8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.1, June 6-8, 2007, Cancún, Mexico

SCALING AND DISCRIMINATION ISSUES IN MONITORING, FAULT DETECTION AND DIAGNOSIS

Swanand R. Khare [*] **Vinay A. Bavdekar** [**]
Sachin C. Kadu [**,****] **Ketan P. Detroja** [***]
Ravindra D. Gudi [**,1]

[*] *Department of Electrical Engineering, IIT Bombay, Mumbai, India*
[**] *Department of Chemical Engineering, IIT Bombay, Mumbai, India*
[***] *ABB Corporate Research, Bangalore, India*
[****] *Reactor Projects Division, B.A.R.C.,Mumbai, India.*

Abstract: The performance of data based monitoring algorithms is crucially dependent on the ability to discriminate between patterns of normal and fault data. In this paper, we analyze discriminatory properties of PCA, FDA and nonlinear scaled version of PCA algorithm proposed by (Ding *et al.*, 2002). We demonstrate improved discriminatory performance of the nonlinearly scaled PCA over traditional algorithms like PCA and FDA. The scaling and discrimination issues have been analyzed for each of the above algorithms using normal and fault data generated from the bench-marked Tennessee Eastman (TE) problem. The TE problem is used to highlight the superiority of the nonlinear scaled PCA (SPCA) over PCA and FDA.

Keywords: PCA, FDA, scaled PCA, iterative scaled PCA, process monitoring, fault detection

1. INTRODUCTION

The operation of a process unit in safe zone is essential to avert any kind of loss in terms of productivity or economics. The broad scope of fault detection and diagnosis (FDD), which includes various malfunctions such as process degradation, parameter changes, failures of process units or sub-units, pose challenges for optimal operation of the plant. The ever increasing complexity of process units has made it difficult to simultaneously analyze online data and ensure that all the variables are being maintained at their safe limits (Venkatsubramanian *et al.*, 2003). It is, therefore, necessary to develop mathematical or statistical algorithms that can facilitate early detection and isolation of faults.

Of the various quantitative methods available for fault diagnosis, those based on historical plant data are preferred (Kresta *et al.*, 1991), as they require minimal prior knowledge of the plant. Multivariate statistical techniques, that fall under the class of data-based methods, are capable of reducing the dimensionality of the data and capturing the features in the reduced dimensions (Kresta *et al.*, 1991). Principal component analysis (PCA) is a standard unsupervised technique used for pattern classification. The coordinate transforma-

[1] Corresponding Author; Email: ravigudi@iitb.ac.in, Tel#: 91-22-2576-7204, Fax#: 91-22-2572-6895

tion is linear and spans the space of maximum variance, thus capturing maximum possible information in reduced dimensions. However, the linear nature of coordinate transformation limits the ability of PCA to classify patterns generated by nonlinear processes. Fisher discriminant analysis (FDA) tries to address this shortcoming of PCA by incorporating a priori knowledge of all the faults that are present in the data. Although the coordinate transformation is still linear, the use of knowledge about number of faults enhances the discrimination between various clusters. However, as is the case with all supervised methods, the implementation of FDA requires prior classification or labeling of samples to clusters, a scenario which is relatively difficult to achieve. Moreover, classification of novel faults is also an important issue in practice.

There are several attributes that an algorithm needs to have, for efficient FDD. Firstly, the large number of process variables may pose a problem of information overload and hamper the real time diagnosis of plant operations. Thus, one of the primary requirements of any FDD algorithm is to be able to extract maximum possible information into a significantly lower dimensional space. Secondly, the various online measurements bring in useful signatures related to the plant operation. Typically, normal operations and various types of faults generate different signatures of the process variables (Detroja *et al.*, 2006). These signatures can be classified into clusters that represent their modes of operations, e.g. normal or fault (Detroja *et al.*, 2006). Thus, another desirable feature in an FDD algorithm is the ability to discriminate between various patterns present in the data and classify them accordingly. Conventionally, FDD algorithms such as PCA use statistical methods that are based on normal data alone. However, data collected during faults can also help in the discrimination/classification task. Towards this end, supervisory training algorithms such as FDA (He *et al.*, 2005), have been used to include information that the fault situations bring. As mentioned earlier, these supervisory methods however, require additional effort related to labeling the samples during the training/model building step. Thus, another desirable feature in the FDD algorithm is to have an inherent discriminative ability that can classify the data for different modes of operation. Another important aspect related to the data sets of normal and fault modes is the relative size of the data sets. Fault data are relatively smaller in size when compared with normal operating data and therefore the latter significantly influence the model characteristics in algorithms such as PCA and FDA. Thus towards enhanced discrimination, the FDD algorithm should also accommodate unbalanced data sets of different sizes. To summarize, an FDD algorithm should be able to– a) achieve dimensionality reduction, i.e. extract maximum information into a lower dimensional space, b) have inherent discriminative ability to distinguish between various signatures, like normal data, actuator failure, sensor failure, etc. c) accommodate unbalanced data sets, that is, smaller data sets for faults relative to normal data set.

