Product Transfer Between Plants Using Historical Process Data

Christiane M. Jaeckle and John F. MacGregor
Dept. of Chemical Engineering, McMaster University, 1280 Main St. West, JHE-374, Hamilton, Ontario, Canada, L8S 4L7

Based on the concepts laid out in an earlier article (Jaeckle and MacGregor, 1998), this paper defines the problem of moving the production of a particular product grade from a plant A to another plant B when both plants have already produced a similar range of grades. Since the two plants may differ in size, configurations, and so on, the process conditions required to produce any given product grade may be very different in the two plants. How historical process data on both plants may be utilized to assist in this problem is investigated. A multivariate latent variable method is proposed that uses data from both plants to predict process conditions for plant B for a grade previously produced only in plant A. The approach is illustrated by a simulation example.

Introduction

Industrial processes often produce a variety of different grades of a product (e.g., a polymer) by simply changing the operating conditions and the ratio of ingredients in the feed to a continuous process or in the recipe of a batch process. The quality specifications for the different grades are typically measured by a set of quality variables such as tensile strength or melt flow index. When the production of a certain grade is transferred to a second process or another plant, it is very important to achieve identical-grade quality from both production sites. This ensures that customers who receive the product will not notice any difference between the grades from different plants.

Consider two similar plants, A and B, at two different locations. Both plants produce products out of the same product family (e.g., LDPE film grades or polymer latex coatings). They may even have some overlap in their product lines, that is, some of the grades may already be produced in both plants. The necessity often arises for a grade that is currently only manufactured in plant A also to be produced in plant B with the same quality specifications. Another problem is where the same grade is already being produced in both plants, but the product quality from the two plants is different enough that customers receiving the grade from both plants notice a performance difference. There is then a need to adjust the conditions of one of the plants to eliminate this difference.

Although making products from the same family and based on the same physical principles, the two plants might be of different size and configuration, or use different technology. They also might have different raw material suppliers, cooling water temperatures, among other things. Therefore, in order to produce identical quality on a particular grade, the operating strategies and conditions for production of this grade will look different for the two plants. The following questions are dealt with in this article. How can we use the information on existing product grades manufactured in both plants A and B, and on the new grade in plant A to find the corresponding process conditions needed to manufacture the new grade in plant B? If both plants are currently producing the same grade, how can we use the data from both plants A and B to adjust the conditions of plant B in order to align their product qualities more closely?

Different approaches could be considered. If a detailed mechanistic model for plant B existed, then an optimization approach could be used to find the new conditions. Alternatively, an experimental program employing designed experiments could be carried out on plant B. If historical process data exist for the two plants, however, it may be beneficial to explore those first before resorting to more expensive studies. A databased approach will in general take less time, be more cost effective, and will reveal what one might be able to ac-

Correspondence concerning this article should be addressed to J. F. MacGregor.
Present address of C. M. Jaeckle: Wacker Polymer Systems, Postfach 1560, 84483 Brughausen, Germany.

complish based on existing process data from the two plants. Depending on the available data, it may directly lead to new process conditions for the second plant that yield the desired-grade quality, or it may at least provide a good starting point for such new conditions, which can then be refined through further experimentation.

We propose a databased solution to the problem using latent variable concepts. The idea can be seen as an extension to the problem of product grade design (within one plant) presented by Jaeckle and MacGregor (1998) and applied to industrial data in Jaeckle and MacGregor (2000). The current problem could be solved considering only data from plant B for which the new process conditions are sought, and applying the method of product-grade design suggested in those former articles. In this case, however, data on plant A also exist as does knowledge on how the desired grade is manufactured there. Therefore, we investigate approaches that also make use of this additional information, available in the data for plant A.

The methodology can be applied to either continuous processes or batch processes. In general, data for this problem are assumed to consist of (a) the values for the variables describing the quality of each product grade, and (b) the corresponding process operating conditions for each grade. The latter are typically characterized by the set of process variables that can be independently manipulated. For a continuous process, these consist of the steady-state operating points of feed rates, temperatures, and so forth. For a batch process they include the recipes for the different grades—that is, the different amounts of ingredients charged to the reactor, as well as, feed rate profiles (if semibatch operation is used) and operating profiles of process variables such as temperature or agitator speed throughout the batch history.

If the m manipulated variable values and k quality values for a grade are arranged as two row vectors, then those row vectors for all n product grades can be used to construct two matrices X $(n \times m)$ and Y $(n \times k)$, where X contains the operating conditions for all the grades and Y the corresponding qualities. Each row in X and Y corresponds to one grade.

The relationship between operating conditions and grade qualities within plant A can then be approximated by a model $Y_A = f_A(X_A)$. In the same way a model can be built for plant B $[Y_B = f_B(X_B)]$. The objective is to predict process conditions $x^T_{B_{\text{pred}}}$ for plant B that result in the grade quality $y^T_{A_{\text{des}}}$, which so far has been only achieved by plant A—by using the historical data X_A, X_B, Y_A, Y_B.

In many cases there are fewer quality variables than manipulated process variables, and the mathematical problem of finding new process conditions is underdetermined. However, those physical variables are typically not independent of each other, but are often highly correlated. It is therefore important to maintain the covariance structure present in the historical data from plant B when predicting new process conditions for plant B.

This idea leads to the concept of latent variable methods where the correlated physical variables are transformed into a smaller set of independent latent variables. Using such latent variables serves to predict process conditions that are consistent with the operating strategies of the plant from the past. This helps to prevent process upsets when testing the new conditions. It also ensures that the new operating conditions maintain the structure of the historical data, and therefore lie within the feasible operating region of plant B. A good overview of multivariable latent variable methods, such as principal components analysis (PCA), principal components regression (PCR), and partial least squares or projection to latent structures (PLS) is given in Geladi and Kowalski (1986), Wold et al., (1987a), Martens and Naes (1991), and Burnham et al. (1999).

