Application of Dimensionality Reduction Methods to a Set of Copolymer Reactor Data

B. Abraham, G.M. Merola,
A. Pendilis and T. Duever
University of Waterloo

RR-99-05

July 1999

APPLICATION OF DIMENSIONALITY REDUCTION METHODS TO A SET OF COPOLYMER REACTOR DATA

B. Abraham, G.M. Merola, A. Penlidis and T. Duever
University of Waterloo

Abstract

In this paper we briefly review dimensionality reduction methods such as principal components analysis, canonical correlation, partial least squares and reduced rank regression. We also consider some recent techniques such as maximum overall redundancy (MOR), and weighted MOR. All these methods are applied to a set of copolymer reactor data.
Keywords: Canonical correlation regression, maximum overall redundancy, partial least squares, principal component regression, reduced rank regression, polymerization.

1. Introduction

Many units in the chemical processing industries processes such as chemical reactors are equipped with sensors connected to computers that can provide hundreds of measurements taken on many process variables ($\boldsymbol{x}$) every few seconds and on output characteristics ($\boldsymbol{y}$) sometimes less frequently. The variables are often highly correlated as well. The availability of the $\boldsymbol{x}$ measurements can be used to monitor the process itself and also as a diagnostic tool for detecting causes of out-of-control values of the $\boldsymbol{y}$-variables. The structure of such data calls for an approach which looks for a lower dimensional subspace in which the process can be monitored and from which $\boldsymbol{y}$ can be predicted. We refer to this as dimensionality reduction approach. In this approach the $\boldsymbol{y}$ variables are predicted by latent variables in $\boldsymbol{x}$ (linear combinations of the $\boldsymbol{x}$-variables).

In section 2, we briefly outline the usual dimension reduction methods (DRMs). Section 3 gives an application of these DRMs to data from a copolymer reactor. Section 4 gives some concluding remarks.

2. Dimensionality Reduction Methods (DRMs)

Let $\boldsymbol{X}$ be an $(n \times p)$ matrix containing n independent measurements on p process (explanatory) variables and $\boldsymbol{Y}$ be an $(n \times q)$ matrix of n independent observations on q response (quality) variables. We assume that the columns of these matrices are mean centred and that the data follow the multivariate linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E} \tag{2.1}$$

where B is the $(p \times q)$ matrix of regression coefficients and $\boldsymbol{E}$ is a matrix of errors.

DRMs typically predict the responses ($\boldsymbol{Y}$) from a subspace of the set of regressors ($\boldsymbol{X}$). Thus linear combinations of the x-variables (latent variables) are used to predict $\boldsymbol{Y}$. There are several methods available and we now briefly consider some of them. The DRMs that we consider determine a set of d ($\leq p$) orthogonal latent variables, $\boldsymbol{T}$, which form a subspace of $L(\boldsymbol{X})$, the space spanned by the columns of $\boldsymbol{X}$.

2.1 Principal Components Regression (PCR)

Let $\boldsymbol{T} = \boldsymbol{X}\boldsymbol{A}$ where $\boldsymbol{A} = (\boldsymbol{a}_1, ..., \boldsymbol{a}_d)$ is a $(p \times d)$ matrix of coefficients. Consider the orthogonal projection of X onto the subspace spanned by the columns of T. This is obtained by minimizing

$$\| \boldsymbol{X} - \boldsymbol{T}\boldsymbol{P} \|^2 = \| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X} \|^2 \tag{2.2}$$

where $\boldsymbol{P}$ is the projection of $\boldsymbol{X}$ onto the subspace. Subject to the condition that the d latent variables are orthogonal, ie. $\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}$, the expression (2.2) is minimized w.r.t. the matrix $\boldsymbol{A}$ by the eigen-vectors of $\boldsymbol{X}'\boldsymbol{X}$ corresponding to the d largest eigen-values (see Rao (1964)). These latent variables are referred to as the principal components (PC) of $\boldsymbol{X}$.