This paper addresses the above requirements in data-based monitoring algorithms. We specifically analyze an alternate nonlinear scaling of the PCA algorithm as proposed by (Ding *et al.*, 2002) towards addressing the above problems. (Ding *et al.*, 2002) introduce a nonlinear scaling to PCA that leads to self-aggregation of the data points into distinct clusters. We show that this aspect enhances the inherent discrimination ability of PCA and aids the FDD task. Another feature of the nonlinear scaled PCA (SPCA) is its ability to classify unbalanced clusters. Unbalanced clusters occur when there is a substantial difference in the number of data points of various clusters. As mentioned earlier, it is common to have a large data set for normal operating conditions, but the length of the data set for a fault condition is much less. This causes the large sized cluster to overshadow the small sized clusters, thus making it difficult to distinguish between various patterns. We show that for the FDD task, use of SPCA provides enhanced discrimination even in the presence of unbalanced clusters. The Tennessee Eastman (TE) problem (Downs and Vogel, 1993), that has become a benchmark problem for various FDD related studies, has been chosen to analyze the above mentioned algorithms and to highlight the various features of the same. Two measures that quantify the performance of the algorithms are introduced, viz, – a) Extent of Separation (EoS)–the ability of an algorithm to discriminate between various classes, i.e. the ability to minimize overlaps and b) Extent of Aggregation (EoA)– the ability of the algorithm to aggregate various data points of a particular cluster and make the cluster as compact as possible. We analyze the performance of PCA, FDA and SPCA using the above metrics and show the superiority of SPCA over other algorithms using data for the TE problem.

The paper is organized as follows– Section 2 gives a brief introduction of existing monitoring algorithms. Section 3 presents the results of the case-study on the TE problem and discusses the performance of the monitoring techniques. Finally, the paper is concluded by a critical evaluation of the obtained results.

2. OVERVIEW OF PCA, FDA AND SPCA

Let $X_{n\times m}$ denote the matrix of training data of n samples and m process variables. This data matrix contains lot of useful information about plant and can be used for further analyses. This matrix can be decomposed or factored such that it spans lower dimensional space, thus facilitating easier analysis.

2.1 PCA

PCA is a dimensionality reduction technique in which the data are projected into lower dimensional space that characterizes state of the process. Under the assumption of statistical independence of observed samples, PCA decomposes X to reveal the relationships among the process variables. The objective function for PCA involves maximization of variance explained in X by each direction. For instance, the first direction is obtained as a solution of following optimization problem in the linear space $t_1 = Xp_1$,

$$\max_{p_1}(t_1^T t_1) = p_1^T X^T X p_1 \tag{1}$$

such that $p_1^T p_1 = 1$. It is shown that solution of above problem can be posed as an eigenvalue eigenvector problem as

$$X^T X p_1 = \lambda p_1. \tag{2}$$

Thus, PCA decomposes the variance of X into scores matrix T and loading matrix P as

$$X = TP^T. \tag{3}$$

Orthogonality of the loading vectors p_i ensures that the above decomposition is unique. In order to capture the variation in data and to minimize noise, the loading vectors corresponding to first k largest eigenvalues are retained. Thus, process monitoring is done in a lower dimensional space. Note that,

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E, \tag{4}$$

where, E contains noise information which can not be captured in the first k components. Generally, choice of k is done using scree plot of eigenvalue numbers versus eigenvalues.