The outline of the article is as follows. The different data matrices that are assumed to be available for this problem are introduced in the next section. This leads to the question of which product grades should be included when building a model that is used for the current purpose. The section on selection of grades addresses this issue and describes a strategy for selecting the grades that are benefical to solving the problem. Afterwards the pretreatment of the data is outlined. This is followed by the description of an approach that can be applied to combine the data from two plants in order to derive plant B process conditions for a grade so far produced only in plant A. The approach is illustrated by a simulation example. A few additional issues and ideas are discussed in the final section.

Although a few key points are repeated here, the reader is referred back to the earlier article by Jaeckle and MacGregor (1998) for a general philosophy on databased approaches to this type of problem.

Data Structure

Before looking at possible solutions, we discuss the structure of the data that one would typically expect in this kind of problem. We assume data for a series of grades belonging to the same type of product to be available *for both plants.*

Each plant is characterized by a matrix X containing the average process conditions of the plant for each grade produced in it, and a matrix Y that contains the corresponding quality measurements for those grades. Obviously each plant can have produced a different number of grades out of that product family. This number is denoted by n_A and n_B, repectively, and marks the number of rows in the data matrices X and Y for the two plants. In the matrices for one plant each row corresponds to a particular grade.

In case there are common grades between the two plants, we denote the matrix that contains the quality specifications for those common grades by Y_c, to contrast with grades that are produced either only in plant A, Y_{Au}, or only in plant B, Y_{Bu}, where "c" stands for "common" and "u" for "unique". (Grades that are *supposed* to be identical in both plants, but do not turn out that way, belong to matrices Y_{Au} and Y_{Bu}.)

The $(n_A \times k)$ matrix Y_A is constructed as $Y_A = \begin{bmatrix} Y_{Au} \\ Y_c \end{bmatrix}$ and for now contains all the grades that have been produced in plant A. (In the section on selection of grades we test whether some grades may have to be eliminated from the initial Y_A in order to benefit the databased approach to this problem.) In the same way matrix Y_B consists of all the grades that are available from plant B, where $Y_B = \begin{bmatrix} Y_c \\ Y_{Bu} \end{bmatrix}$. These Y-matrices are assumed to contain the same number of columns (k) with

each of the k quality variables corresponding to the same columns in Y_A and Y_B.

The process conditions for both plants resulting in the respective grades that are unique to each plant are denoted by X_{Au}, X_{Bu}. Their rows relate to those of Y_{Au} and Y_{Bu}. The process conditions for the common grades are summarized in matrix X_{Ac} for plant A, in X_{Bc} for plant B such that their respective rows correspond to the grades in Y_c.

The process data for each plant can be combined to a $(n_A \times m_A)$ matrix $X_A = \begin{bmatrix} X_{Au} \\ X_{Ac} \end{bmatrix}$ and a $(n_B \times m_B)$ matrix $X_B = \begin{bmatrix} X_{Bc} \\ X_{Bu} \end{bmatrix}$. Since the two plants are expected to be dissimilar in several ways, some of the process variables used in X_A may differ from those used in X_B. One can also expect the *number* of process variables in the two plants to be different $(m_A \neq m_B)$.

Figure 1 shows these different data blocks. The question that arises is: How can the different blocks be used in predicting plant B process conditions for the desired grade $y^T_{A_{des}}$, which so far has only been produced in plant A?

Obviously, the *main* information on how to find such conditions for plant B is contained in $X_B = \begin{bmatrix} X_{Bc} \\ X_{Bu} \end{bmatrix}$ and $Y_B = \begin{bmatrix} Y_c \\ Y_{Bu} \end{bmatrix}$. However, we already know how to produce $y^T_{A_{des}}$ in plant A, and how the process conditions in both plants (X_{Ac}, X_{Bc}) are related to the quality measurements of prior common grades (Y_c). Adding this information on X_{Ac}, $x^T_{A_{des}}$, and $y^T_{A_{des}}$ to the information on plant B can be beneficial in dealing with this problem.

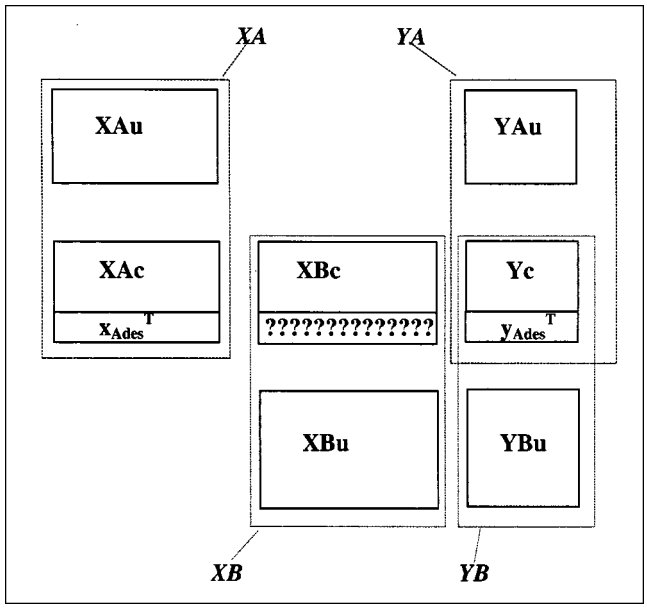

Figure 1. Data structure for process conditions and resulting quality in two plants A and B.

Process conditions for plant B that need to be determined are denoted by ???.

The data on the "unique" grades in plant A (X_{Au}, Y_{Au}) are useful in helping to better define the latent variable space of X_A, Y_A. To help with prediction of plant B conditions, however, only those grades of Y_{Au} that are consistent with the covariance structure of the plant B grades $\left(Y_B = \begin{bmatrix} Y_c \\ Y_{Bu} \end{bmatrix}\right)$ should be retained. This is discussed further in the next section.

