Monitoring of Metallurgical Process Plants by Using Biplots

C. Aldrich
Dept. of Chemical Engineering, University of Stellenbosch, Stellenbosch, Matieland, 7602, South Africa

S. Gardner and N. J. Le Roux
Dept. of Statistics and Actuarial Science, University of Stellenbosch, Stellenbosch, Matieland, 7602, South Africa

DOI 10.1002/aic.10170
Published online in Wiley InterScience (www.interscience.wiley.com).

Although principal component analysis has been applied widely for monitoring plant performance in a broad range of industrial processes, the use of accompanying biplots has not received similar attention. Moreover, principal component analysis is a linear technique that tends to break down when processes exhibit significant nonlinear behavior. In this paper biplot methodology is introduced. This methodology allows for projecting high-dimensional data to a low-dimensional subspace that can be visualized by a human operator. It provides management with sophisticated tools, highly graphical in nature, to extract information regarding variation in process variables, correlations among these variables, as well as class separation, taking into account the multidimensional character of the data. Biplot methodology can also be applied to data sets exhibiting nonlinear behavior. As is shown by way of two case studies, operating regions can be quantitatively explored by superimposing alpha bags on biplots, which facilitates the automatic detection and visualization of process disturbances. © 2004 American Institute of Chemical Engineers *AIChE J,* 50: 2167–2186, 2004
Keywords: biplot, canonical variate analysis, principal component analysis, alpha bag

Introduction

Although modern instrumentation allows large volumes of process data to be accumulated, these data are of limited use without appropriate processing, especially with respect to the discovery of abnormal events derived from the interaction between variables, tracking of process drift, and so forth. Initial research focused on the development of statistical quality control methods concerned with the monitoring of product quality data only, at the expense of variables related to process conditions, where special events may be reflected as well. The latter variables are often measured considerably more frequently (and more accurately) than the process quality variables. Moreover, the measurement of process quality may only occur after the product has been made, that is, too late for control purposes.

In the last few decades, several multivariate statistical methods have been developed to identify correlations between variables and to exploit process information for monitoring and control purposes (Kourti and MacGregor, 1995; Martin et al., 1996, 1999; Wise et al., 1990). The superimposition of two univariate control charts on top of each other illustrates the advantages of multivariate statistical process control over univariate control, as shown in Figure 1. Broken lines indicate the individual lower and upper control limits (LCL_1, UCL_1) and (LCL_2, UCL_2) of each of the variables x_1 and x_2. The approach outlined in Figure 1a is not particularly accurate with correlated variables, given that the redundancy of the variables is not exploited. In contrast, Figure 1b shows the true control region of the two variables, where the correlation between the two

Correspondence concerning this article should be addressed to C. Aldrich at ca1@sun.ac.za

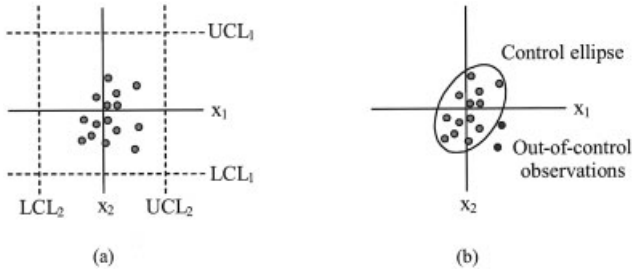

Figure 1. (a) Two superimposed univariate process control charts (dashed lines), and (b) a control chart taking the correlation between the variables into account (bold lines).

Out-of-control observations (indicated by black circles) fall within the acceptable range of each variable and can be detected only by considering both variables simultaneously.

variables is reflected by the aspect ratio of the principal axes of the control ellipse. Note that two of the observations that were considered to be in control in Figure 1a, are shown to be out of control in Figure 1b.

Principal component analysis (PCA) and related techniques are commonly used to construct charts like the one shown in Figure 1b, as discussed by Kresta et al. (1991), Nomikos and MacGregor (1995), Amirthalingam and Lee (1997), Kosanovich et al. (1996), Kourti and MacGregor (1995), Jia et al. (1998), Jaeckle and MacGregor (1998), Seasholtz (1999), and Zhang et al. (1996, 1997, 1999), among many others. PCA effectively reduces the observation space, which facilitates process failure and disturbance diagnosis through the capture of correlations between variables and characterizing normal operating conditions.

A random sample of n observations on p process control variables can be represented as a data matrix $X : n \times p$. If the units of measurement of the p variables are commensurable, a PCA is performed on the covariance matrix associated with X, else the correlation matrix of X is to be used. The biplot, as the simultaneous display of the rows and columns of a data matrix, was introduced by Gabriel (1971). A new perspective on the traditional biplot of Gabriel is offered by Gower and Hand (1996). This perspective provides a unified approach to PCA biplots based on Pythagorean distance; canonical variate analysis (CVA) biplots based on Mahalanobis distance; nonlinear biplots based on Euclidean embeddable distances; and generalized biplots for use with both continuous and categorical variables. Gardner (2001) extends the biplot methodology of Gower and Hand by considering a wide variety of applications in statistical discrimination and classification. This leads to discriminant analysis by means of PCA biplots, CVA biplots, nonlinear biplots, and generalized biplots. Classification regions defined for linear discriminant analysis (LDA) are applied in the CVA biplot leading to discriminant analysis using biplot methodology. Situations where the assumptions of LDA are not met are considered by Gardner (2001) and various existing alternative discriminant analysis procedures are formulated in terms of biplots. Apart from PCA biplots, quadratic discriminant analysis (QDA), flexible discriminant analysis (FDA), and discriminant subspace (DSM) biplots are defined and constructed and their usage is illustrated. It is demonstrated that biplot methodology naturally provides for managing categorical and continuous variables simultaneously. Gardner (2001) argues that in situations occurring in practice where existing discriminant analysis procedures based on distances from means fail, discriminant analysis based on distances from principal curves (a form of nonparametric principal components also a form of conditional means) should be considered. She was able to demonstrate the superiority of biplot classification procedure based on principal curves in several instances. Robust PCA and CVA biplots restricting the role of influential observations on biplot displays are also considered.

Commenting on the desirability of dimension reduction in discriminant analysis, Schott (1993) held the view that such dimension reduction not only might prevent an increase in the probability of misclassification attributed to the presence of unnecessary dimensions but "First of all, when reduction to 2 or 3 dimensions is possible one can get graphical representations which then allow visualization of differences between populations and the location of new observations relative to the population." According to Chambers et al. (1983), "there is no single statistical tool that is as powerful as a well-chosen graph." In a review of available graphical methods for exploring multivariate data Everitt (1994) adds "there are many patterns and relationships that are easier to discern in graphical displays than by any other data analysis method."

In this communication the use of biplot methodology in monitoring plant performance is illustrated by considering two case studies. The usefulness of biplot methodology as an analytical and graphical tool for describing and displaying multidimensional variation, interrelationships, class separation and overlap, outlying observations, and classifying new observations will be demonstrated. Although biplot methodology has much more to offer, this demonstration will be restricted to PCA and CVA biplots only. Before attending to the case studies, a brief review of the main features of PCA and CVA biplots is given.

Biplot Methodology

PCA biplot

The biplot was introduced by Gabriel (1971) as a graphical display consisting of a vector for each row and a vector for each column of a matrix of rank two. In this display any element of the matrix is represented by the inner product of the vectors corresponding to its row and column. The biplot can be related to PCA and involves the singular value decomposition of a matrix. Gabriel (1972) demonstrated its usefulness by a detailed exposition of several examples. Any matrix of rank $k > 3$ can be approximated by a matrix of rank 2 or 3. Indeed, such low rank approximations are in practice often adequate or even better than those of higher rank (see Caussinus and Ferré, 1992). According to Gabriel (1999), weighted least squares is the most common and mathematically tractable method of lower rank approximation. He showed how (1) observations represented in a data matrix, (2) variances and covariances, and (3) pairwise dissimilarities between rows of a data matrix can be displayed in biplots in two or three dimensions. Depending on the matrix to be biplotted, its centering, weight matrices used, and the chosen metric, various types of biplots can be constructed.

Gabriel's biplot proved to be extremely useful in displaying important features of a data matrix and was soon put to use in

many fields of application, including industrial processes where Sparks et al. (1997) showed how the PCA biplot can be used to monitor multivariate process data. By using the biplot dynamically, they are able to display concisely but accurately changes in location, variation, and correlation structure of their multivariate process data.

As a diagnostic tool, the biplot is often used for detecting outliers. Contrary to this, robust PCA biplots are also proposed by Daigle and Rivest (1992) and Choi and Huh (1996) to provide biplot displays that are resistant to influential observations.

Although the traditional Gabriel biplot can be used—and is widely applied in practice—as a graphical aid, the ideas of Gower (1995) have much to offer when applying biplot methodology for monitoring multidimensional plant performance. These ideas culminated in the monograph by Gower and Hand (1996), which introduced a new philosophy in biplot methodology. In this new perspective biplots are regarded as the multivariate analogs of scatterplots, allowing for visual appraisal of the structure of the data in a few dimensions. Biplot axes are used to relate the plotted points to the original variables, as is the case in ordinary scatterplots. Gower and Hand comment that scatterplots "are not only easy to produce, but they also have the merit of being straightforward to interpret, requiring very little, if any, formal training." By extending the scatterplot principles to multivariate displays, these biplots ease interpretation and are accessible to nonstatistical audiences.

The data matrix $X: n \times p$ gives the coordinates of the n observations in the space $\mathcal{R}^p$. The principal component method chooses the r-dimensional subspace $\mathcal{L}$ that is best fitting in the least-squares sense such that the matrix $\hat{X}: n \times p$ minimizes $\| X - \hat{X} \| = \text{trace}[(X - \hat{X})'(X - \hat{X})]$ for all matrices $\hat{X}: n \times p$ of rank at most r. The p-dimensional observations are projected onto the subspace $\mathcal{L}$ and the subspace for which the sum of squared distances from the sample points to their projections is a minimum is chosen as the space of best fit. It can be proved that the subspace $\mathcal{L}$ passes through the centroid of the sample points.

Let $L: p \times r$ be a matrix with orthonormal columns forming a basis for $\mathcal{L}$. The projections of the rows of X onto $\mathcal{L}$ are given by XLL' with residuals $X(I - LL')$. The principal components of X form the columns of the matrix L that minimizes $\| X(I - LL') \|$. Writing the singular value decomposition $X'X = V\Lambda V'$ with $V'V = \mathrm{I}$, Gower and Hand (1996) show that $L = V_r$, the first r columns of V.

Because V_r forms an orthonormal basis for $\mathcal{L}$, these principal axes define a natural set of orthogonal coordinate axes that could be used as scaffolding for plotting the sample points in the biplot. Relative to the principal axes, the coordinates of the projections of the sample points are given by $Z = XV_r$. Gower and Hand (1996) use the terms *interpolation* and *prediction* for the relationships between the scaffolding and the original variables. Instead of showing the scaffolding axes in the biplot, the variables are represented either by calibrated interpolation axes or (preferably) by calibrated prediction axes. Interpolation is defined as finding, for a given or new sample point $\mathbf{x}^*: p \times 1$, its representation in the space $\mathcal{L}$, that is $\mathbf{z}^*: r \times 1$ in terms of the biplot scaffolding. On the other hand, prediction is inferring the values of the original variables $\hat{\mathbf{x}}^*: p \times 1$ for a point $\mathbf{z}^*: r \times 1$ in $\mathcal{L}$ given in terms of the biplot scaffolding. In the case of interpolation the projection of $\mathbf{x}^*: p \times 1$ onto $\mathcal{L}$, in terms of the basis for $\mathcal{L}$ (scaffolding) is given by $\mathbf{z}^{*\prime} = \mathbf{x}^{*\prime}V_r$, whereas for prediction, because $\mathcal{L} \subset \mathcal{R}$, the given point $\mathbf{z}^*: \mathrm{r} \times 1 \in \mathcal{R}$. Therefore, $\mathbf{z}^*: r \times 1$ can be written in terms of the basis for $\mathcal{R}$ as $\mathbf{x}^*: p \times 1$. Because the point $\mathbf{x}^* \in \mathcal{L}$, it will project onto itself, giving $\mathbf{x}^{*\prime} = \mathbf{x}^{*\prime}V_r(V_r'V_r)^{-1}V_r'$. Because $\mathbf{z}^{*\prime} = \mathbf{x}^{*\prime}V_r$ and because the columns of V_r are orthonormal, the prediction of $\mathbf{z}^*$ is given by $\hat{\mathbf{x}}^{*\prime} = \mathbf{z}^{*\prime}V_r$. From this it follows that the coordinates of the interpolant $\mathbf{z}^*: r \times 1$ of $\mathbf{x}^*: p \times 1$ in terms of the basis for $\mathcal{R}$, are given by $\mathbf{x}^*V_rV_r'$.

The relationship $\mathbf{z}^{*\prime} = \mathbf{x}^{*\prime}V_r$ can be used to derive p interpolation biplot axes (one for each variable) in $\mathcal{L}$. Furthermore, these interpolation axes can be calibrated in the original scales of measurement. However, Gardner (2001) argues that the use of interpolation biplot axes is limited to graphical interpolation of (new) sample points that can be done efficiently with any suitable computer program. When inspecting any graphical representation, the values of the axes are usually used to read off the values of the variables. This is closely related to Gower and Hand's concept of prediction: inferring the values of the original variables for points in $\mathcal{L}$. Prediction biplot axes must therefore be found to facilitate reading off values for the original variables. Prediction is the inverse of interpolation. In the full p-dimensional space, the interpolant (projection) $\mathbf{u}^*: p \times 1$ of $\mathbf{x}^*: p \times 1$ is given by $\mathbf{u}^{*\prime} = \mathbf{x}^{*\prime}V$. Inverting the latter expression leads to $\mathbf{x}^{*\prime} = \mathbf{u}^{*\prime}V'$. When $\mathbf{u}^{*\prime}$ lies in $\mathcal{L}$, only the first r components, $\tilde{\mathbf{z}}^*: r \times 1$ can be nonzero, so that the prediction is given by $\hat{\mathbf{x}}^{*\prime} = \tilde{\mathbf{z}}^{*\prime}V_r'$ This equation allows for constructing prediction biplot axes, calibrated in the original scales of measurement.