PCR uses an appropriate number of these PC's to predict $\boldsymbol{y}$. It should be noted that the PC's are not obtained using an optimality criterion for the prediction of $\boldsymbol{y}$ but for the reconstruction of $\boldsymbol{X}$ (prediction of $\boldsymbol{x}$).

2.2 Canonical Correlation Regression (CCR)

In this procedure pairs of latent variables $(\boldsymbol{r}_j, \boldsymbol{t}_j) = (\boldsymbol{Y}\boldsymbol{d}_j, \boldsymbol{X}\boldsymbol{a}_j)$ are obtained by maximizing the objective function

$$(\boldsymbol{a}_j'\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{d}_j)^2/[\boldsymbol{a}_j'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_j\boldsymbol{d}_j'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{d}_j] \tag{2.3}$$

with respect to $\boldsymbol{d}_j$ and $\boldsymbol{a}_j$ such that $\boldsymbol{a}_j'\boldsymbol{a}_j = 1 = \boldsymbol{d}_j'\boldsymbol{d}_j$ and $\boldsymbol{a}_j'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_i = 0 \quad i < j$ (see Hotelling (1936), Magnus and Neudecker (1988)). For predicting $\boldsymbol{y}$ the first d latent variables $\boldsymbol{T} = \boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_d)$ are used as regressors. From the objective function it is clear that subsets of latent variables chosen this way are not optimal for the prediction of the $\boldsymbol{y}$ variables.

2.3 Reduced Rank Regression (RRR)

The linear model in the latent variables $\boldsymbol{T}$ can be written as

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{Q} + \boldsymbol{E}^* \tag{2.4}$$

where $\boldsymbol{Q}$ is a $(d \times q)$ matrix of regression coefficients. Taking $\boldsymbol{T} = \boldsymbol{X}\boldsymbol{A}$ as before

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{Q} + \boldsymbol{E}^* = \boldsymbol{X}\boldsymbol{M} + \boldsymbol{E}^*. \tag{2.5}$$

The linear relationship between the explanatory variables $\boldsymbol{X}$ and the responses $\boldsymbol{Y}$ is expressed by the $(p \times d)$ matrix $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{Q}$ of rank d. Thus, the use of DRMs in prediction can be regarded as a regression with a rank deficient matrix of coefficients. Model (2.5) is known as the reduced rank regression (RRR) model.

The residual sum of squares (RSS) for this model can be written as

$$\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{Y} \|^2 . \tag{2.6}$$

This is to be minimized with respect to $\boldsymbol{A}$ subject to the condition $\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}$. It is known (see Rao (1964), Izenman (1975)) that the optimal latent vectors for this RRR model are given by the eigen-vectors associated with the d largest eigen-values obtained from

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\boldsymbol{Y}'\boldsymbol{T} = \boldsymbol{T}\boldsymbol{\Phi} \tag{2.7}$$

where $\boldsymbol{\Phi}$ is the diagonal matrix of eigen-values. Suppose now that $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{B}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ is the ordinary least square estimate of $\boldsymbol{Y}$, then $\hat{\boldsymbol{Y}}'\boldsymbol{T} = \boldsymbol{Y}'\boldsymbol{T}$ and hence (2.7) becomes

$$\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'\boldsymbol{T} = \boldsymbol{T}\boldsymbol{\Phi}. \tag{2.8}$$

This means that the latent variables obtained are the principal components of the orthogonal projection of $\boldsymbol{Y}$ onto $\boldsymbol{X}$. It should be noted that these latent vectors are obtained from an optimality criteria (minimization of RSS) for the prediction of $\boldsymbol{y}$.

2.4 Partial Least Squares (PLS)

Wold (1982) proposed an algorithm (which is known as PLS) to compute pairs of latent variables by maximizing their covariance. Several versions of this algorithm have been proposed, see, for example, Gelaldi and Kowalski (1986), Hoskuldsson (1988), Helland (1988), Nomikos and MacGregor (1993), de Jong (1993) and Schmidli (1995). Merola (1998) gave an improved algorithm which is given below.

