category of fault. Single- line to ground fault, line to line fault are exactly located where the fault has occurred. The total zone identification is 91.67%.

VI. CONCLUSION

Authors have been successful in locating faults in a standard IEEE 13- node distribution system using PSCAD. It is obvious from the literature that the feeder is an unbalanced and therefore there has been problem in locating faults since there exists large variation in the current magnitude. A valuable methodology has been presented. The approach is based on statistical modeling of the samples obtained after been clustered by the use of FCM. Results obtained after the clusters are administered to expectation – maximization algorithm for the given feeder is promising It should be kept in mind that the proposed algorithm has not been applied to current samples of distribution system as per the literature available till date. Also, IEEE

13- node has not been considered extensively as the reference system for collecting samples in order to locate faults.

REFERENCES

[1] *IEEE guide for Protective Relay Applications to Distribution Line,* IEEE Std. C37.230, 2007.

[2] Yuan Liao, "Fault location using unsynchronized voltage measurements during fault," Electric Power Components and Systems,vol. 12, pp. 1283–1293, 2006.

[3] R. Das, M. Sachdev, T. Sidhu, "A fault locator for radial sub transmission and distribution lines," Proceeding of IEEE Power Engineering Society Summer Meeting, Seattle,vol. 1, pp. 443- 448, 2000.

[4] J. Mora, J. Meléndez, et al, "An overview to fault location methods in distribution system based on single end measures of voltage and current", Proceeding of ICREPQ'04 – International Conference on Renewable Energy and Power Quality, Barcelona, April 2004.

[5] M. Mirzaei, M.Z. A Ab Kadir, E. Moazami, H. Hizam, A Review of Fault Location Methods for Distribution Power System, Australian Journal of Basic and Applied Sciences, vol. 3, pp. 2670 – 2676, 2009.

[6] T. Short, Electrical Power Distribution Handbook, CRC Press, 2003.

[7] J. Zhu, D. Lubkeman, A. Girgis, "Automated Fault location and diagnosis on electric power distribution feeders," IEEE Transactions on Industrial Applications, vol. 2, no. 2, pp. 801 – 809, 2002.

[8] J. Mora - Florez, et a, "Comparism of impedance based fault location methods for power distribution system," Electric Power System Research, vol. 78, no. 4, pp. 657 – 666, 2008.

[9] Pratul Arvind, Rudra Prakash Maheshwari, " A Gabor Filter Based Approach for Locating faults in Distribution System," Proceedings of IEEE International Conference on Control, Robotics and Cybernetics, vol. 2, pp. 309-313, March 2011.

[10] S. Herraiz, J. Meléndez, G. Ribugent,et al, "Application for Fault Location in Electrical Power Distribution Systems," Proceeding of 9th International Conference on Electrical Power Quality and Utilisation, 9 -11 October 2007.

[11] D. Thukaram, et al, "Artificial neural network and support vector machine approach for locating faults in radial distribution systems," IEEE Transaction on Power Delivery, vol. 2,no. 20, pp. 710–720, April 2005.

[12] U. D. Dwivedi, S. N. Singh, S. C. Srivastava, "A Wavelet Based Approach for Classification and Location of Faults in Distribution Systems", Proceeding of IEEE INDICOM, Kanpur, vol.2,pp.488-493, Dec 2008.

[13] J. Mora - Florez, et a, "k –mean algorithm for locating faults in power system," Electric Power System Research, vol. 79, no. 4, pp. 714 – 721, 2009.

[14] IEEE Distribution Planning Working Group Report, "Radial distribution test feeders," IEEE Transaction on Power System, Vol. 6, No. 3, pp. 975 – 985, August 1991.

[15] "PSCAD, User Guide," Manitoba Research Centre, Canada.

[16] Pratul Arvind, Rudra Prakash Maheshwari, "Simulation of IEEE 13-node feeder and Its wavelet decomposition," Proceedings of ICOPS'10 in International Journal of Emerging Technologies and Applications in Engineering Technology and Sciences (IJ-ETA-ETS), pp. 423 – 430, December, 2010.

[17] Dunn, J. C., A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Taylor & Francis Journal of. Cybernetics and Systems, 3 (3) (1974) 32–57.

[18] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, *Plenum Press*, 1981, New York.

[19] J. Hair, R. Anderson, R. Tatham, W. Black, Multivariable Data Analysis, Prentice Hall, Madrid, 1999.

[20] Haojun Sun, S. Wang and Q. Jiang, "FCM-Based Selection Algorithm for Determining the Number of Clusters," *Elsevier J. Pattern Recognition*, vol. 37, pp. 2027-2037, 2004

[21] M. Jordan, R. Jacobs, "Hierarchical mixture of experts and the EM algorithm, IEEE Transactions on Neural Computation, vol.6, pp. 181-214, 1994