Proceedings of the 9th International Symposium on
Dynamics and Control of Process Systems (DYCOPS 2010),
Leuven, Belgium, July 5-7, 2010
Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, Ilse Smets (Eds.)

Pre-analysis of Multi-batch Bioprocesses Data with Finite Mixture Models in the Reduced Feature Subspace

Weilu Lin * **Elaine Martin** ** **Gary Montague** **

* *State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China (e-mail: wllin@ecust.edu.cn)*
** *School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne, NE1 7RU (e-mail: e.b.martin@ncl.ac.uk, gary.montegue@ncl.ac.uk)*

Abstract: Multi-batch bioprocesses data, unlike the data from other industries, are highly correlated due to the operation characteristics of the industry. In this work, pairwise Fisher discriminant analysis (FDA) is successfully utilized to reveal the similarity between two batches. In order to handle the mixture pattern for the data projected into the reduced feature subspace represented by the first several generalized eigenvectors, the finite Gaussian mixture model is adopted here to calculate the confidence region of each mixture. There are several challenges facing application engineers when estimate finite mixture models (FMMs), such as initialization of the expectation-maximization (EM) algorithm and determination of number of mixtures. In this work, an initialization method based on the uniform prior distribution assumption and a new method to determine the number of components of FMMs based on estimated density histogram are proposed. The utility of the proposed method has been demonstrated in simulation studies. Combined with the pairwise FDA, the method has been successfully applied to a large scale multi-batch bioprocess data set.

Keywords: finite mixture models, Gaussian mixtures, Fisher discriminant analysis, histogram density

1. INTRODUCTION

One features of the day-to-day operation in the biopharmaceutical industry is that the cell culture goes through several passages before being feeded into the production bioreactor. Furthermore, at each passage, the cell is cultivated under batch or fed batch operation mode undergoing different or same operation durations. The main purpose of this work is to apply multivariate statistical analysis methods to multi-batch bioprocess data, at present the work is focusing on batch processes operating at the same scale with different bioreactors. Facing with the large scale industrial data set at different operational scales and from different bioreactors, the pre-analysis of the data is an important step to understand the statistical correlation of the process. Fisher discriminant analysis (FDA), as one of pattern recognition methods, has been applied to the process data analysis (Chiang et al., 2001). When apply FDA to industrial data sets, after projecting the benchmark data into the reduced feature subspace some data sets shows multi-mixture characteristics which may correspond to the multiple operation regions. The objective of this paper is to address the multi-mixture pattern using finite mixture models (FMMs) (McLachlan and Peel, 2000). To deal with the algorithmic difficulty coming with FMMs, in this work, we address the initialization of the algorithm and the determination of number of unknown components with proposed approaches. Furthermore, the similarity between comparative and benchmark batches in the featured space is defined. As shown in the following sections, the proposed approach works well in simulation studies and is successfully applied to a large scale industrial multi-batch bioprocesses data set.

The rest of the paper is organized as follows: In section 2, the pairwise FDA is introduced as one of batch data pre-analysis methods. Section 3 introduces finite mixture models along with an improved expectation-maximization (EM) algorithm. Methods for pre-analysis of industrial data in the reduced featured subspace is introduced in Section 4. The simulation studies of the proposed FMMs algorithm are shown in Section 5. The utility of the method is demonstrated with a large scale multi-batch bioprocess in section 6. Section 7 ends the paper by conclusions.

2. FISHER DISCRIMINANT ANALYSIS METHOD

In multi-batch bioprocesses operation, after the collection of operation data, the first challenge faced the application engineer is to pre-analyze the data to explore the statistical structure of different batches, i.e., the batch-to-batch correlation information, before utilizing multi-way data analysis methods such as multi-way principle component analysis and multi-way partial least squares (Nomikos and MacGregor, 1995; Kourti and MacGregor,

1995). Discriminant analysis, as the method to explore embedded patterns from different data sets, can be applied to the data pre-analysis applications. In this research, aimed at exploring the batch-to-batch correlation, pairwise FDA is introduced. The objective of the pairwise discriminant analysis for the multi-batch processes can be stated as follows,

Given two data sets, $X_1 \in R^{N_1 \times d}$ and $X_2 \in R^{N_2 \times d}$, such that N_1 and N_2 stand for number of samples and d stands for number of variable, the objective of discriminant analysis is to obtain the matrix, $A \in R^{d \times k}$, such that $k \leq d$ and the projections of two data sets onto to the reduced feature space represented by the first k column vectors of A (loading directions of pair-wise FDA) maximize the difference between these two data sets.