Let us denote by $\boldsymbol{T}_{(k-1)}$ the latent variables in the $\boldsymbol{X}$-space after $(k-1)$ pairs of latent variables have been determined. Then

$$\boldsymbol{X}^{(k)} = \boldsymbol{X} - \boldsymbol{T}_{(k-1)}(\boldsymbol{T}'_{(k-1)}\boldsymbol{T}_{(k-1)})^{-1}\boldsymbol{T}'_{(k-1)}\boldsymbol{X}$$

represent the residuals from the $\boldsymbol{X}$ matrix at this stage and is refered to as the deflated $\boldsymbol{X}$-matrix. Then the coefficients of the k-th pair of latent variables are obtained by maximizing

$$\boldsymbol{d}'_k\boldsymbol{Y}'\boldsymbol{X}^{(k)}\boldsymbol{a}_k \text{ such that } \boldsymbol{a}'_k\boldsymbol{a}_k = 1 = \boldsymbol{d}'_k\boldsymbol{d}_k. \tag{2.9}$$

The k-th pair of latent variables is given by

$$\boldsymbol{r}_k = \boldsymbol{Y}\boldsymbol{d}_k \text{ and } \boldsymbol{t}_k = \boldsymbol{X}^{(k)}\boldsymbol{a}_k. \tag{2.10}$$

The algorithm then becomes:

0) Initialize $\boldsymbol{F}_0 = \boldsymbol{X}$.

1) svd$(\boldsymbol{F}'_{j-1}\boldsymbol{Y}) = \boldsymbol{J}\boldsymbol{\Delta}\boldsymbol{\Gamma}'$; $Svd(\boldsymbol{X})$ = singular values of the matrix $\boldsymbol{X}$.

2) $\boldsymbol{c}_j = \boldsymbol{j}_1$ (first column of $\boldsymbol{J}$)

3) $\boldsymbol{t}_j = \boldsymbol{F}_{j-1}\boldsymbol{c}_j$

4) $\boldsymbol{c}_j \leftarrow [\boldsymbol{c}_j/\sqrt{\boldsymbol{t}'_j\boldsymbol{t}_j}]\sqrt{n}$;

5) $\boldsymbol{t}_j \leftarrow [t_j/\sqrt{\boldsymbol{t}'_j\boldsymbol{t}_j}]\sqrt{n}$

6) $\boldsymbol{F}_j = (\boldsymbol{I}_n - \frac{\boldsymbol{t}_j\boldsymbol{t}'_j}{n})\boldsymbol{F}_{j-1}$

7) If ($\| \boldsymbol{F}_j \|^2 < \epsilon$) go to 8
else go to 1

8) qr$(\boldsymbol{N}) = \boldsymbol{Q}\boldsymbol{R}$
$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R}^{-1}$

After the coefficient $\boldsymbol{A}$ is computed, $\hat{\boldsymbol{M}}_k$ is computed as

$$\hat{\boldsymbol{M}}_k = \boldsymbol{a}_k\boldsymbol{t}'_k\boldsymbol{Y} + \hat{\boldsymbol{M}}_{k-1}, \quad k = 1, ..., d, \ \hat{\boldsymbol{M}}_0 = \boldsymbol{0}$$

2.5 Maximum Overall Redundancy (MOR)

In the context of multivariate process control it is important to predict $\boldsymbol{y}$ as well as to reconstruct the $\boldsymbol{X}$ matrix from the latent subspace. However, there is a trade off between these two objectives and PLS is a compromise between these without showing any particular optimality. Now we like to obtain a set of latent variables which meet both of these objectives. Earlier we saw that

(i) the first d principal components of $\boldsymbol{X}$ are the d latent variables in $L(\boldsymbol{X})$ giving the best reconstruction of $\boldsymbol{X}$.

(ii) the first d RRR-latent variables (PC of $\hat{\boldsymbol{Y}}$) are optimal for predicting $\boldsymbol{y}$.

