Product Design through Multivariate Statistical Analysis of Process Data

Christiane M. Jaeckle and John F. MacGregor
Dept. of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada, L8S 4L7

A methodology is developed for finding a window of operating conditions within which one should be able to produce a product having a specified set of quality characteristics. The only information assumed to be available is that contained within historical data on the process obtained during the production of a range of existing product grades. Multivariate statistical methods are used to build and to invert either linear or nonlinear empirical latent variable models of the existing plant operations to obtain a window of operating conditions that are capable of yielding the desired product and that are still consistent with past operating procedures and constraints. The methods and concepts are illustrated using a simulated high-pressure tubular reactor process for producing low-density polyethylene.

Introduction

Processes can often be operated over a range of conditions to produce various grades of a product (such as a polymer). Each grade has a set of quality specifications such as melt flow index, tensile strength and so on. Often, one wants to find process operating conditions needed to achieve a new product grade having a modified set of product quality properties. An example is shown in Figure 1 where the product quality is characterized by number and weight average molecular weight (MWn and MWw). Points 1 to 9 (o) represent average qualities of nine different grades of polymer, which have been produced in the past. Operating strategies are therefore known for these grades. On the other hand, points 10 to 15 (*) mark new polymer grades which would be desirable to achieve, but where the corresponding operating conditions are still unknown. If one has a good theoretical model of the process, then an appropriate optimization algorithm can be used to find those conditions within the system constraints that yield the desired product. If such a model is not available, then designed experiments can be run on the process and response surface methods used to move the process to the desired conditions. However, even before one performs experiments, there exists information that is available within the historical process data obtained from past operations over the range of the existing product grades.

This article looks at multivariate statistical approaches to analyzing historical plant data and estimating process operating conditions, which should yield a product with the desired properties. Very little has been published on this topic. One of the few publications is that of Moteki and Arai (1986). They used a combination of multivariate statistical methods (PCA) and theoretical models to analyze historical operating

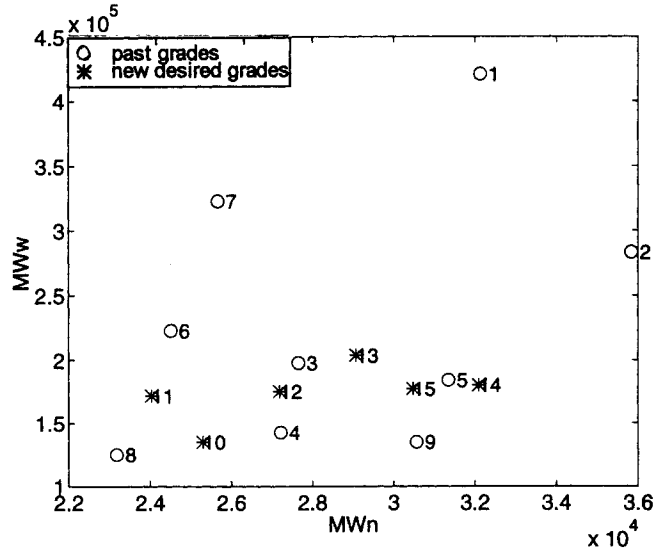

Figure 1. Quality data on 9 grades produced in the past (o) and 6 new desired grades (*).

Correspondence concerning this article should be addressed to J. F. MacGregor.

data on an LDPE process, and inferred conditions that led to new film and injection molding grades of polyethylene. They even claimed to be able to take a competitor's product, which their plant had not produced before, by finding the operating conditions for *their* process that lead to the same quality. Unfortunately, they give few details on their approach.

The objective of this article is to examine an approach to extract the necessary information from past data, and to specify a feasible target region within which the plant operating conditions should lie in order to produce the desired quality. Since they are being derived through a data-based approach, any new conditions are limited by the range and structure of the historical data. It is possible that there exists a set of operating conditions which are very different from any conditions applied in the past and which would lead to the desired quality. However, such conditions cannot be found by this empirical approach as it can only return conditions having the same range and structure as past conditions.

The problem outlined above addresses both continuous and batch processes. For continuous processes, the historical database is assumed to contain data taken at steady-state operating points corresponding to different product grades. For batch processes, the historical database will contain recipes (that is, amounts of each ingredient charged to the reactor initially), as well as feedrate profiles for various ingredients (if semi-batch operation is used), and operating profiles on process variables such as temperature and agitator speed throughout the batch histories. In this article we present a simulation example for a continuous process. We assume that data is available for: (a) the values of the quality variables of the produced grades (such as molecular weights, melt flow index, and so on); (b) the operating conditions for those grades. The latter are typically characterized by the values of the manipulated variables of the process.

Industrial processes are often operated over a large range of conditions to produce the full scope of products. Figure 2 shows the range of linear low-density polyethylene (LLDPE) products in the market. The behavior over this entire range of products is typically nonlinear. However, here we will be mostly concerned only with using the existing data on multiple grades of a *single* product type to find operating conditions for a new grade within that same product family. For example, we may be interested in producing a new film grade of LLDPE using data only from within the existing film grade region of Figure 2. Figure 1 serves as an example for such data. The corresponding process behavior may be approximately linear over this region. Therefore, the essential concepts and methodology are first developed using linear models and the relatively straightforward extension to nonlinear models is considered later.

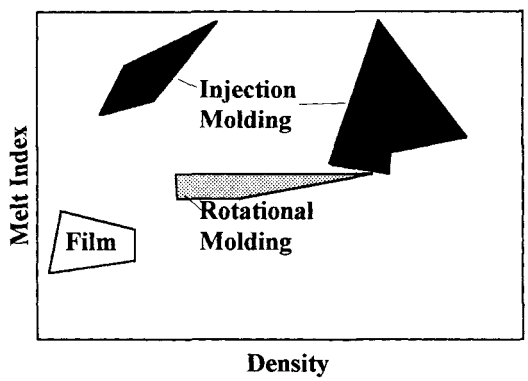

Figure 2. Linear low density polyethylene (LLDPE) products.

Given a set of historical data on various product grades consisting of measured product quality (Y), and the corresponding process conditions (X), a model can be built to characterize the behavior of the process data (X) and the relationship between X and Y.

The problem at hand is then to find new process conditions x^T_{pred} for a desired product quality y^T_{des}. Thus, we have to use some form of model inverse to predict a set of manipulated variables x^T_{new} from a specified set of quality variables. A major problem with this approach is the fact that we have typically more manipulatable variables than independent quality variables. Most commonly, one will therefore have to predict a larger number of manipulated variables from a smaller number of quality variables. This results in an underdetermined equation system, which has an infinite number of solutions x^T_{new}. However, not all of these solutions are acceptable. We must constrain the solution to be physically feasible and consistent with the sets of process conditions from the past, thereby ensuring that the new conditions x^T_{pred} can be implemented.

The following sections outline a methodology which aims at finding at least a region of process conditions that are physically feasible, and consistent with past operating strategies, and that are expected to yield the desired product qualities. Latent variable modeling methods (such as principal components regression (PCR) and partial least squares (PLS)) provide models for both the X-space (process variables) and the Y-space (quality and productivity variables). Simultaneously modeling both spaces is shown to be essential in this model inversion problem. The standard regression model which models the Y-space only through regression methods such as ordinary least squares (OLS) is shown to be inadequate for this problem. Also, contrasted with these approaches is a method based on estimating missing data using conditional expectations.

The methodology is illustrated by a simulated example of a high-pressure tubular low density polyethylene (LDPE) reactor.

Methodology

In the following sections all the manipulated and quality variables are first mean-centered and scaled, and then arranged as the columns in an X and Y-matrix, respectively. In general we assume that there are more manipulated variables than quality variables, and, hence, that X has more columns than Y. Both matrices have the same number of rows as each row corresponds to one particular grade.

Selection of quality variables

In specifying settings for the quality variables for a new grade y^T_{des}, one first has to make sure that the production of this new grade is possible within the existing process. Although there may be several variables available to describe the quality of a grade, they are often not independent from

each other and are usually highly correlated. Specifying independent values for number average and weight average molecular weights, viscosity, and so on will very likely result in a product that is physically not feasible to produce within the existing process, even if each single variable has a reasonable value. In other words, it is not possible to design just any combination of quality properties. The correlation structure among the quality variables from the existing data has to be respected. To ensure this, one can build a principal component model (Wold et al., 1987; Jackson, 1991) using the entire set of quality data. The principal component score vectors (columns of T) in this model form a set of orthogonal vectors, which fully describe the structure of the historical grade qualities. The number of significant principal components will indicate the number of independent quality variables that can be specified. T can be interpreted as a transformation of *all* quality variables into a smaller set of independent variables. It is therefore reasonable to use the T-score matrix as "quality matrix" Y in the design problem. (The set of physical quality variables (Y) could simply be replaced by the latent variables (T) in the methodology that is developed in this article without any further changes.) However, people are usually more comfortable working with "real" physical variables rather than latent variables. To select the most complimentary subset of physical quality variables that best capture the information in the "independent variable subspace," the selective PCA method (Roffel et al., 1989; Kettaneh-Wold et al., 1994) can be applied. These physical variables could then be used in the quality matrix Y.

With either approach, one still has to make sure that the new set of quality specifications (y_{des}^T) is physically feasible, and obeys a covariance structure similar to the one present in the qualities from the past. This requires: (a) a small squared prediction error (SPE) for the new point; (b) score values that lie within a 95% or 99% control ellipse of the PCA model (Kourti and MacGregor, 1996). A useful consequence resulting from the described selection of quality variables is the fact that the Y-matrix will be in general well conditioned. The benefits of this will be seen later in the model inversion.

Standard regression model and its generalized inverse

Under some suitable transformation of the Ys and Xs, the relationship between the product quality (Y) and the operating conditions (X) can often be reasonably well approximated by the linear model

$$\underset{(n\times k)}{Y} = \underset{(n\times m)}{X} \cdot \underset{(m\times k)}{\beta} + \underset{(n\times k)}{E} \tag{1}$$

where the rows in X and Y correspond to manipulated and quality variables in mean centered and autoscaled form.