Ideally one would prefer a method that employs all the relevant information in the data blocks. However, there are some simple approaches that use only a subset of information:

1. $X_B \rightarrow Y_B$: Using PCR or PLS to model the relationship between the plant B process conditions and the product quality in plant B, one of the inversion approaches of Jaeckle and MacGregor (1998) could be applied, and process conditions for the new grade predicted.

2. $X_{Ac} \rightarrow X_{Bc}$: If there are enough common grades between the two plants, and especially if at the same time the two plants *are* very similar, then a latent variable model (e.g., PLS) relating X_{Ac} and X_{Bc} could be used to predict a new set of process conditions for $y^T_{A_{des}}$ for plant B from the corresponding process conditions in plant A. But especially when parts of the equipment in the two plants or the operating strategies are very different, it is likely that process conditions in plant A can predict only *some* features of the corresponding process conditions in plant B. In other words, X_{Ac} will most likely *not* project into *all* of the directions spanned by X_B. [A possible combination of the previous two approaches is suggested in Jaeckle (1998).]

3. $\begin{pmatrix} X_{Ac} & X_{Bc} & Y_c \\ \cdots & \cdots & \cdots \\ x^T_{A_{des}} & ??? & y^T_{A_{des}} \end{pmatrix}$

Process conditions from both plants for only the common grades and the corresponding quality measurements can be combined in one matrix. A model can be fit to describe the variation within this data structure. The plant B process conditions for $y^T_{A_{des}}$ can then be treated as missing data, and their expected values, given the values for the quality measurements $y^T_{A_{des}}$ and the process conditions for plant A $x^T_{A_{des}}$, can be calculated based on the model. This approach would be analogous to the Conditional Expectation approach presented in Jaeckle and MacGregor (1998) for a single plant.

Since the preceding approaches use only part of the information available, they will not be discussed further. The approach outlined later on in this article is an attempt to use as much of the relevant information as possible. In particular, it will use X_B, Y_B, and selected grades from Y_A to build an extended PCR-model. The part of this model that corresponds to plant B is then inverted for a desired grade from plant A, $y^T_{A_{des}}$, to predict plant B process conditions for this grade.

Selection of Grades

There is no guarantee that all the grades that have been produced by plant A can be produced by plant B, even if there exist some common grades between the plants. Only grades from plant A that conform with the covariance structure of the quality data for plant B grades will be of support

in designing process conditions in plant B. The quality matrices Y_A and Y_B each span a space within which the respective plant has been producing grades so far. A databased approach should only attempt to predict conditions for a new grade if it lies within the same space spanned by the *Y*-matrix for that plant.

In order to select the grades from plant A that might be produced in plant B, and that may prove useful in building a model, we need to compare the spaces that are spanned by Y_A and Y_B, respectively. Several situations can be distinguished.

1. Both spaces, Y_A and Y_B, have the same dimension and overlap completely. In this case, all the grades in Y_A can be included in the model. Any grade contained in Y_A might reasonably be expected to be produced in plant B, except perhaps for grades that lie far from any of the existing Y_B grades. However, whether they look feasible, that is, whether the model still has validity in that area, and whether an attempt to implement them seems promising or not, should be apparent from the set of solutions obtained for $x_{B_{pred}}^T$.

2. Y_B shows a higher dimensionality than Y_A, but the space of Y_A is contained with the Y_B-space. In this case, the same comments just made for case 1 will again hold.

3. A more general case is one where the spaces of Y_A and Y_B intersect, but where part of the Y_A-space is *not* contained within Y_B. This implies that there are some grades in Y_A that have a quality structure that is substantially different from those that have so far been produced in plant B (Y_B). Obviously, the available data for plant B do not justify the prediction of process conditions for production of such grades in plant B, even if they were possible. Hence it is necessary to exclude those grades in Y_A that have this different structure. Once this is done, we will again be in the situation described under case 1 or 2.

4. There is no overlap between the two spaces Y_A and Y_B. In this case, no solution is possible based only on the empirical data.

We use the following method to compare the two spaces and to select suitable plant A grades that lie within the space of Y_B. A PCA model is build on Y_B to describe the quality space spanned by the grade data of plant B. If the dominant variation in the n_B grades of plant B can be captured by using F principal components, then the F-dimensional space spanned by these grades is defined by

$$\hat{Y}_B = T_B P_B^T,$$

where T_B is a $(n_B \times F)$ matrix whose columns consist of the values of the first F principal components (scores) for the n_B grades of the Y_B matrix, and P_B is a $(k \times F)$ matrix of loading weights showing how the principal component scores are computed from the original Y_B matrix ($T_B = Y_B P_B$). All the plant A grades have to be tested against this model, that is, they are projected onto the model space of Y_B and their squared prediction error (SPE) is calculated. The projection of plant A grade i is computed as

$$t_{A_i}^T = y_{A_i}^T P_B, \tag{1}$$

where y_{A_i} is the $(k \times 1)$ vector of quality variables for grade i of plant A, and t_{A_i} is the $(F \times 1)$ vector of projected score locations for grade i of plant A on the F-dimensional PCA model plane of plant B grades.

The SPE is calculated for each grade i as

$$\mathrm{SPE}_i = \left(y_{A_i} - \hat{y}_{A_i}\right)^T \left(y_{A_i} - \hat{y}_{A_i}\right), \quad \text{where} \quad \hat{y}_{A_i}^T = t_{A_i}^T P_B^T. \tag{2}$$

By comparing those SPEs to the SPEs of the model grades Y_B, one can detect the grades in Y_A with a much larger SPE. These grades do not lie in the space of the Y_B grades, and therefore should be excluded from the Y_A data used for the modeling. Unfortunately, in most data sets there will not be enough grades available in plant B to allow for estimation of a confidence limit for the SPE, as is done in Nomikos and MacGregor (1995). Therefore, the decision whether or not to include a certain grade in Y_A might have to be based on one's own judgment on the acceptable size of the SPE for the Y_A grades, relative to those for the Y_B grades.

