ELSEVIER

Chemometrics and Intelligent Laboratory Systems 50 (2000) 243–252

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis

Leo H. Chiang, Evan L. Russell, Richard D. Braatz *

Large Scale Systems Research Laboratory, Department of Chemical Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Box C-3, Urbana, IL 61801, USA

Accepted 1 November 1999

Abstract

Principal component analysis (PCA) is the most commonly used dimensionality reduction technique for detecting and diagnosing faults in chemical processes. Although PCA contains certain optimality properties in terms of fault detection, and has been widely applied for fault diagnosis, it is not best suited for fault diagnosis. Discriminant partial least squares (DPLS) has been shown to improve fault diagnosis for small-scale classification problems as compared with PCA. Fisher's discriminant analysis (FDA) has advantages from a theoretical point of view. In this paper, we develop an information criterion that automatically determines the order of the dimensionality reduction for FDA and DPLS, and show that FDA and DPLS are more proficient than PCA for diagnosing faults, both theoretically and by applying these techniques to simulated data collected from the Tennessee Eastman chemical plant simulator.

Keywords: Fault diagnosis; Process monitoring; Pattern classification; Discriminant analysis; Chemometric methods; Fault detection; Large scale systems; Multivariate statistics; Dimensionality reduction; Principal component analysis; Discriminant partial least squares; Fisher's discriminant analysis

1. Introduction

Large amounts of data are collected in many chemical processes. The data can be analyzed to determine whether or not a fault has occurred in the process, where a *fault* is defined as abnormal process behavior whether associated with equipment failure, equipment wear, or extreme process disturbances. This task of determining whether a fault has occurred is called *fault detection*, whereas *fault diagnosis* is the task of determining which fault has occurred. The proficiency of fault detection and diagnosis can be improved by using dimensionality reduction techniques such as principal component analysis (PCA), discriminant partial least squares (DPLS), and Fisher's discriminant analysis (FDA).

Chemical engineers in academia and industry have applied PCA for abstracting structure from multidimensional chemical process data [1]. PCA deter-

* Corresponding author. Tel.: +1-217-333-5073; fax: +1-217-333-5052.

E-mail address: braatz@uiuc.edu (R.D. Braatz).

0169-7439/00/$ - see front matter
PII: S0169-7439(99)00061-1

mines the most accurate lower dimensional representation of the data in terms of capturing the data directions that have the most variance. The resulting lower dimensional models have been used for detecting out-of-control status and for diagnosing disturbances leading to the abnormal process operation [2–7]. Several applications of PCA to real chemical data have been conducted at DuPont and other companies over the past 6 years, with much of the results published in conference proceedings and journal articles [8–11]. Several academics have performed similar studies based on data collected from computer simulations of processes [3–6,12–15].

FDA provides an optimal lower dimensional representation in terms of discriminating among classes of data [16,17], where for fault diagnosis, each class corresponds to data collected during a specific known fault. Although FDA has been heavily studied in the pattern classification literature and is only slightly more complex than PCA, its use for analyzing chemical process data is not described in the literature. This is interesting, since FDA should outperform PCA when the primary goal is to discriminate among faults. We suspect that part of the reason that FDA has been ignored in the chemical process control literature is that more chemical engineers read the statistics literature (where PCA is dominant) than the pattern classification literature (where FDA is dominant).

Discriminant Partial Least Squares (DPLS), also known as Discriminant Projection to Latent Structures, is a data decomposition method for maximizing covariance between predictor (independent) block X and predicted (dependent) block Y for each component, where the predicted variables are dummy variables (1 or 0) where '1' indicates an in-class member while '0' indicates a non-class member [18–20]. DPLS computes a lower dimensional representation, which maximizes the covariance between variables in that space and the predicted variables [11]. Several researchers have applied DPLS and PCA to small-scale classification problems and showed that DPLS improved class separation over PCA [18,21]. In general, fewer factors are needed in DPLS to give the same level of prediction successes [21].

PCA, FDA, and DPLS and their application to fault diagnosis are described next. An information criterion for FDA and DPLS is developed for determining the order of the dimensionality reduction without cross-validation. Then the proficiency of these techniques for fault diagnosis are evaluated by application to data collected from the Tennessee Eastman chemical plant simulator.