Fisher discriminant analysis is a linear discriminant analysis methods, the pairwise FDA is to find optimal directions that maximally separate the two data sets. Mathematically, it can be stated as follows (Anderson, 1984),

$$S_b A = S_w A \Lambda \tag{1}$$

such that

$$S_b = \sum_{i=1}^{2} n_i (\bar{x_i} - \bar{x})(\bar{x_i} - \bar{x})^T \tag{2}$$

$$S_w = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - \bar{x_i})(x_{ij} - \bar{x_i})^T \tag{3}$$

are the sampled between groups covariance matrix, S_b, and the sampled within groups covariance matrix S_w, respectively. $\bar{x}$ is the overall mean of two data sets, $\bar{x}_i$ is the mean vector of the i^{th} data set, and x_{ij} is the j^{th} sample of the i^{th} variable. Equation (1) can be regarded as the symmetric general eigenvalue problem see (Golub and Loan, 1996) with $A \in R^{d \times k} (k \leq d)$ as the generalized eigenvectors and the diagonal matrix $\Lambda \in R^{k \times k}$ as the generalized eigenvalues. For each batch data, the projection to the first k FDA directions can be calculated as follows,

$$T_1^{FDA} = X_1 A \tag{4}$$

$$T_2^{FDA} = X_2 A \tag{5}$$

where X_1 and X_2 are benchmark data and comparative data, respectively. T_1^{FDA} and T_2^{FDA} are FDA score matrices, notice that the latter one can be calculated on-line. Compare to PCA, the FDA does not need to scale the data since it utilizes the with-in and between group sampled covariance matrices as indicated in equation (1).

3. FINITE MIXTURE MODELS

Finite mixture models (FMMs), as one of parametric probabilistic models, has been applied to processes modelling and monitoring recently (Ou and Martin, 2008; Yu and Qin, 2008). It assumes that process data is generated from a finite mixture model as follows,

$$p(\mathbf{x}|\theta) = \Sigma_{i=1}^{K} \alpha_i p_i(\mathbf{x}|\theta_i) \tag{6}$$

such that $\mathbf{x} \in R^d$ is a random variable generated from $p(\mathbf{x}|\theta)$, $p_i(\mathbf{x}|\theta_i)$ is the i^{th} mixture's probability function with parameter θ_i, α_i is the mixing probabilities that satisfy $\sum_{i=1}^{K} \alpha_i = 1$, K is the number of mixtures, θ is the parameters of FMMs composed of a_i, θ_i for $i = 1, \cdots, K$. The FMMs has the potential to handle processes undergoing multiple operation regions, which can be regarded as the tradeoff between linear and nonlinear modelling approaches. Due to the complex correlation and the curse of dimensions of the industrial data, the direct application of FMMs to process monitoring and modelling is difficult. However, when combined with dimension reduction method such as FDA or principle component analysis (PCA), it can catch the multi-mode nature of the process in the reduced subspace. In this work, the FMMs are adopted to handle that data with mixture characteristics in the reduced feature subspace represented by the first several loading directions of FDA or PCA.

There are several challenges facing application engineers when use FMMs (McLachlan and Peel, 2000; Figueiredo and Jain, 2002). Firstly, as the FMMs are traditionally obtained by the expectation-maximization (EM) algorithm which is an iterative algorithm highly depended on the initial value. Furthermore, the choose of number of mixtures for data with unknown number of mixtures is still an unsolved problem. Finally, the EM algorithm occasionally converges to the boundary of the parameter space, i.e., ends up with one or several $\alpha_i = 0$ along with almost singular covariance matrices. In this work, several methods are introduced as tentative approaches to resolve these problems.