The simplest and probably best known way to estimate β in the linear empirical model (Eq. 1) is ordinary least squares (OLS). The regression parameters $\hat{\beta}_{ij}$ are obtained by

$$\hat{\beta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y.$$

The first difficulty with OLS arises when X is ill-conditioned. This often occurs with plant operating data, because many of the manipulated variables are moved together in a correlated manner. The inversion of $(X^T \cdot X)$ can then be a problem, and OLS will give estimates of β that have extremely large variances [$\text{Var}(\hat{\beta}_{\text{OLS}}) = (X^T \cdot X)^{-1} \cdot \sigma$]. To help overcome this, one can use regularized least-squares solutions such as ridge regression (Draper and Smith, 1981), which give biased solutions and are much better conditioned. If the number of variables in X (m) is greater than the number of grades in the historical data (n), then an OLS solution does not exist.

Assuming the model (Eq. 1) and parameter estimates $\hat{\beta}$, new process conditions x_{new}^T have to be predicted from the desired quality specifications y_{des}^T such that

$$\underset{(1\times k)}{y_{\text{des}}^T} = \underset{(1\times m)}{x_{\text{new}}^T} \cdot \underset{(m\times k)}{\hat{\beta}} . \tag{2}$$

If $k \geq m$, x_{new}^T can be estimated by OLS or a regularized regression method such as ridge regression. However, in general, there exist more process conditions than quality variables ($m > k$). In this case, since $\hat{\beta}$ is not square and $m > k$, solving Eq. 2 for x_{new}^T is an underdetermined equation system which has an infinite number of solutions. Therefore, a commonly used estimate $\hat{x}_{\text{new}}^T$ is the least-squares solution of Eq. 2 which has the minimum 2-norm

$$\hat{x}_{\text{new}}^T = y_{\text{des}}^T \cdot \left(\hat{\beta}^T \cdot \hat{\beta}\right)^{-1} \cdot \hat{\beta}^T \tag{3}$$

$(\hat{\beta}^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T$ is the generalized inverse or pseudo-inverse of $\hat{\beta}$. The solution (Eq. 3) ensures that $\hat{x}_{\text{new}}^T$ has the following properties

$$\min\|x_{\text{new}}^T - \hat{x}_{\text{new}}^T\|_2 \quad \text{and} \quad \min\|\hat{x}_{\text{new}}^T\|_2 .$$

However, there are some serious problems with the use of this standard regression model (Eq. 1), and its inversion (Eq. 3). Since this model (Eq. 1) does not contain any information about the covariance structure within X, the relationships among the manipulated variables that were present during the production of previous grades are ignored. As a consequence, the solution given by Eq. 3 will not respect those previous structural relationships when solving for the new conditions x_{new}^T. Therefore, this formulation based on the standard regression model (Eq. 1) cannot guarantee consistency of the new conditions with past operating procedures. This problem is illustrated in the example.

Latent variable models and their generalized inverses

Multivariate statistical projection techniques work with data matrices by extracting the main underlying orthogonal directions of variation in correlated data. They summarize the information within the large number of physical variables in a usually much smaller number A of independent latent variables (Geladi and Kowalski, 1986). Principal components analysis (PCA), for example, transforms the $(n \times m)$ data matrix X containing m highly correlated manipulated variables into an $(n \times A)$ matrix T containing only A independent latent variables, which are linear combinations of the original manipulated variables. The weights for these linear combina-

tions are stored in a matrix V_A, which serves to transform T into X and vice versa

$$X = T \cdot V_A^T + E, \tag{4}$$

where E is the error or residual matrix, and V_A is an $(m \times A)$ matrix with orthonormal columns, where $A \leq m$. PCA can also be interpreted in terms of the singular value decomposition of X where only the A dominant singular directions are retained:

$$X = U_A \cdot \Sigma_A \cdot V_A^T + E.$$

U_A has orthonormal columns and Σ_A is a diagonal matrix with the ath element being the ath largest singular value of X. By comparison with Eq. 4, the score matrix T can be expressed as

$$T = U_A \cdot \Sigma_A.$$

In effect, the columns of U_A are the normalized score vectors.

Both OLS and PCR are linear regression methods. In OLS the columns of Y are regressed onto the columns of X. In PCR the columns of Y are regressed onto the columns of the latent variables (scores) of X (T in Eq. 4), that is

$$\hat{Y} = T \cdot B \quad \text{where} \quad B = (T^T \cdot T)^{-1} \cdot T^T \cdot Y \tag{5}$$

$$\hat{X} = T \cdot V_A^T \tag{6}$$

or

$$\hat{Y} = U_A \cdot \Sigma_A \cdot B \quad \text{where} \quad B = \Sigma_A^{-1} \cdot U_A^T \cdot Y \tag{7}$$

$$\hat{X} = U_A \cdot \Sigma_A \cdot V_A^T \tag{8}$$

Equations 6 and 8 provide a model for the covariance structure of X. While the A columns of U_A are orthogonal and of unit length, the entries in Σ_A and the rows in V_A^T contain the features of the variance and correlation structure present in X. In the product design problem being treated in this work, it would be desirable to impose this same covariance structure onto the conditions x_{new}^T for the new product. This would make these new operating conditions consistent with the past ones. Thus, the new conditions should conform with the model (Eq. 8) as

$$x_{\text{new}}^T = u_{\text{new}}^T \cdot \Sigma_A \cdot V_A^T.$$

Accordingly, it is the A elements of u_{new}^T that have to be estimated. This can be done using Eq. 7

$$\underset{(1\times k)}{y_{\text{des}}^T} = \underset{(1\times A)}{u_{\text{new}}^T} \cdot \underset{(A\times A)}{\Sigma_A} \cdot \underset{(A\times k)}{B}. \tag{9}$$

This equation has the same structure as Eq. 2. However, while in Eq. 2 one has to find m elements for x_{new}^T, here one only has to find $A < m$ elements for u_{new}^T, that is, the equation system has been reduced in dimension. The problem is redefined to estimate A latent variables for the process conditions from k quality variables. Consequently, we can distinguish between three cases in Eq. 9.

(1) $k > A$: The model inversion corresponds to a projection from a high dimensional (k) space to a lower dimensional (A) space. It can be performed as a simple least-squares projection.

$$\hat{u}_{\text{new}}^T = y_{\text{des}}^T \cdot B^T \cdot (B \cdot B^T)^{-1} \cdot \Sigma_A^{-1}$$

$$\hat{x}_{\text{new}}^T = \hat{u}_{\text{new}}^T \cdot \Sigma_A \cdot V_A^T$$

$$= y_{\text{des}}^T \cdot B^T \cdot (B \cdot B^T)^{-1} \cdot V_A^T$$

(2) $k = A$: In this case an exact inversion from one k- or A-dimensional basis to the other is possible.

$$\hat{u}_{\text{new}}^T = y_{\text{des}}^T \cdot B^{-1} \cdot \Sigma_A^{-1}$$

$$\hat{x}_{\text{new}}^T = \hat{u}_{\text{new}}^T \cdot \Sigma_A \cdot V_A^T$$

$$= y_{\text{des}}^T \cdot B^{-1} \cdot V_A^T$$

(3) $k < A$: This is the most common situation. Although the number of variables to be predicted has been reduced from m to A, still a projection from a lower (k) to a higher (A) dimensional hyperplane is required. This is obviously the worst of the three cases. It occurs when the effective rank of the X-matrix (A) is larger than the number of independent Y-variables (k), that is, when there are variations in some directions of the process operating space (X) which have little influence on the particular set of quality variables under consideration (Y).

For cases (1) and (2), there exists a unique solution of minimum squared error. In the third underdetermined case, however, the set of solutions is infinite.

In analogy to Eq. 3 we use the pseudo-inverse for this case (3) to invert Eq. 9.

$$\hat{u}_{\text{new}}^T = y_{\text{des}}^T \cdot \left(B^T \cdot \Sigma_A^2 \cdot B\right)^{-1} \cdot B^T \cdot \Sigma_A \tag{10}$$

The new process conditions can now be calculated as

$$\hat{x}_{\text{new}}^T = \hat{u}_{\text{new}}^T \cdot \Sigma_A \cdot V_A^T \tag{11}$$

$$= y_{\text{des}}^T \cdot M_{\text{PCR}}^T \tag{12}$$

where

$$M_{\text{PCR}}^T = \left(B^T \cdot \Sigma_A^2 \cdot B\right)^{-1} \cdot B^T \cdot \Sigma_A^2 \cdot V_A^T \tag{13}$$

$$= \left(Y^T \cdot U_A \cdot U_A^T \cdot Y\right)^{-1} \cdot Y^T \cdot U_A \cdot U_A^T \cdot X.$$

Note that the matrix being inverted in Eq. 13 is well-conditioned since it depends solely on Y and U_A. The columns in Y are chosen as an independent set of the quality variables, and the columns of U_A are orthogonal by definition.

Equation 11 reflects the way in which PCR and its inverse impose the covariance structure of the past operating conditions onto the new conditions $\hat{x}_{\text{new}}^T$: $\hat{u}_{\text{new}}^T$, which is obtained from Eq. 10 and is multiplied by Σ_A and V_A^T which contain

the variance and correlation structure of the original data X. The new $\hat{x}^T_{new}$, therefore, follows the same covariance structure as the past operating conditions. There is no equivalence to this in the approach based on the standard regression model (Eq. 1), since unlike the latent variable methods (PCR, PLS) in the standard regression model (Eq. 1), no model for X is proposed. (The only model is for the correlation structure between X and Y.)

Equation 12 shows that $\hat{x}^T_{new}$ is a linear combination of the rows of M^T_{PCR} which itself is a linear combination of the A rows in V^T_A. While the past operating conditions (rows in X) fall in the space spanned by the A rows of V^T_A, the $\hat{x}^T_{new}$ obtained from Eq. 12 fall only into the k-dimensional space (which in fact is a subspace of V^T_A) defined by the rows of M^T_{PCR}. In other words, Eq. 12 constructs only *that part* of the new operating conditions which corresponds to the projection of the past process conditions X onto the M^T space. This is the part of the new operating conditions which, according to the model, affects the product quality. However, there are features in the process space X which do not affect this particular set of quality characteristics y^T_{des} and which are not accounted for in $\hat{x}^T_{new}$. We will denote this space by x^T_{null}. These x^T_{null} may still be important operating features which should be respected in any new operating conditions x^T_{pred}. They lie within the A-dimensional V^T_A-space that defines past process operations (X), but outside the subspace of M^T, and they do not affect Y.