One of the major shortcomings of PCA is that it has limited discriminative ability. For example, clusters of similarly behaving samples are plotted together on scores plot which may cause overlap between the clusters. The objective function for PCA does not bring out possibility of discrimination. FDA is more suited for discrimination purpose, as its objective function explicitly incorporates discrimination.

2.2 FDA

FDA takes into account information between the classes as well as information within the classes. Given h distinct classes in the data points, FDA determines maximum of $h-1$ discriminating directions such that scatter between different classes is maximized and scatter within each class is minimized. It is to be noted that, in FDA, number of classes in data needs to be known a priori.

Let n denote number of samples, m number of measurement variables, h number of classes, n_j number of samples in class j and x_i, i^{th} row of X. Now various quantities are defined to understand objective function of FDA. The total scatter matrix S_t is defined as

$$S_t = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \tag{5}$$

where $\bar{x}$ is total mean vector

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{6}$$

Let $\mathcal{X}_j$ be the data in class j. Within scatter matrix for class j, S_j, is defined as,

$$S_j = \sum_{x_i \in \mathcal{X}_j}(x_i - \bar{x}_j)(x_i - \bar{x}_j)^T \tag{7}$$

where, $\bar{x}_j$ is mean vector for class $\mathcal{X}_j$ defined as,

$$\bar{x}_j = \frac{1}{n_j}\sum_{x_i \in \mathcal{X}_j} x_i. \tag{8}$$

The within class scatter matrix, S_w, is

$$S_w = \sum_{j=1}^{m} S_j \tag{9}$$

and between class cluster, S_b, is

$$S_b = \sum_{j=1}^{h} n_j(x_i - \bar{x})(x_i - \bar{x})^T. \tag{10}$$

It can be shown that $S_t = S_w + S_b$. The objective function of FDA is posed as follows.

$$J(v) = \max_{v\neq 0}\frac{v^T S_b v}{v^T S_w v}. \tag{11}$$

One can show that a vector v that maximizes (11) satisfies eigenvalue eigenvector problem,

$$S_w^{-1} S_b v = \lambda v. \tag{12}$$

In equation (11), if $S_w = I$, then objective function of FDA reduces to that of PCA (Chiang

et al., 2001). As seen in the above discussion, FDA objective function explicitly incorporates discrimination in samples. An issue that arises at this point is: Would it be possible to obtain the same extent of discrimination as that of FDA, but without requiring supervisory training. SPCA attempts to answer this question.

2.3 SPCA

In SPCA (Ding *et al.*, 2002), a new nonlinear scaling that promotes *self-aggregation*, is introduced. Nonlinear scaling of principal components is done to project the data on lower dimensional space where samples self-aggregate into distinct clusters. A similarity metric quantifies association among data objects. A nonlinear scaling of the similarity matrix is introduced in SPCA. The similarity matrix $W = [w_{ij}]$ chosen here, measures correlation between process variables.

$$w_{ij} = \exp(r(i,j))/c \tag{13}$$

where, $r(i,j)$ is Pearson's correlation coefficient between variables X_i and X_j. Division by c is to ensure that w_{ij} are not too small. In case when the correlations are small then w_{ij} will be further smaller if not properly scaled by c. The scaling factor $D = (d_i)$ is a diagonal matrix defined as,

$$d_i = \sum_{j=1}^{m} w_{ij}. \tag{14}$$

We next define,

$$W = D^{1/2}(D^{-1/2}WD^{-1/2})D^{1/2}. \tag{15}$$

The standard PCA is then applied on scaled component $\hat{W} = D^{-1/2}WD^{-1/2}$, which yields,

$$W = D(\sum_{k} q_k \lambda_k q_k^T)D \tag{16}$$

where, $q_k = D^{-1/2}p_k$ are called the scaled principal components. These are obtained by solving following eigenvalue decomposition problem as

$$(D^{-1/2}WD^{-1/2})p = \lambda p. \tag{17}$$

Equation (17) achieves an important feature of discrimination as shown in (Ding *et al.*, 2002) which can be stated as follows: When there are no overlaps among k clusters in regular Euclidean space, the scaled k principal components $(q_1, q_2, \cdots, q_k) = Q_k$ get the same maximum eigenvalue equal to 1. In the SPCA space spanned by Q_k, all the objects within same cluster self-aggregate into a single point. However, when overlaps between different clusters are present, samples within same cluster tend closer to each other in SPCA space than in Euclidean space.