Merola (1998) has shown that, if the two objectives are equally important, then the corresponding latent variables are given by the eigen-solution to

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'(\boldsymbol{Y}\boldsymbol{Y}^{T}+\boldsymbol{X}\boldsymbol{X}')\boldsymbol{T}=\boldsymbol{T}\boldsymbol{\Lambda} \tag{2.11}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigen-values and $(\boldsymbol{X}'\boldsymbol{X})^{-}$ is a generalised inverse of $\boldsymbol{X}'\boldsymbol{X}$. Now equation (2.11) can be written as

$$(\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'+\boldsymbol{X}\boldsymbol{X}')\boldsymbol{T}=\boldsymbol{T}\boldsymbol{\Lambda} \tag{2.12}$$

where $\hat{\boldsymbol{Y}}$ is the projection of $\boldsymbol{Y}$ onto $L(\boldsymbol{X})$. Thus the resulting latent variables are the eigen-vectors corresponding to the d largest eigen-values of $\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'+\boldsymbol{X}\boldsymbol{X}'$, which is the sum of the matrices generating the latent variables in RRR and PCA respectively. We refer to this procedure as the Maximum Overall Redundancy (MOR).

Weighted MOR

We can generalize the MOR procedure to obtain a set of latent variables which are the eigen-vectors of

$$(1-\boldsymbol{\alpha})\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'+\boldsymbol{\alpha}\boldsymbol{X}\boldsymbol{X}' \tag{2.13}$$

which is a convex combination of the matrices in RRR and PCA. This will be referred to as the Weighted MOR (WMOR) method. It is easy to see that $\alpha = 0, 0.5, 1$ correspond to RRR, MOR and PCR solutions, respectively. As α becomes smaller the prediction of $\boldsymbol{y}$ gets more weight and vice versa. Now we consider some special weighting schemes. Let $\ell(\boldsymbol{y}) = \boldsymbol{tr}[(\boldsymbol{Y}'\boldsymbol{Y})^2]$, $\ell_1(\boldsymbol{y}) = \boldsymbol{tr}(\boldsymbol{Y}'\boldsymbol{Y})$, $\ell(\hat{\boldsymbol{y}}) = \boldsymbol{tr}[(\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}})^2]$, $\ell_1(\hat{\boldsymbol{y}}) = \boldsymbol{tr}(\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}})$, $\ell(\boldsymbol{x}) = \boldsymbol{tr}[(\boldsymbol{X}'\boldsymbol{X})^2]$, and $\ell_1(\boldsymbol{x}) = \boldsymbol{tr}(\boldsymbol{X}'\boldsymbol{X})$.

Based on these we define

(i) $\alpha_1 = \sqrt{\ell(\boldsymbol{y})}/(\sqrt{\ell(\boldsymbol{y})}+\sqrt{\ell(\boldsymbol{x})})$, (ii) $\alpha_2 = \ell_1(\boldsymbol{y})/(\ell_1(\boldsymbol{y})+\ell_1(\boldsymbol{x}))$,
(iii) $\alpha_3 = \sqrt{\ell(\hat{\boldsymbol{y}})}/(\sqrt{\ell(\boldsymbol{x})}+\sqrt{\ell(\hat{\boldsymbol{y}})})$, and (iv) $\alpha_4 = \ell_1(\hat{\boldsymbol{y}})/(\ell_1(\hat{\boldsymbol{y}})+\ell_1(\boldsymbol{x}))$ (2.16)

As can be seen, the weights are based on the norms of $\boldsymbol{Y}$, $\hat{\boldsymbol{Y}}$ and $\boldsymbol{X}$. The procedure

corresponding to α_i may be referred to as $WMOR_i$ $(i = 1, 2, 3, 4)$. In the next section we apply the different DRMs to a set of data. For comparing different methods we consider a measure which is called a Redundancy Index (RI) and is defined as

$$RI = tr(\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}})/tr(\boldsymbol{Y}'\boldsymbol{Y}). \tag{2.14}$$

The numerator is the sum of the 'regression sums of squares' for all the $\boldsymbol{y}$ variables and the denominator is the sum of the corresponding 'total sums of squares'. Thus it can be interpreted as an overall measure similar to R^2. When all the variables are standardized to unit norm, RI simplifies to be the average of all the individual R^2 for the $\boldsymbol{y}$-variables.