This suggests that any feature lying in the space of x^T_{null} can be added to $\hat{x}^T_{new}$, where x^T_{null} lies in the space which is orthogonal to M and spans the remaining $(A-k)$ dimensions of the A-dimensional X-space. Adding this component should not alter the relationship between $x^T_{pred} = \hat{x}^T_{new} + x^T_{null}$ and y^T_{des}. In other words, x^T_{null} lies in the remaining $(A-k)$ dimensions of V^T_A that are not spanned by M^T. Since this component x^T_{null} has no effect on Y, there is some degree of freedom with respect to its magnitude and direction *within* that $(A-k)$-dimensional space. The only limitation is the constraint on x^T_{pred} to be statistically consistent with past operating conditions. By choosing a whole range of different x^T_{null} within the 2σ-range of X in that additional space, it is possible to look at a "window" of process conditions $x^T_{pred} = \hat{x}^T_{new} + x^T_{null}$ that would, in principle, yield the same y^T_{des}. Using some process knowledge, the most promising set of new process conditions x^T_{pred} can be selected. In this way, when $k < A$, one can only present a window of possible solutions by inverting the model from y^T_{des}.

Figure 3 serves as an example to illustrate this graphically. The X-matrix in this example contains three manipulated variables which are shown as coordinates. However, the X-data fall only into a two-dimensional (2-D) space indicated by the gray plane. This plane represents the main underlying directions in X. The actual area the data fall into is determined by their covariance structure and may be approximated by an ellipse in that plane. The lengths of the ellipse axes are proportional to the singular values in Σ_A. In this figure, the Y-matrix is assumed to consist of a single quality variable only, and thus $k = 1$. Accordingly, the corresponding M-space is 1-D as well, and, as a subspace of the X-space, it defines a line in the plane. This line is the locus of all $\hat{x}^T_{new}$-predictions obtained from different y^T_{des}. Additional components x^T_{null} can be added to $\hat{x}^T_{new}$ such that the whole 2-D X-space is spanned. Resulting $x^T_{pred} = \hat{x}^T_{new} + x^T_{null}$ for one y^T_{des} are shown as white dots on a second line in Figure 3. Any point on this second line (x^T_{null}-line) corresponds to process conditions that are constructed such that their correlation structure is consistent with the historical X-data *and* result in the desired y^T_{des}. To ensure that we restrict ourselves to x^T_{pred} that also lie *within the range* of past operating procedures, different x^T_{pred} along the line can be transformed into the score space: $t^T_{pred} = x^T_{pred} \cdot V_A$, and then checked *where* in the T-space t^T_{pred} lies compared to the t^T_i of the past conditions. If t^T_{pred} does not fall within the cluster of past t^T_i-s, it should not be included in the set of conditions since it represents an extrapolation beyond the historical data region.

For $A - k > 1$, the approach will give *regions* of possible operation; the process conditions x^T_{pred} from those regions should all yield the specified properties y^T_{des}. It is from points in this region that the experienced process operators and engineers will have to select one which they deem most suitable to implement in the plant.

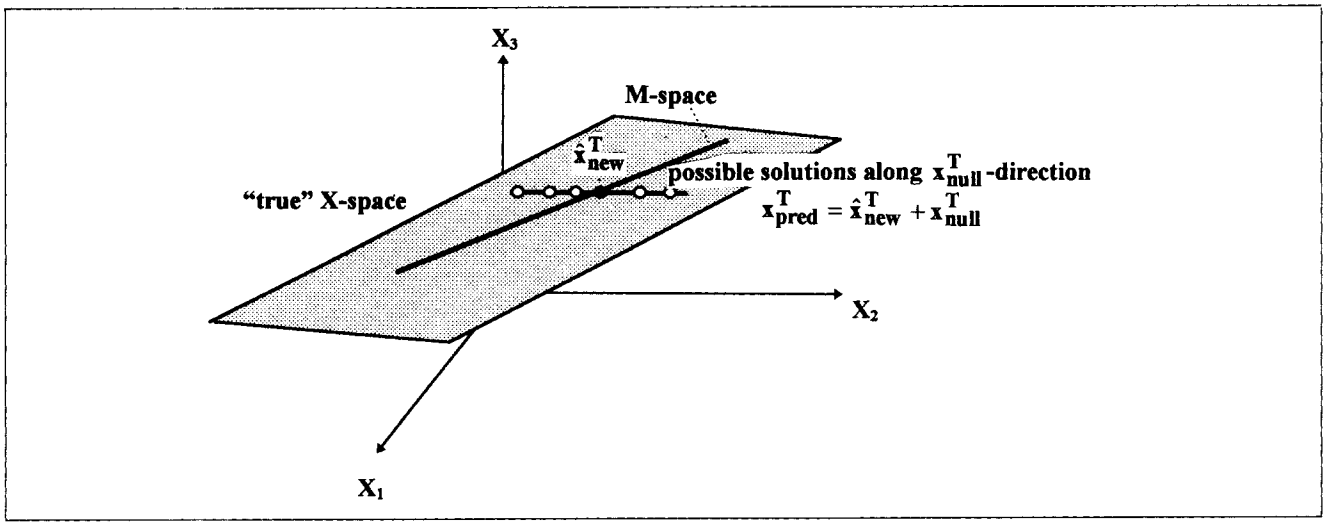

Figure 3. Spaces for $\hat{x}^T_{new}$ and x^T_{null} when $m = 3$, $A = 2$ and $k = 1$.

Other multivariate latent variable methods

Apart from PCR, there exists a whole range of other multivariate statistical modeling techniques which theoretically could be used and inverted for this problem. Common to all of them is the representation of data matrices by latent variables that summarize certain types of variation. PCR focuses on the high variance directions in X, canonical correlation regression (CCR) focuses only on the high correlation directions between X and Y, reduced rank regression (RRR) focuses on high variance directions in Y, and projection to latent structures (PLS) lies somewhere in between by simultaneously explaining variation in X, as well as the correlation with Y. Burnham et al. (1996) discuss the objective function frameworks behind all these methods.

For the problem at hand which tries to construct a x^T_{new} from a given y^T_{des}, the key is to have a model for the structure of the data in X, as well as for the relationship between X and Y. PLS and PCR are therefore preferred over methods like OLS, CCR and RRR, which provide no model for X.

In most cases PCR and PLS have been found to give similar results. In general, of course, the method giving a smaller number of latent variables is preferable in order to keep $(A-k)$ as small as possible. $((A-k)$ is the number of directions in X, which are not captured by the M-space and in which x^T_{null} has freedom to move.) In this sense there may be cases where a PLS model is beneficial compared to a PCR model or vice versa, depending on which method uses fewer components to explain the greatest amount of both X and Y.

The methodology for the inversion of PLS models is identical to that for PCR described earlier. The PLS model and its inverse are given in the section on a common framework. In the extension to nonlinear systems (discussed later in this article), a nonlinear PLS formulation is used.

Conditional expectation approach

Although this approach looks at the problem from a quite different angle, it nevertheless leads to results that are comparable to and supporting of those obtained in the previous sections.

Past operating conditions X and the corresponding product qualities Y are combined into a single matrix, and are looked at as samples from a multivariate normal distribution. The unknown operating conditions x^T_{new} that would correspond to the newly specified quality y^T_{des} are interpreted as missing data within a new row added to the existing $X|Y$ data matrix, c.f.

$$[X|Y] = \begin{pmatrix} X & Y \\ \vdots & \vdots \\ x^T_{new} & y^T_{des} \end{pmatrix}.$$

The missing values x^T_{new} are calculated as the conditional expectations given the values for the "observed" variables y^T_{des}. (In our problem this converts to: Given the quality values y^T_{des}, what is the conditional expectation for x^T_{new} based on the mean and the covariance matrix of the combined data set $[X|Y]$?) For mean centered X and Y, the covariance matrix can be written as

$$\mathrm{cov}([X|Y]) = \begin{pmatrix} X^T \cdot X & X^T \cdot Y \\ Y^T \cdot X & Y^T \cdot Y \end{pmatrix}$$

and the conditional expectation for x^T_{new} given the values of y^T_{des} is (Johnson and Wichern, 1988; Nelson et al., 1996)

$$\hat{x}^T_{new} = y^T_{des} \cdot (Y^T \cdot Y)^{-1} \cdot Y^T \cdot X = y^T_{des} \cdot M^T_{cond}. \tag{14}$$

This expression is equivalent to an inverse regression where X is regressed onto the columns of Y rather than vice versa. We can see that conceptually this expression is similar to the one for $\hat{x}^T_{new}$ from PCR-inversion (see Eq. 12). The only difference is that in PCR Y and X are replaced by their projections onto U_A.

Equation 14 shows that again $\hat{x}^T_{new}$ can fall only into a k-dimensional subspace of the space spanned by the past operating conditions X. However, in contrast to the PCR and PLS model inversions of the previous sections, we again do not have a model for the structure in X from this approach. It is therefore not possible to find that $(A-k)$-dimensional hyperplane for x^T_{null} in the X-space within which alternative choices of x^T_{pred} will have the same predicted quality y^T_{des}.

Common Framework

For the methods described above the new operating conditions x^T_{pred} consist of two parts:

$$x^T_{new} = y^T_{des} \cdot M^T_{Method}$$

where

$$M^T_{Method} = (\tilde{Y}^T \cdot \tilde{Y})^{-1} \cdot \tilde{Y}^T \cdot \tilde{X}$$

and an additional component x^T_{null} which complements x^T_{new}, and which is derived solely from information contained within X and does not affect Y.

$\tilde{X}$ and $\tilde{Y}$ are different for each method, but they can always be interpreted as a projection of X and Y onto some space. Table 1 gives the $\tilde{X}$ and $\tilde{Y}$ for each method. The similarity of PCR and PLS is obvious in the expressions: In both cases X and Y are projected onto an $(n \times A)$ matrix (U_A and U_A) with orthonormal columns which span the column space of $\hat{X}$. In PCR these columns stem from the singular value decomposition in X, in PLS they are derived from eigenvalue decompositions on the residual covariance matrices in both spaces (Höskuldsson, 1988).