Although a Y_A grade may have a small SPE, the location of its projection into the space of the PCA-model (t_{A_i}-scores of Eq. 1) may still fall far from the locations of the scores of the Y_B grades ($t_{B_i}^T = y_{B_i}^T P_B$) in that same space. In fact, it is possible that *all* the selected Y_A grades form a cluster in the *t*-space that is separate from the cluster of Y_B grades. This may happen when there are no common grades between the plants, and when the means of Y_A and Y_B differ a lot from each other. In such a case, the Y_A grades should be used with caution.

Scaling of the Data

The columns in X_B and Y_B are mean centered and scaled to unit variance. Since selected grades from plant A must be lined up with Y_B later on, Y_A is also centered and scaled using the mean and the standard deviation of the Y_B data. Because the approach discussed later in this article does not use the data in X_A at all, one does not need to be concerned with centering and scaling the X_A data. Scaling both Y_A and Y_B with the mean and standard deviation of Y_B offers the following benefits.

1. The common grades between Y_A and Y_B still have equal values after scaling and adjusting for the mean, and therefore have the same PCA-scores as well. A PCA model on the scaled and mean adjusted $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ will thus be more focused on the common space.

2. This common mean centering and scaling of Y_A and Y_B ensures that the relationship between the two spaces is not altered. In other words, if the unscaled Y_A, Y_B fall into the same space or intersect in a certain way, then so will the scaled matrices when scaled and mean adjusted as just suggested.

Approach Using an Extended PCR Model

In this section the concepts and motivation for the methodology are presented before the mathematical equations for it are given.

The suggested approach to this problem is closely related to the PCR inversion approach in Jaeckle and MacGregor (1998). It is based on two data matrices: X_B $(n_B \times m_B)$ and $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ $((n_A + n_B) \times k)$. Here, X_B contains the adjustable variables of plant B that make up the process conditions for the different grades. Since plant A already produces the desired product grade $y^T_{A_{des}}$, its physical feasibility is ensured. We also assume that it is possible to produce $y^T_{A_{des}}$ in plant B, as it belongs to the grades that have a reasonably small SPE with respect to a model on historical plant B grades, and have PCA score values that lie within the range of the Y_B grades (see section on selection of grades). The quality matrix $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ can contain any available quality measures that are considered vital for the final product, and are available for both plants.

Because the grades in Y_A have been selected to lie mostly in the same space as the Y_B grades, the PCA model on $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ should be effectively the same as that on Y_B alone. The benefit of using the augmented matrix is that the use of more grade data should lead to a more robust model, in particular in the area that is common to Y_A and Y_B grades.

Since the data matrices X_B and $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ are almost invariably less than full statistical rank (i.e., most variation lies in subspaces of these matrices), latent variable methods are used (Burnham et al., 1999). The problem thus becomes one of estimating the latent variable values required for the product, and then computing the actual manipulated variables from those. Assume that the process variable space X_B and the quality space $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ have effective dimensions B and C, respectively, where in general $B \geq C$.

We begin with singular-value decompositions on X_B and $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$, where the first B and C significant components, respectively, are retained. They serve as approximations for X_B and $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$:

$$\underset{(n_B \times m_B)}{\hat{X}_B} = \underset{(n_B \times B)}{U_B} \cdot \underset{(B \times B)}{\Sigma_B} \cdot \underset{(B \times m_B)}{V_B^T}$$

$$\underset{(n \times k)}{\begin{bmatrix} \hat{Y}_A \\ \hat{Y}_B \end{bmatrix}} = \underset{(n \times C)}{U_y} \cdot \underset{(C \times C)}{\Sigma_y} \cdot \underset{(C \times k),}{V_y^T} \tag{3}$$

where B is the number of latent directions in X_B, C is the number of latent directions in $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$, $n = n_A + n_B$; and U_B and U_y can be interpreted as the normalized latent variables that span X_B and $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$, respectively.

To relate the latent variables of X_B to the latent variables of Y_B, we first need to separate the matrix U_y into the two parts that correspond to Y_A and Y_B, respectively:

$$U_y = \begin{bmatrix} U_{y_A} \\ U_{y_B} \end{bmatrix}.$$

The latent structure of the grades in Y_B is represented by U_{yB}, which is of dimension $(n_B \times C)$.

In order to express U_{y_B} in terms of the latent variables of X_B, we must rotate and stretch the latent variables U_B until they line up as closely as possible with the latent variables U_{y_B}. In other words, we are looking for a matrix R_B such that:

$$\underset{(n_B \times C)}{U_{y_B}} = \underset{(n_B \times B)}{U_B} \cdot \underset{(B \times C)}{R_B}. \tag{4}$$

An estimate for R_B can be obtained from the least squares solution to Eq. 4:

$$\underset{(B \times C)}{\hat{R}_B} = \underset{(B \times n_B)}{U_B^T} \cdot \underset{(n_B \times C)}{U_{y_B}}.$$

The preceding is basically principal components regression (PCR) where the principal components of the Y_B-space are regressed onto the principal components of the X_B space.

The Y-scores for the grade $y_{A_{des}}$ that is to be produced in plant B are already available as a row in U_{yA}, since $y^T_{A_{des}}$ is a row in the Y_A-matrix, and as such part of the model (Eq. 3). An advantage of using the combined matrix $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ is that Y-scores for a grade are the same whether a grade is produced in plant A or B, that is, $u^T_{y_{B_{des}}} = u^T_{y_{A_{des}}}$.