2. Methods

2.1. PCA

PCA is an optimal dimensionality reduction technique in terms of capturing the variance of the data. PCA determines a set of orthogonal vectors, called *loading vectors*, which can be ordered by the amount of variance explained in the loading vector directions. Given n observations of m measurement variables stacked into a training data matrix $X \in \mathscr{R}^{n\times m}$, the loading vectors can be calculated via the Singular Value Decomposition (SVD)

$$\frac{1}{\sqrt{n-1}} X = U\Sigma V^T \tag{1}$$

where $U \in \mathscr{R}^{n\times n}$ and $V \in \mathscr{R}^{m\times m}$ are unitary matrices and the diagonal matrix $\Sigma \in \mathscr{R}^{n\times m}$ contains the nonnegative real *singular values* of decreasing magnitude ($\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$). The loading vectors are the orthonormal column vectors in the matrix V, and the variance of the training set projected along the ith column of V is equal to σ_i^2.

2.1.1. Fault detection

Normal operations can be characterized by employing Hotelling's T^2 statistic [2]:

$$T^2 = \mathbf{x}^T P \Sigma_a^{-2} P^T \mathbf{x} \tag{2}$$

where P includes the loading vectors associated with the a largest singular values, Σ_a contains the first a rows and columns of Σ, and $\mathbf{x}$ is an observation vector of dimension m. Given a number of loading vectors, a, to include in Eq. (2), the threshold can be calculated for the T^2 statistic using the probability distribution

$$T_\alpha^2 = \frac{(n^2-1)a}{n(n-a)} F_\alpha(a, n-a) \tag{3}$$

where $F_\alpha(a, n-a)$ is the upper $100\alpha\%$ critical point of the F-distribution with a and $n-a$ degrees of freedom [22,23]. A value for the T^2 statistic, Eq. (2), greater than the threshold given by Eq. (3) indicates that a fault has occurred.

The portion of the measurement space corresponding to the lowest $m-a$ singular values can be monitored by using the Q statistic [12,24].

$$Q = \mathbf{r}^T\mathbf{r}, \quad \mathbf{r} = (I - PP^T)\mathbf{x}, \tag{4}$$

where $\mathbf{r}$ is the residual vector. Since the Q statistic does not directly measure the variations along each loading vector but measures the total sum of variations in the space corresponding to the lowest $m-a$ singular values, the Q statistic does not suffer from an over-sensitivity to inaccuracies in the lower singular values [24].

The threshold for the Q statistic can be computed from its approximate distribution [24]

$$Q_\alpha = \theta_1\left[\frac{c_\alpha\left(2\theta_2 h_0^2\right)^{1/2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right]^{1/h_0} \tag{5}$$

where $\theta_i = \sum_{j=\alpha+1}^{n}\sigma_j^{2i}, h_0 = 1 - (2\theta_1\theta_3)/(3\theta_2^2)$, and c_α is the normal deviate corresponding to the upper $(1-\alpha)$ percentile.

2.1.2. Reduction order

A key step in a dimensionality reduction technique is to determine the order of the reduction, that is, its dimensionality. There exist several techniques for determining the number of loading vectors, a, to maintain in the PCA model [12,25–28]. Parallel analysis determines the dimensionality of the PCA model by comparing the singular value profile to that obtained by assuming independent measurement variables. The dimension is determined by the point at which the two profiles cross. This approach is particularly attractive since it is intuitive, easy to automate, and performs well in practice.

2.1.3. Fault diagnosis

Several researchers have proposed techniques to use principal component analysis for fault diagnosis. The simplest approach is construct a single PCA model and define regions in the lower dimensional space which classifies whether a particular fault has occurred [29]. This approach is unlikely to be effective when a significant number of faults can occur [7]. Another approach is to compute the group of process variables which make the greatest contributions to the deviations in the squared prediction error and the scores [3]. Although such information can narrow down the search for an assignable cause of abnormal behavior, it will not unequivocally diagnose the cause. A related approach is to construct separate PCA models for each process unit [12]. A fault associated with a particular process unit is assumed to occur if the PCA model for that unit indicates that the process is out-of-control. Again, this approach can narrow down the cause of abnormal process operations, it will not unequivocally diagnose the cause. This distinguishes these *fault isolation* techniques (which are based on non-supervised classification) from the fault diagnosis techniques (which are based on supervised classification) of interest here. Diagnosis approaches particular to sensor faults [30,31] will also not be considered further here because the focus is on more general types of faults.