Example: Simulated LDPE Polymerization Process

This section illustrates these concepts using as a process a fundamental simulation model of a high-pressure tubular reactor system for the production of low density polyethylene (Kiparissides et al., 1993; MacGregor et al., 1994).

The simulated system consists of two jacketed reactor zones with feed of ethylene, solvent and initiator both to the first zone and between the first and second zones.

Table 1. Common Framework for the Methods: $x^T_{new} = y^T_{des} \cdot M^T_{Method}$

	$M^T_{Method} = (\tilde{Y}^T \cdot \tilde{Y})^{-1} \cdot \tilde{Y}^T \cdot \tilde{X}$	$\tilde{Y}$	$\tilde{X}$	Comments
M^T_{PCR}	$(Y^T \cdot U_A \cdot U_A^T \cdot Y)^{-1} \cdot Y^T \cdot U_A \cdot U_A^T \cdot X$ $= (B^T \cdot \Sigma_A^2 \cdot B)^{-1} \cdot B^T \cdot \Sigma_A^2 \cdot V_A^T$	$U_A^T \cdot Y$	$U_A^T \cdot X$	
M^T_{PLS}	$(Y^T \cdot \mathrm{U_A} \cdot \mathrm{U_A^T} \cdot \mathrm{Y})^{-1} \cdot Y^T \cdot \mathrm{U_A} \cdot \mathrm{U_A^T} \cdot \mathrm{X}$ $= (Q \cdot \mathrm{S_A^2} \cdot Q^{\mathrm{T}})^{-1} \cdot Q \cdot \mathrm{S_A^2} \cdot \mathrm{P^T}$	$\mathrm{U_A^T} \cdot \mathrm{Y}$	$\mathrm{U_A^T} \cdot X$	Similar to PCR.
M^T_{OLS}	$(Y^T \cdot U \cdot \Sigma^{-2} \cdot U^T \cdot Y)^{-1} \cdot Y^T \cdot U \cdot \Sigma^{-2} \cdot U^T \cdot X$ $= (\hat{\beta}^T \cdot \hat{\beta})^{-1} \cdot \hat{\beta}^T$	$\Sigma^{-1} \cdot U^T \cdot Y$	$\Sigma^{-1} \cdot U^T \cdot X$	Does not maintain the covariance structure of X. No x^T_{null} since no model for X.
M^T_{Cond}	$(Y^T \cdot Y)^{-1} \cdot Y^T \cdot X$	Y	X	Inverse regression. No x^T_{null} since no model for X.

The five measured quality and productivity variables are: conversion (Conv), number average molecular weight (MWn), weight average molecular weight (MWw), long chain branching (LCB), and short chain branching (SCB).

Six manipulated variables are available: inlet temperature of the feed (Tin), pressure in the reactor (P), initiator feed rate to both zones (Fi1, Fi2), solvent flow to both zones as % of ethylene (Cs1, Cs2).

More details on this process and its simulation can be found in Kiparissides et al. (1993) and MacGregor et al. (1994).

Generation of data

Although six manipulated variables are available, they are not set independently. Except for some added random error, Fi2 and Cs2 (initiator flow rate and solvent concentration in the feed to zone 2) are kept at the same levels as their counterparts in the feed to zone 1. This leaves only four variables to be moved. Their settings have been chosen such that they are varied in a correlated manner, and the resulting X-matrix has only three (two major and one minor) independent underlying directions. 15 points representing different grades of LDPE have been simulated in this way. Out of these 15 grades, 9 have been chosen to represent the average process conditions and quality of previously produced grades. These 9 points (1-9 in Figure 1) form the historical database for the multivariate statistical model. The quality values of the remaining 6 grades serve as quality specifications that one may want to produce. Hence, we pretend not to know the settings for the manipulated variables of these 6 points.

Case studies

The method of data generation described above ensures that the new quality specifications are (a) physically feasible and (b) consistent with grades produced in the past. We start by finding out how many and which quality variables we should include in the Y matrix. PCA on the matrix containing all five quality variables shows that after the first three principal components 99% of the overall variation is explained. Consequently, we are looking for the three most significant quality and productivity variables. Selective PCA (Roffel et al., 1989) determines these as MWw, MWn, and Conv in this order. These three variables account for 96% of the variation in all five variables. MWn and MWw alone account for 76.25% of the total variation.

In the following studies, X consists of the values for the six manipulated variables for the nine grades that have been produced in the past. Two different Y-matrices are considered: Y3 = [Conv MWn MWw] and Y2 = [MWn MWw] for each of the nine grades.

Table 2 shows how much variation is explained by PCR in two models: X and $Y3$, and X and $Y2$. In both cases, the number of principal components in the PCR-model is chosen as $A = 3$.

Case 1: Y3 = [Conv MWn MWw]. The use of the standard regression model and OLS is problematic even in this small example since we have to invert an ill conditioned $X^T \cdot X$. However, OLS was performed using MATLAB to obtain values for $\hat{\beta}$. The variances of these OLS estimates were very large.

In PCR on the other hand, we encounter here the special case where the number of Y-variables equals the number of principal components in the model ($k = A$). For the inversion of Eq. 9 this means that we can directly invert B and calculate $\hat{x}^T_{new} = y^T_{des} \cdot B^{-1} \cdot V_A^T$. Since for $k = A$ the M-space spans all the main directions in X, there are no orthogonal directions x^T_{null}. We therefore come up with one unique set of manipulated variables for each desired new grade.

The resulting process conditions from PCR and OLS inversion were applied to the full nonlinear simulated process and the corresponding quality values obtained. The results are shown in Figures 4, 5, and 6. The points denoted by a (*) are the desired qualities, (o) denotes the qualities achieved from OLS-inversion, and (x) the qualities achieved from PCR-inversion. If one compares the estimates of the quality (MWn, MWw) and productivity (Conv) variables obtained from the OLS- and PCR-inversion, apart from one exception (grade 13) PCR achieves products much closer to the desired ones than OLS. Whereas Figures 4 and 5 show the three measured responses (Conv, MWn and MWw) that are included in Y, Figure 6 displays the two quality variables which are not present in Y (LCB and SCB), but nevertheless very important

Table 2. Percentage of Variation Explained by PCR in X and Y

No. of Principal Components	1	2	3	4	5	6
X	63.70	96.36	99.45	99.94	100	100
Case 1: Y3	23.51	68.07	94.05	96.22	97.58	99.10
Case 2: Y2	29.03	95.35	97.04	98.51	98.74	99.36

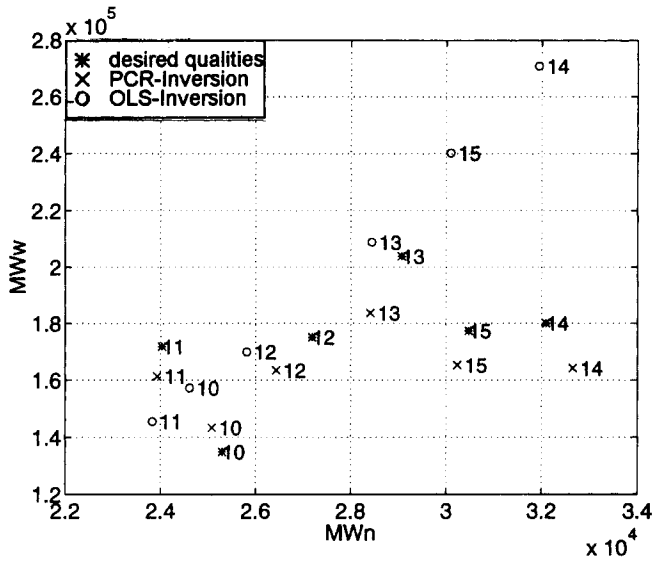

Figure 4. PCR and OLS results for number and weight average molecular weight, when Y = [Conv MWn MWw].

quality characteristics for the product. We see that the process conditions obtained from the PCR latent variable model yielded values for the LCB and SCB properties that were much closer to the actual (but unknown) values in the desired product, than did the conditions obtained from inverting the OLS model. This is one of the benefits of constructing process conditions that are consistent with past operating strategies. Other unmeasured quality variables will maintain their same historical relationship with those that are measured.

Remark: In this case, LCB and SCB are not specified in the Y-matrix, so realistically there are no desired values for them. However, we did simulate grades 10-15 in a way consistent with the grades 1-9 which are used for the model building, and, therefore, we have values for LCB and SCB for these grades.

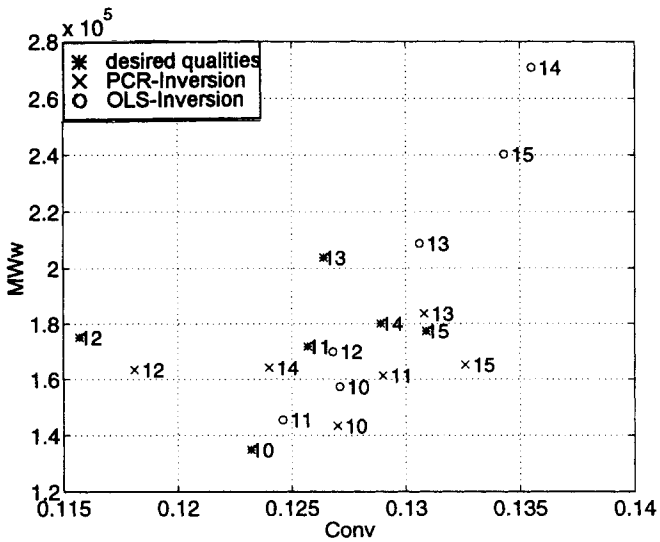

Figure 5. PCR and OLS results for conversion and weight-average molecular weight, when Y = [Conv MWn MWw].