In general, the interpolation and prediction biplot axes have different directions, but Gower and Ngouenet (2004) show that in the case of an additive distance metric, the directions of the interpolation and prediction biplot axes will coincide because the distances for each variable separately are Euclidean. The interpolation axes can therefore be used as prediction axes, fitted with a different set of prediction markers.

The aim of PCA was outlined above as choosing the r-dimensional subspace $\mathcal{L}$, which is best fitting in the least squares sense and the residuals were given as $X(I - V_rV_r')$. The sum of squared residuals is given by $\| X(I - V_rV_r') \| = \text{tr}\{X(I - V_rV_r')X'\} = \text{tr}[\Lambda - \Lambda_{(p)r}] = \sum_{j=r+1}^{p} \lambda_j$, given that $(I - V_rV_r')$ is idempotent and the columns of V are orthogonal. The matrix $\Lambda_{(p)r}$ is a $p \times p$ diagonal matrix with the r largest eigenvalues of $X'X$ on the diagonal and the remaining diagonal elements equal to zero. The *quality* of the display can therefore be measured by the ratio

$$\frac{\sum_{j=1}^{r} \lambda_j}{\sum_{j=1}^{p} \lambda_j}$$

Apart from measuring the overall quality of the display, Gower and Hand (1996) suggest a measure of the quality of the display of each of the original p variables. Because V is an orthogonal matrix, its rows have unit sums-of-squares and it follows that the sums-of-squares of the first r columns (V_r) measure the adequacy of representation for each variable. Algebraically this is given by the $p \times p$ diagonal matrix $\text{diag}(V_rV_r')$ where the ith diagonal element indicates the *adequacy* of the representation for the ith variable. In an exact represen-

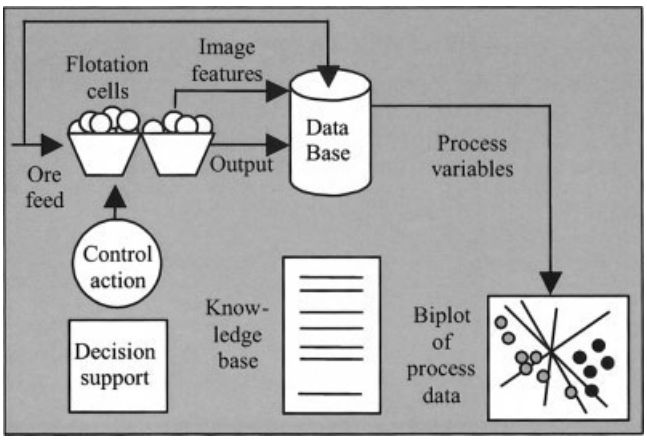

Figure 2. Monitoring of flotation cells with process maps.

Measurements on the ore feed, features extracted from digitized images of the flotation froths, and other process variables are stored on-line, from where selected variables are viewed on a process map by an operator controlling the process.

tation, this diagonal matrix would be the identity matrix I_p, whereas in all approximate representations the diagonal elements will vary between 0 and 1.

In the Gower and Hand biplot the sample points are represented by the rows of the interpolated samples matrix $Z = XV_r$ and the variables are represented by the biplot axes. The directions of these prediction biplot axes are given by the rows of the matrix V_r. In this display the matrix elements themselves and the Pythagorean distances between sample points are approximated. As will be demonstrated in Case Study 1, this display provides the process control manager with a powerful tool for inspecting multidimensional data. Although the variances, covariances, and correlations are not approximated directly, the directions of the biplot axes give an indication of the correlations. Gower and Hand remark that complete correlation between two variables induces a loss of rank, and thus an automatic dimension reduction, which manifests itself in coincidence between the corresponding projected biplot axes. High correlations induce similar effects so that a small angle between two biplot axes still suggests a high correlation between the corresponding variables. If wanted, however, Gardner (2001) shows that with a few alterations a biplot in terms of Gower and Hand's presentation can be found where the matrix elements are approximated and the cosines of the angles between the biplot axes approximate the correlations.

The CVA biplot

Often, when monitoring plant performance, it is of importance to distinguish between various classes of multidimensional observations or making a decision about the class origin of a new observation. The PCA biplot described above does not maximally separate classes but the unifying approach of Gower and Hand (1996) to biplot methodology allows for extending the methodology to CVA biplots for this purpose. CVA aims to find the linear combination of the predictor variables $X\boldsymbol{\beta}$ that maximizes the ratio of the between class to within class variance, that is, the ratio $\boldsymbol{\beta}'\hat{\Sigma}_B\boldsymbol{\beta}/\boldsymbol{\beta}'\hat{\Sigma}_W\boldsymbol{\beta}$ with respect to $\boldsymbol{\beta}$, where $\hat{\Sigma}_B$ and $\hat{\Sigma}_W$ denote the between classes and within classes covariance matrices, respectively. The scaling of $\boldsymbol{\beta}$ is arbitrary, so to obtain a unique result, the constraint $\boldsymbol{\beta}'\hat{\Sigma}_W\boldsymbol{\beta} = 1$ is chosen. A set of p solutions, $(p - 1)$ of them suboptimal, can be found forming the columns of the matrix $B : p \times p$, where the ith column $\boldsymbol{\beta}_i$ is the solution of the two-sided eigenvalue problem satisfying $B'\hat{\Sigma}_W B = 1$ and $\hat{\Sigma}_B B = \hat{\Sigma}_B B\Lambda$, with Λ a diagonal matrix containing the eigenvalues sorted in descending order.

Let the rows of the matrix $\bar{X} : J \times p$ represent the class mean vectors $\bar{\mathbf{x}}_j'$ with $j = 1, 2, \ldots, J$, then $U = \bar{X}B$ is a transformation to a new set of means, known as the canonical means. Because $UU' = \bar{X}'\hat{\Sigma}_W^{-1}\bar{X}$, this indicates that the Mahalanobis distances between the class means in $\bar{X}$ are Pythagorean distances in the canonical space. If this Pythagorean canonical space is denoted by $\mathcal{R}$, the canonical means can be represented in an r-dimensional subspace $\mathcal{L}$. These canonical means can be related to the original variables in the CVA biplot. Similar to the PCA biplot, the canonical means are represented in $\mathcal{L}$ by $\bar{Z} = \bar{X}B_r$. From this it follows that the interpolation of $\mathbf{x}^* : p \times 1$ onto $\mathcal{L}$, in terms of the basis for $\mathcal{L}$ (scaffolding) is given by $\mathbf{z}^{*\prime} = \mathbf{x}^{*\prime}B$. Because prediction is the inverse of interpolation it follows that the interpolant (projection) $\mathbf{u}^* : p \times 1$ of $\mathbf{x}^* : p \times 1$ in the full p-dimensional space, is given by $\mathbf{u}^{*\prime} = \mathbf{x}^{*\prime}B$. Inverting this equation leads to $\mathbf{x}^{*\prime} = \mathbf{u}^{*\prime}B^{-1}$. When $\mathbf{u}^{*\prime}$ lies in $\mathcal{L}$, only the first r components $\tilde{\mathbf{z}}^* : r \times 1$ can be nonzero, so that the prediction is given by $\hat{\mathbf{x}}^{*\prime} = \tilde{\mathbf{z}}^{*\prime}B^r$ with B^r the first r rows of B^{-1}.

The directions of the interpolation biplot axes are given by the rows of B_r. Prediction biplot axes are found similar to those of the PCA biplot, but it has to be kept in mind that the columns of B_r are not orthogonal, leading to axes where the direction of the kth prediction biplot axis is given by the kth row of the matrix $[\text{diag}^{-1}(B^{r\prime}B^r)]B^{r\prime}$. As is the case with PCA biplots, these prediction axes are calibrated in terms of the original measurement scales. However, the interpolation and prediction biplot axes have different directions. The PCA biplot is a special case of the CVA biplot when $\hat{\Sigma}_W = I_p$, resulting in B being an orthogonal matrix with $B^r = B_r'$

Transforming to canonical means leads to a powerful reduction in the dimensionality of the problem: Flury (1997) proved that although p solutions can be obtained for the two-sided eigenvalue problem satisfying $B'\hat{\Sigma}_W B = I$ and $\hat{\Sigma}_B B = \hat{\Sigma}_W B\Lambda$, only the first $\min(p, J - 1)$ of these will be informative, leading to the subspace $\mathcal{L}$ being of dimension $\min(p, J - 1)$ in the case of CVA biplots.

The CVA biplot methodology described above is closely related to linear discriminant analysis (LDA). Ignoring prior

Table 1. Means and Variances of Feature Variables in the Platinum Froth Flotation Data Set

Class	SNE	ENTROPY	INERTIA	LOCHOM	GLLD
A ($N = 99$)					
Mean	0.539	−0.388	−0.867	0.501	−0.347
Variance	0.130	0.304	0.380	1.137	1.060
B ($N = 99$)					
Mean	0.742	−0.717	−0.233	−0.246	0.456
Variance	0.128	0.452	0.259	1.022	1.027
C ($N = 99$)					
Mean	−1.369	1.165	1.225	−0.228	−0.179
Variance	0.165	0.416	0.358	0.454	0.513

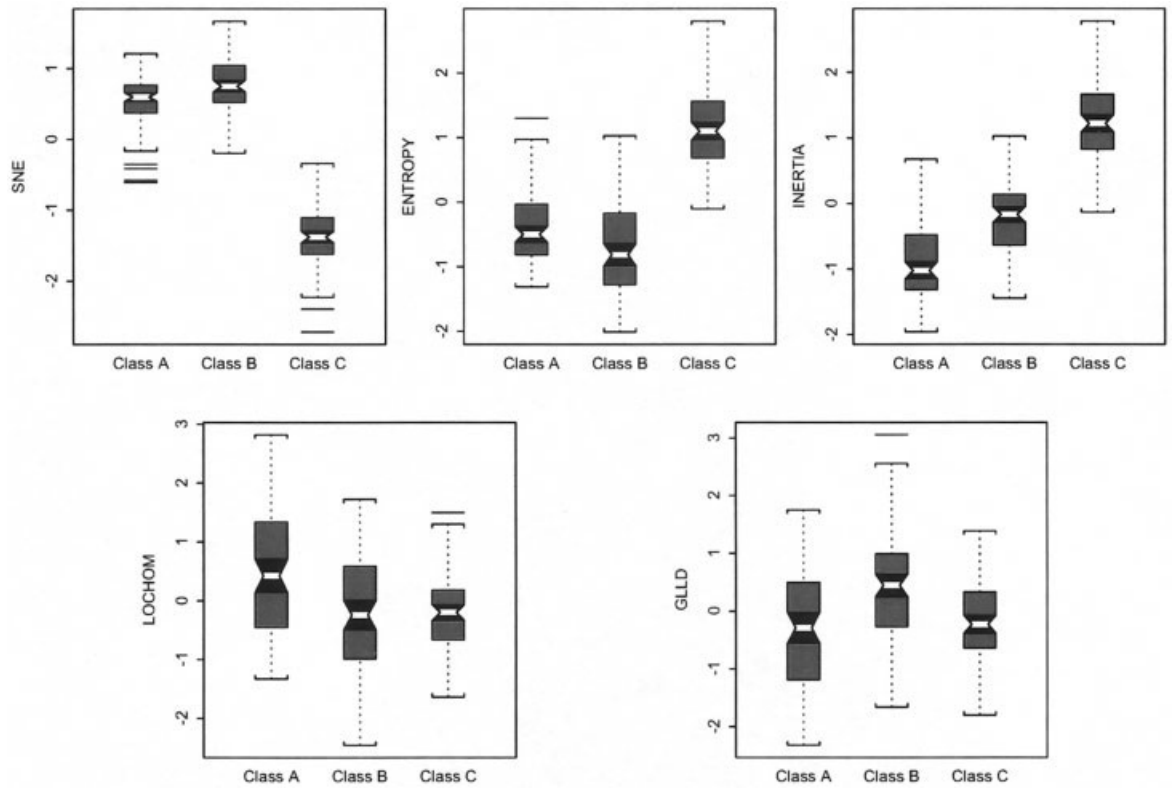

Figure 3. Notched boxplots of all variables in Platinum Froth Flotation data set.

probabilities, Gardner (2001) defined a classification region for class j as follows:

Definition 1: Classification Region. The classification region for class j is defined as $C_j = \{\mathbf{u} \in \mathcal{R}^K : \|\mathbf{u} - \bar{\mathbf{u}}_j\| < \|\mathbf{u} - \bar{\mathbf{u}}_h\|$ for all $h \neq j\}$ for $j = 1, 2, \ldots, J$, where $\mathcal{R}^K$ is the canonical space, $\bar{\mathbf{u}}_j$ is the jth canonical mean, and $\|\mathbf{u} - \bar{\mathbf{u}}_j\|$ is the Pythagorean distance between $\mathbf{u}$ and $\bar{\mathbf{u}}_j$.

The CVA biplot is an r-dimensional representation of the class means after a transformation to the canonical space. In this transformation the first r columns of the matrix B are used to transform the matrix of class means $\bar{X}$ in the observation space to $\bar{Z}_{\mathrm{CVA}} = \bar{X}B_{\mathrm{LDA},r}$ in the canonical space. Because this transformation is linear the data matrix X can likewise be transformed into a matrix $Z_{\mathrm{CVA}} = XB_{\mathrm{LDA},r}$. Gardner (2001) has plotted Z_{CVA} and $\bar{Z}_{\mathrm{CVA}}$ simultaneously for a CVA biplot, showing the spread of the observations around their mean values. Furthermore an algorithm is provided for equipping CVA biplots with classification regions.

Alpha bags

The CVA biplot where Z_{CVA} and $\bar{Z}_{\mathrm{CVA}}$ are simultaneously plotted provides for visual appraisal of the degree of overlap between classes. However, what is needed is a measure of describing and quantifying overlap in two-dimensional (2-D) space. Gardner (2001) proposes alpha bags for this purpose.

Table 2. Correlation Matrix of Feature Variables in the Platinum Froth Flotation Data Set

	SNE	ENTROPY	INERTIA	LOCHOM	GLLD
SNE	1.000	−0.894	−0.833	0.001	0.275
ENTROPY	−0.894	1.000	0.746	0.293	−0.532
INERTIA	−0.833	0.746	1.000	0.013	−0.288
LOCHOM	0.001	0.293	0.013	1.000	−0.940
GLLD	0.275	−0.532	−0.288	−0.940	1.000

In univariate data the boxplot is a useful tool to summarize the location, spread, skewness, and outliers in the data. The bagplot proposed by Rousseeuw et al. (1999) is a bivariate generalization of the boxplot. The univariate boxplot is based on ranks. For the construction of the bagplot the univariate rank concept is generalized to the concept of halfspace location depth of a point relative to a bivariate data set (see Rousseeuw et al., 1999).