A PCA approach which can handle general multiple faults is to develop a separate PCA model based on data collected during each specific fault situation, and then apply the Q [32], T^2 [4], or other statistics [4–7] applied to each PCA model to predict which fault or faults most likely occurred. This approach is essentially a combination of principal component analysis and discriminant analysis [5].

2.2. Fisher's discriminant analysis

For fault diagnosis, data collected from the plant during specific faults is categorized into classes, where each class contains data representing a particular fault. FDA is an optimal dimensionality reduction technique in terms of maximizing the separability of these classes. It determines a set of projection vectors that maximize the scatter between the classes while minimizing the scatter within each class.

Stacking the training data for all classes into the matrix $X \in \mathscr{R}^{n\times m}$ and representing the ith row of X

with the column vector $\mathbf{x}_i$, the *total-scatter matrix* is [16,17]

$$S_t = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (6)$$

where $\bar{\mathbf{x}}$ is the *total mean vector* whose elements correspond to the means of the columns of X. Let the matrix X_i contain the rows of X corresponding to class i, then

$$S_i = \sum_{\mathbf{x}_i \in X_i} (\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T \quad (7)$$

is the *within-scatter matrix* for class i where $\bar{\mathbf{x}}_i$ is the mean vector for class i. Let c be the number of classes, then

$$S_w = \sum_{i=1}^{c} S_i \quad (8)$$

is the *within-class-scatter matrix*, and

$$S_b = \sum_{i=1}^{c} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (9)$$

is the *between-class-scatter matrix*, where n_i is the number of observations in class i. The total-scatter matrix is equal to the sum of the between-scatter matrix and the within-scatter matrix [16],

$$S_t = S_b + S_w. \quad (10)$$

Assuming invertible S_w, the FDA vectors are determined by computing the singularities of the optimization problem

$$\max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^T S_b \mathbf{v}}{\mathbf{v}^T S_w \mathbf{v}} \quad (11)$$

(equations for the case of non-invertible S_w are provided elsewhere [33–35]). The FDA vectors are equal to the generalized eigenvectors of the eigenvalue problem:

$$S_b \mathbf{w}_i = \lambda_i S_w \mathbf{w}_i \quad (12)$$

where the eigenvalues λ_i indicate the degree of overall separability among the classes. Because the direction and not the magnitude of $\mathbf{w}_i$ is important, the norm is usually chosen to be $\|\mathbf{w}_i\| = 1$.

2.2.1. Fault diagnosis

While numerous researchers have developed techniques based on first constructing PCA models based on data collected for each fault class, and then applying some form of discriminant analysis or related approach to diagnose faults [4–7,32], the FDA approach simultaneously uses all of the data to obtain a single lower dimensional model used to diagnose faults. The lower dimensional representation provided by FDA can be employed with discriminant functions, such as the T^2 statistic, to diagnose faults. FDA can be used to detect faults by including a class of data collected during normal process operation.

2.2.2. Reduction order

Akaike's information criterion (AIC) is a well-known method for selecting the model order for system identification [36]. The AIC contains an error term and a term which penalizes the model complexity. A strength of the AIC is that it relies only on information in one set of data (the training data), unlike cross validation which requires either additional data or a partitioning of the original data set into two sets. We propose to determine the order of the FDA model by computing the dimensionality, a, which minimizes the information criterion

$$f(a) + \frac{a}{\bar{n}} \quad (13)$$

where $f(a)$ is the misclassification rate for the training set by projecting the data onto the first a FDA vectors and $\bar{n}$ is the average number of observations per class. Eq. (13), which is similar in form to the AIC, appears to be reasonable since the penalty term scales relatively well with the error term. This is confirmed later by application.

2.3. Discriminant partial least squares

DPLS is a dimensionality reduction technique for maximizing covariance between the predictor (independent) block X and the predicted (dependent) block Y for each component [18–20,37]. DPLS models the relationship between X and Y using a series of local least-squares fits. In DPLS, the training data for p classes are stacked into the data matrix $X \in \mathscr{R}^{n \times m}$, where $q_1 + q_2 + \ldots + q_p = n$ and q_i is

the number of observations for class i. There are two methods, known as PLS1 and PLS2, to model Y. The predicted block $Y \in \mathscr{R}^{n \times p}$ in PLS2 is

$$Y = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad (14)$$

where each column in Y corresponds to a class. Each element of Y is filled with either *one* or *zero*. The first q_1 elements of column 1 are filled with a '1', which indicates that the first q_1 rows of X are data from fault 1. In PLS1, the algorithm is run p times, each with the same X, but for each separate column of Y in Eq. (14). This results in one model for each class.