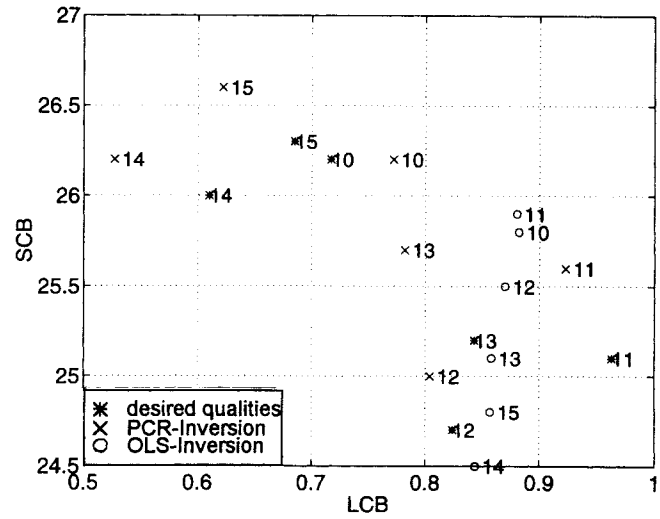

Figure 6. PCR and OLS results for long and short chain branching, when Y = [Conv MWn MWw].

Case 2: Y2 = [MWn MWw]. As discussed earlier, a PCA analysis on all five quality variables indicated there are *three* underlying directions among these variables which account for the total variation within them. Thus, ideally, one would want to specify three independent quality variables. However, in many processes only a few correlated properties are actually measured on a product, and these may be insufficient to span the entire quality space of a product. In this example, by specifying MWn and MWw only two out of the three directions are being set, and about 24% of the total quality variation is not accounted for.

For the PCR-inversion, this case provides a different situation from case 1: the number of Y-variables ($k = 2$) is less than the number of independent directions in X ($A = 3$). Consequently, the resulting $\hat{x}^T_{\text{new}}$ conditions will lie only in a 2-D subspace of X. The additional direction, which is complementing the M space (see Figure 3), must be calculated. Projecting the X-data onto this direction gives an estimate of the amount of variation in X that exists along that direction. Five different vectors x^T_{null} (within 2σ of the rows of X along this direction) were added to each $\hat{x}^T_{\text{new}}$. This gives a set of five suggestions $x^T_{\text{pred}} = \hat{x}^T_{\text{new}} + x^T_{\text{null}}$ for each new grade. From the five suggestions for each grade a subset was selected such that their $t^T_{\text{pred}} = x^T_{\text{pred}} \cdot V_A$ lay within the cluster of t^T_i-s from the past grades. For example, for grade 11 only two of the five suggestions had appropriate t-values and were implemented.

In Figure 7 two of the resulting manipulated variables (Cs1 and P, inlet solvent concentration to zone 1 and inlet pressure) are plotted vs. each other for the various sets of conditions. The points corresponding to the same grade lie on a line, and the lines for all the grades are parallel to each other. This is not surprising since it reflects the fact that the additional component x^T_{null} is free only to move along $(A - k = 1)$ directions in this case. The lines in Figure 7 are the projection of this direction onto the Cs1-P-plane.

Again, the new process conditions were implemented using the detailed fundamental simulation process, and the result-

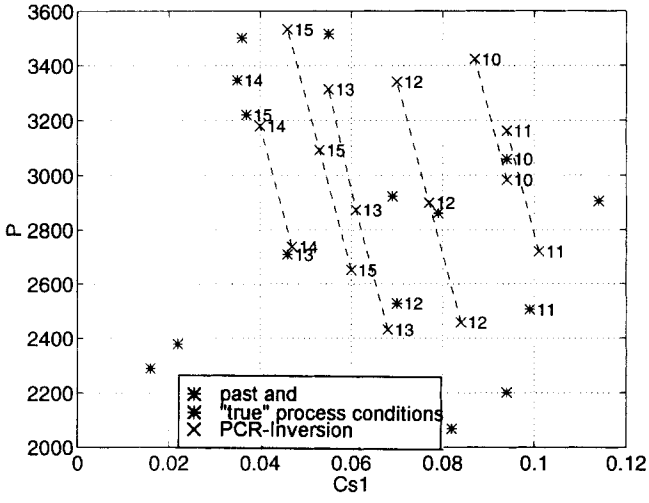

Figure 7. PCR (a line of) inversion for Y2 = [MWn MWw].

PCR gives process conditions for each grade.

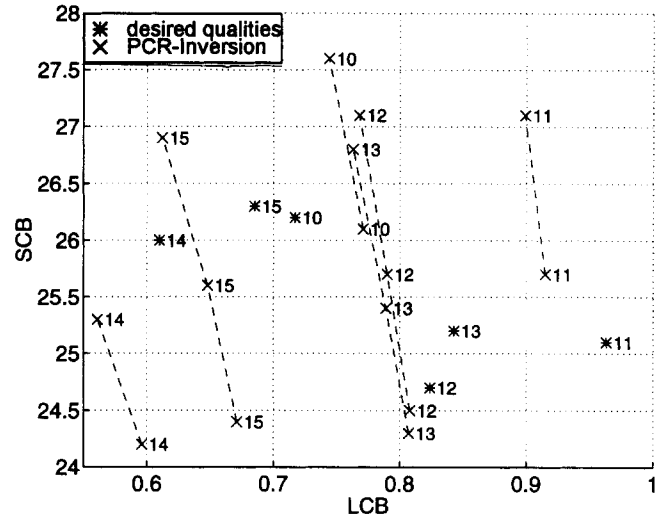

Figure 9. PCR results for long and short chain branching, when Y = [MWn MWw].

ing qualities plotted in Figures 8 and 9. In general, the grades obtained by implementing the new process conditions come close to the desired ones as in case 1. Looking at these PCR-results in Figures 8 and 9, we see another interesting but expected feature: While *measured* qualities (MWn and MWw) resulting from different x_{pred}^T suggestions for one grade (Figure 8) are almost the same, the resulting *unmeasured* qualities (LCB and SCB in Figure 9) display much more variation. (For each grade, the resulting unmeasured qualities are connected by dashed lines in Figure 9.) This can be explained by the random component x_{null}^T in x_{pred}^T. It was shown that according to the model this component does not affect the measured qualities Y (MWn and MWw), however, in this process it does affect the unmeasured qualities (LCB and SCB).

The conditional expectation approach was also applied to this problem. The results are shown in Figure 10 together with the PCR results given previously in Figure 8. One can see that the results of this approach are very close to those obtained by the PCR model inversion. However, this conditional expectation approach gives only a single point for each specified product, when in fact there is a whole line of equivalent process conditions as given by PCR. Results obtained using a PLS latent variable model were almost identical to those shown for PCR and are not displayed here. Inversion of the standard regression model (Eq. 1) fitted by OLS again gives very poor results in this case. One of the problems with this OLS approach is illustrated in Figure 11. This figure shows the values for the two initiator flow rates (Fi1 and Fi2) to the two zones as they have been designed by OLS-inversion (o) and by PCR-inversion (x) for the different grades. The way the process has been run in the past, the two rates are nearly identical at all times. However, since the OLS re-

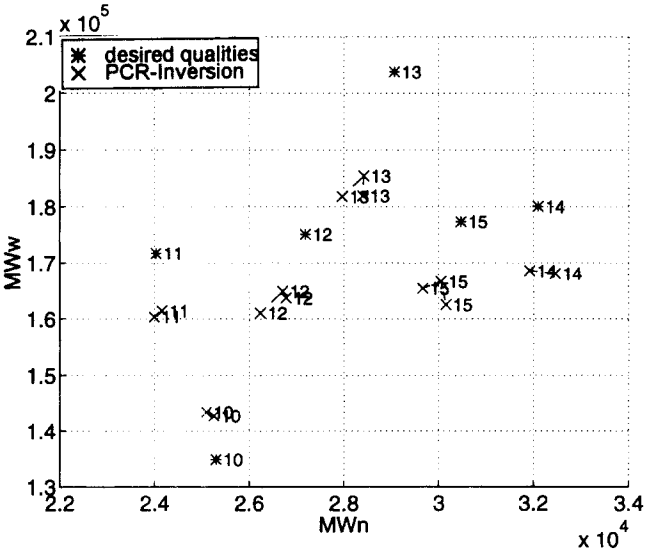

Figure 8. PCR results for number and weight-average molecular weight, when Y = [MWn MWw].

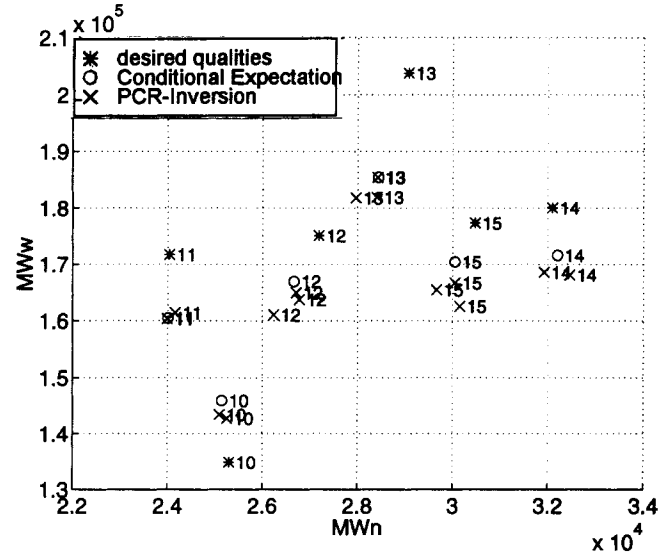

Figure 10. Conditional expectation and PCR results when Y = [MWn MWw].