3. Analysis of Copolymer Reactor Data

We have some data available from a copolymer reactor model which simulates a chemical reaction with 5 input variables. These inputs are:

x_1: MMA, flow rate of first monomer (methyl methacrylate)

x_2: STY, flow rate of second monomer (styrene)

x_3: INI, initiator flow rate x_4: TOL, solvent flow rate

x_5: TEMP, temperature (degrees Kelvin). Flow rates are in mol/min.

Such a chemical reaction requires a certain amount of time to stabilize from the time it is started. The simulator gives measures on 9 different responses. We consider 5 of these

y_1: CPC, copolymer composition y_2: CR, radical concentration

y_3: RMW, accumulated molecular weight y_4: RP, polymerization rate

y_5: Z, weight conversion

The process was activated following the prescription given in Table 3.1

Table 3.1: Specification of the simulated reaction.

Name	Value	Tolerance
MMA	0.08725	±0.008725
STY	0.08170	±0.00817
INI	0.02	±0.002
TOL	0.1758	±0.0175
TEMP	333.0	±0.1

These specifications represent the process under "normal" operating conditions. The simulator reads the values of the 5 input variables and gives the corresponding readings of responses.

3.1 Data and Initial Analysis

The process was simulated by letting the recipe specified in Table 3.1 run for 200 minutes to reach steady state. After that pseudo-random noises were added to the 5 input variables every 8 minutes and the outputs were measured correspondingly. The noises were generated as multinormal, each with mean zero and standard deviation equal to $\frac{1}{3}$ of the tolerance interval shown in Table 3.1. We ran the process to obtain 150 observations of the process under random fluctuations of the inputs, with the same error structure.

Prior to the analysis, the variables have been mean centered, that is the mean has been subtracted from each column.

Table 3.2 gives the eigen-values of the covariance matrix of the $\boldsymbol{x}$ variables (scaled and unscaled) and the cumulative proportion of variance explained.

Table 3.2 Eigen-values and cumulative variance explained for the $\boldsymbol{X}$ matrix.

Eig-val $\boldsymbol{X}$	Cum. var.	Eig-val $\boldsymbol{X}$ scaled	Cum. var.
15.91	0.99	1.450	0.29
0.0048	0.99	1.230	0.53
0.0014	0.99	0.988	0.73
0.0008	1	0.767	0.88
0.0000	1	0.564	1

The effect of standardization is obvious from the table. The value and separation between eigen-values changes drastically. Based on the eigen-values of the unscaled $\boldsymbol{X}$ one may decide

the rank of $\boldsymbol{X}$ to be less than full. On the other hand, the correlation matrix indicates that the matrix has full rank, as it is the case here. The difference is due to the fact that some of the input variables have small readings due to the units they are measured in.

The Principal Component Analysis of the $\boldsymbol{X}$ matrix can be synthesized by the squared correlations between the principal components and the $\boldsymbol{x}$ variables. Tables 3.3 and 3.4 show these correlations for the unscaled and scaled variables.

Table 3.3: Squared correlations between the unscaled $\boldsymbol{x}$ variables and the principal components of the $\boldsymbol{X}$ matrix.

$Corr^2$	1st PC	2nd PC	3rd PC	4th PC	5th PC
MMA	0.00	0.01	0.74	0.25	0.00
STY	0.01	0.02	0.50	0.48	0.00
INI	0.06	0.00	0.06	0.02	0.85
TOL	0.02	0.98	0.00	0.00	0.00
TEMP	1.00	0.00	0.00	0.00	0.00

Table 3.4: Squared correlations between the scaled $\boldsymbol{x}$ variables and the principal components of the scaled $\boldsymbol{X}$ matrix.