The matrices X and Y are autoscaled. The matrix X is decomposed into a score matrix $T \in \mathscr{R}^{n \times a}$ and a loading matrix $P \in \mathscr{R}^{m \times a}$, where a is the PLS component (order), plus a residual matrix $E \in \mathscr{R}^{n \times m}$ [38]

$$X = TP^T + E \quad (15)$$

In PLS2, Y is decomposed into a score matrix $U \in \mathscr{R}^{n \times a}$, a loading matrix $Q \in \mathscr{R}^{p \times a}$, plus a residual matrix $F^* \in \mathscr{R}^{n \times p}$

$$Y = UQ^T + F^* \quad (16)$$

The estimated Y is related to X through the score matrix T:

$$Y = TBQ^T + F \quad (17)$$

where F is the prediction error matrix. The matrix B is selected such that the induced 2-norm of F (the maximum singular value of F [39], $\|F\|_2$, is minimized [13]. In PLS1, similar steps are performed, resulting in

$$y_i = T_i B_i q_i^T + f_i \quad (18)$$

where $y_i \in \mathscr{R}^n$ is the ith column of Y, $T_i \in \mathscr{R}^{n \times a}$ is the score matrix, $B_i \in \mathscr{R}^{a \times a}$ is the regression matrix, $q_i^T \in \mathscr{R}^a$ is the loading vector, and $f_i \in \mathscr{R}^n$ is the prediction error vector. Since there are p columns in Y, the range of i is from 1 to p.

The most popular algorithm used to compute the parameters of Eqs. (17) and (18) in the calibration step is known as Non-Iterative Partial Least Squares (NIPALS) [11,38].

2.3.1. Fault diagnosis

After the parameters have been determined, the predicted block Y_{train1} of the training set using PLS1 and the predicted Y_{train2} of the training set using PLS2 are calculated for all orders. In general, the rows of Y_{train1} and Y_{train2} will not have the form [0,0,0, ..., 1, ..., 0,0]; discriminant analysis is needed to predict class c_k at each observation k. One common approach is to define c_k to be the column index whose element has maximum value among row k. This approach works well in the ideal case; that is, the classification is not overestimated nor underestimated (overestimation means the score of an in-class member > 1 and the score of a non-class member > 0, while underestimation means the score of an in-class member < 1 and the score of a non-class member < 0). This approach also works well if *all* of the scores are overestimated or underestimated [20].

However, if some of the scores are underestimated while others are overestimated, the above approach can give poor results. A method to solve this problem is to take account of the underestimation and overestimation of Y into a second cycle of PLS algorithm [20]. For PLS1 and PLS2, the NIPALS is run for the second time by replacing Y by Y_{train2} and y_i by the ith column of Y_{train1} in Eqs. (17) and (18), respectively. To distinguish between the *normal* PLS method and this *adjusted* method, PLS1 and PLS2 are denoted as PLS1_{adj} and PLS2_{adj}, respectively. Here the orders for all PLS models are determined using the proposed criterion (13).

Although numerous researchers have proposed fault diagnosis algorithms based on PCA, the PCA objective of capturing the most variance is not directly related to the objective of fault diagnosis. As such, the resulting lower dimensional space may contain little of the information required to discriminate

among various faults [21]. Since DPLS exploits fault information when constructing its lower dimensional model, it would be expected that DPLS can provide better fault diagnosis than PCA. Since the FDA objective directly coincides with the objective of fault diagnosis, it would be expected to outperform both DPLS and PCA. Using the proposed FDA statistics, and the FDA and DPLS order selection criterion (13), these theoretical predictions are should to be valid for the Tennessee Eastman Industrial Challenge Problem.