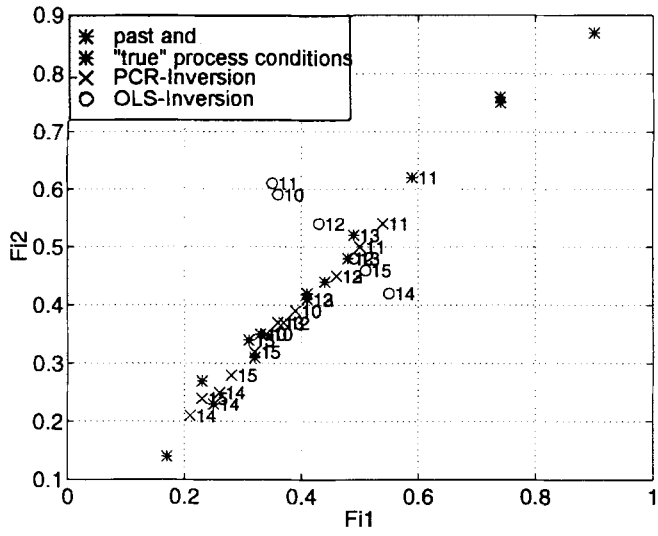

Figure 11. Process operating conditions from OLS- and PCR inversion when Y = [MWn MWw].

gression model (Eq. 1) does not provide a model for the X-space, it cannot and does not respect the existing correlation between the two rates, whereas it can be seen that the results from PCR-inversion continue to obey the correlation structure of the past data.

Since the conditions obtained from the OLS regression model inversion are well outside those ever experienced in the historical database, one could not expect them to produce the desired product, and, indeed, the product qualities obtained by implementing those conditions are very far from those specified.

Extension to Nonlinear Systems

In some situations process nonlinearities may be important. Examples might include products whose grades cover a wide range of product properties. Another situation is where data on grades from several product groups may have to be combined. For example, in Figure 2 one might have to combine data from the two groups of injection molding grades and the rotational molding grades in order to design conditions for a new injection molding grade lying between the two existing groups. However, since the number of grades available is often small there are also dangers in using empirical nonlinear models for this inversion problem. Some of these will be discussed following the extension of the methodology to nonlinear processes, and their illustration using the LDPE example.

As should be obvious now from the previous treatment of linear systems, the essential elements of any model (linear or nonlinear) for this model inversion problem, is that it provides a model for the X space as well as the relationship between the X and Y spaces. For nonlinear modeling approaches, this rules out most standard nonlinear regression and neural network type of models. One of the few existing nonlinear approaches that accomplish this dual feat are nonlinear PLS latent variable models although nonlinear PCR type models could also be proposed. Because of the few data points available, complex nonlinear models may lead to significant overfitting and have poor inversion properties. Therefore, in this article we employ only quadratic nonlinear PLS, although the method is easily extendable to any nonlinear structure.

Nonlinear PLS and inversion

The most common approach to nonlinear PLS (NPLS) modeling (Wold et al., 1989; Wold, 1992) represents X and Y as linear latent variable spaces

$$\hat{X} = T \cdot P^T$$
$$\hat{Y} = Z \cdot Q^T$$

and introduces a nonlinear inner relationship between the latent variables z_a of the Y space and t_a of the X space such as the quadratic relationship

$$z_a = c_{0_a} + c_{1_a} \cdot t_a + c_{2_a} \cdot t_a^2 \qquad a = 1, \ldots, A$$

where t_a and z_a are the score vectors for the ath dimension, and c_{i_a} are scalar coefficients. The algorithm and more details on the NPLS model can be found in Wold et al. (1989).

The requirements for the design of new product quality specifications are the same as in the linear case: For a desired new grade y_{des}^T, we want to obtain new process conditions x_{new}^T such that the established *nonlinear* model still holds. Since consistency with past operating strategies is required, the new process conditions must again maintain the same covariance structure as the data from the past.

We express the PLS model for the X-space as

$$\hat{X} = U_A \Sigma_A \cdot P^T$$

where U_A has orthonormal columns, and Σ_A and P^T contain the variance and correlation information of the past data X. We therefore postulate for the new process conditions

$$\hat{x}_{\text{new}}^T = \hat{u}_{\text{new}}^T \cdot \Sigma_A \cdot P^T \tag{15}$$

and

$$\hat{t}_{\text{new}}^T = \hat{u}_{\text{new}}^T \cdot \Sigma_A \cdot$$
$$= \left[\hat{u}_{\text{new}_1} \cdot \sigma_1 \quad \hat{u}_{\text{new}_2} \cdot \sigma_2 \ \ldots \ \hat{u}_{\text{new}_A} \cdot \sigma_A \right]$$

This enforcement of consistency, that is, of maintaining the historical covariance structure, reduces the problem from estimating m correlated process variables x_j $(j = 1, \ldots, m)$ to the estimation of A independent latent variables $\hat{u}_{\text{new}_a}$ $(a = 1, \ldots, A)$, where $A < m$. The problem that remains is to find these latent values $\hat{u}_{\text{new}_a}$ such that the resulting product quality is the desired y_{des}^T. However, y_{des}^T has to fit the model as well, that is

$$y_{\text{des}}^T = \hat{z}_{\text{new}}^T \cdot Q^T, \tag{16}$$

where for each element a in $\hat{z}_{\text{new}}^T$

$$\hat{z}_{\text{new}_a} = c_{0_a} + c_{1_a} \cdot \hat{u}_{\text{new}_a} \cdot \sigma_a + c_{2_a} \cdot \hat{u}_{\text{new}_a}^2 \cdot \sigma_a^2. \tag{17}$$

In order to combine Eqs. 16 and 17, we first rewrite the vector Eq. 16 in terms of single elements $y_{j_{des}}$

$$y_{j_{des}} = \sum_{a=1}^{A} \hat{z}_{new_a} \cdot q_{ja} \quad \text{for} \quad j = 1, \ldots, k$$

where q_{ja} is the (j,a)-th element of Q.

$y_{j_{des}}$ can now be expressed as

$$y_{j_{des}} = \sum_{a=1}^{A} \left(c_{0_a} + c_{1_a} \cdot \hat{u}_{new_a} \cdot \sigma_a + c_{2a} \cdot \hat{u}^2_{new_a} \cdot \sigma_a^2 \right) \cdot q_{ja} \quad \text{for } j = 1, \ldots, k. \quad (18)$$

This finally states our problem: We have to solve a set of k Eqs. 18 to obtain A values $\hat{u}_{new_a}$. They can then be substituted into equation Eq. 15

$$\hat{x}^T_{new} = \hat{u}^T_{new} \cdot \Sigma_A \cdot P^T$$

to give process conditions that are: (1) designed to give the desired product properties y^T_{des}; (2) consistent with past operating strategies.

For one new grade, y^T_{des} (Eq. 18) is a set of k quadratic equations in A unknowns $\hat{u}_{new_a}$. Depending on the number (k) of equations and (A) of variables, a solution can be found using different numerical techniques. In the case of a square system $(A = k)$, we can apply the Multivariate Newton's Method to find roots of the system. An overdetermined system where $A < k$ is not very likely in this type of problem. It would imply that there is significant variation in the product quality (Y) that cannot be explained by the process variables present in X either because X does not contain *all* the variables that affect Y or because the noise ratio in Y is very high. Either case would make it fairly impossible to "design" the desired quality y^T_{des} by estimating the proper values for the process variables in X. However, a mathematical solution for this overdetermined system can be found, for example, by using a least-squares technique to find the A unknowns minimizing the error sum of squares in Eqs. 18.

As before, the most common case is one where the process variables vary in more independent directions than the number of quality variables specified in Y: $A > k$. Such an underdetermined nonlinear equation system can again have an infinite number of solutions which, according to the model, will all result in the desired y^T_{des}.

Solving Eq. 18 is an optimization problem and many algorithms are available to find solutions. One possible way is to minimize

$$f(\hat{u}_{new})^T \cdot f(\hat{u}_{new})$$

using different starting values, where

$$f^T \cdot f = \sum_{j=1}^{k} \left[y_{j_{des}} - \sum_{a=1}^{A} \left(c_{0_a} + c_{1_a} \cdot \hat{u}_{new_a} \cdot \sigma_a + c_{2_a} \cdot \hat{u}^2_{new_a} \cdot \sigma_a^2 \right) \cdot q_{ja} \right]^2. \quad (19)$$

Table 3. Percentage of Variation Explained by NPLS in X and Y

No. of Principal Components	1	2	3	4	5	6
X	55.73	96.28	99.44	99.93	100	100
Case 1: Y3	35.93	70.65	95.80	96.73	98.63	99.00

An underdetermined system $(A > k)$ can be accommodated by fixing a grid of values for $A - k$ unknowns $\hat{u}_{new_a}$, and solving through optimization for the remaining k $\hat{u}_{new_a}$. Connecting the solutions defines a curve or surface of possible solutions. Only those solutions for u^T_{new} that fall within the region of u^T values from the historical grades will be retained for investigation. Thus, for $A \geq k$ one will obtain a window of possible operating conditions all of which are predicted to yield y^T_{des}.

Case study on LDPE-polymerization

This study deals with the exact same process and data sets as described in the example above.

Case 1: Y3 = [Conv MWn MWw]. A nonlinear PLS model with a quadratic inner relation was calculated for the 9 historical grades. Results are shown in Table 3. Comparison to the linear PLS model (not displayed) shows that the two models explain about the same amount of variation in each dimension. Furthermore, the loading vectors of both models are very similar which suggests that the individual latent variables describe the same type of variation in both models. Consequently, the NPLS model will use 3 latent variables as well. Compared to the linear PCR model (see Table 2), NPLS explains the same amount of variation in X and a little bit more in Y. In Table 4 the coefficients for the first three inner relations of the NPLS model suggest mostly linear behavior for the first 2 dimensions. Only the third dimension has a significant square term. We may conclude that the data show only slightly nonlinear behavior, and that a linear model is perhaps adequate. However, the nonlinear approach using the NPLS model introduced above, is carried out for illustrative purposes.