To sharpen the clusters, self-aggregation process can be repeated. This is done as follows. Only the first k terms are chosen from equation (15) and λ's are set to obey $\lambda_1 = \lambda_2 = \cdots \lambda_k = 1$. Then

$$S = D(\sum_{j=1}^{k} q_j \lambda_j q_j^T)D = DQQ^T D \tag{18}$$

where, $Q = [q_1, q_2, \cdots, q_k]$. This lower dimensional projection contains the essential cluster structure. Combining this cluster structure with original similarity matrix as $S^{(2)} = 0.5DQQ^T D + 0.5W$ and applying SPCA on $S^{(2)}$ leads to further aggregation. For further details on SPCA reader is directed to the work by (Ding *et al.*, 2002). The nonlinear scaling in SPCA enables aggregation of points belonging to the same cluster. The scaling also helps in reducing the intra-cluster distances, which further enhances segregation. In contrast, the discriminative ability of FDA is because of its supervisory nature. Therefore, based on some statistic, if the discriminative ability of SPCA is comparable to FDA, the former method would be preferred because it is an unsupervised algorithm.

3. RESULTS

The earlier sections briefly reviewed the data based methods used for discrimination of data. This section discusses the performance of the above methods for a benchmark case-study. Two scalar measures– a) extent of separation (EoS) which quantifies the extent of overlaps between clusters and b) extent of aggregation (EoA) which quantifies the closeness of data points in the same cluster are computed for comparing the performance of different methods. The scores matrices obtained for different clusters after projecting data in lower dimensional space are used to calculate EoS and EoA. The EoS and EoA are computed using following equations–

$$\text{EoS} = \frac{\text{trace}(S_b)}{\max(\text{eig}(S_b))} \quad \text{and} \tag{19}$$

$$\text{EoA} = \frac{\text{trace}(S_w)}{\max(\text{eig}(S_w))}, \tag{20}$$

where, S_b and S_w are between class and within class scatter matrices respectively. They are computed using the scores for different operating conditions. The trace of the matrix is representative of total scatter in dimensions (PCs) that are retained and it is normalized by the largest eigenvalue of respective matrix. Because of normalization, the values of EoS and EoA are always greater than or equal to one. EoS being extent of separation, higher the value of EoS better is the separation. It is clear from Eq. 19 that if EoS is

close to one, then all the information is captured by the first direction, and we get very poor separation. If EoS is large then there is significant separation in all the principal directions. On the other hand, EoA being a measure of aggregation, we want it to be closer to one. From Eq. 20 it is clear that if first PC explains almost all the aggregation of the data, then data are very compact and further helps in the separation of one cluster from another.

Data from the Tennessee Eastman (TE) process (Downs and Vogel, 1993) simulation were used as a case-study. The process has a total of 52 variables (41 measured + 11 manipulated) available for monitoring. 21 different types of faults (IDV-1 to IDV-21) are pre-programmed and can be introduced during the plant operation. In the case study, three classes of operation are considered– a) normal operation, b) fault IDV-1 and c) fault IDV-2. The data for all these were downloaded from http://brahms.scs.uiuc.edu.

Table 1 shows the performance of the above clustering algorithms when used to classify training data. From table 1 it is clear that as we move from PCA to SPCA and iterative SPCA, the EoS increases while the EoA decreases. It is therefore clear that the nonlinear scaling of principal components promotes self-aggregation. The aggregation can be improved by increasing the number of iterations in iterative SPCA. Table 2 summarizes the discriminative ability of these algorithms when the training data sets are unbalanced. It is evident from the table that SPCA and iterative SPCA perform better in case of unbalanced clusters. To generate these unbalanced clusters, only 25% of samples in fault clusters were considered in model building stage. This is an advantage of SPCA and iterative SPCA. The scores plots, using unbalanced data sets, for the above mentioned FDD techniques (Fig. 1 to Fig. 6) also verify the results in the tables. From the figures it is clear that iterative SPCA gives results that are comparable to FDA, but without requiring the labeling and a priori classification necessary for FDA.