3. Application

The process simulator for the Tennessee Eastman (TE) Industrial Challenge Problem was created by the Eastman Chemical Company to provide a realistic industrial process for evaluating process control and monitoring methods [40]. The TE process simulator has been widely used by the process monitoring community as a source of data for comparing various approaches [5,6,12,14,41–43]. The plant simulator is based on an actual chemical process where the components, kinetics, and operating conditions were modified for proprietary reasons (see Fig. 1). The gaseous reactants A, C, D, and E and the inert B are fed to the reactor where the liquid products G and H are formed. The reactions in the reactor are irreversible, exothermic, and approximately first-order with respect to the reactant concentrations. The reactor product stream is cooled through a condenser and then fed to a vapor–liquid separator. The vapor exiting the separator is recycled to the reactor feed through the compressor. A portion of the recycle stream is purged to keep the inert and byproducts from accumulating in the process. The condensed components from the separator (Stream 10) are pumped to the stripper. Stream 4 is used to strip the remaining reactants in Stream 10, and is combined with the recycle stream. The products G and H exiting the base of the stripper are sent to a downstream process which is not included in this process. The simulation code allows 21 preprogrammed major

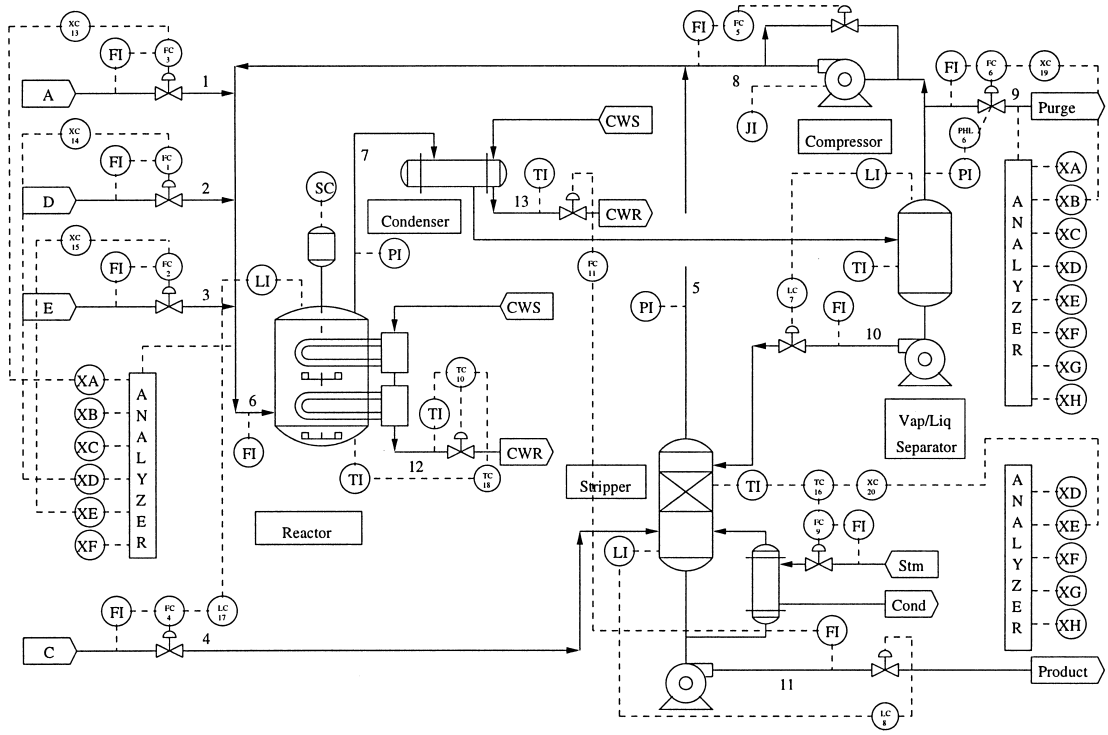

Fig. 1. A diagram of the Tennessee Eastman process simulator.

process faults, as shown in Table 1. The plant-wide control structure recommended in Lyman and Georgakis [44] was implemented to generate the closed loop simulated process data for each fault.

The training and testing data sets for each fault consisted of 500 and 960 observations, respectively. Note that only the training data was used in model order selection-the testing data is used to see how well the methods performed, and to determine the effectiveness of the model order selection criterion (13). Each data set started with no faults, and the faults were introduced 1 and 8 simulation hours into the run, respectively, for the training and testing data sets. All the manipulated and measurement variables except for the agitation speed of the reactor's stirrer for a total of $m = 52$ variables were recorded. The data was sampled every 3 min, and the random seed (used to specify the stochastic measurement noise and disturbances) was changed before the computation of the data set for each fault. Twenty-one testing sets were generated using the preprogrammed faults (Fault 1-21). In additional, one testing set (Fault 0) was generated with no faults. The data were scaled in the standard manner before the application of PCA, FDA, and DPLS. That is, the sample mean was subtracted from each variable, and then divided by its standard deviation. All the training and testing data sets have been made available at http://brahms.scs.uiuc.edu.