Definition 2: Halfspace Location Depth (Tukey, 1975). The halfspace location depth of a point θ, in two dimensions, relative to a bivariate data cloud $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, denoted by ldepth($\theta$, X), is the smallest number of $\mathbf{x}_i$ contained in any closed halfplane with boundary line through θ.

In one dimension, the location depth of a point is defined as the minimum number of data points, when the data set is ordered on the real line, to the left or to the right of the point, respectively (Donoho and Gasko, 1992). Intuitively it can be visualized as the smallest "distance" of the point from the "edge" of the data set.

In two dimensions the concept of halfspace location depth can be visualized by a line extending beyond the maximum and minimum values of the two variables drawn through a point θ. Rotating this line through 360°, the number of points on the same side of the line is counted at each rotated position. The halfspace location depth is the smallest count of data points for any position of the line in the 360° rotation. For calculation purposes an efficient algorithm, $O(n \log n)$ computations, is given by Ruts and Rousseeuw (1996). The halfspace location

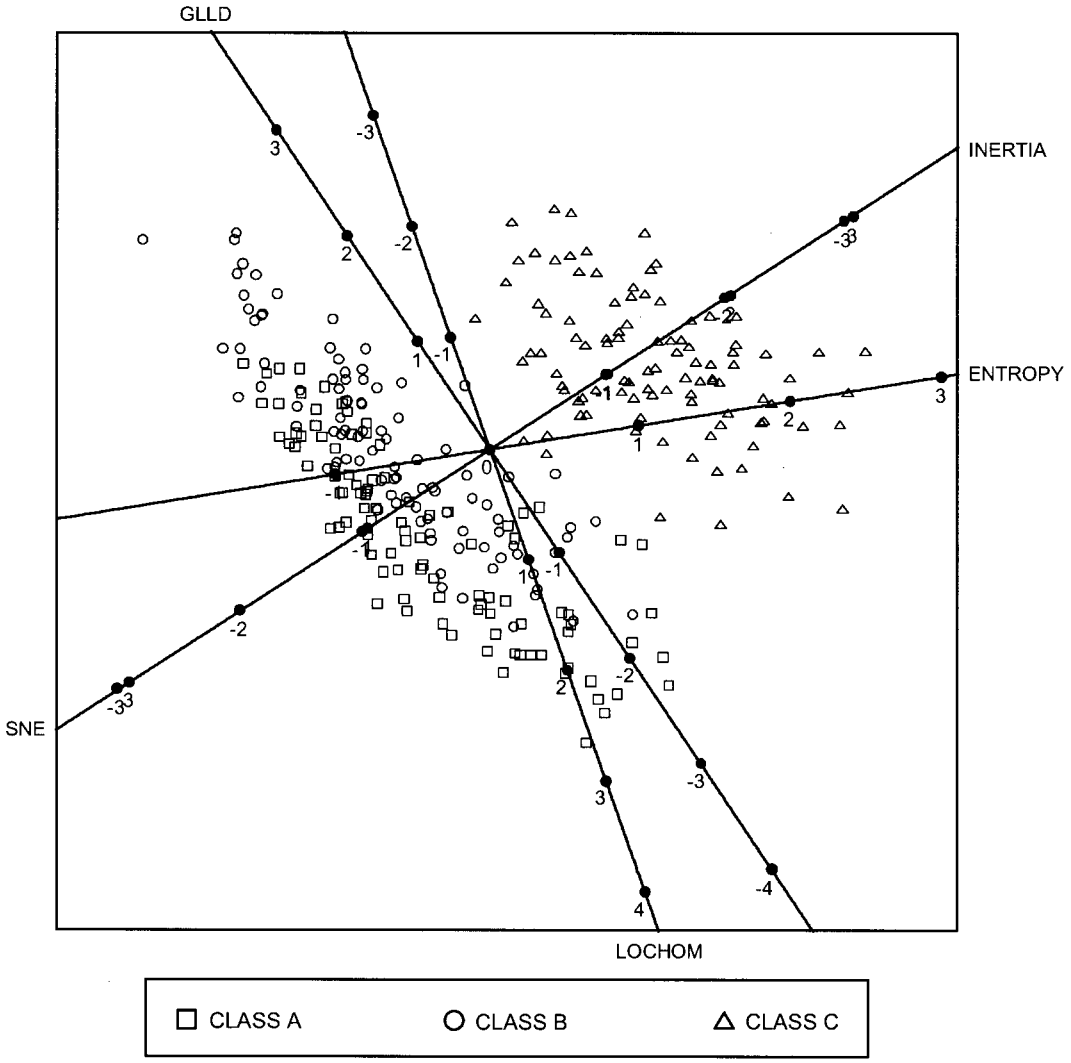

Figure 4. PCA biplot of the Platinum Froth Flotation data set.

The original variables (five features extracted from froth images) are shown on different axes superimposed on a principal component analysis biplot of the data. Correlations between variables are easily discernible from the angles between the axes of the original variables (small angles represent large correlations and vice versa). In addition, any given sample point can be related to the original observations by perpendicular projections on the variable axes.

depth is the basis for the definition of a depth median and depth region used in the construction of the bagplot and alpha bags.

Definition 3: Depth Region, D_k (Ruts and Rousseeuw, 1996). The depth region D_k is the set of all θ with ldepth(θ, X) $\geq k$. The depth regions are convex polygons such that D_{k+1} is contained in D_k but they do not necessarily coincide with the convex layers resulting from convex hull peeling, given that the latter process first removes the outer layers before fitting the next convex hull.

Construction of the bag of the bagplot is accomplished as follows: Let $\#D_k$ be the number of data points contained in the depth region D_k, then the value k for which $\#D_k \leq \left\lfloor \frac{n}{2} \right\rfloor \leq \#D_{k-1}$ is determined. These depth regions are used to interpolate linearly between D_k and D_{k-1}, relative to the depth median T^*, to obtain the bag B of $n/2$ observations with the largest depth. The bag obtained in this way is a convex polygon. By using the interpolation algorithms provided in the *bagplot* function of Rousseeuw and Ruts (1997), Gardner (2001) constructs depth contours enclosing the exact proportion of bivariate data points specified by the user. The resulting contours are called α-bags and are proposed as graphical summaries of bivariate data points. Instead of finding a bag B containing $n/2$ observations, the 50% cut-off is replaced by a value of α ranging between 0 and 100. Typically a value of 90 or 95 will be useful for enclosing a cluster of observations, excluding the 10 or 5% of the observations at the extremes of the cluster. Comparing the α-bag to depth contours or convex peeling, this method allows for controlling the number of observations outside the enclosure. The following two case studies illustrate the usefulness of PCA biplots, CVA biplots, and α-bags in industrial plant performance.

Case Study 1: Platinum Froth Flotation Plants

Froth flotation is a process of major economic importance, mainly as a result of its efficiency in the treatment of large ore

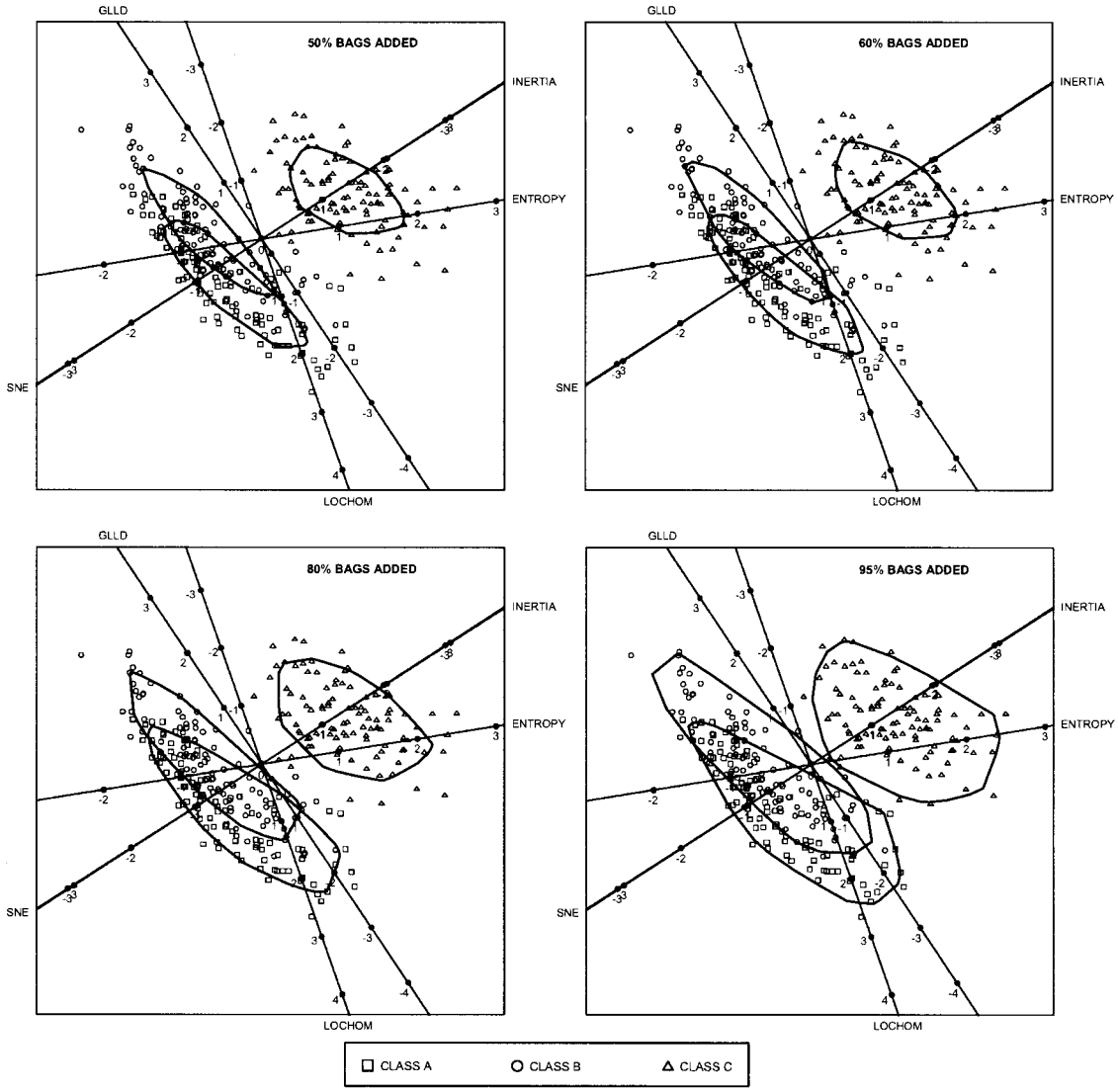

Figure 5. PCA biplot of the Platinum Froth Flotation data set, similar to the one shown in Figure 4, with alpha bags superimposed on each froth class.

The percentages associated with the bags indicate the percentage of observations contained in each bag and also give an indication of the separability of the classes. For example, with 95% bags, it can be seen that classes B (open circles) and C (open triangles) barely touch each other (see panel lower right) because the inner 95% of the samples in each class do not overlap.

tonnage and its relative ease of implementation in industry. The process is complex, and in practice the control of industrial flotation plants is often based on the visual appearance of the froth phase, and depends on the experience and ability of a human operator. A standard control cycle constitutes the fixing of initial set points, a settling period for transient dynamics to subside, a period of measurement and evaluation, and a final estimate of appropriate set points. Some studies have indicated that operators often tend to make the periods for settling and measurement too small. Apart from these aspects, the inexperience or inability of the operator can have a further significant impact on the control of the plant. Consequently, optimal control is not usually maintained, especially where incipient erratic behavior in the plant is difficult to detect.

Process and image variables that are measured on-line are stored in a data base, from where a predefined set of process variables are used to construct a process map (see Figure 2), which can be interpreted by an experienced process operator. Alternatively, the map is used by some form of a knowledge-based or decision support system to assist the operator in the final control decision. Although some work is currently being done to assess the potential of closed-loop model-based control systems, such systems are not presently used in industry, especially because flotation processes are characterized by considerable disturbances associated with variations in the ore feed.