Table 1. Table of EoS and EoA for test data in case of balanced clusters

	FDA	PCA	SPCA	ISPCA(1 itr)	ISPCA(2 itrs)	ISPCA(5 itrs)
EoA	1.463	1.719	1.578	1.461	1.391	1.255
EoS	1.291	1.044	1.015	1.020	1.025	1.041

Table 2. Table of EoS and EoA for test data in case of unbalanced clusters

	FDA	PCA	SPCA	ISPCA(1 itr)	ISPCA(2 itrs)	ISPCA(5 itrs)
EoA	1.765	1.662	1.554	1.516	1.446	1.254
EoS	1.001	1.020	1.024	1.040	1.048	1.051

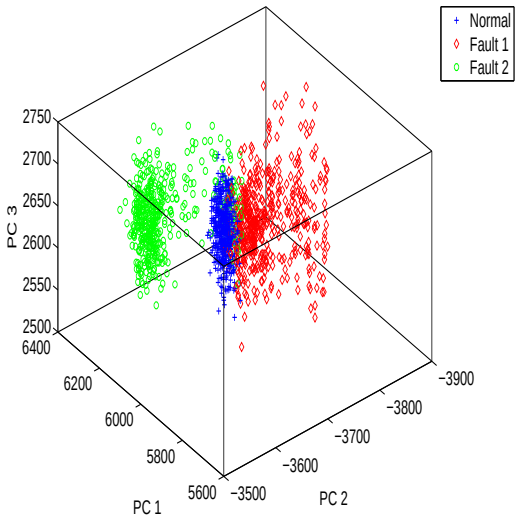

Fig. 1. Scores plot of PCA on TE data for unbalanced clusters

4. CONCLUSION

A nonlinear scaled version of PCA is shown in this paper to have enhanced discriminative ability relative to the traditional PCA algorithm. The discriminative ability of SPCA is comparable to that of FDA, despite the former being an unsupervised algorithm, while the latter is a supervised one. It has also been shown that SPCA has a better performance when the data sets are unbalanced. The superiority of SPCA has been demonstrated in our paper using the benchmark TE problem. Further issues related to the nonlinear scaling in terms of its ability to classify nonlinear relationships as well as its applicability for the dynamic case is an issue for the further exploration.

5. ACKNOWLEDGMENTS

The authors would like to acknowledge the interactions and discussions with fellow students in the course CL 650: Fault Detection and Diagnosis (Fall 2006) at the Department of Chemical Engineering, Indian Institute of Technology, Mumbai.

REFERENCES

Chiang, L.H., E. L. Russell and R.D. Braatz (2001). *Fault Detection and Diagnosis in Industrial Systems*. Springer Verlag. London.

Detroja, K. P., R.D. Gudi and S.C. Patwardhan (2006). A possibilistic clustering approach to novel fault detection and diagnosis. *J. Proc. Cont.* **16**, 1055–1073.

Ding, C., X. He, H. Za and H. Simon (2002). Unsupervised learning: Self-aggregation in scaled principal component space. *Proceedings of the 6th European Conference on Principles*