Table 1
Process faults for the Tennessee Eastman process simulator

Variable	Description	Type
IDV(1)	A/C Feed Ratio, B Composition Constant (Stream 4)	Step
IDV(2)	B Composition, A/C Ratio Constant (Stream 4)	Step
IDV(3)	D Feed Temperature (Stream 2)	Step
IDV(4)	Reactor Cooling Water Inlet Temperature	Step
IDV(5)	Condenser Cooling Water Inlet Temperature	Step
IDV(6)	A Feed Loss (Stream 1)	Step
IDV(7)	C Header Pressure Loss — Reduced Availability (Stream 4)	Step
IDV(8)	A, B, C Feed Composition (Stream 4)	Random Variation
IDV(9)	D Feed Temperature (Stream 2)	Random Variation
IDV(10)	C Feed Temperature (Stream 4)	Random Variation
IDV(11)	Reactor Cooling Water Inlet Temperature	Random Variation
IDV(12)	Condenser Cooling Water Inlet Temperature	Random Variation
IDV(13)	Reaction Kinetics	Slow Drift
IDV(14)	Reactor Cooling Water Valve	Sticking
IDV(15)	Condenser Cooling Water Valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed at the steady state position	Constant Position

The overall misclassification rate for each measure when applied to all disturbances of the testing set are listed in Table 2. As anticipated by comparing their objectives, FDA produced the lowest overall misclassification rate, followed by DPLS, and PCA.

Plots of the misclassification rates of PLS and FDA as a function of model order (Fig. 2a–c) indicate that FDA with any order greater than 10 outperforms all of the PLS methods, with most of the separation between fault classes by FDA being provided by the first 13 generalized eigenvectors. This indicates that the superior fault diagnosis provided by FDA is inherent and not due to the model order selection criteria used in this study (Eq. 13). Fig. 2a–c also indicate that the AIC (Eq. 13) does a good job in selecting the model order for both FDA, PLS, and adjusted PLS. For FDA, the AIC captures the shape and slope of the misclassification rate curve for the testing data. In Fig. 2b–c, the AIC curve nearly overlaps the classification rate curves for PLS2 and adjusted PLS2, which indicates that the AIC will give similar model orders as cross-validation in these cases. For PLS1 and adjusted PLS1, the AIC does not overlap with the classification rate curves, but does have a minimum at approximately the same order as where the misclassification rate curves for the testing data flatten out. Again, this indicates that the AIC provides good model orders for the PLS1 methods.

Table 2
Overall misclassification rates and orders for the various models

Method	Basis	Misclassification rate	Order
PCA	T^2	0.742	4–16*
PCA	Q	0.609	–
PCA	T^2 and Q	0.667	–
PLS1	–	0.565	13
PLS2	–	0.567	45
$PLS1_{adj}$	–	0.576	16
$PLS2_{adj}$	–	0.574	41
FDA	T^2	0.206	51

*Order selection varies for each fault class.

All PLS methods gave similar overall misclassification rates (see Table 2). For a fixed model order, the PLS1 methods almost always gave better fault diagnosis than the PLS2 methods (see Fig. 2b–c). The performance of the PLS1 methods was also less sensitive to order selection than the PLS2 methods, and with the AIC resulting in lower model orders (see Table 2). Fig. 2d indicates that the adjusted PLS methods provided more consistent misclassification rates for various faults than the non-adjusted DPLS methods. For example, 7 of 21 classes had misclassification rates between 0.90 to 1.00 using PLS1 and PLS2, respectively. However, only 2 of 21 classes fit in the range using $PLS1_{adj}$ and $PLS2_{adj}$ and the highest misclassification rate was 0.93. Since the overall misclassification rates for the methods are similar, this indicates that for the non-adjusted PLS methods do better for some faults, and the adjusted PLS methods do better for others.