In this case study, five image features are projected to a process chart by use of biplots. These gray level dependency matrix features were based on those originally proposed by Haralick et al. (1973), Haralick (1979), and Sun and Wee (1983), for the characterization of textures (Muhamad and Deravi, 1992; Siew et al., 1988). The SNE (small number

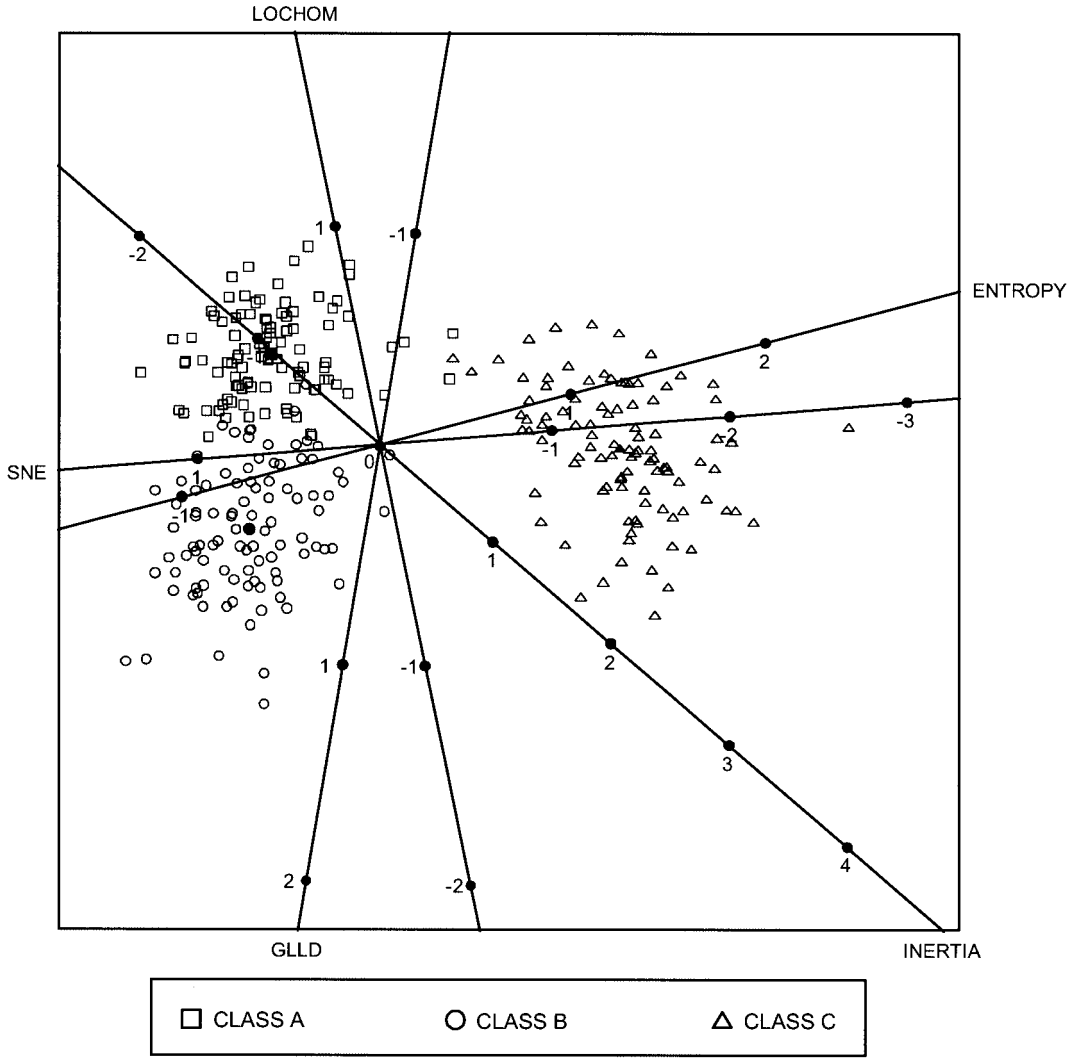

Figure 6. CVA biplot of the Platinum Froth Flotation data set.

The original variables (five features extracted from froth images) are shown on different axes superimposed on a canonical variate biplot of the data. Class means are indicated by solid symbols. Variables contributing to the separation between classes are easily discernible. The angles between the axes provide indications of the correlations between the original variables (small angles represent large correlations and vice versa). In addition, any given sample point can be related to the original observations by perpendicular projections on the variable axes.

emphasis) feature defined below is based on a 2-D array Q, where $Q(r, s)$ can be considered as frequency counts of grayness variation of a processed image. The dimensions of the array are $r \times s$, where r represents the number of gray levels and s is the number of possible neighbors to a pixel in an image

$$\text{SNE} = \sum_r \sum_s [Q(r, s)/s^2] \Big/ \sum_r \sum_s [Q(r, s)] \qquad (1)$$

The SNE is a rough measure of the average bubble size of the froth; that is, the higher the SNE, the smaller the bubbles. In contrast to the first feature, derived from a *neighboring* gray level dependency matrix (Sun and Wee, 1983), the following features defined in Eqs. 2–5 are derived from *spatial* gray level dependency matrices (Haralick, 1979; Haralick et al., 1973). The features can be computed (for positive integers d, a) by counting the number of times that the difference between each element in the image function $f(i, j)$ and its neighbors is equal to or less than a at a certain distance d

$$\text{ENTROPY} = -\sum_i \sum_j [f(i, j, d, a)\log\{f(i, j, d, a)\}] \qquad (2)$$

The entropy is a measure of the complexity of the image; that is, a complex image tends to have a higher entropy than that of a simple one

$$\text{INERTIA} = \sum_i \sum_j [(i - j)^2 f(i, j, d, a)] \qquad (3)$$

The inertia is a measure of the number of local variations in the image. Therefore an image with a large number of local variations will have a larger inertia value

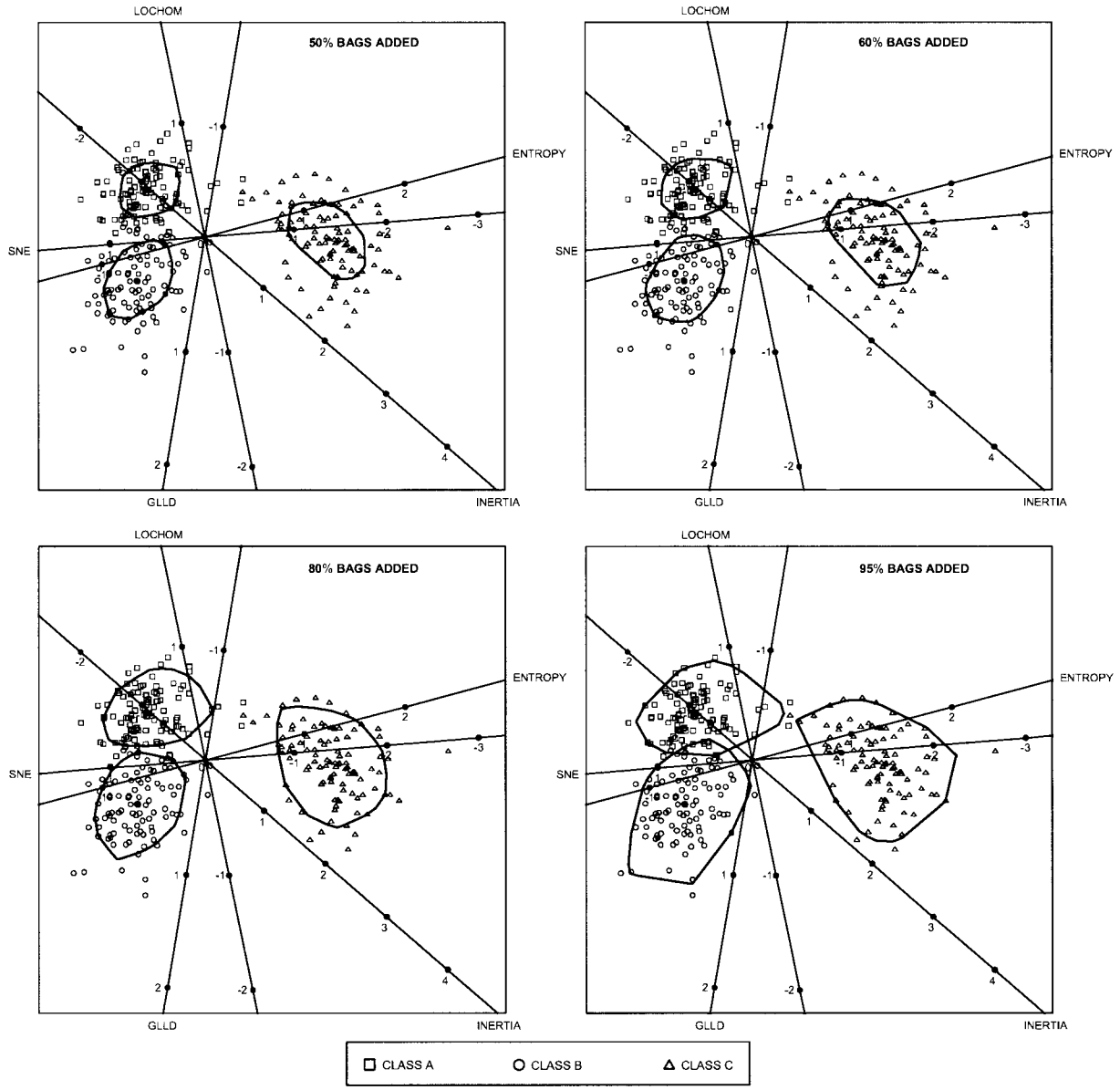

Figure 7. CVA biplot of the Platinum Froth Flotation data set, similar to the one shown in Figure 6, but with alpha bags superimposed on each froth class.

The percentages associated with the bags indicate the percentage of observations contained in each bag and also give an indication of the separability of the classes. For example, with 95% bags, it can be seen that the innermost 95% of observations in classes B (open circles) and C (open triangles) do not overlap (see panel lower right).

$$\text{LOCHOM} = \sum_i \sum_j [f(i, j, d, a)/\{1 - (i + j)\}^2] \qquad (4)$$

The local homogeneity (LOCHOM) is a measure of the tendency of similar gray levels to be neighbors. The distance metric d implicit in the preceding equation can be defined by (Haralick, 1973) $d\{(k, l), (m, n)\} = \max\{|k - m|, |l - n|\}$, where (k, l) and (m, n) are the coordinates of two pixels separated by distance d

$$\text{GLLD} = \sum_i \sum_j [(i - \mu_x)(j - \mu_y)\{f(i, j, d, a)/\sigma_x\sigma_y\}] \qquad (5)$$

In Eq. 5, μ_x and σ_x are, respectively, the mean and standard deviation of the row sums of the matrix; and μ_y and σ_y are the mean and standard deviation of the column sums. This is a measure of the gray level linear dependencies (GLLD) in the image. A normalization factor may be applied to either the feature matrix or to each of the feature parameters. Moolman (1995) and Moolman et al. (1998) give an extended review of these parameters and their use in the characterization of flotation froth images.

The Platinum Froth Flotation data set consists of a matrix with 497 rows representing the plant observations/cases and 6 columns representing the above five feature variables as well as

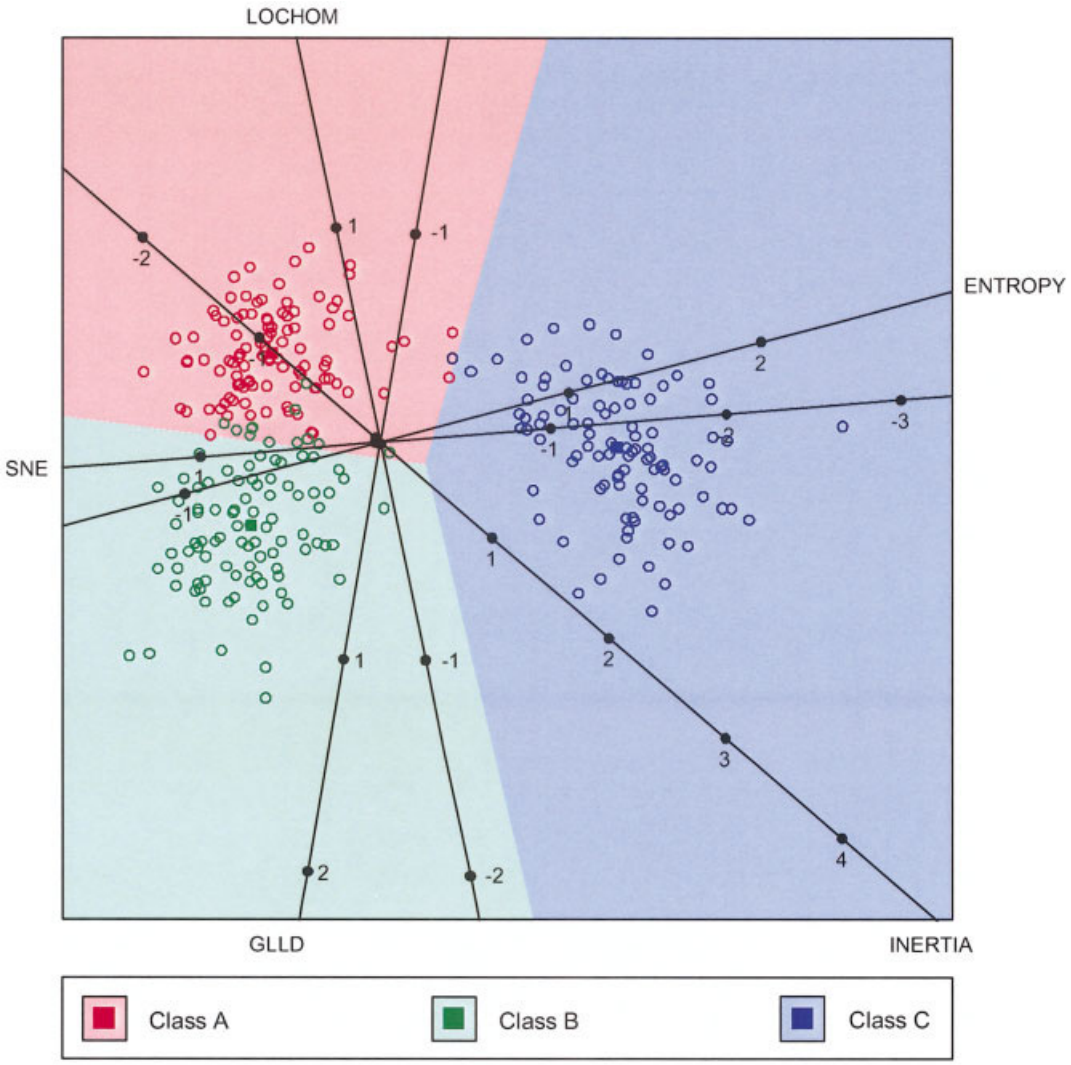

Figure 8. CVA biplot of the Platinum Froth Flotation data set with linear classification regions added.

a class variable. The class is determined by a plant expert, with B the desired class representing conditions of optimal stability. Class A represents too stable a froth, preventing optimal recovery of the concentrate, whereas too unstable a froth having a similar net effect is denoted by class C. The means and variances of these three classes for each of the feature variables are given in Table 1.

In Figure 3 notched boxplots for all five feature variables are displayed. The notches correspond to approximate 95% confidence intervals for the respective medians (see McGill et al., 1978). Therefore if notches do not overlap it is an indication that a null hypothesis of no difference between class medians should be rejected at an approximate 5% significance level.