$Corr^2$	1st PC	2nd PC	3rd PC	4th PC	5th PC
MMA	0.26	0.28	0.07	0.36	0.03
STY	0.46	0.14	0.15	0.04	0.20
INI	0.55	0.05	0.05	0.20	0.15
TOL	0.07	0.27	0.50	0.12	0.04
TEMP	0.11	0.49	0.20	0.05	0.14

For the unscaled variables the first PC consists of temperature. This is to be expected since this variable has the largest variance. For the same reason the solvent (TOL) corresponds to the second principal component and the initiator (INI) to the last. The third and fourth principal components are combinations of the two monomers, that have, roughly, the same variance. PCA on the scaled variables gives a completely different set up. The PC's for the scaled variables cannot be identified with any of the original variables.

Canonical Correlation Analysis (CCA)

We consider CCA to describe the linear relationships between the explanatory variables and the responses. CCA is particularly powerful for detecting outliers with respect to lin-

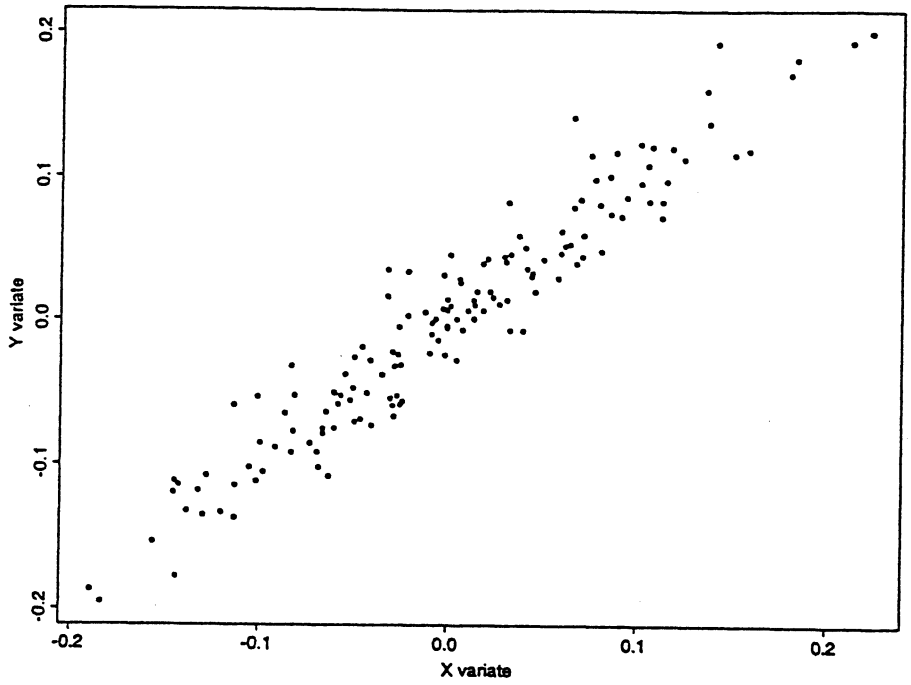

Figure 3.1: First pair of Canonical Correlation variables.

ear dependences between the two spaces. Since CCA is invariant to changes of scale, we perform the analysis on the variables standardized to unit length. The squared Cannonical Correlations (CC) are the following:

$$\rho_1^2 = 0.958,\ \rho_2^2 = 0.526,\ \rho_3^2 = 0.344,\ \rho_4^2 = 0.118,\ \rho_5^2 = 0.017$$

These show that there is only one very highly collinear direction common to the two sets of data. Figures 3.1 and 3.2 show the scatter plots of the first three canonical variates; in the second plot the points are labeled for ease of identification. There does not seem to be outliers or influential points in these directions. The weights (that is, the coefficients of the latent variables standardized to unit length) for the canonical variates in the $\boldsymbol{X}$ space are given in Table 3.5.