3.1 Expectation-maximization algorithm

Given N samples of random variable, $\mathbf{x}$, denoted as $X \in R^{N \times d}$, the likelihood function can be defined as the joint density function of X.

$$L(X|\theta) = p(x^1, x^2, \cdots, x^N|\theta) \tag{7}$$

where x^i is the i^{th} sample of the X. If the samples are mutually independent, the *log* likelihood function can be regarded as the summation of the *log* finite mixture density as follows,

$$logL(X|\theta) = \sum_{i=1}^{N} log \sum_{j=1}^{K} \alpha_j p_j(x^i|\theta_j) \tag{8}$$

The objective of EM algorithm is to estimate, θ, and to specify number of components K, given the data matrix X and the initial guess of the FMMs' parameters. If one assumes that each sample, x^i, generated from only one mixture density, $p_m(\cdot|\theta_m)$, an auxiliary binary vector $z^i \in R^{K \times 1}$, can be introduced such that if x^i is generated from m^{th} mixture, $z_m^i = 1$, otherwise $z_m^i = 0$. After introducing z^i into equation (8), the log likelihood can be expressed as follows,

$$logL(X, Z|\theta) = \sum_{i=1}^{N} \sum_{j=1}^{K} z_j^i log[\alpha_j p_j(x^i|\theta_j)] \tag{9}$$

Based on this formulation, the EM algorithm is an iterative computation of maximum-likelihood estimates when the

observations can be viewed as incomplete data (Dempster et al., 1977).

On the other hand, instead of assuming that one sample comes from only one mixture density, given the estimated parameters of FMMs, $\hat{\theta}$ at the t^{th} iteration, one can calculate the posteriori probability of i^{th} sample generating from the m^{th} mixture, w_m^i, based on Bayes law as follows,

$$w_m^i = p(z_m^i = 1|x^i, \hat{\theta}(t)) \tag{10}$$

$$= \frac{p(x^i, z_m^i = 1, \hat{\theta}(t))}{p(x^i, \hat{\theta}(t))} \tag{11}$$

$$= \frac{\alpha_m p_m[x^i|\hat{\theta}_m(t)]}{\sum_j^K \alpha_j p_j[x^i|\hat{\theta}_j(t)]} \tag{12}$$

After introducing w_m^i into equation (8), the log likelihood can be obtained as follows,

$$logL(X, W|\hat{\theta}(t)) = \sum_{j=1}^{K}\sum_{i=1}^{N} w_j^i log[\alpha_j p_j(x^i|\hat{\theta}_j(t))] \tag{13}$$

$$- \sum_{j=1}^{K}\sum_{i=1}^{N} w_j^i log w_j^i \tag{14}$$

The calculation of posteriori probability, equation (12), is the *expectation* step of the algorithm, while the estimation of α and θ at the $(t+1)^{th}$ iteration is the *maximization* step of the EM algorithm, which can be expressed as follows,

$$\alpha_m(t+1) = \frac{\sum_{i=1}^N w_m^i}{N} \quad for\ m = 1, \cdots, K. \tag{15}$$

$$\hat{\theta}(t+1) = \arg\max_{\theta}[logL(X, W|\hat{\theta}(t))] \tag{16}$$

where equation (16) is the maximum-likelihood (ML) estimator of θ. When the m^{th} mixture is the Gaussian probability density distribution as follows,

$$p(\mathbf{x}|\theta_m) = \frac{(2\pi)^{\frac{d}{2}}}{\sqrt{|\Sigma_m|}} exp\{-\frac{1}{2}(\mathbf{x} - \mu_m)^T \Sigma_m^{-1}(\mathbf{x} - \mu_m)\} \tag{17}$$

the ML estimator of θ at the $(t+1)^{th}$ iteration can be calculated as follows (Harville, 1997),

$$\hat{\mu}_m(t+1) = (\sum_{i=1}^{N} w_m^i)^{-1} \sum_{i=1}^{N} x^i w_m^i$$

$$\hat{\Sigma}_m(t+1) = \frac{\sum_{i=1}^N (x^i - \hat{\mu}_m(t+1))(x^i - \hat{\mu}_m(t+1))^T w_m^i}{(\sum_{i=1}^N w_m^i)^{-1}}$$

for $m = 1, \cdots, k$.