Perusal of Figure 3 shows that too stable a froth (class A) has significantly higher ENTROPY values, significantly lower INERTIA values, significantly higher LOCHOM values, and significantly lower GLLD values than conditions of optimal stability (class B). On the other hand, too unstable a froth (class C) differs from class B by significantly lower SNE, higher ENTROPY, higher INERTIA, and lower GLLD values. However, Table 2 shows the five feature variables not to be uncorrelated. Thus any decision based on univariate treatment of the

Table 3. Classification Matrix Obtained by Applying CVA Biplot Classification to the Platinum Froth Flotation Data Set

	Observations Belonging to		
Observations Classified to	Class A	Class B	Class C
Class A	98	11	1
Class B	1	88	0
Class C	0	0	98
Total	99	99	99

Table 4. Classification Matrix Obtained by Applying QDA Biplot Classification to the Platinum Froth Flotation Data Set

	Observations Belonging to		
Observations Classified to	Class A	Class B	Class C
Class A	96	3	0
Class B	2	96	0
Class C	1	0	99
Total	99	99	99

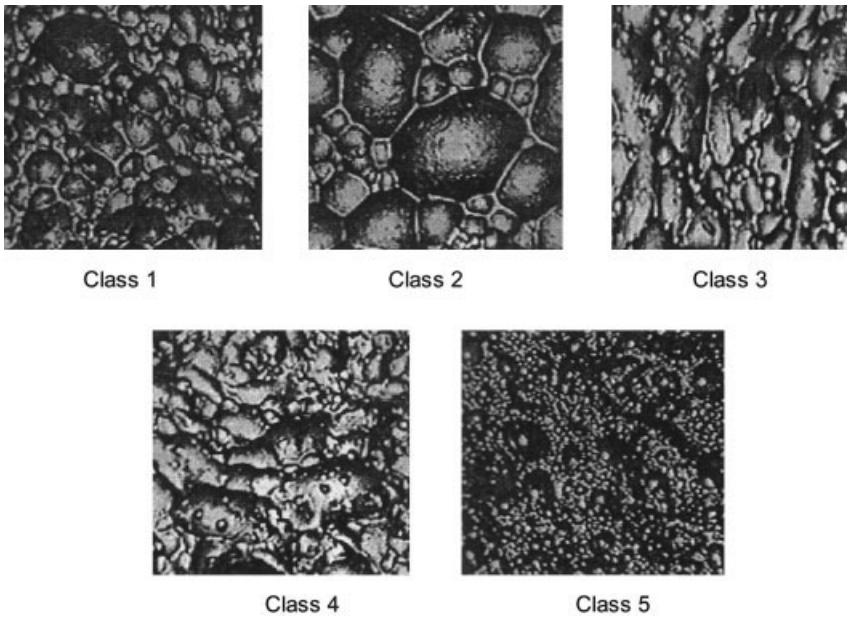

Figure 9. Typical examples of froth images of classes 1–5 in the copper flotation plant.

data set is in danger of being misleading. Decisions regarding differences among the three classes should take cognizance of the dependence structure in the 5-D feature variable space.

It follows from Figure 3 that the five feature variables are expressed in comparable units of measurement. Therefore, a PCA can be performed on the covariance matrix and an associated PCA biplot constructed to assist in describing the variation in the 5-D Platinum Froth Flotation data set. Because the first two principal components account for 93.58% of the variation in the data, the PCA biplot in Figure 4 captures almost all the variation among the data in the full 5-D observational space.

Because the five axes representing the feature variables in the biplot displayed in Figure 4 are calibrated in the original units, the biplot can be used to determine values of a data point for all feature variables by projecting orthogonally onto each axis successively. The adequacies associated with these five axes range from 0.3120 to 0.5274. Therefore, all five axes are reasonably well represented in the biplot space. It is clear from the figure that classes B and A have the largest variation along the GLLD and LOCHOM axes. The shape of the cloud of class C points differs from that of classes A and B, indicating differences in its variation pattern from those of the latter two classes. Furthermore class C is well separated from classes A and B. It is also clear that this is mainly attributed to relatively larger ENTROPY and INERTIA values and smaller SNE values.

Although the PCA biplot discussed here is constructed to explain variance patterns in particular, it also conveys information about the correlation among the variables. The strong negative correlations between LOCHOM and GLLD, between INERTIA and SNE, and between SNE and ENTROPY result in small angles between the respective axes. The relatively small angle between the ENTROPY and INERTIA axes together with calibrations increasing in the same directions points to a large positive correlation. The almost zero correlation between some of the variables; for instance SNE and LOCHOM leads to relatively large angles (almost rectangular in some instances) between the respective axes.

Turning once again to the overlap among the three classes it seems that, although classes A and B exhibit a considerable amount of overlap, it does appear as if class B is concentrated in a region falling between classes A and C. This is true of variables ENTROPY, INERTIA, and SNE.

Figure 5 shows the PCA biplot of Figure 4 but with 50, 60, 80, and 95% bags superimposed. The nature as well as the degree of overlap among the three classes can now be appraised: Class C is almost completely separated from classes A and B; the 50% bags of classes A and B already have a considerable amount of overlap, although some differences mainly along the LOCHOM and GLLD axes are apparent.

To maximize differences among the three classes a CVA biplot can be constructed. Because there are three classes, the canonical space is of dimension two, irrespective of the number of variables. Classification in this two-dimensional space together with the resulting biplot will be optimal. The three classes appear to be well separated in the CVA biplot in Figure 6. In addition all axes are equipped with scales pertaining to the original observational space. The mean values of all classes (the solid symbols in Figure 6) for all variables can be accurately obtained from the biplot axes, as is evident by comparing these values with the actual observed values given in Table 1.

Figure 7 contains the CVA biplot of Figure 6, but with selected α-bags added. For example, with 95% bags, it can be seen that class B (open circles) and class C (open triangles) can be classified with at least 95% accuracy, given that the two bags barely touch each other (see lower right panel). Also, it is clear from Figure 7 that the three classes are well separated, with a small degree of overlap between classes A and B evident in the case of the 95% bags. The variables contributing mainly to the separation can also easily be established.

The CVA biplot can also be equipped with classification regions. Classification regions are useful for classifying future observations of unknown origin by interpolating them

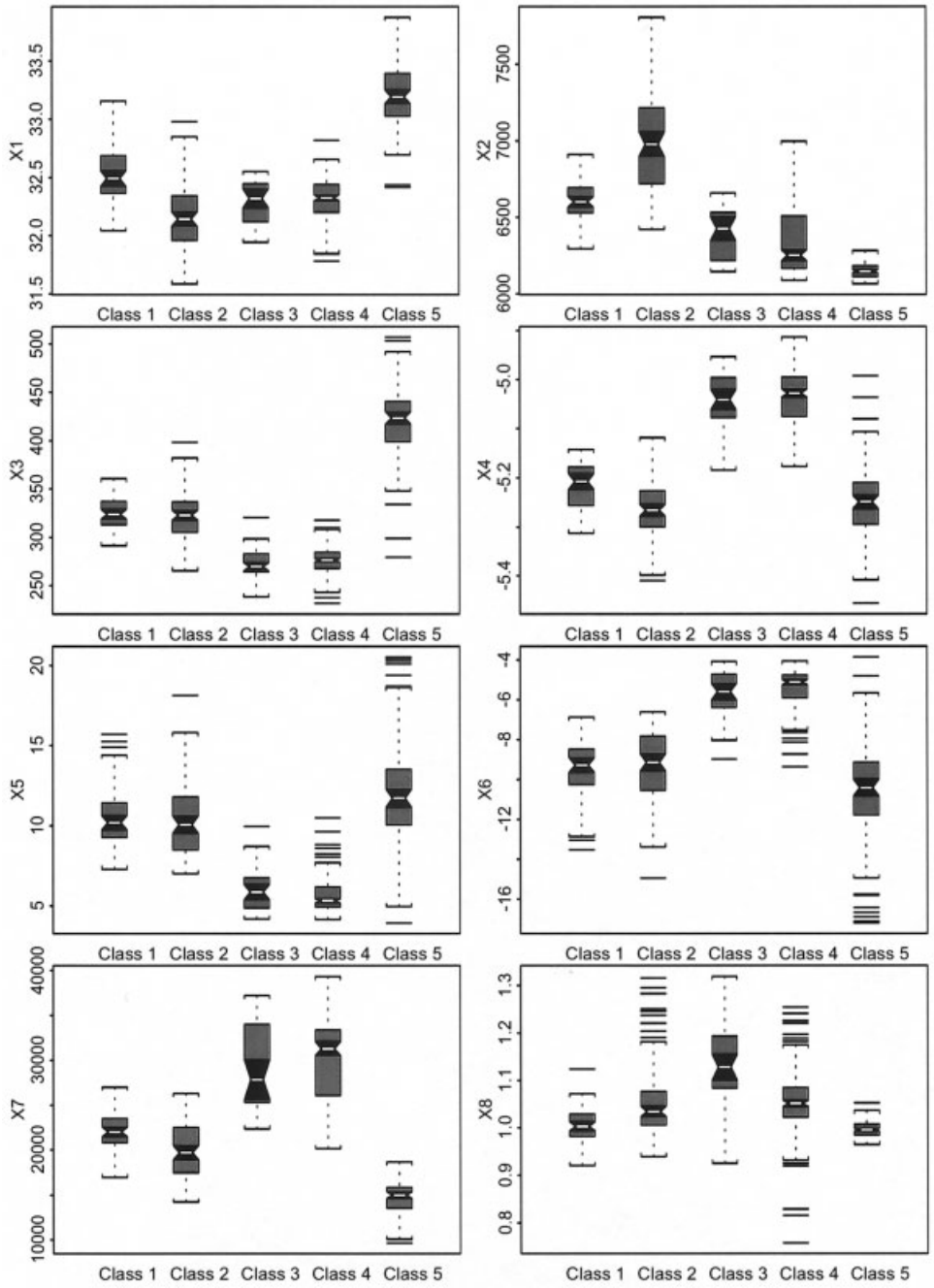

Figure 10. Notched boxplots of all feature variables in Copper Froth Flotation data set showing univariate differences among the five classes of froths shown in Figure 9.

into the biplot. This is best done by a computer program using interpolation axes discussed in the section describing biplot methodology. Figure 8 shows a CVA biplot of the Platinum Froth Flotation data set with classification regions added. The original observations are also shown.

It follows from Table 3 that the CVA biplot classification procedure results in an apparent error rate of 4.04%. This can be improved by applying the quadratic discriminant analysis (QDA) biplot procedure as proposed by Gardner (2001). The latter procedure results in an apparent error rate of only 2.02% (see Table 4).

The classification regions obtained with the QDA biplot classification procedure, not shown here, are linear and similar to the classification regions in Figure 8. Although it is well known that the apparent error rate is optimistically biased (for example, McLachlan, 1992) it is used here for comparison purposes only.

Case Study 2: Copper Froth Flotation Plants

In the second case study, data obtained from a South African copper flotation plant are considered. The plant treats nearly

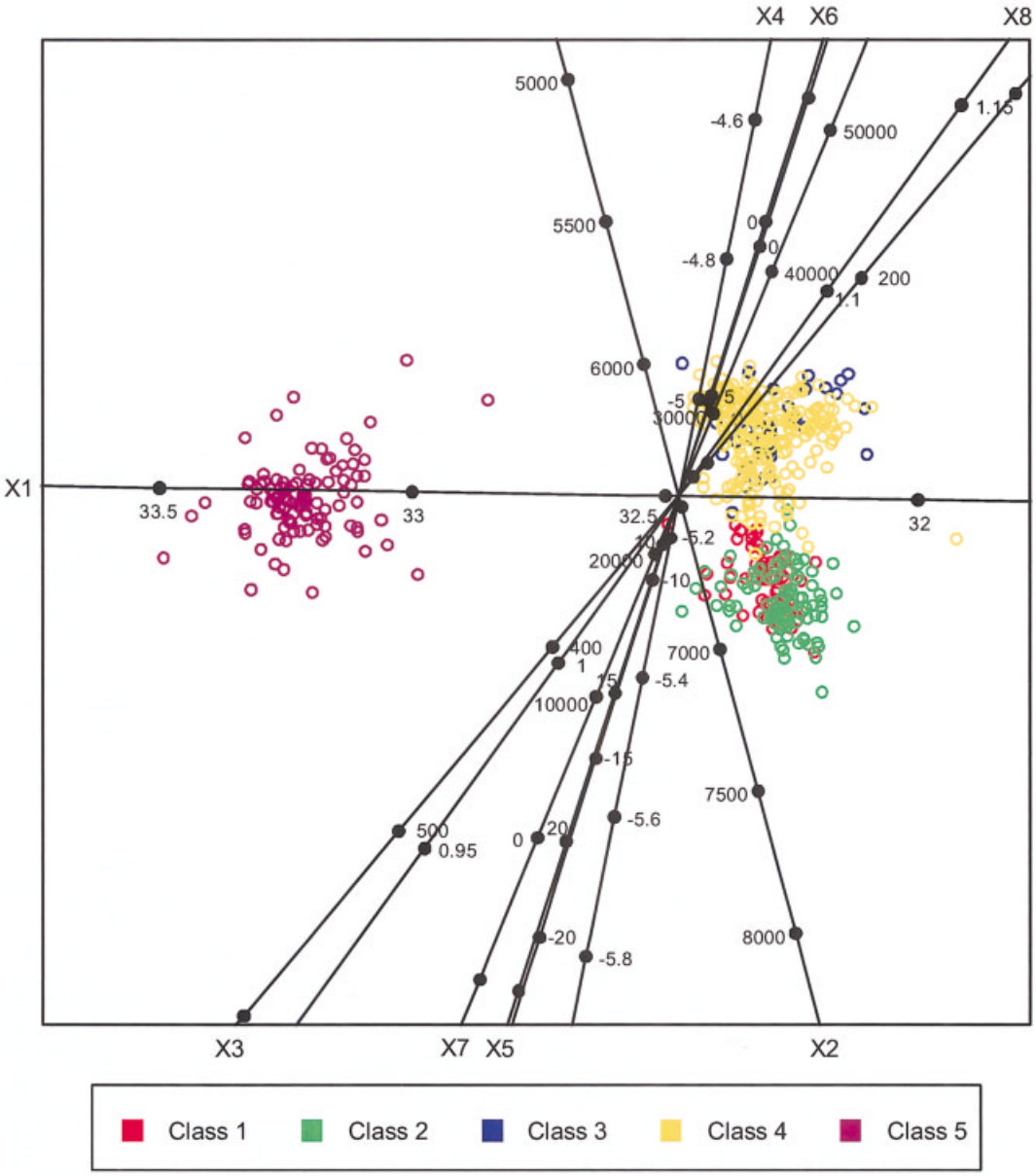

Figure 11. CVA biplot of Copper Froth Flotation data set.