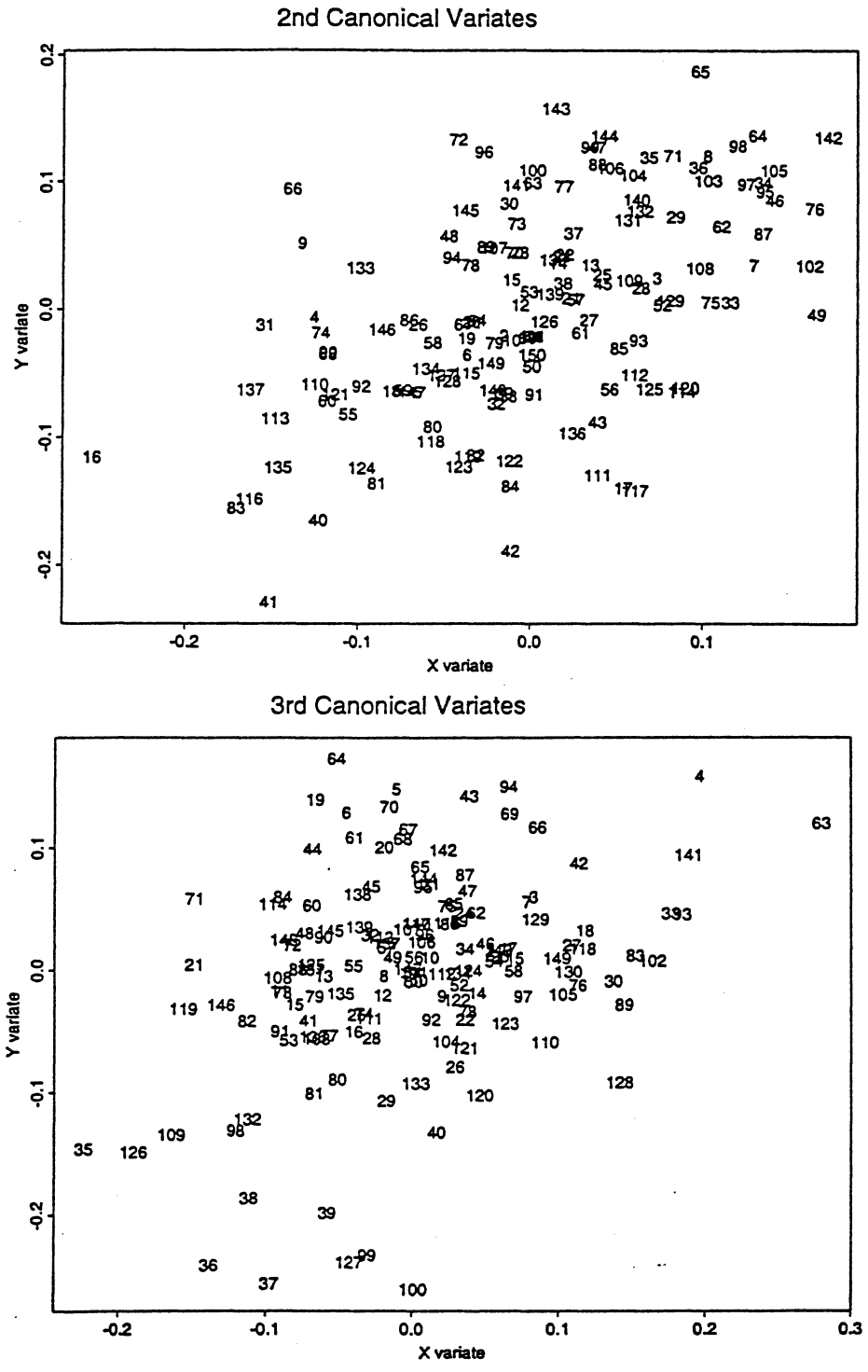

Figure 3.2: Plots of the second and third pairs of Canonical Correlation of variables.