The problem that remains is to transform these values from the latent space of Y_B to the latent space of X_B. The key here is Eq. 4, which can be written for the desired grade as

$$\underset{(1 \times C)}{u^T_{y_{B_{des}}}} = \underset{(1 \times B)}{u^T_{B_{pred}}} \cdot \underset{(B \times C)}{\hat{R}_B}. \tag{5}$$

For $B > C$ we have an underdetermined equation system with C equations and B unknowns that has an infinite set of solutions $u^T_{B_{pred}}$ with minimum error. Each of the solutions can be decomposed into the minimum norm solution

$$\hat{u}^T_{B_{new}} = u^T_{y_{B_{des}}} \cdot \left(\hat{R}_B^T \cdot \hat{R}_B \right)^{-1} \cdot \hat{R}_B^T, \tag{6}$$

and a null component $u^T_{B_{null}}$ so that $u^T_{B_{pred}} = \hat{u}^T_{B_{new}} + u^T_{B_{null}}$. Whereas the preceding inversion (Eq. 6) of the PCR model provides the best solution that lies in that subspace of $\hat{X}_B$, which according to the model affects the product quality, the null components lie in the remaining subspace of $\hat{X}_B$ that does not affect the product quality. This latter subspace is

characterized by the following relationship:

$$u_{B_{\text{null}}}^T \cdot \hat{R}_B \cdot \Sigma_y \cdot V_y^T = 0. \tag{7}$$

In other words, any vector $u_{B_{\text{null}}}^T$ that lies in the left null space of $\hat{R}_B$ fulfills Eq. 7. This null space is defined with the help of the singular-value decomposition of $\hat{R}_B$ by

$$\underset{(B\times C)}{\hat{R}_B} = \underset{(B\times C)(B\times(B-C))}{\left[G_1 \vdots G_2\right]} \cdot \underset{(B\times C)}{\Sigma_{\hat{R}_B}} \cdot \underset{(C\times C).}{V_{\hat{R}_B}^T}$$

G_2 in this equation denotes the left null space of $\hat{R}_B$. Hence, $u_{B_{\text{null}}}^T$ can be any vector in the space spanned by the rows of G_2^T:

$$\underset{(1\times B)}{u_{B_{null}}^T} = \underset{[1\times(B-C)]}{\lambda^T} \cdot \underset{[(B-C)\times B]}{G_2^T}$$

where λ^T is an arbitrary row vector of the proper dimension.

Although mathematically $u_{B_{\text{null}}}^T$ is not only of arbitrary direction within G_2^T, but also of arbitrary magnitude, for the current problem the only solutions that are acceptable are those within the range of the historical data. In other words any $u_{B_{\text{pred}}}^T = \hat{u}_{B_{\text{new}}}^T + u_{B_{\text{null}}}^T$ must fall close to the cluster of the existing U_B-rows in order to guarantee a solution that is compatible with the past plant B operating strategies that are represented in X_B. In practice various values of λ are assumed, and those solutions falling close to existing U_B values are retained as possible solutions.

The reconstruction of values for the manipulated variables is achieved in the same fashion as described in Jaeckle and MacGregor (1998). By imposing the variance and covariance information contained in Σ_B and V_B^T the values for the scaled variables are obtained:

$$x_{B_{\text{pred}}}^T = \left(\hat{u}_{B_{\text{new}}}^T + u_{B_{\text{null}}}^T\right) \cdot \Sigma_B \cdot V_B^T.$$

Multiplying with the standard deviation and adding the mean of X_B will give the actual values for the manipulated X variables.

We again obtain not just a single solution, but a whole *region* of operating conditions for plant B, which according to the model, should result in the same grade quality as $y_{A_{\text{des}}}^T$ from plant A. Engineering judgment would then have to be used to select a reasonable condition from within this window, perhaps one providing the most economical operation. Iterations between implementing the predicted conditions, rebuilding the model with the newly obtained data point, and reinverting may also be necessary to fine-tune the quality values. Alternatively a designed experiment may be performed near this location to accomplish the final adjustments.

To keep the mathematical outline of the approach simple the following issue has been left aside so far: The inversion in Eq. 6 requires the $(B\times C)$ matrix $\hat{R}_B$ to be of full rank C $(C \le B)$. Then $\hat{R}_B$ is calculated as

$$\underset{(B\times C)}{\hat{R}_B} = \underset{(B\times n_B)}{U_B^T} \cdot \underset{(n_B\times C)}{U_{yB}},$$

and therefore $\hat{R}_B$ is of rank C unless at least one of the columns in U_{y_B} lies in the null space of U_B. However, this would suggest that some of the variation in Y_B is caused by an "event" that is not present in the variation in X_B. In other words, there is variation in the quality of the grades that is not caused by the manipulated variables contained in X_B. Obviously, it is not possible in such a case to change the grade quality in the desired way by only changing the manipulated variables that are present in X_B. Such a data set would not be suitable for a databased approach, and hence is not considered any further.

Simulation Example

It would have been desirable to present industrial applications of this methodology much as was done on the product design problem (Jaeckle and MacGregor, 2000). However, obtaining such confidential data from a company on their operating conditions for all grades at two or more plants, and then evaluating the results by producing the new product, was not possible at this time. Therefore, in this section we evaluate the approach using simulated data that have characteristics expected to be seen in practice.

The simulated data set is presented in the Appendix in Tables A1 and A2. It consists of data on 8 process variables (X_B) and 5 quality variables (Y_B) for 12 grades from a plant B and for 9 grades from plant A. The quality data for plant A are collected in a 9×5 matrix $Y_{A_{\text{initial}}}$ to indicate that the grades have not been tested yet to see whether they are similar in structure to the Y_B grades or whether they must be excluded in Y_A. The first two grades of Y_B and $Y_{A_{\text{initial}}}$ are common to both plants. Details on how the data were simulated are given in the Appendix.

Table 1 shows the cumulative variance explained by separate PCA-models on X_B, Y_B, and $Y_{A_{\text{initial}}}$. (All matrices are mean centered and scaled to unit variance.) Based on the results in Table 1, the matrices X_B, Y_B, and $Y_{A_{\text{initial}}}$ appear to be spanned by four, three, and three latent variables respectively.