The PCA-based measures produced higher misclassification rates than the PLS and FDA measures. The main reason for this is that PCA reduces the dimensionality of each class by using the information in only one class but not the information from all the classes. The measures for PCA based on the residual space (Q-statistic) performed better than the measures based on the score space. There are two reasons for most of the discriminatory power being in the residual space. First, the residual space is larger than the score space (see Table 2). Just by the larger size of the residual space, it can contain more discrimina-

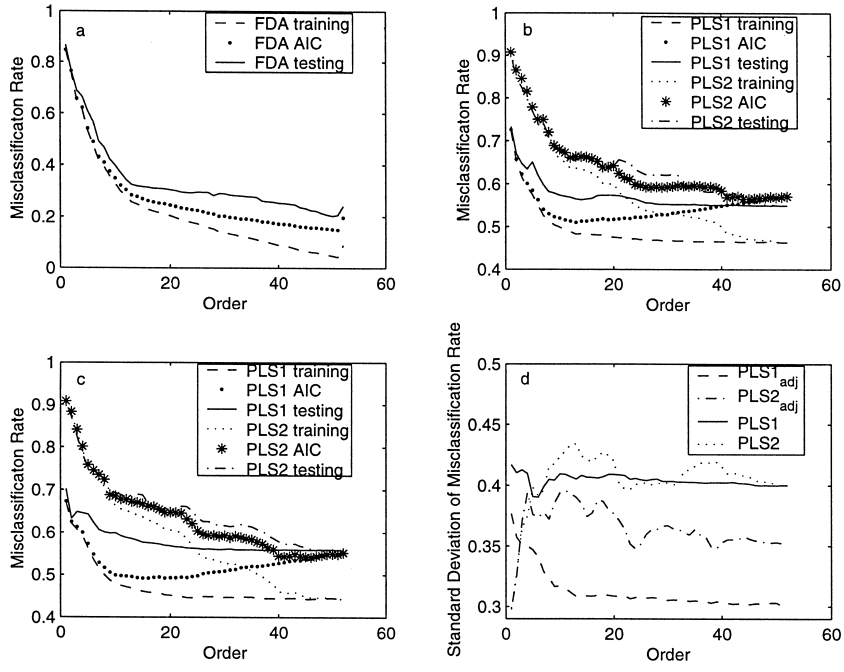

Fig. 2. The overall misclassification rates for the training and testing sets and the information criterion (AIC) for various orders using FDA, PLS1, PLS2, $PLS1_{adj}$, and $PLS2_{adj}$ and the standard deviation of misclassification rates for the testing set for various orders using PLS1, PLS2, $PLS1_{adj}$, and $PLS2_{adj}$.

tory power. Secondly, different faults tend to create different states, and the statistical properties typical of the residual space enables it to contain more discriminatory power.

4. Conclusions

Fisher discriminant analysis and discriminant PLS were shown to be better dimensionality reduction techniques than principal component analysis for fault diagnosis. Although numerous researchers have developed techniques for using PCA to diagnose faults, it is not well suited because it does not take into account the information between the classes when determining the lower dimensional representation. FDA provides an optimal lower dimensional representation in terms of maximizing the separation amongst several classes. The projection vectors are ordered in terms of maximizing the scatter between the classes while minimizing the scatter within each class. In discriminant PLS, the covariance between the predictor block (data from all classes) and predicted block (representation of class membership) are maximized for each factor. Information between classes are used when determining each factor. A model selection criterion for FDA and discriminant PLS were proposed based on the Akaike information criterion. The techniques were applied to data collected from the Tennessee Eastman chemical plant simulator, where FDA performed the best, followed by DPLS and PCA.

Acknowledgements

This work was supported by International Paper.