3.2 Initialization of EM algorithm

As an iterative algorithm, the EM method has the initialization problem, i.e., it could end up the local optimal for the inappropriate initial guess of the parameters. Figueiredo and Jain (Figueiredo and Jain, 2002) discuss the issue and propose the method for initial values for the covariance matrices of FMMs with Gaussian mixtures. In this work, we adopt their approach for the initialization of covariance matrices while use histogram to choose the initial values of the mean vectors. For the Finite Gaussian mixture models, the approach is very intuitive: the initial values of mean vectors should be in the neighborhood of high histogram density region of the data if the data have mixture characteristics. Given N samples of random variable, $\mathbf{x}$, denoted as $X \in R^{N\times d}$ and the hyper-rectangle B_k of the size $h_1 \times h_2 \times \cdots \times h_d$ in R^d where h_k is calculated with equation (24), the histogram probability density can be defined as

$$\hat{p}(\mathbf{x}) = \frac{\nu_k}{N h_1 h_2 \cdots h_d} \quad for \ \ \mathbf{x} \in B_k \ where \ \sum_k \nu_k = N$$

for a generic hyper-rectangular bin, B_k, contains ν_k points.

The initial values of the mean vectors of the EM algorithm are chosen as center points of the bins with higher than uniform distribution histograms as follows,

$$\hat{p}(\mathbf{x}) \geq \frac{N}{\text{Total number of bins}} \tag{18}$$

The initial values of α_m are calculated as

$$\alpha_m = \frac{1}{\text{Number of initial mixtures}}. \tag{19}$$

3.3 Determination of Number of Mixtures

Another challenge facing the engineer when apply FMMs is how to determine the number of mixtures. In recent years, several criterions have bee proposed which include minimum massage length (MMLC) (Figueiredo and Jain, 2002), informational complexity (ICOMP) (Bozdogan, 1993), normalized entropy (NEC) (Biernacki et al., 1999), Laplace-empirical (LEC) criterions (McLachlan and Peel, 2000) along with the traditional Bayesian inference (Schwart, 1978) (BIC) and Akaike's information (Akaike, 1974) (AIC) criterions. The traditional AIC criterion can be formulated as follows,

$$AIC(\hat{\theta}) = logL(X, W|\hat{\theta}) + 2 \times k_m \tag{20}$$

where k_m is the number of independently adjustable parameters of FMMs, i.e., in Gaussian finite mixture case is

$$k_m = K[\frac{d(d+3)}{2} + 1] - 1 \tag{21}$$

where K is the number of mixtures, d is the dimension of random variable, $\mathbf{x}$.

In this work, we propose a new criterion based on the difference between the histogram of the over-fitted finite Gaussian mixture models and the histogram of the data. The histogram difference criterion (HDC) assumes that the finite Gaussian mixture with unknown number of component can be approximated with over-fitting of the data (Miloslavsky and van der Laan, 1978), which is quite intuitive given that the over-fitted model approximates the probability density function well enough. Mathematically, the HDC can be formulated as follows,

$$HDC(\hat{\theta}) = \|vec(M_{HDC})\| \tag{22}$$

where $vec(\cdot)$ is the vectorization operator, $\|\cdot\|$ is the two-norms operator, M_{HDC} is a multi-dimension array defined

as the difference of histogram density at pre-defined bins, B_k, as

$$M_{HDC} = \frac{1}{N_c}\hat{p}_2(\mathbf{x}) - \hat{p}_1(\mathbf{x}) \tag{23}$$

where $\hat{p}_1(\mathbf{x})$ is the histogram calculated based on data, $\hat{p}_2(\mathbf{x})$ is the histogram calculated based on N_c times Monte Carlo simulated data generated from estimated FMMs with different number of components. As will show in the simulation studies, the HDC performs better than AIC when data have complicate mixture patterns.