The first step is to test which of the grades in $Y_{A_{\text{initial}}}$ lie in the same space as the grades of Y_B. For this purpose $Y_{A_{\text{initial}}}$ is mean centered using the mean of the Y_B data and then scaled by the standard deviations of the columns of Y_B. Each of these mean-centered and scaled $Y_{A_{\text{initial}}}$ grades is then projected onto the 3-dimensional PCA model for Y_B to obtain their scores and their squared prediction errors (SPE) (see Eqs. 1 and 2).

Figure 2 shows the SPE values for the grades in Y_B denoted by $(*)$ and the SPE values for $Y_{A_{\text{initial}}}$ grades denoted by $(\bigcirc)$. The SPEs of $Y_{A_{\text{initial}}}$ grades 1, 2, 4, 5, 6, 7, and 8 are within the range of those obtained for the Y_B grades. (Remember that 1 and 2 are the common grades, and hence are

Table 1. Percentage of Variation Explained by PCA in X_B, Y_B, and $Y_{A_{\text{initial}}}$

	No. of Components			
	1	2	3	4
PCA on X_B	35. 96	63. 59	85. 93	95. 44
PCA on Y_B	46. 70	80. 90	99. 08	99. 82
PCA on $Y_{A_{\text{initial}}}$	51. 83	76. 77	99. 97	99. 99

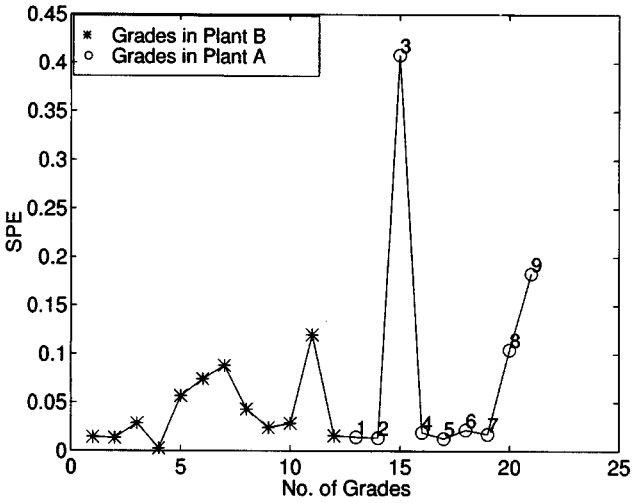

Figure 2. Squared prediction errors for 12 plant B grades (∗) and 9 plant A grades (∘).
The first two grades from both plants are common.

expected to have the same SPE in both matrices.) Grade 3 of plant A with its large SPE should obviously not be used in Y_A, since it clearly has characteristics not seen in the grades produced so far in plant B. The SPE for grade 9 in plant A is a bit larger than any SPE of the Y_B grades. The contribution plot (MacGregor et al., 1994) for this SPE (Figure 3) shows that most of this error is due to variables 2 and 3. We will assume that we are not extremely concerned with these two particular variables and include grade 9 in Y_A, as well as the selected grades (1, 2, 4, 5, 6, 7, 8) with smaller SPEs.

With the combined matrix $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$ and X_B, the extended PCR model is built as described in the previous section. Table 2 shows the variance explained in Y_B by the model, when four latent directions are used in X_B for different numbers

Table 2. Variation in Y_B by Extended PCR Model Using First Four Latent Variables in X_B and Different Numbers of Components for Y_B

	1	2	3	4	5
Y_B	41. 64%	61. 93%	91. 39%	91. 62%	91. 63%

of latent directions in Y_B. These results show that a three-component model for Y_B is sufficient.

To evaluate the methodology, the plant B process conditions for the six grades in Y_A that are not common with Y_B (i.e., grades 4, 5, 6, 7, 8 and 9) are obtained through inversion of the model, as explained in the previous section. Since there is no real process behind this simulation, ''implementation'' of these conditions can only be achieved by resimulating using all six predicted conditions $\hat{x}^T_{B_{\text{new}}}$, where no null component is added. According to the model, a null component would not affect the quality anyhow. After this ''implementation'' of $\hat{x}^T_{B_{\text{new}}}$ for the six Y_A grades, the achieved $Y_{B_{\text{achieved}}}$ are compared to the desired qualities in Y_A in Figures 4 and 5. A selective PCA analysis (Roffel et al., 1989) on Y_B indicates that the variables Y_4, Y_3, and Y_5 are the most significant variables of all five, and account for 98% of the variability.

Obviously, the achieved results are very close to the desired ones, as expected, due to the simulated nature of the data. However, we do not claim that this example *proves* much in terms of performance of the method. In fact, the performance depends on the particular data set in a way similar to that discussed in the article on product design (Jaeckle and MacGregor, 1998). The example here serves merely to illustrate the application of the methodology.

Discussion and Conclusions

Based on the problem of product design treated in Jaeckle and MacGregor (1998), this article addresses the related problem of transferring production of a product grade from one plant, A, to another plant, B. The same methodology can be applied to adjust the operating conditions of two plants

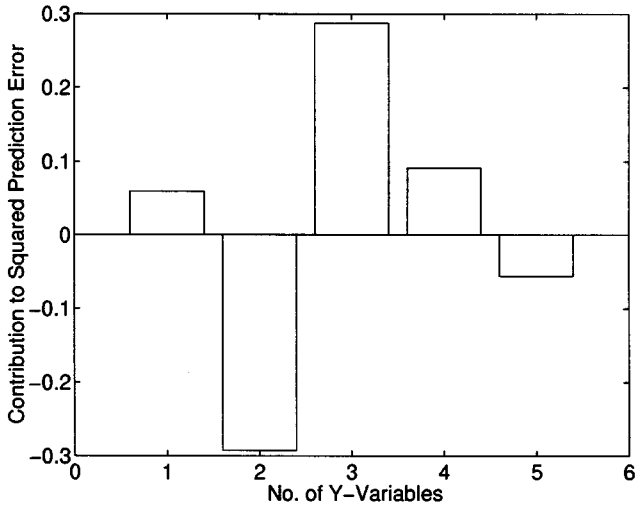

Figure 3. Contributions of the *Y*-variables to the SPE of grade 9 in Y_A.