The original variables (X_1– X_8) are shown as different axes superimposed on a canonical variate analysis biplot of the data. Variables contributing to the separation between classes are easily discernible. The angles between the axes provide indications of the correlations between the original variables (small angles represent large correlations and vice versa). In addition, any given sample point can be related to the original observations by perpendicular projections on the variable axes.

80,000 tons of relatively coarse ore (80–300 μm) per day. The principal copper minerals in the ore are bornite and chalcopyrite, whereas chalcocite, valleriite, and cubanitite are also present in lower concentrations. The feed to the plant has an approximate copper assay of 0.5%, where it is concentrated to about 36% after flotation, which is enhanced by two reagents only: a collector and a frother.

The appearance of the froth plays a significant role in the operation of the plant. Unlike the froths considered in the previous case study, the color of the froth is an indication of the loading of the valuable mineral (copper), whereas the texture of the froth is influenced by the froth pH and the types of minerals being floated. Simultaneous interpretation of these features is very difficult without the use of digital image analysis. In this particular case, five froth structures were considered, briefly summarized as follows.

- *Froth 1: Ideal froth structure*. The bubbles are heavily loaded with minerals and display a characteristic reflective spot with a watery sheen on top of each bubble. This is the desired class of froth, to be maintained by means of reagent flow rate and pulp level control.
- *Froth 2: Polyhedral bubbles*. The froth is deep and well drained with adequate separation between minerals and gangue. The depth of the froth phase depends on the pulp level in particular.
- *Froth 3: Ellipsoidal bubbles*. An excessive tough froth

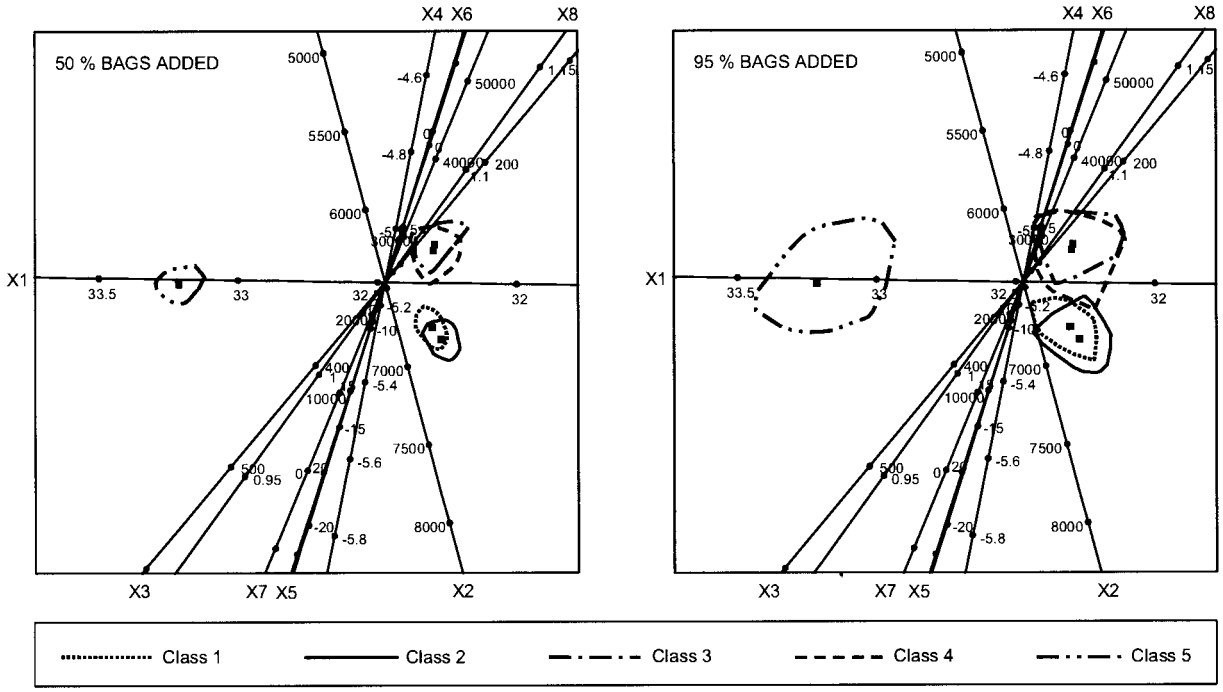

Figure 12. CVA biplot of the Copper Froth Flotation data set similar to the one shown in Figure 11, but with alpha bags superimposed on each froth class.

The percentages associated with the bags indicate the percentage of observations contained in each bag and also give an indication of the separability of the classes. For example, with 95% bags, it can be seen that the innermost 95% of observations in classes 2 and 3 are completely separated (see right panel).

that shows resistance to flow, caused by (among other phenomena) too low a pulp level, too high a pulp density, or by the character of the minerals being floated.

- *Froth 4: Irregular froth structure.* An excessively stable froth that has an adverse effect on the concentration of the valuable mineral.
- *Froth 5: Depleted froth.* A shallow froth with a low mineral concentration, characterized by numerous small spherical bubbles.

As can be seen from Figure 9, the deep dry froths of classes 1 and 2 are very similar in appearance. Likewise, the tough sticky froths of classes 3 and 4 are almost indistinguishable in appearance, with the main difference being the ellipsoidal bubbles typical of the class 3 froths.

Qualitative differences and similarities among classes 1 through 5 in Figure 9 are obvious. However, the question is: How can these differences and similarities be quantified keeping in mind the multidimensional character of the underlying process variables? This question is addressed by means of CVA biplot methodology.

Eight features similar to those described in the previous case study were extracted from digitized images of the froths. The first four were neighboring gray level dependency matrix features (LNE, NNU, SM, and ϵ_N), whereas the last four were spatial gray level dependency matrix features (ENERGY, ENTROPY, INERTIA, and RA).

Although it is difficult to attach physical meaning to the features, LNE, NNU, and ϵ_N are measures of the coarseness of the texture, but it is difficult to explain which specific textural characteristics they represented (Siew et al., 1988; Sun and Wee, 1983). SM is a measure of the homogeneity of the image. These features are defined as (see Case Study 1)

$$\text{LNE} = \sum_r \sum_s [s^2 Q(r,s)] \Big/ \sum_r \sum_s [Q(r,s)] \tag{6}$$

$$\text{NNU} = \sum_s \left[\sum_r Q(r,s) \right] \Big/ \sum_r \sum_s [Q(r,s)] \tag{7}$$

$$\text{SM} = \sum_r \sum_s [Q(r,s)]^2 \Big/ \sum_r \sum_s [Q(r,s)] \tag{8}$$

$$\varepsilon_N = \sum_r \sum_s [Q(r,s) \log Q(r,s)] \Big/ \sum_r \sum_s [Q(r,s)] \tag{9}$$

The spatial gray level dependency matrix feature RA, not explicitly described here, was defined ad hoc as a measure of the variation in the spatial gray level dependency matrix features with the rotational direction of the image (0, 45, 90, and 135°). The other spatial gray level dependency matrix features are defined as follows

$$\text{Energy } (E) = \sum_i \sum_j [f(i,j,d,a)]^2 \tag{10}$$

This is a measure of the homogeneity of the image. For an image that is not homogeneous, the matrix will have a large number of small off-diagonal entries, and hence the energy (E) will be small. The diagonal and region close to the diagonal represent transitions between similar gray levels. The entropy (ϵ_S) and inertia (I) features are described by Eqs. 2 and 3. For convenience, the eight features LNE, NNU, SM, ϵ_N, RA, ENERGY, ENTROPY, and ENERTIA are henceforth referred to as $X_1, X_2, \ldots, X_8$, respectively.

The Copper Froth Flotation data set differs in two important respects from the data set considered in Case Study 1: The eight feature variables considered here are not in commensurable units and the class variable differentiates between five classes. Therefore, we restrict our attention to CVA biplots only.

Perusal of the boxplots in Figure 10 shows differences, as well as similarities, among the five classes but, because these univariate plots do not take into account the dependency structure of the variables, they cannot be used to infer how the five classes are separated in the eight-dimensional observation space. The CVA biplots in Case Study 1 were optimal for discriminating between the three classes considered there in the sense that all differences among the three classes were channeled into the 2-D canonical space. Transforming the observation space of the Copper Froth Flotation data set to the canonical space reduces the problem to four dimensions. The CVA biplot methodology proposed by Gardner (2001) allows for a discriminant analysis of the five groups in all four dimensions of the canonical space but the graphical display of the biplot is limited to at most three dimensions. Although three dimensional biplots can be constructed (see, for example, Gardner, 2001) it will be argued here that 2-D CVA biplots can be used effectively for exploring the four-dimensional canonical space of the Copper Froth Flotation data set. Similar to the PCA biplot the quality of display for a CVA biplot reflects the ratio of the variance represented by the two eigenvectors forming the scaffolding for the biplot to the total variance in the canonical space. The quality of the 2-D CVA biplot space of Figure 11 is an excellent 91.7%.

It is clear from Figure 11 that class 5 is well separated from the other four classes, whereas classes 1 and 2 have a considerable overlap but are apart from classes 3 and 4. The latter two classes exhibit considerable overlap. Note that the axes representing the feature variables are calibrated according to the measurement scales in the original observation space. Not only can these axes be used to determine the class means for all variables, but variables contributing to overlap or separation can easily be discerned. To quantify the overlap, α-bags can be added to the biplot. Figure 12 is an example where 50 and 95% bags are fitted to the CVA biplot of Figure 11. Only class means are shown in Figure 12.

Figure 12 suggests the following: Complete separation of class 5 from the other classes; almost complete separation between classes 1 and 2 taken together and classes 3 and 4 taken together; almost complete overlap between the deepest 50% observations of classes 3 and 4 with slightly less overlap between classes 1 and 2. The overlap between the latter classes can easily be more accurately quantified by considering other α-bags as well.

The confusion matrix given in Table 5 shows the results of the biplot discriminant analysis in two dimensions. The classification results given in this table are in agreement with the biplots of Figures 11 and 12. The difficulty of distinguishing in two dimensions between classes 1 and 2 as well as between classes 3 and 4 is highlighted. An overall apparent error rate of 0.4105 clearly exposes the insufficiency of the 2-D CVA biplot classification procedure for the Platinum Copper Flotation data set.

Applying the CVA biplot classification procedure in three dimensions leads to an improvement of the apparent error rate to 0.3320. The resulting classification results given in Table 6 disclose that this improvement is mainly the result of a better distinction between classes 1 and 2, whereas classes 3 and 4 remain almost inseparable. Therefore, application of the biplot classification procedure in the full 4-D canonical space should be considered.

Table 7 contains the classification results when the biplot classification procedure is applied in the full canonical space. Classes 3 and 4 are now much better separated, although the overall apparent error rate is still 0.1610 but cannot be improved by considering higher dimensions.

Unfortunately, human visual perception is limited to three dimensions. Therefore, although biplot classification procedures can be carried out in any dimension, visual displays are

Table 5. Classification Matrix Obtained by Applying CVA Biplot Classification in Two Dimensions to the Platinum Copper Flotation Data Set

Observations Classified to	Observations Belonging to				
	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	33	33	1	14	0
Class 2	22	63	0	1	0
Class 3	0	0	18	113	0
Class 4	0	0	20	82	0
Class 5	0	0	0	0	97
Total	55	96	39	210	97

Table 6. Classification Matrix Obtained by Applying CVA Biplot Classification in Three Dimensions to the Platinum Copper Flotation Data Set

Observations Classified to	Observations Belonging to				
	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	53	18	1	6	0
Class 2	2	78	0	5	0
Class 3	0	0	18	113	0
Class 4	0	0	20	86	0
Class 5	0	0	0	0	97
Total	55	96	39	210	97

Table 7. Classification Matrix Obtained by Applying CVA Biplot Classification in Four Dimensions to the Platinum Copper Flotation Data Set

Observations Classified to	Observations Belonging to				
	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	53	18	0	8	0
Class 2	2	78	0	4	0
Class 3	0	0	27	36	0
Class 4	0	0	12	162	0
Class 5	0	0	0	0	97
Total	55	96	39	210	97

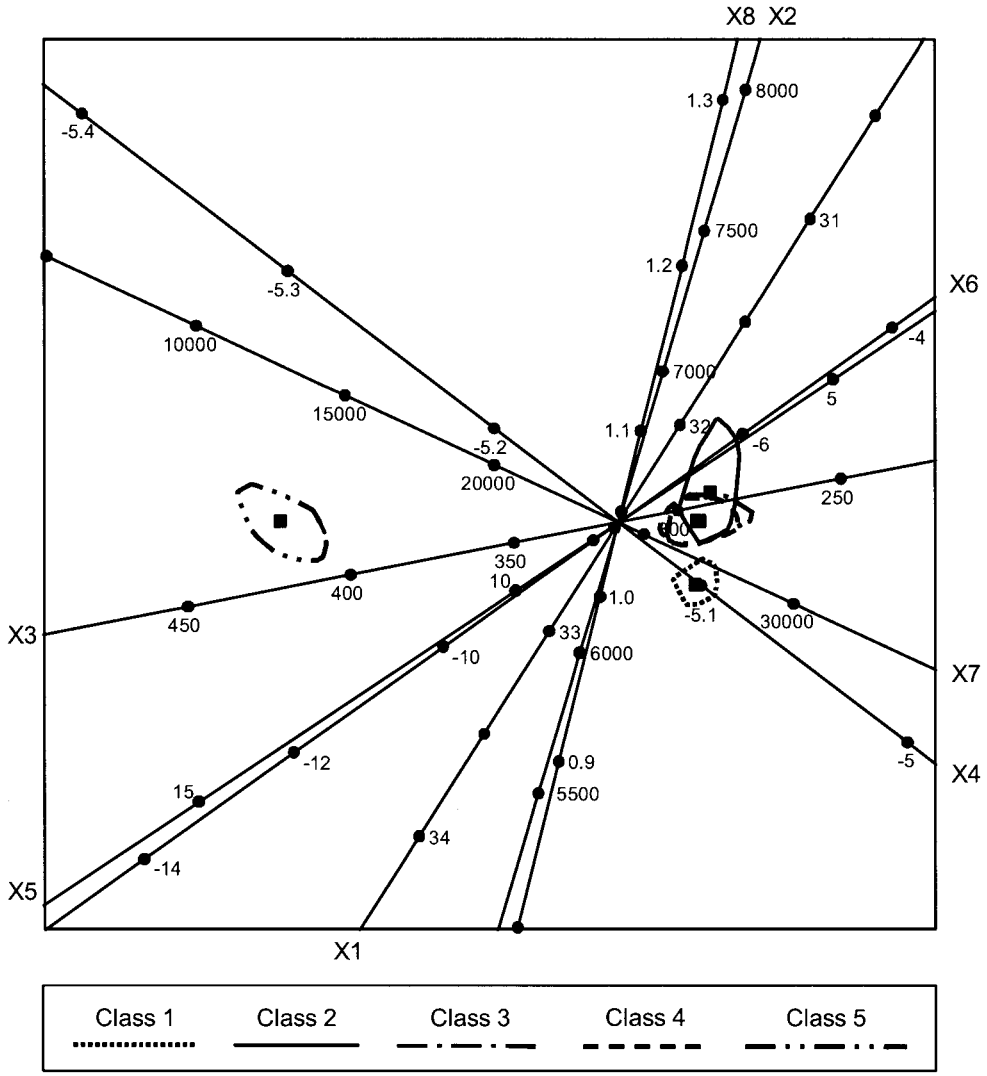

Figure 13. Biplot of the Copper Froth Flotation data set using eigenvectors 1 and 3 as scaffolding, with 50% bags added.