3.4 Proposed algorithm to estimate FMMs

At the first step, the histogram of the data in the reduced feature space is generated. The size of bins is calculated based on the multivariate normal distribution with diagonal covariation matrix as follows (Scott, 1992),

$$h_k = 3.5\hat{\sigma}_k N^{1/(2+d)} \text{ for } k = 1, \cdots, d \tag{24}$$

where $\hat{\sigma}_k$ is the estimated standard deviation of k^{th} variable. The initial mean vectors are chosen based on Eq. 18, the initial mixing probabilities of FMMs are calculated based on Equation (19). Assume initial diagonal covariance matrix, the k^{th} diagonal term is calculated as follows (Figueiredo and Jain, 2002),

$$\hat{\sigma}_k = \frac{1}{10d} trace(S_x) \text{ for } k = 1, \cdots, d \tag{25}$$

where S_x is the sampled covariance matrix of the data. After obtaining the initial guess of FMMs parameters, the EM algorithm is adopted to iteratively estimated the parameters until they converge. Based on the estimation of the parameters of FMMs with k_m components and the definition of local density for each mixtures at certain confidence level as,

$$\hat{\rho}_k = \frac{\text{Number of samples inside the } k^{th} \text{ ellipse}}{\text{Area of the } k^{th} \text{ ellipse}} \tag{26}$$

for $k = 1, \cdots, m$. one can delete one mixture at one time. After taking account of local density of the k^{th} mixture through a appropriate parameter, β, one avoids deleting mixtures with high local density and small mixing probabilities (i.e. one deletes the mixture with the minimal α_m as defined in equation (15) such that $\hat{\rho}_k > \beta$) as shown in the following section.

After obtaining the estimated parameters with the number of mixtures from k_{max} to k_{min}, one can plot the HDC for $m = k_{min}, \cdots, k_{max}$ and find the appropriate order of FMMs when HDC starts to increase.

4. PRE-ANALYSIS PATTERNS IN THE FEATURED SUBSPACE

Combined with PCA, FMMs have been successfully utilized to monitor the industrial processes recently (Choi et al., 2004; Thissen et al., 2005). It has been shown that when benchmark data representing normal operation condition (NOC) have mixture characteristics in the first two PCs subspace, the confidence density contour obtained by FMMs is more accurate. In this work, as shown in the industrial case study (Lin, 2008), when apply FDA to multi-batch bioprocess data some batches show mixture patterns in the first several discriminant directions. The pre-analysis procedure in this paper is as follows: Firstly, FDA is applied to discriminate the difference between the benchmark batch and the comparative batch. Data from both batches are projected onto the first several loading directions of FDA. Secondly, the pattern projected onto one pair of FDA directions, for example, the first two FDA directions, is captured by FMMs. The similarity between the comparative and benchmark data is defined as follows,

$$S_{bc} = \frac{N_1}{N_{total}} \tag{27}$$

where N_1 is the number of samples in comparative batch falls into at least one confidence contour defined by the FMMs model estimated with the benchmark batch, and N_{total} is the total number of samples in comparative batch. If certain sample in the comparative batch falls into at least one Gaussian mixture contour of the FMMs estimated from the benchmark data, one can conclude that the sample falls into the confidence density contour of the benchmark batch. Follows the same procedure, one can obtain the similarity of comparative batch using equation (27). For multi-batch bioprocess data, we adopt the pairwise FDA and calculate the similarity for each pair of batches as the measurement of statistical correlation of two batches (Lin, 2008).

5. SIMULATION CASES STUDY

The simulation example is a four-component finite Gaussian mixture model (Figueiredo and Jain, 2002) with different mixing probabilities as $\alpha_1 = \alpha_2 = \alpha_3 = 0.3$ and $\alpha_4 = 0.1$, respectively. The mean vectors of it are $\mu_1 = \mu_2 = [-4, -4]^T$, $\mu_3 = [2, 2]^T$ and $\mu_3 = [-1, -6]^T$, while the covariance matrices are:

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \Sigma_2 = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}; \Sigma_4 = \begin{bmatrix} 0.125 & 0 \\ 0 & 0.125 \end{bmatrix}$$

respectively. As we can see in Fig. 1 that for this simulation example there are mixtures with the same mean but different covariance matrices further more there is one mixture inside another mixture. 900 samples are generated from this example, the hesitation results are shown in Fig. 1. The *AIC* and *HDC* indices are shown in Fig. 2 and Fig. 3, respectively. The results shown that the EM algorithm sequentially deletes one mixture at one time. HDC correctly indicates four component FMMs, while AIC indicates three component FMMs which is a underfitting model of the example. The result demonstrates the capability of HDC under the complicate mixture distribution situation and indicates that it may be more suitable for real data sets.