In order to obtain process conditions for six new grades (see example above) the 3×3 nonlinear equation system was inverted using a least-squares algorithm to minimize $f^T \cdot f$ according to Eq. 19. No multiple solutions were found. For each of the six new desired grades, the resulting estimated process conditions were implemented on the fundamental simulation and the corresponding product qualities obtained. Plots on the desired and achieved product qualities can be seen in Figures 12 and 13. In terms of the two molecular weight properties MWn and MWw (Figure 12), NPLS achieves qualities that are closer to the desired values in all but one case

Table 4. Inner Relationships for the First 3 Principal Components (Case 1)

Principal Component No.	Quadratic Inner Relationship
1	$\hat{z}_1 = -0.25 + 0.56 \cdot t_1 + 0.09 \cdot t_1^2$
2	$\hat{z}_2 = -0.16 + 0.66 \cdot t_2 + 0.07 \cdot t_2^2$
3	$\hat{z}_3 = -0.12 + 2.12 \cdot t_3 + 0.71 \cdot t_3^2$

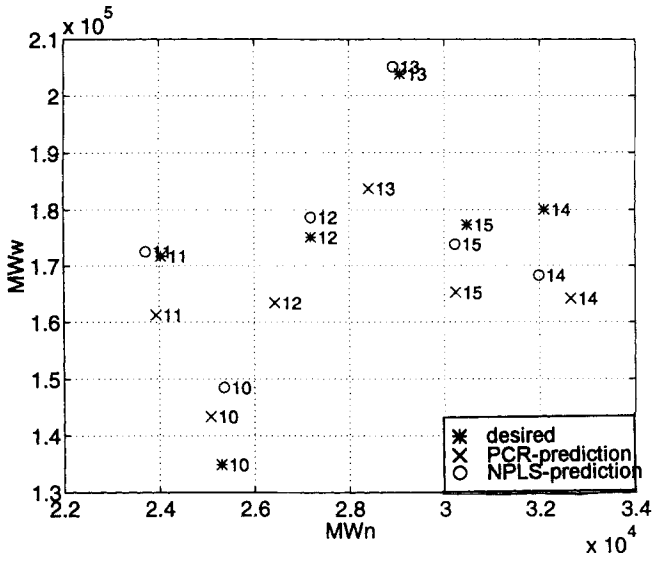

Figure 12. NPLS and PCR results for number and weight-average molecular weight when Y = [Conv MWn MWw].

(grade 10). However, NPLS does not perform as well with respect to conversion (Figure 13), where PCR-grades are closer to the desired values in most cases.

To obtain a better feel for the overall closeness to the desired results for all 6 grades, Table 5 shows the Mahalanobis distances between desired y^T_{des} and actually achieved grade quality y^T_{ach}

$$\left(y^T_{\text{des}} - y^T_{\text{ach}}\right)\cdot[\text{Cov}(Y)]^{-1}\cdot\left(y^T_{\text{des}} - y^T_{\text{ach}}\right)^T \qquad (20)$$

where $[\text{Cov}(Y)]^{-1}$ is the inverse of the covariance matrix of the historical Y-data. Except for grades 11 and 14, PCR-inversion shows better results than NPLS-inversion. Therefore, we conclude that although the nonlinear model may provide a slightly better fit for the historical data set than the linear

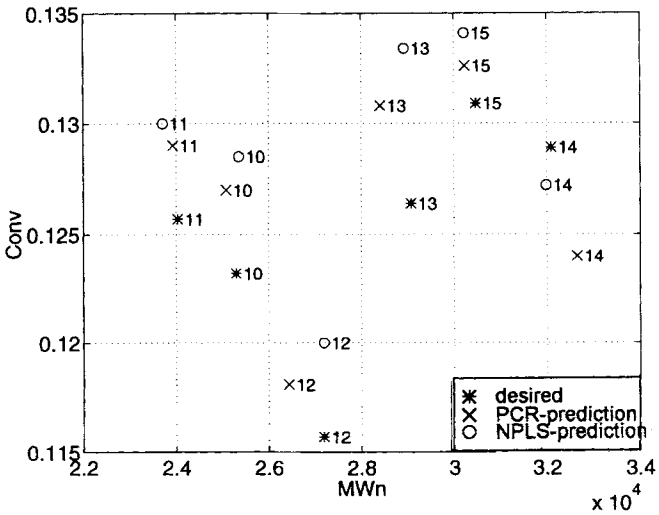

Figure 13. NPLS and PCR results for number average molecular weight and conversion when Y = [Conv MWn MWw].

Table 5. Mahalanobis Distances to Desired Grades Based on the Specified Y Variables (Case 1): $(y^T_{\text{des}} - y^T_{\text{ach}})\cdot[Cov(Y)]^{-1}\cdot(y^T_{\text{des}} - y^T_{\text{ach}})^T$

Inversion	Gr. 10	Gr. 11	Gr. 12	Gr. 13	Gr. 14	Gr. 15
PCR	0.096	0.149	0.086	0.277	0.162	0.061
NPLS	0.199	0.144	0.151	0.414	0.024	0.094

Table 6. Percentage of Variation Explained by NPLS in X and Y

No. of Principal Components	1	2	3	4	5	6
X	46.49	96.35	99.44	99.93	100	100
Case 2: Y2	58.93	97.53	99.29	99.63	99.84	99.95

PCR model, this does not necessarily lead to better inversion results, as we can see in this example.

Case 2: Y2 = [MWn MWw]. In this case only two of the previous 3 quality variables are assumed to be available for inclusion in Y: Y2 = [MWn MWw]. Table 6 shows the amount of variation explained in the resulting NPLS-model. With respect to the corresponding linear PCR-model (refer to Table 2), a slight improvement can be observed for the amount of Y explained by the NPLS model.

The coefficients for the inner relationships of the 3 dominant principal components can be seen in Table 7. The first dimension shows a slight curvature between the latent vectors t_1 and $\hat{z}_1$, as illustrated in Figure 14.

For this case of only $k = 2$ Y variables, we have 2 nonlinear equations but $A = 3$ latent variables $\hat{u}_{\text{new}_a}$. Defining a grid from -1 to 1 (stepwidth 0.1) for $\hat{u}_{\text{new}_1}$, at each grid point the resulting minimization problem $f^T\cdot f$ for $\hat{u}_{\text{new}_2}$ and $\hat{u}_{\text{new}_3}$ was solved according to Eq. 19. Figure 15 shows historical and predicted u-values in the u_1–u_3 plane. Indicated by (x) are the u values of the actual grades 10–15 as they were simulated. (In a "real" example we would not know these latter values.) Different predictions u for each grade are denoted by (o), and connected by a dashed line to indicate solution curves. The u-values of the historical grades used in the model are shown as (*). The loci of solutions for grades 11 to 15 are nearly linear, while for grade 10 the locus of solutions is parabolic. However, the strongly curved part of the solution locus for grade 10 is outside of the range of the historical u-data, and should not be considered for possible implementation.

As before, the corresponding process conditions for these $\hat{u}_{\text{new}}$-vectors were computed, implemented and the results compared to the results from the linear PCR-inversion. For this underdetermined case, we have several process conditions for each grade, all of which should yield the same values of MWn and MWw, but in practice lead to slightly vary-

Table 7. Inner Relationships for the First 3 Principal Components (Case 2)

Principal Component No.	Quadratic Inner Relationship
1	$\hat{z}_1 = -0.23 + 0.64\cdot t_1 + 0.10\cdot t_1^2$
2	$\hat{z}_2 = -0.11 + 0.52\cdot t_2 + 0.04\cdot t_2^2$
3	$\hat{z}_3 = -0.01 + 0.43\cdot t_3 + 0.04\cdot t_3^2$

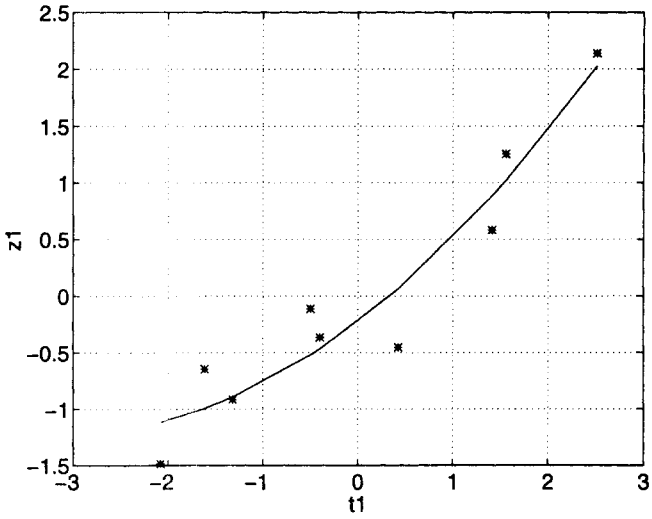

Figure 14. Inner relationship between *X*- and *Y*-scores for the first latent NPLS-dimension.

ing values (see, for example, Figure 8 for results from the linear PCR approach). The reason for this is mismatch between the empirical model (here PCR or NPLS) and the fundamental simulation model. Therefore, we compare for each grade the results that came closest to the desired MWn and MWw values in each model inversion. The Mahalanobis distances between desired and achieved grades are calculated to find these "closest" grades for each method. The results in Table 8 reveal that NPLS performs much better than linear PCR for this *Y*-space.

The results are also displayed in Figure 16. We can see that the grades obtained from NPLS inversion are indeed very close to the desired ones. These better results for NPLS when only MWn and MWw are considered in *Y* might have been expected from the previous case, where the nonlinear model clearly fit these properties better.

Since there are many risks involved with the use of nonlinear models for this inversion problem, they should be used

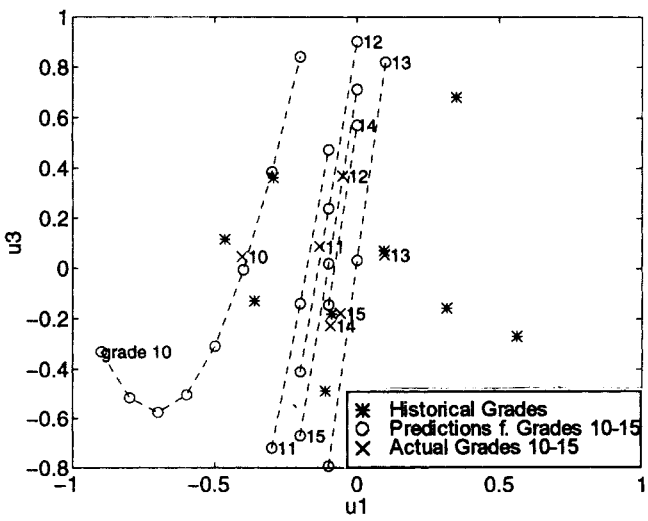

Figure 15. *u*-values of different predictions for grades 10–15, the actual *u*-values, and the *u*-values of the historical grades used for modeling.