References

[1] J.V. Kresta, T.E. Marlin, J.F. MacGregor, Can. J. Chem. Eng. 69 (1991) 35–47.
[2] T. Kourti, J.F. MacGregor, J. Quality Technol. 28 (1996) 409–428.
[3] J.F. MacGregor, Proc. of the IFAC Conference on Advanced Control of Chemical Processes, Pergamon Press, New York, 1994, pp. 427–435.
[4] A.C. Raich, A. Cinar, Proc. of the IFAC Conf. on Advanced Control of Chemical Processes, Pergamon, New York, 1994, pp. 427–435.
[5] A.C. Raich, A. Cinar, Chemometrics and Intelligent Laboratory Systems 30 (1995) 37–48.
[6] A.C. Raich, A. Cinar, AIChE J. 42 (1996) 995–1009.
[7] J. Zhang, E. Martin, A.J. Morris, Proc. of the American Control Conf., IEEE Press, Piscataway, NJ, 1995, pp. 751–755.
[8] K.A. Kosanovich, M.J. Piovoso, K.S. Dahl, J.F. MacGregor, P. Nomikor, Proc. of the American Control Conf., IEEE Press, Piscataway, NJ, 1994, pp. 1294–1298.
[9] M.J. Piovoso, K.A. Kosanovich, R.K. Pearson, Proc. of the American Control Conf., Piscataway, IEEE Press, NJ, 1992, pp. 2359–2363.
[10] M.J. Piovoso, K.A. Kosanovich, Int. J. Control 59 (1994) 743.
[11] B.M. Wise, N.B. Gallagher, J. Process Control 6 (1996) 329–348.
[12] D.M. Himes, R.H. Storer, C. Georgakis, Proc. of the American Control Conf., IEEE Press, Piscataway, NJ, 1994, pp. 1279–1283.
[13] M.H. Kaspar, W.H. Ray, AIChE J. 38 (1992) 1593–1608.
[14] W. Ku, R.H. Storer, C. Georgakis, Chemometrics and Intelligent Laboratory Systems 30 (1995) 179–196.
[15] H. Tong, C.M. Crowe, AIChE J. 41 (7) (1995) 1712–1722.
[16] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[17] R. Hudlet, R. Johnson, in: J. Van Ryzin (Ed.), Classification and clustering, Academic Press, New York, 1977, pp. 371–394.
[18] B.K. Alsberg, R. Goodacre, J.J. Rowland, D.B. Kell, Analytica Chimica Acta 348 (1997) 389–407.
[19] M. Defernez, K. Kemsley, Trends in Analytical Chemistry 16 (1997) 216–221.
[20] J. Nouwen, F. Lindgren, W. Karcher, B. Hansen, H.J.M. Verharr, J.L.M. Hermens, Environ. Sci. Technol. 31 (1997) 2313–2318.
[21] E.K. Kemsley, Chemometrics and Intelligent Laboratory Systems 33 (1996) 47–61.
[22] J.F. MacGregor, T. Kourti, Control Engineering Practice 3 (1995) 403–414.
[23] N.D. Tracy, J.C. Young, R.L. Mason, J. Quality Control 24 (1992) 88–95.
[24] J.E. Jackson, G.S. Mudholkar, Technometrics 21 (1979) 341–349.
[25] J.L. Horn, Psychometrica 30 (2) (1965) 179–185.
[26] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.
[27] S. Wold, Technometrics 20 (1978) 397–405.
[28] W.R. Zwick, W.F. Velicer, Psychological Bulletin 99 (3) (1965) 432–442.
[29] B.M. Wise, N.L. Ricker, D.F. Veltkamp, Upset and sensor failure detection in multivariate processes,Technical report, Eigenvector Research, Manson, WA, 1989.
[30] R. Dunia, S.J. Qin, T.F. Edgar, T.J. McAvoy, AIChE J. 42 (1996) 2797–2812.
[31] A. Negiz, A. Cinar, Proc. of the American Control Conf. Piscataway, IEEE Press, NJ, 1992, pp. 2364–2368.
[32] W. Ku, R.H. Storer, C. Georgakis, In AIChE Annual Meeting, 1993. Paper 149g.
[33] Y.Q. Cheng, Y.M. Zhuang, J.Y. Yang, Pattern Recognition 25 (1992) 101–111.

[34] Z.Q. Hong, J.Y. Yang, Pattern Recognition 24 (1991) 317–324.
[35] Q. Tian, J. Opt. Soc. Am. A 5 (1988) 1670–1672.
[36] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[37] B.K. Alsberg, D.B. Kell, R. Goodacre, Analytical Chemistry 70 (1998) 4123–4133.
[38] P. Geladi, B.R. Kowalski, Analytica Chimica Acta 185 (1986) 1–17.
[39] G.H. Golub, C.F. van Loan, Matrix Computations, Johns Hopkins Univ. Press, Baltimore, MD, 1983.
[40] J.J. Downs, E.F. Vogel, Comput. Chem. Eng. 17 (1993) 245–255.
[41] G. Chen, T.J. McAvoy, J. Process Control 8 (1997) 409–420.
[42] C. Georgakis, B. Steadman, V. Liotta, Proc. of the 13th IFAC World Congress, IEEE Press, Piscataway, NJ, 1996, pp. 97–101.
[43] A.C. Raich, Proc. of the 13th IFAC World Congress, IEEE Press, Piscataway, NJ, 1996, pp. 283–288.
[44] P.R. Lyman, C. Georgakis, Comput. Chem. Eng. 19 (1995) 321–331.