6. INDUSTRIAL CASE STUDY

In this section, two batches of fermentation data from two different bioreactors are analyzed. There are ten measurements that include desolved oxygen tension (DOT), pH,

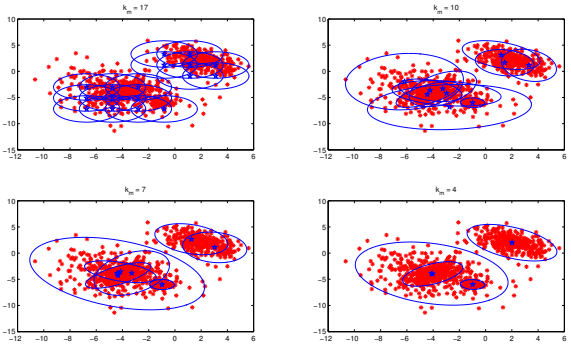

Fig. 1. The result of the estimation of the four-component Gaussian mixture model with the EM algorithm ($\beta = 1.5$).

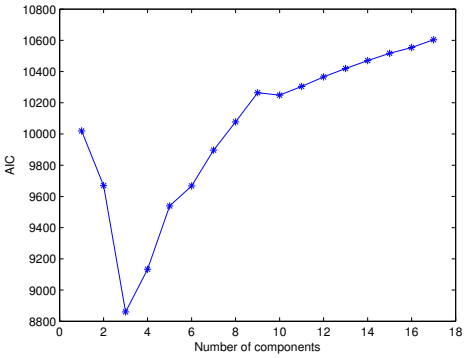

Fig. 2. The Akaike information criterion indices for the four-component Gaussian mixture model with different numbers of components.

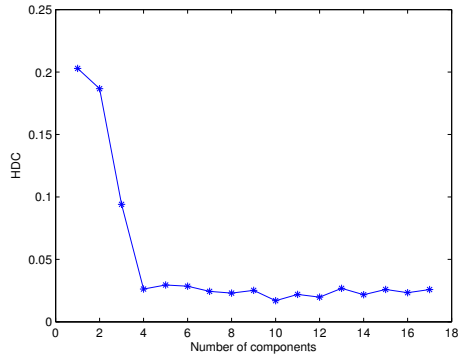

Fig. 3. The difference histogram criterion indices for the four-component Gaussian mixture model with different numbers of components.

temperature, air flow rate, *etc.*. The data are not shown here for proprietary reason. The data are analyze with proposed method for demonstration purpose. These two batches are not scaled before the application of pair-wise FDA. The data are projected into the feature subspace defined by the first two FDA directions before the estimation of FMMs.

The AIC and HDC indices are shown in Fig. 5 and Fig. 6, respectively. While AIC indicates three mixtures, HDC

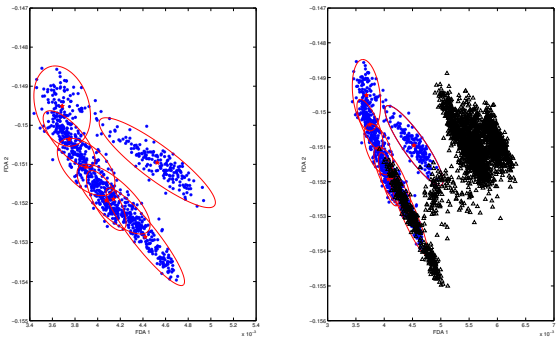

Fig. 4. The FMMs estimation result for two batches data come from different bioreactors ($\beta = 2$).

indicates a 7-component FMM which is adopted here. The projected data in the reduced feature subspace are shown in Fig. 4, the figure on the left hand side is the estimated FMMs with 95% confidence contour, while the figure on the right hand side shows the data from the comparative batch in the featured subspace defined by the benchmark data. As one can see, the comparative batch partly falls into the confidence region of the benchmark data with $S_{bc} = 0.12$.