possible in at most three dimensions. However, it can be demonstrated how 2-D biplots can be employed to probe into higher-dimensional space: Consider the biplot in Figure 11. What is seen is the view from a third axis in a direction perpendicular to the two perpendicular axes forming the scaffolding of the display plane. Now imagine viewing from the second axis perpendicular in the direction of a plane resulting from axis 1 in its original position and axis 3. From this direction classes 1 and 2 appear to be separated, one lying on top of the other. This view can be shown in a biplot with scaffolding the first eigenvector held fixed in its original position but replacing the second eigenvector with the third. Another possibility is a biplot scaffolding where the second eigenvector is kept fixed while replacing the first with the third. This scaffolding is analogous to an observer looking from the first dimension of the canonical space perpendicular to the plane resulting from the second and the third dimensions. Figures 13 and 14 contain these biplots, respectively. The quality of the display space drops from the 91.7% of Figure 11 to 83.1% in Figure 13 and sharply to 18.1% in Figure 14.

In both Figures 14 and 15 the 50% bags of classes 1 and 2 do not overlap. Note however, that the overall separation is worse. This is in particular true in the case of the biplot in Figure 15 where the separation found between classes 1 and 2 in the third dimension is at the expense of class 5 overlapping with classes 3 and 4 in that dimension. Note that projecting the class means onto the horizontal axis in Figure 13 results in exactly the same positions as projecting the class means onto the horizontal axis of the original biplot in Figure 11. Similarly, projecting class means onto the vertical axis of Figure 14 results in the identical positions of the projections of the class means onto the vertical axis of the original Figure 11 biplot.

Now, imagine standing at a viewpoint on the second dimension of the 4-D canonical space looking perpendicular at the 3-D subspace formed by the mutually orthogonal first, third, and fourth dimensions. Although this is perceptually an impossibility for human beings, mathematically, and hence in biplot methodology this is a perfectly legitimate operation. In biplot methodology the above is achieved by a biplot constructed on a scaffolding consisting of the first eigenvector in its original

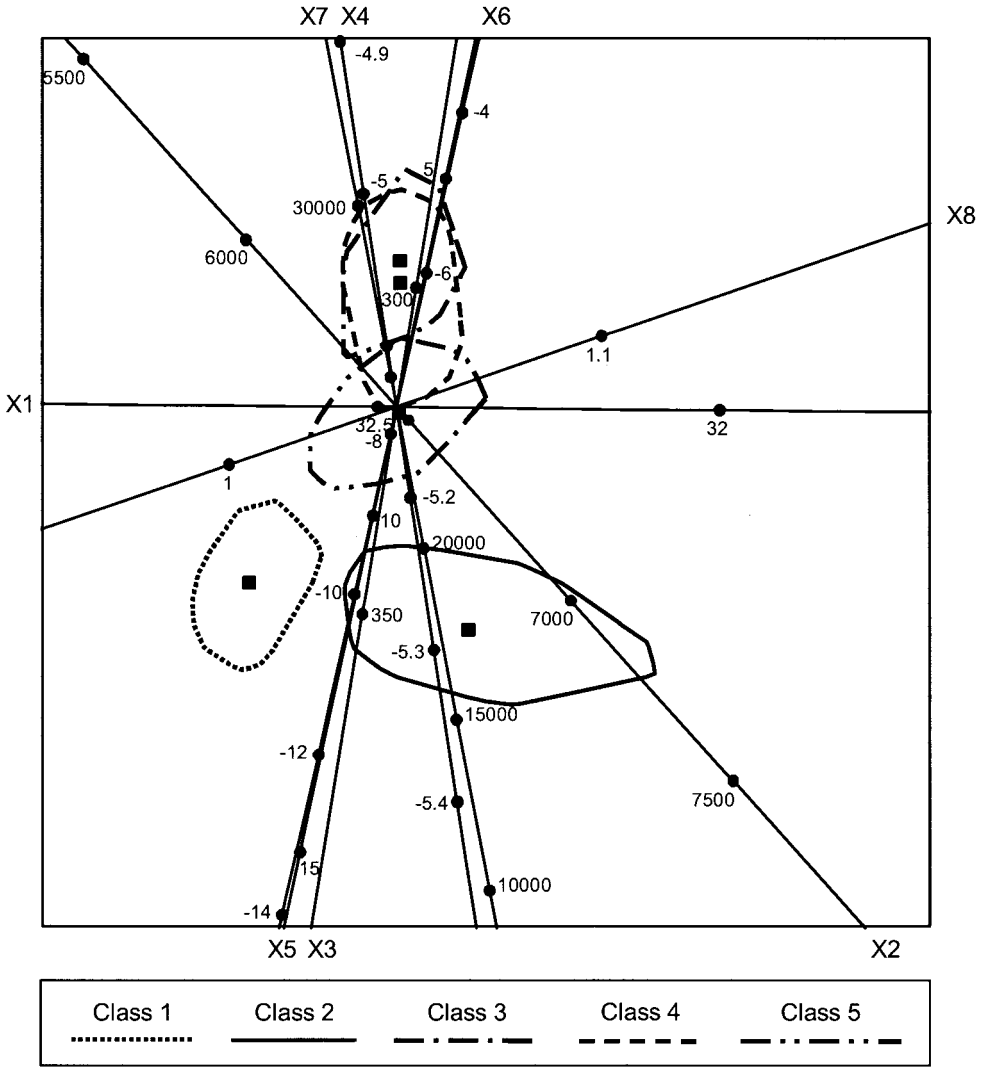

Figure 14. Biplot of the Copper Froth Flotation data set using eigenvectors 3 and 2 as scaffolding, with 50% bags added.

position and the second eigenvector replaced by the fourth eigenvector or equivalently the second eigenvector fixed in its original position with the fourth eigenvector substituted in place of the first. This is shown in Figure 15 (with quality of display 81.9%) and Figure 16 (with a quality of display of 16.9%), respectively.

Conclusions

In this paper biplots resulting from the modern perspective of Gower and Hand (1996) were introduced. In particular, possibilities from this perspective of PCA and CVA biplots in monitoring metallurgical process plants were demonstrated in two case studies. Focus was on describing multidimensional variation and the separation of different classes of observations. It was shown how biplots can be used to monitor process conditions providing management with information to act upon. Although biplot techniques are highly graphical by nature, one of its strengths is that quantitative information of all variables is simultaneously provided by the calibrated axes. These calibrations are in the measurement scales of the original observations. All interpoint distances in the biplot are assessed (quantitatively) in terms of ordinary Pythagorean (Euclidean) distances. Moreover, the overlap and separation among different classes or clusters of classes are quantitatively expressed in terms of the largest value of α (between 0 and 100), resulting in complete separation. Although PCA and CVA biplots are widely applicable, the validity of conclusions drawn from them is dependent on an adequate representation of the multidimensional variation in the 2-D biplot space as expressed quantitatively by the quality of the display. However, we have seen that in the case of CVA biplots, as long as the number of different classes is not much larger than three, the quality of the display is generally high.

The versatile PCA biplot can also be used to good purpose in monitoring product quality. Indeed, Gardner (2001) constructs a quality index (QI) biplot based on a PCA biplot. In this QI biplot a multidimensional industrial quality target is displayed together with quality regions demarcating prescribed

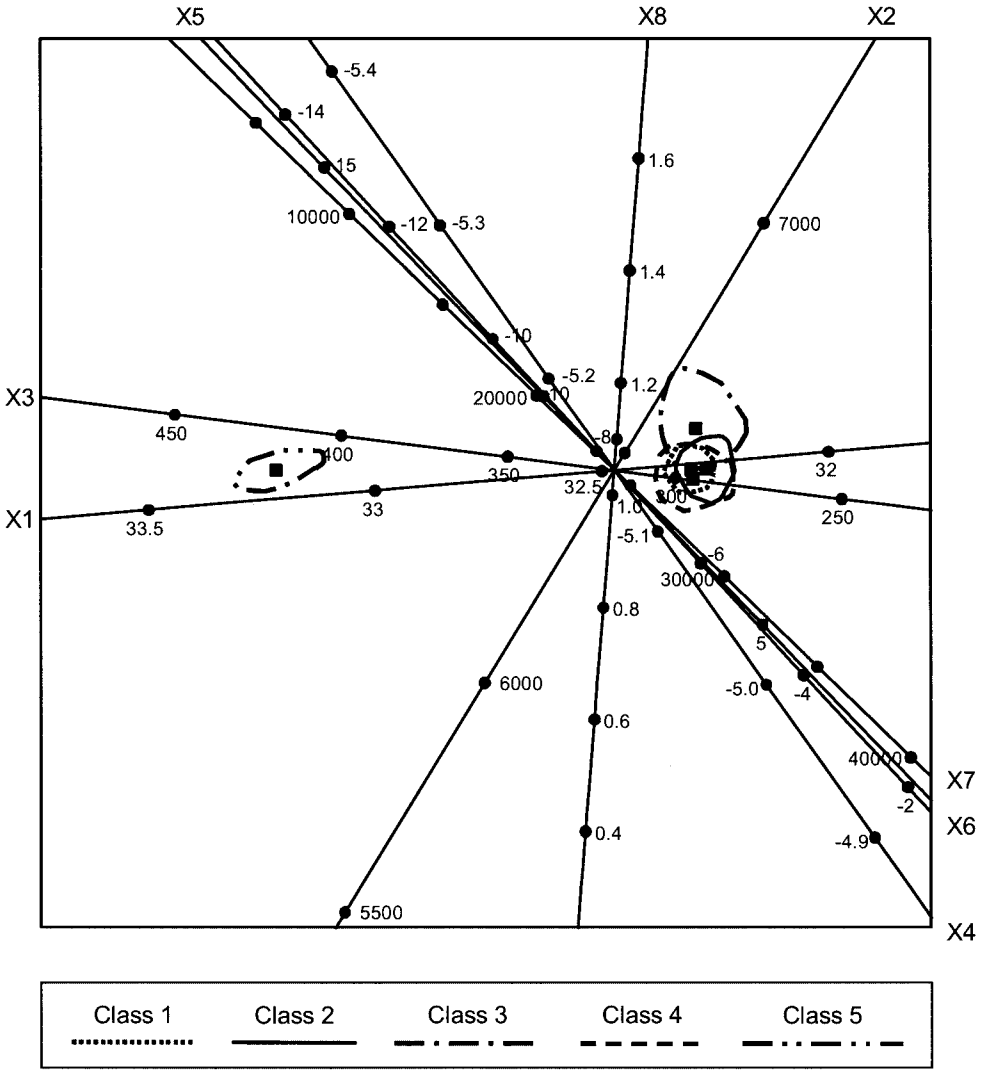

Figure 15. Biplot of the Copper Froth Flotation data set using eigenvectors 1 and 4 as scaffolding, with 50% bags added.

departures from the target as well as calibrated axes representing the variables influencing the quality of the manufactured product. The QI biplot is updated in real time thus monitoring the status of product quality continuously taking into account not only the multidimensional character of the product but interrelationships among process variables as well.

The Gower and Hand perspective permits different distance metrics to be used, leading to nonlinear biplots and generalized biplots in which both continuous and discrete data are allowed to be featured. Gardner (2001) considers discrimination and classification problems from the viewpoint of biplot methodology. She was able to extend this methodology for use in data where linear and nonlinear relationships exist. Thus, biplots and the accompanying methodology provide for investigating many different aspects of multidimensional process control.