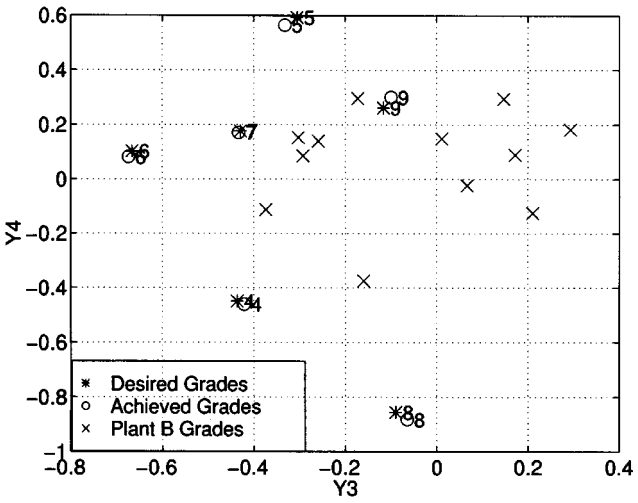

Figure 4. Qualities for Y_B grades, desired Y_A grades, and Y_A grades achieved in plant *B*; Y_3–Y_4.

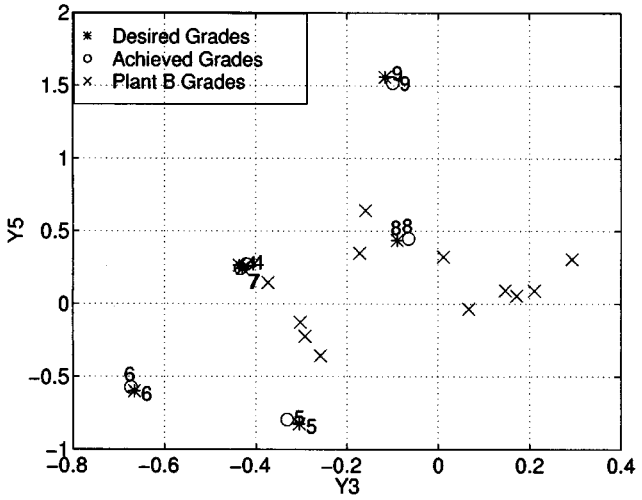

Figure 5. Qualities for Y_B grades, desired Y_A grades, and Y_A grades achieved in plant B; $Y_3 - Y_5$.

producing a common grade so that the product quality produced in the two plants become more similar.

Since for this problem, data sets from *two* plants are available, the structure of these data sets and the relationships between them are very important. It is of particular interest to compare the quality spaces from the two plants. Whether process conditions in plant B for a desirable plant A grade $y^T_{A_{\text{des}}}$ can indeed be derived through a databased approach hinges largely on the relationship between the qualities produced in the two plants so far:

1. The model for plant B must still be valid in that area of the score space into which $y^T_{A_{\text{des}}}$ falls.
2. The resulting process conditions for plant B must still be feasible and must not hit any hard constraints.

Of course, there are cases where, due to the different equipment and configuration of the two plants, production of $y^T_{A_{\text{des}}}$ may not be possible in plant B at all. On the other hand, even if it was theoretically possible to produce $y^T_{A_{\text{des}}}$ in plant B, a databased approach cannot forsee this unless $y^T_{A_{\text{des}}}$ lies within the structure and range of the plant B grades present in the historical data. In other words, the typical limitations inherent to any databased approach apply here as well. One cannot extrapolate too far, nor predict process conditions that are not consistent with the operating strategies that have been recorded for this plant in the past. The empirical models should also explain a large part of the variation in all the data matrices (X_B, Y_B, Y_A), in order to be reliably used for this inversion. Furthermore, it is essential that the measured quality variables truly represent the product quality in its final application.

The approach described in this article is only one of several possible approaches. It can be seen as an extended PCR model with inversion, where the quality space of plant B is supported by suitable plant A grades. A test has been developed that selects those plant A grades that have a latent structure similar to the grades in Y_B. Combining those plant A grades with Y_B enhances the model on the quality structure, especially in the quality area that is common to both plants.

The inversion of the model follows the same concepts as in Jaeckle and MacGregor (1998), where first predictions for the normalized *latent variables* of new process conditions are obtained. Imposing the variance and correlation structure of the plant B process variables (X_B) onto these values then results in the settings for the physical manipulated variables. This ensures that the new process conditions will maintain exactly the same relationships among each other as already present in existing grades.

If the process conditions span a higher dimensional space than the quality measurements, as is often the case, the solution can again be presented as a window of operating conditions that—according to the model—all lead to the same desired quality $y^T_{A_{\text{des}}}$.

The information used in the approach consists of all the plant B data and the quality data of plant A grades that have a covariance structure similar to the plant B grades. No use was made of the process conditions in plant A (X_A). Future work might investigate the usefulness of this information and ways of incorporating it into the solution. A couple of approaches that used all the data from the two plants (Figure 1) have been considered by Jaeckle (1998). They were based on multiblock latent variable methods (Wangen and Kowalski, 1988; Wold et al., 1996, 1987b; Westerhuis et al., 1998), but were intuitive in nature, and so not included in this article.