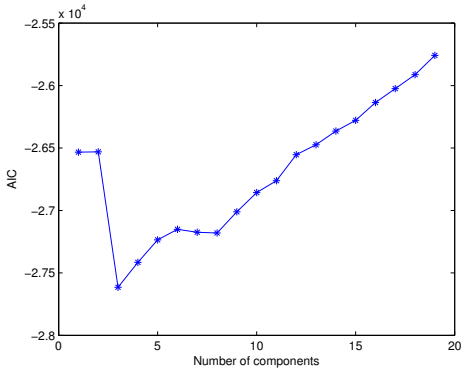

Fig. 5. The Akaike information criterion indices for two fermentation batches from different bioreactors case at different numbers of components of FMMs.

7. CONCLUSIONS

In this work, the pair-wised Fisher discriminant analysis is utilized to pre-analyze the multi-batch data from the pharmaceutical industry. Patterns in the reduced feature subspace defined by the first several FDA directions are captured by finite Gaussian mixture models. The high density regions of the data are chosen as the initial mean vectors of the FMMs. The histogram density difference of the sampled data and of estimated FMMs is adopted here as a criterion to choose the number of components of the FMMs. The simulation and industrial studies demonstrate the utility of the proposed method for estimating FMMs, the histogram difference criterion (HDC) shows advantage over Akaike information criterion (AIC).

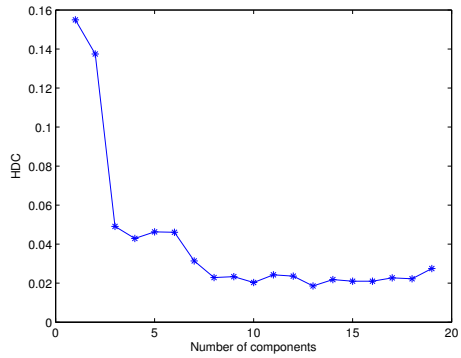

Fig. 6. The difference histogram criterion indices for two fermentation batches from different bioreactors case at different numbers of components of FMMs.

ACKNOWLEDGEMENT

The authors acknowledge the GlaxoSmithKline Co. for providing the industrial data.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Auto. Cont.*, 19, 716–723.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Inc., second edition.

Biernacki, C., Celeux, G., and Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, 20, 267–272.

Bozdogan, H. (1993). Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the inverse-Fisher information matrix. In O.Opiz, B.Lausen, and R.Klar (eds.), *Information and Classification*, 40–54.

Chiang, L., Russell, E., and Braatz, R. (2001). *Fault detection and diagnosis in industrial systems*. Springer-Verlag, Inc.

Choi, S.W., Park, J.H., and Lee, I. (2004). Process monitoring using a gaussian mixture model via principal component analysis and discriminant analysis. *Computer Chem. Eng.*, 28, 1377–1387.

Dempster, A., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39, 1–38.

Figueiredo, M. and Jain, A.K. (2002). Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24, 381–396.

Golub, G. and Loan, C.V. (1996). *Matrix Computations*. Johns Hopkins, third edition.

Harville, D. (1997). *Matrix algebra from a statistician's perspective*. Springer, Inc.

Kourti, T. and MacGregor, J.F. (1995). Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Lab Systems*, 28, 3–21.

Lin, W. (2008). Highlights of multivariate statistical analysis result of data set from GlaxoSmithKline. In *SysMA Technique Report*. Newcastle University, UK.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. John Wiley and Sons, Inc.

Miloslavsky, M. and van der Laan, M.J. (1978). Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Annals of Statistics*, 6, 461–464.

Nomikos, P. and MacGregor, J.F. (1995). Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Lab Systems*, 30, 97–108.

Ou, X. and Martin, E. (2008). Batch process modelling with mixtures of Gaussian processes. *Neural Compuat Application*, 17, 471–479.

Schwart, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Scott, D. (1992). *Multivariate density estimation: Theory, practice, and visualization*. John Wiley and Sons, Inc.

Thissen, U., Swierenga, H., de Weijer, A.P., Wehrens, R., Melssen, W.J., and Buydens, L.M.C. (2005). Multivariate statistical processes control using mixture modelling. *J. of Chemometrics*, 19, 23–31.

Yu, J. and Qin, S. (2008). Multimode process monitoring with bayesian inference-based finite gaussian mixture models. *AIChE J.*, 54, 1811–1829.