Notation

B = $p \times p$ matrix, where the ith column $\hat{\boldsymbol{\beta}}_i$ is the solution of the two-sided eigenvalue problem satisfying $B'\hat{\Sigma}_W B = I$ and $\hat{\Sigma}_B B = \hat{\Sigma}_W B\Lambda$, with Λ a diagonal matrix containing the eigenvalues sorted in descending order

B^r = first r rows of B^{-1}

C_j = classification region for jth class

D_k = depth region, which is the set of all θ with ldepth$(\theta, X) \geq k$

$\#D_k$ = the number of data points contained in the depth region D_k

ENERGY = energy, an SGLDM feature, defined by Eq. 10

ENTROPY = entropy measure of the SGLDM method, a measure of the complexity of an image, defined by Eq. 2

$f(i, j, d, a)$ = probability density function of the spatial gray level dependency matrix (SGLDM) method, with i, j the coordinates of a pixel element in the image, d a distance metric, and a representing a direction parameter

GLLD = gray level linear dependencies in an image, defined by Eq. 5

INERTIA = a measure of the number of local variations in the image, defined by Eq. 3

L = $p \times r$ matrix with orthonormal columns forming a basis for $\mathcal{L}$

$\mathcal{L}$ = r-dimensional subspace of $\mathcal{R}^p$ used for the biplot representation

ldepth(θ, X) = halfspace location depth of a point θ, relative to a bivariate data set X, is the smallest number of points contained in any closed halfplane with boundary line through θ

LNE = large number emphasis measure of NGLDM method, a measure of the coarseness of the image, defined by Eq. 6

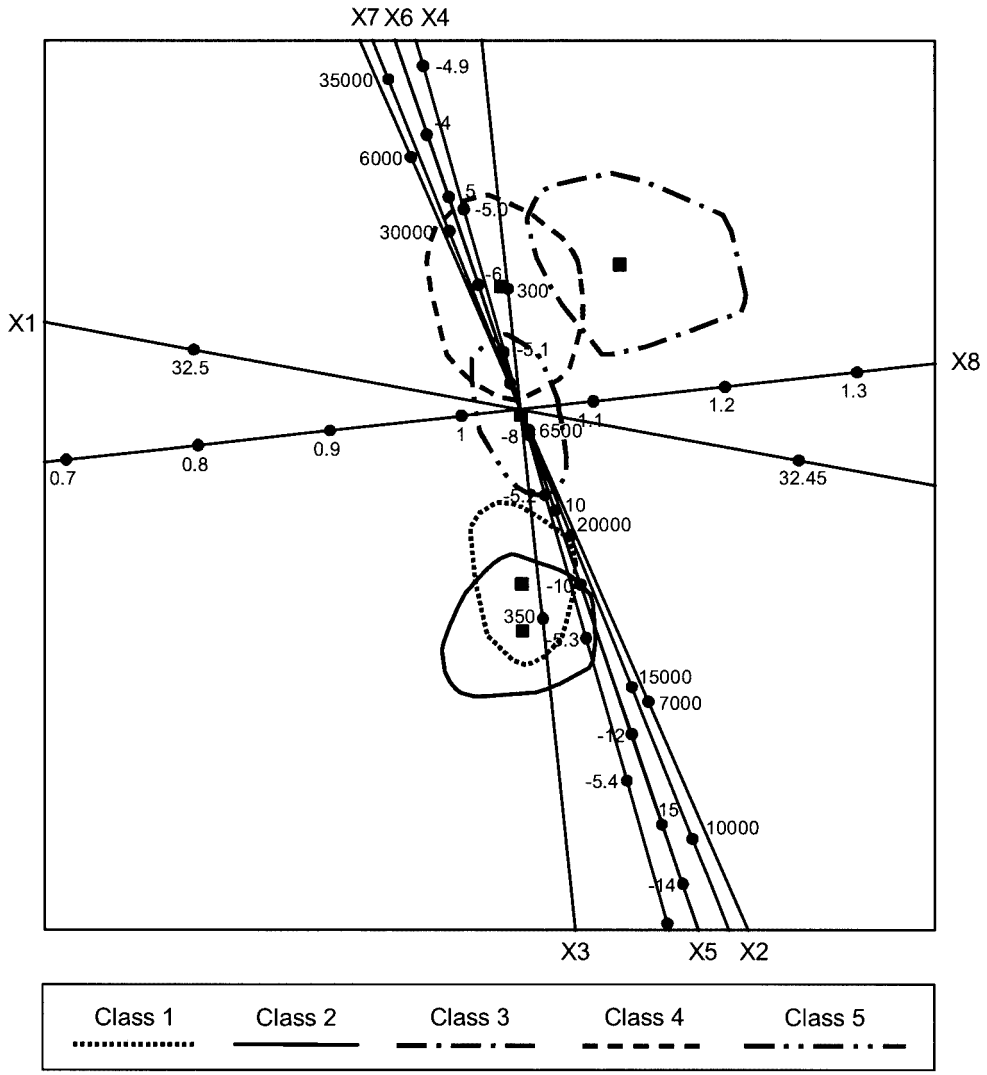

Figure 16. Biplot of the Copper Froth Flotation data set using eigenvectors 4 and 2 as scaffolding, with 50% bags added.

LOCHOM = a measure of the tendency of similar gray levels in an image to be neighbors, defined by Eq. 4
NNU = number nonuniformity measure of NGLDM method, a measure of the coarseness of the image, defined by Eq. 7
$Q(r, s)$ = feature matrix of nearest gray level dependency matrix (NGLDM) method, where r represents the number of gray levels and s the number of neighbors of a pixel
$\mathcal{R}^p$ = full p-dimensional space, for PCA this is the space of the data matrix $X : n \times p$, for CVA this is the full canonical space of $\bar{X}B : J \times p$
RA = ratio of features of SGLDM measured in different directions
SM = second moment measure of NGLDM method, defined by Eq. 8
SNE = small number emphasis, defined by Eq. 1
T^* = the depth median of a bivariate set of data points
U = interpolant in the full space $\mathcal{R}^p$
$\mathbf{u}^*$ = $p \times 1$ vector, interpolant of $\mathbf{x}^*$ in the full space $\mathcal{R}^p$
$\bar{\mathbf{u}}_j$ = $p \times 1$ vector, interpolant of the jth canonical mean $\bar{\mathbf{x}}_j$ in the full space $\mathcal{R}^p$
V = $p \times p$ orthonormal matrix for constructing principal components of X
V_r = $n \times r$ matrix containing the first r columns of V
X = $n \times p$ data matrix with n observations on p variables (typically the original process variables)
$\mathbf{x}^*$ = $p \times 1$ vector of coordinates of new sample point
$\bar{X}$ = $J \times p$ matrix, the rows of which represent the class mean vectors
$\bar{\mathbf{x}}_j$ = $p \times 1$ mean vector of jth class
$\hat{X}$ = $n \times p$ matrix with n observations and the predicted values of the p variables
$\hat{\mathbf{x}}^*$ = $p \times 1$ vector of predicted values for the point z^* in the biplot space
$\tilde{\mathbf{z}}^*$ = $r \times 1$ vector, if $\mathbf{u}^*$: $p \times 1$ lies in $\mathcal{L}$, only the first r elements are nonzero and therefore $\mathbf{u}^{*\prime} = [\tilde{\mathbf{z}}^{*\prime} \quad \mathbf{0}']$
Z = $n \times r$ matrix, interpolant of X in the biplot space $\mathcal{L}$
$\bar{Z}$ = $J \times r$ matrix, interpolant of the class means $\bar{X}$ in the biplot space $\mathcal{L}$
$\mathbf{z}^*$ = $r \times 1$ vector interpolant of $\mathbf{x}^*$ in the biplot space $\mathcal{L}$

Greek letters

$\boldsymbol{\beta}$ = $p \times 1$ vector of coefficients defining a linear combination of the predictor variables in a data set, so that the combination maximizes of the between-class to within-class variance

ϵ_N = entropy measure of the NGLDM method, a measure of the complexity of an image, defined by Eq. 9
θ = a point in two dimensions
Λ = $p \times p$ diagonal matrix consisting of the eigenvalues of $X'X$
λ_j = jth eigenvalue of $X'X$
$\Lambda_{(p)r}$ = $p \times p$ diagonal matrix with the r largest eigenvalues of $X'X$ on the diagonal and the remaining diagonal elements equal to zero
μ_x = mean of the sums of the rows of a matrix
μ_y = mean of the sums of the columns of a matrix
σ_x = standard deviation of the sums of the rows of a matrix
σ_y = standard deviation of the sums of the columns of a matrix
Σ_W = within-class covariance matrix
Σ_B = between-class covariance matrix

Literature Cited

Amirthalingam, R., and J. H. Lee, "Subspace Identification Based Inferential Control of a Continuous Pulp Digester," *Comput. Chem. Eng. Suppl.*, **21**, S1143 (1997).

Caussinus, H., and L. Ferré, "Comparing the Parameters of a Model for Several Units by Means of Principal Components Analysis," *Comput. Stat. Data Anal.*, **13**, 269 (1992).

Chambers, J. M., S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical Methods for Data Analysis*, Wadsworth, Belmont, CA (1983).

Choi, Y.-S., and M.-H. Huh, "Resistant Singular Value Decomposition and Its Statistical Applications," *J. Korean Stat. Soc.*, **25**, 49 (1996).

Daigle, G., and L. P. Rivest, "A Robust Biplot," *Can. J. Stat.*, **20**, 41 (1992).

Donoho, D. L., and M. Gasko, "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness," *Ann. Stat.*, **20**, 1803 (1992).

Everitt, B. S., "Exploring Multivariate Data Graphically: A Brief Review with Examples," *J. Appl. Stat.*, **21**, 63 (1994).

Flury, B. A., *First Course in Multivariate Statistics*, Springer-Verlag, New York (1997).

Gabriel, K. R., "The Biplot Graphical Display of Matrices with Application to Principal Component Analysis," *Biometrika*, **58**, 453 (1971).

Gabriel, K. R., "Analysis of Meteorological Data by Means of Canonical Decomposition and Biplots," *J. Appl. Meteorol.*, **11**, 1071 (1972).

Gabriel, K. R., "Biplots and Their Use for Modeling, Statistical Modelling," Proc. of the 14th International Workshop on Statistical Modelling, H. Friedl, A. Berghold, and G. Kauermann, eds., Graz, Austria, pp. 192–198 (1999).

Gardner, S., "Extensions of Biplot Methodology to Discriminant Analysis with Applications of Non-Parametric Principal Components," Unpublished PhD Thesis, Department of Statistics and Actuarial Science, University of Stellenbosch, Stellenbosch, South Africa (2001).

Gower, J. C., "A General Theory of Biplots," *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, ed., Clarendon Press, Oxford, UK, pp. 283–303 (1995).

Gower, J. C., and D. J. Hand, *Biplots*, Chapman & Hall, London (1996).

Gower, J. C., and R. F. Ngouenet, "Nonlinearity Effects in Multidimensional scaling," *J. Multivariate Anal.*, in press (2004).

Haralick, R. M., "Statistical and Structural Approaches to Texture," *Proc. IEEE*, **67**(5), 786 (1979).

Haralick, R. M., K. Shanmugan, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst., Man, Cybern.*, **SMC-3**(6), November, 610 (1973).

Jaeckle, C., and J. F. MacGregor, "Product Design through Multivariate Statistical Analysis of Process Data," *AIChE J.*, **44**(5), 1105 (1998).

Jia, F., E. B. Martin, and A. J. Morris, "Non-Linear Principal Component Analysis for Process Fault Detection," *Comput. Chem. Eng. Suppl.*, **22**, S851 (1998).

Kosanovich, K., K. S. Dahl, and M. J. Piovoso, "Improved Process Understanding Using Multiway Principal Component Analysis," *Ind. Eng. Chem. Res.*, **35**, 138 (1996).

Kourti, T., and J. F. MacGregor, "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods," *Chemom. Intell. Lab. Syst.*, **28**, 3 (1995).

Kresta, J., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35 (1991).

Martin, E. B., A. J. Morris, and C. Kiparissides, "Manufacturing Performance Enhancement through Multivariate Statistical Process Control," *Annu. Rev Control*, **23**, 35 (1999).

Martin, E. B., A. J. Morris, and J. Zhang, "Process Performance Monitoring Using Multivariate Statistical Process Control," *Syst. Eng. Automation*, **143**(2), 132 (1996).

McGill, R., J. W. Tukey, and W. A. Larsen, "Variations of Box Plots," *Am. Statistician*, **32**, 12 (1978).

McLachlan, G. J., *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York (1992).

Moolman, D. W., "The Interpretation of Flotation Froths by Using Computer Vision and Connectionist Networks," PhD Thesis, University of Stellenbosch, Stellenbosch, South Africa (1995).

Moolman, D. W., J. S. J. van Deventer, and C. Aldrich, "The Interrelationship between Flotation Variables and Froth Appearance," *Frothing and Flotation II*, J. S. Laskowski and E. T. Woodburn, eds., Gordon & Breach Science Publishers, New York, Chapter 8, pp. 245–273 (1998).

Muhamad, A. K., and F. Deravi, "Neural Networks for Texture Classification," Proc. of 4th International Conference on Image Processing and Its Applications, April 7–9, Maastricht Exhibition and Convention Centre, The Netherlands (1992).

Nomikos, P., and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995).

Rousseeuw, P. J., and I. Ruts, "The Bagplot: A Bivariate Box-and-Whiskers Plot," Technical report, Department of Mathematics and Computer Science, Universitaire Instelling Antwerpen, Antwerp. Available at http://win-www.uia.ac.be/u/statis/ (1997).

Rousseeuw, P. J., I. Ruts, and J. W. Tukey, "The Bagplot: A Bivariate Boxplot," *Am. Statistician*, **53**, 382 (1999).

Ruts, I., and P. J. Rousseeuw, "Computing Depth Contours of Bivariate Point Clouds," *Comput. Stat. Data Anal.*, **23**, 153 (1996).

Schott, J. R., "Dimensionality Reduction in Quadratic Discriminant Analysis," *Comput. Stat. Data Anal.*, **16**, 161 (1993).

Seasholtz, M. B., "Making Money with Chemometrics," *Chemom. Intell. Lab. Syst.*, **45**, 55 (1999).

Siew, L. H., R. M. Hodgson, and E. J. Wood, "Texture Measures for Carpet Wear Assessment," *IEEE Trans. Pattern Anal. Machine Intell.*, **10**(1), 92 (1988).

Sparks, R., A. Adolphson, and A. Phatak, "Multivariate Process Monitoring Using the Dynamic Biplot," *Int. Stat. Rev.*, **65**, 325 (1997).

Sun, C., and W.G. Wee, "Neighboring Gray Level Dependence Matrix for Texture Classification," *Comput. Vision Graphics Image Process.*, **23**, 341 (1983).

Tukey, J. W., "Mathematics and the Picturing of Data," *Proc. Int. Congress Mathematicians*, **2**, 523 (1975).

Wise, B. M., N. L. Ricker, D. F. Veltkamp, and B. R. Kowalski, "A Theoretical Basis for the Use of Principal Component Models for Monitoring Multivariate Processes," *Process Control & Quality*, **1**, 41 (1990).

Zhang, J., E. B. Martin, and A. J. Morris, "Process Monitoring Using Non-Linear Principal Component Analysis," Proc. of the IFAC World Congress, International Federation of Automatic Control, San Francisco, CA (1996).

Zhang, J., E. B. Martin, and A. J. Morris, "Process Monitoring Using Non-Linear Statistical Techniques," *Chem. Eng. J.*, **67**, 181 (1997).

Zhang, J., A. J. Morris, E. B. Martin, and C. Kiparissides, "Estimation of Impurity and Fouling in Batch Polymerisation Reactors through the Application of Neural Networks," *Comput. Chem. Eng.*, **23**, 310 (1999).

Manuscript received Sep. 19, 2002, and revision received Nov. 19, 2003.