Table 8. Mahalanobis Distances to Desired Grades Based on the Specified *Y* Variables (Case 2): $(y_{\text{des}}^T - y_{\text{ach}}^T) \cdot [Cov(Y)]^{-1} \cdot (y_{\text{des}}^T - y_{\text{ach}}^T)^T$

Inversion	Gr. 10	Gr. 11	Gr. 12	Gr. 13	Gr. 14	Gr. 15
PCR	0.0082	0.0144	0.0169	0.0434	0.0133	0.0159
NPLS	0.0004	0.0114	0.0002	0.0029	0.0004	0.0001

with caution. Nonlinear models have more parameters than linear ones. This is a problem in this particular study, since one rarely has more than a small number (usually ≤ 10) of historical grades on which to build the model. In this situation nonlinear models can become very sensitive to an extreme condition. They can also rapidly exhaust all the degrees of freedom in the data set and begin to fit random error. In these situations, although the nonlinear model may fit the data better, they might often yield poorer inverses. Therefore, we recommend that one first uses linear PLS or PCR models using only the historical grade data closest to the new y_{des}^T. If the inner relationship between the scores of the dominant latent vectors in PLS appears reasonably linear, and if the linear models explain nearly the same amount of the X and Y spaces as nonlinear PLS, then we recommend retaining the linear model. Further discussion is given in Jaeckle (1998).

Conclusions

In this article, we have developed a methodology based on latent variable techniques that use historical process data to determine a window of process operating conditions within which a product with desired quality specifications can be produced. It is important not only to design process conditions that will achieve the specified qualities y_{des}^T, but also to ensure that these process conditions retain the same covariance structure as the historical data. This is the key point, which leads to preference of multivariate methods such as PLS and PCR over other methods. The advantage of these multivariate latent variable models arises from the fact that

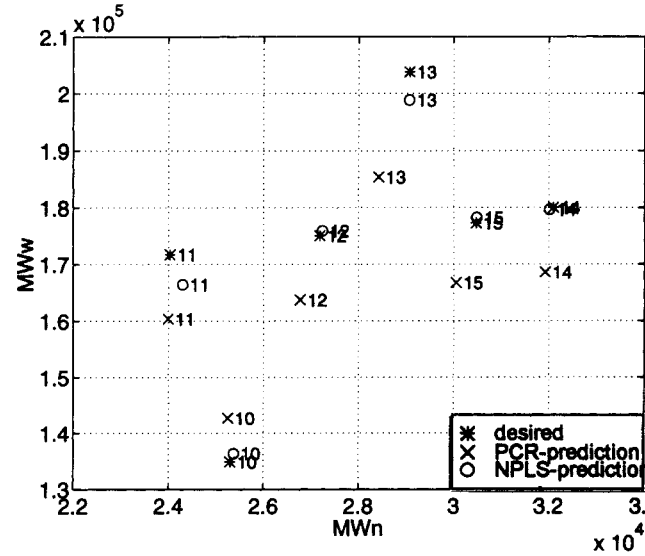

Figure 16. NPLS and PCR results for number and weight-average molecular weight when *Y* = [MWn MWw].

unlike other methods they model the X-space (process conditions), as well as the Y-space (product quality). These techniques reduce the dimensionality of the variable space and, at the same time, lead to better conditioned model inverses which respect past operating strategies.

Any approach based solely on empirical data entails certain limitations: Not every set of process and quality variables is equally suitable for this purpose of product development. Analyzing the model can provide insight into how much information exists in the data. If only about 50% of Y or less is explained due to the lack of information in X, of course, any predictions of process conditions from that model will not be very reliable. Such data do not contain the necessary information for a data-based approach to this problem.

Since we are using empirical models which rely on data from past production, the new quality specifications should lie within the range of past ones, although in some cases slight extrapolations may be possible. For the same reason, it is not possible to predict process conditions which are completely different from any others used before. Even though there may exist other sets of conditions which achieve the desired quality, this methodology cannot find these conditions if they are outside the range and correlation structure of the data and therefore the validity of the model.

The approach discussed here does not necessarily lead to the desired quality in one step. Repeatedly following the sequence of data analysis, running the suggested process conditions, adding the result to the data set, rebuilding the model and reinverting it should lead to the desired result in a few iterations. It is expected that some fine-tuning of the resulting process conditions may require a few additional designed experiments. The process conditions obtained from the inversion should serve as a very good starting point for the experimental design.

Acknowledgments

We would like to express our gratitude to C. Kiparissides for providing us with a version of the LDPE simulation program. Financial support for this work has been granted by the National Science and Engineering Research Council and the Ontario Graduate Scholarship program.

Notation

a = index for PLS dimension
A = number of PCR or PLS dimensions that are used
$B = (A \times k)$ matrix of regression coefficients in PCR
$c_{0_a}, c_{1_a}, c_{2_a}$ = coefficients for quadratic inner relationship in NPLS dimension a
j = index for variables in X or Y
k = number of columns (variables) in Y
m = number of columns (variables) in X
$M^T = (k \times m)$ inversion matrix for a model
n = number of rows (objects) in X, Y (number of grades)
$P = (m \times A)$ matrix of loading vectors for X
$Q = (k \times A)$ matrix of loading vectors for Y
$S_A, \mathbf{S}_\mathbf{A}$ = A-dimensional diagonal matrix containing standard deviations of columns in T
SPE = squared prediction error
t, t_a = score vector for X in PCR or PLS, in PCR or PLS dimension a
t^T = X-scores for one grade
$\hat{t}^T_{\text{new}}$ = prediction of X scores for new grade
$T = (n \times A)$ matrix of X-score vectors
u^T = row vector containing normalized X-scores for one grade
$\hat{u}^T_{\text{new}}, \hat{u}_{\text{new}_a}$ = prediction of normalized X-scores for new grade, ath element of $\hat{u}^T_{\text{new}}$
U = matrix of left singular vectors of X
$U_A, \mathbf{U}_\mathbf{A}$ = $(n \times A)$ matrix of A normalized X-score vectors
V = matrix of right singular vectors of X
$V_A = (m \times A)$ matrix with A orthonormal columns; (first A columns of V)
x^T = row in X = process conditions for one grade
$\hat{x}^T_{\text{new}}$ = part of *predicted* process conditions resulting from inversion
$\hat{X}$ = estimate for X
y^T = row in Y = quality values for one grade
$\hat{Y}$ = estimate for Y
z, z_a = score vector for Y in NPLS, ath element of z
$\hat{z}^T_{\text{new}}, \hat{z}_{\text{new}_a}$ = prediction of Y-scores for new grade, ath element of $\hat{z}^T_{\text{new}}$
$Z = (n \times A)$ matrix of Y-score vectors in NPLS

Greek letters

$\beta = (m \times k)$ matrix of regression coefficients
σ = standard deviation
σ_a = ath diagonal element in Σ_A
Σ = diagonal matrix resulting from SVD on X
$\Sigma_A = (A \times A)$ diagonal matrix

Literature Cited

Burnham, A. J., R. Viveros, and J. F. MacGregor, "Frameworks for Latent Variable Multivariate Regression," *J. Chemometrics*, **10**, 31 (1996).

Draper, N. R., and H. Smith, *Applied Regression Analysis*, Wiley, New York (1981).

Geladi, P., and B. R. Kowalski, " Partial Least-Squares Regression: A Tutorial," *Anal. Chim. Acta*, **185**(1),1 (1986).

Höskuldsson, A., "PLS Regression Methods," *J. Chemometrics*, **2**, 211 (1988).

Jackson, J. E., *A User's Guide to Principal Components*, Wiley, New York, (1991).

Jaeckle, C. M., "Multivariate Statistical Approaches to Product and Process Optimization," PhD Thesis, McMaster Univ., Hamilton, Canada, in press (1998).

Johnson, R. A., and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood Cliffs, NJ (1988).

Kettaneh-Wold, N., J. F. MacGregor, B. Dayal, and S. Wold, "Multivariate Design of Process Experiments (M-DOPE)," *Chemometrics and Int. Lab. Syst.*, **23**, 39 (1994).

Kiparissides, C., G. Verros, and J. F. MacGregor, " Mathematical Modelling, Optimization and Quality Control of High Pressure Ethylene Polymerization Reactors," *J. Macromol. Sci.-Rev. Macromol. Chem. Phys.*, **33**(4), 437 (1993).

Kourti, T., and J. F. MacGregor, "Multivariate SPC Methods for Process and Product Monitoring," *J. Qual. Tech.*, **28**(4), 409 (1996).

Little, R. J. A., and D. B. Rubin, *Statistical Analysis with Missing Data*, Wiley, New York, (1987).

MacGregor, J. F., C. M. Jaeckle, C. Kiparissides, and M. Koutoudi, "Monitoring and Diagnosis by Multi-Block PLS Methods," *AIChE J.*, **40**(5), 826 (1994).

Moteki, Y., and Y. Arai, "Operation Planning and Quality Design of a Polymer Process," *IFAC DYCORD*, Pergamon Press, Bournemouth, UK (1986).

Nelson, P. R., P. A. Taylor, and J. F. MacGregor, "Missing Data Methods in PCA and PLS: Score Calculations with Incomplete Observations," *Chemometrics and Int. Lab. Syst.*, **35**, 45 (1996).

Roffel, J. J., J. F. MacGregor, and T. W. Hofmann, "The Design and Implementation of a Multivariable Internal Model Controller for a Continuous Polybutadiene Polymerization Train," *IFAC DYCORD*, Pergamon Press, Maastricht, Netherlands (1989).

Wold, S., "Nonlinear Partial Least Squares Modelling: 2. Spline Inner Relation," *Chemometrics and Int. Lab. Syst.*, **14**, 71 (1992).

Wold, S., Esbensen, K. and P. Geladi, "Principal Component Analysis," *Chemometrics and Int. Lab. Syst.*, **2**, 37 (1987).

Wold, S., N. Kettaneh-Wold, and B. Skagerberg, "Nonlinear PLS Modelling," *Chemometrics and Int. Lab. Syst.*, **7**, 53 (1989).

Manuscript received Sept. 17, 1997, and revision received Jan. 15, 1998.