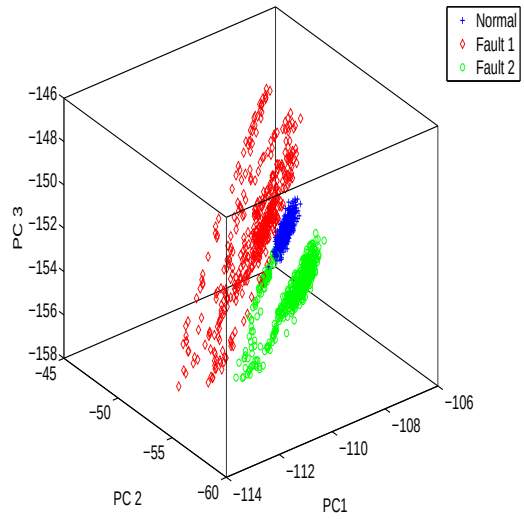

Fig. 2. Classification using scaled PCA

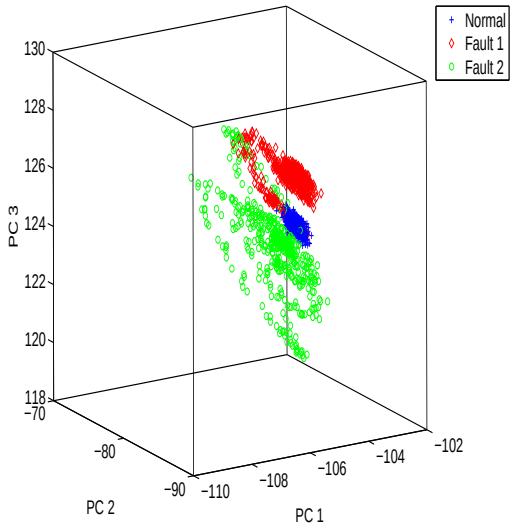

Fig. 3. Classification using iterative SPCA (1 update)

of Data Mining and Knowledge Discovery **2431**, 112–124.

Downs, J. J. and E.F. Vogel (1993). A plant wide industrial process control problem. *Comp. Chem. Engg.* **17(3)**, 245–255.

He, Q. P., S. J. Qinn and J. Wang (2005). A new fault diagnosis method using fault directions in fisher discriminant analysis. *AIChE J.* **51(2)**, 555–571.

Kresta, J.V., J. F. MacGregor and T.E. Marlin (1991). Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng.* **69**, 35–47.

Venkatsubramanian, V., R. Rengaswamy, S. N. Kavuri and K. Yin (2003). A review of process fault detection and diagnosis part i: Quantitative model based methods. *Comp. Chem. Engg.* **27**, 293–311.

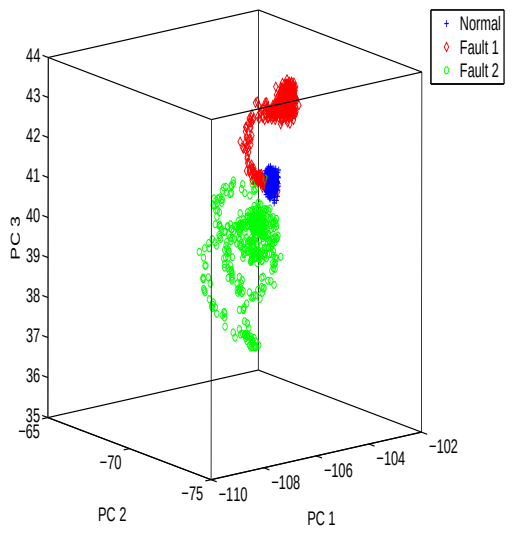

Fig. 4. Classification using iterative SPCA (2 updates)

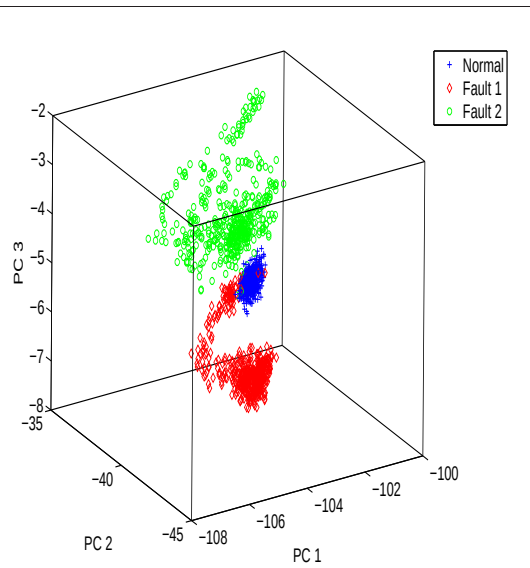

Fig. 5. Classification using iterative SPCA (5 updates)

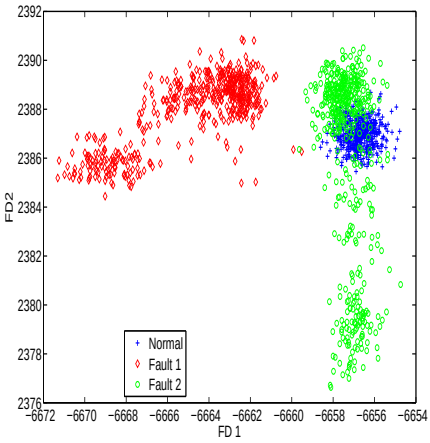

Fig. 6. Fisher directions plot on TE data