Notation

F = effective rank of Y_B
i = index for grades
Σ_B = diagonal matrix of B dominant singular values of X_B
Σ_y = diagonal matrix of C dominant singular values of $\begin{bmatrix} Y_A \\ Y_B \end{bmatrix}$
$u^T_{B_{\text{pred}}}$ = prediction of normalized scores of plant B process conditions

Literature Cited

Burnham, A. J., R. Viveros, and J. F. MacGregor, "Latent Variable Multivariate Regression Modelling," *Chemometrics Intelligent Lab. Syst.*, **48**, 167 (1999).

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

Jaeckle, C. M., "Product and Process Improvement Using Latent Variable Methods," PhD Thesis, McMaster Univ., Hamilton, Ont., Canada (1998).

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

Jaeckle, C. M., and J. F. MacGregor, "Industrial Applications of Product Design through the Inversion of Latent Variable Models," *Chemometrics Intelligent Lab. Syst.*, **50**, 199 (2000).

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).

Martens, H., and T. Naes, *Multivariate Calibration*, Wiley, New York (1991).

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

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

Wangen, L. E., and B. R. Kowalski, "A Multiblock Partial Least Squares Algorithm for Investigating Complex Chemical Systems," *J. Chemometrics*, **3**, 3 (1988).

Table A1. Data Used in the Simulation Example: X_B, Y_B

Grades	X_B								Y_B				
1	−0.0593	−0.0211	0.0615	0.2808	−0.0656	−0.2057	0.0196	0.4516	−0.2366	−0.1005	0.0118	0.1485	0.3231
2	0.1757	0.3035	0.0402	−0.0814	0.0305	−0.3364	0.1263	0.1312	−0.3758	0.0658	0.1475	0.2947	0.0913
3	0.1182	0.0981	0.0839	−0.3737	−0.0630	0.6254	−0.0236	0.1998	−0.4493	−0.1615	−0.1723	0.2958	0.3479
4	−0.0534	0.0583	0.2503	−0.1441	0.2169	0.2082	0.1480	−0.0320	0.1549	0.0805	0.2110	−0.1252	0.0896
5	0.2318	−0.1343	−0.1630	−0.5129	−0.0944	−0.0961	−0.0678	−0.2221	−0.0599	−0.1597	−0.2924	0.0859	−0.2239
6	−0.2223	0.1945	0.1129	0.1716	0.0117	−0.3565	0.0479	0.1163	−0.2206	0.0820	0.2934	0.1816	0.3066
7	0.2658	−0.1752	−0.1682	−0.2625	−0.1108	0.0301	−0.0254	−0.0435	−0.1064	−0.2416	−0.2587	0.1396	−0.3562
8	−0.1773	−0.4899	0.0542	−0.3187	−0.1215	−0.0541	−0.2146	0.0765	0.0986	−0.3924	−0.3739	−0.1129	0.1443
9	0.1301	0.1675	0.1678	−0.2262	0.1950	0.0174	0.1231	−0.1207	0.0029	0.0045	0.0675	−0.0236	−0.0357
10	0.1599	−0.1322	−0.2595	0.0530	−0.2219	−0.1801	−0.1060	0.1215	−0.2037	−0.2006	−0.3029	0.1526	−0.1280
11	0.3205	0.3047	0.1929	−0.1332	0.2654	0.1074	0.2977	0.0229	−0.0459	0.1688	0.1724	0.0892	0.0551
12	−0.3218	−0.3096	0.5135	−0.2815	0.2055	0.0453	0.0173	0.2976	0.2344	−0.2765	−0.1598	−0.3746	0.6405

Westerhuis, J. A., T. Kourti, and J. F. MacGregor, "Analysis of Multiblock and Hierarchical PCA and PLS Models," *J. Chemometrics*, **12**, 301 (1998).

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

Wold, S., S. Hellberg, T. Lundstedt, and M. Sjöström, "PLS Modeling with Latent Variables in Two or More Dimensions," *Proc. Frankfurt PLS-Meeting*, Frankfurt, Germany (1987b).

Wold, S., N. Kettaneh, and K. Tjessem, "Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection," *J. Chemometrics*, **10**, 463 (1996).

Appendix: Data Generation for the Example

A (12×9) orthonormal matrix $Z=[z_1\cdots z_9]$ was used to generate data on 12 grades for plant B as follows. A linear combination of $[z_1\cdots z_5]$ was used to create a (12×5) Y_B-matrix. Linear combinations of $[z_1\ z_2\ z_3\ z_6\cdots z_9]$ were used to create the (12×8) X_B-matrix. In both cases z_1, z_2, z_3 had larger weights in the linear combination than the other z_i. This ensures that the three dominant latent dimensions in X_B and Y_B overlap. The remaining dimensions are supposed to account for noise, and also for directions in X_B that have no effect on the quality space Y_B.

The first two grades of Y_B are common to both plants. Additional plant A grades had to be generated such that at least some of them fall within the Y_B space. This space is spanned by the first three principal components of a PCA-model on Y_B. Linear combinations of these three components result in grades that lie within the Y_B space. A small amount of independent variation was added to some of these grades. Including the common grades, nine grades for plant A were generated. The data are presented in Tables A1 and A2.

Table A2. Data Used in the Simulation Example: $Y_{A_{\text{initial}}}$

Grades	$Y_{A_{\text{initial}}}$				
1	−0.2366	−0.1005	0.0118	0.1485	0.3231
2	−0.3758	0.0658	0.1475	0.2947	0.0913
3	−1.2877	−0.2473	0.3312	0.6936	2.1893
4	0.4053	−0.4166	−0.4367	−0.4487	0.2630
5	−0.5240	−0.0612	−0.3053	0.5927	−0.8251
6	−0.0599	−0.4038	−0.6662	0.1022	−0.5997
7	−0.2983	−0.4194	−0.4294	0.1766	0.2428
8	0.8859	−0.1694	−0.0900	−0.8557	0.4352
9	−0.6848	−0.4659	−0.1165	0.2616	1.5587

Manuscript received Jan. 11, 2000, and revision received Apr. 20, 2000.