Table 3.5: Weights of the CC variates in the $\boldsymbol{X}$ space.

weights	cc.var1	cc.var2	cc.var3	cc.var4	cc.var5
MMA	0.000	0.868	-0.327	0.493	-0.127
STY	0.049	0.444	0.795	-0.394	-0.058
INI	0.055	-0.062	0.373	0.224	-0.899
TOL	-0.076	-0.213	0.331	0.742	0.359
TEMP	0.994	0.026	0.112	-0.026	-0.208

From Table 3.5 it is evident that the first CC variate in the $\boldsymbol{X}$ space is practically TEMP. The second variate is highly correlated with MMA, uncorrelated with TEMP and somewhat correlated with the other 3 components. Also the correlations between the $\boldsymbol{x}$ variables and the canonical variates, given in Table 3.6, confirm that the first canonical (latent) variable consists of temperature.

Table 3.6: Correlation between the original $\boldsymbol{x}$ variables and the canonical variates in the $\boldsymbol{X}$ space

Corr($\boldsymbol{X}$, CC.$\boldsymbol{X}$)	CC.var1	CC.var2	CC.var3	CC.var4	CC.var5
MMA	0.00	0.91	-0.24	0.34	-0.07
STY	-0.07	0.58	0.73	-0.29	0.20
INI	-0.20	-0.25	0.18	0.25	-0.90
TOL	0.06	-0.22	0.49	0.73	0.43
TEMP	1.00	-0.03	-0.01	0.06	0.07

The correlations between responses and CC variates in the $\boldsymbol{X}$ space, given in Table 3.7, help understand which responses are best explained by the CC latent variates.

Table 3.7: Correlation between $\boldsymbol{y}$ variables and the canonical variates in the $\boldsymbol{X}$ space

Corr($\boldsymbol{Y}$, CC.$\boldsymbol{X}$)	CC.var1	CC.var2	CC.var3	CC.var4	CC.var5
CPC	-0.05	-0.17	0.00	-0.06	-0.01
CR	0.93	-0.10	0.04	0.00	0.00
RMW	-0.14	0.13	-0.04	0.08	-0.01
RP	0.94	0.01	-0.05	0.01	0.00
$\boldsymbol{Z}$	0.18	-0.43	-0.08	-0.05	0.01

The first canonical variate in the $\boldsymbol{X}$ space is highly correlated with the responses CR and RP (which are highly correlated between themselves), while the second is only mildly

correlated with the weight conversion $\boldsymbol{Z}$. The last three components are almost uncorrelated with all $\boldsymbol{y}$ variables. This confirms that the linear relationship between the $\boldsymbol{x}$ and $\boldsymbol{y}$ variables is one or at most two dimensional, as the CC coefficients indicated. We could also deduce that the responses CR and RP can be very well predicted by the first CC (that is, by temperature), but y_1, y_3 and y_5 do not seem to have a strong linear relation with any of the $\boldsymbol{x}$ variables.

3.2 Dimensionality Reduction Methods

We now consider the DRMs PLS, MOR, RRR, CCR and PCR for prediction. The methods have been applied both on the raw data and on the data standardized to unit length. We report only the analysis performed on the standardized data for which the different methods are better compared. In addition, we look at WMOR with different weights. Since the variables are scaled to unit length, $\alpha_2 = q/(p+q) = 5/(5+5) = 0.5$. Hence WMOR_2 gives equal weights to the prediction of $\boldsymbol{y}$ and the reconstruction of $\boldsymbol{X}$. Hence WMOR_2 is the same as MOR in this context. It turns out that α_1 and α_3 are very close to 0.5 and $\alpha_4 \approx 0.7$. Hence we only consider WMOR_4 in further discussions.

In order to compare the latent components we examine the correlations between the $\boldsymbol{x}$ variables and the first two latent components obtained with different methods. These are given in Tables 3.8-3.9.

Table 3.8: Correlations between the $\boldsymbol{x}$ variables and the first latent variables for different DRMs

COR	PLS	MOR	WMOR4	RRR	CCR	PCR
MMA	0.11	-0.10	-0.10	-0.09	0.00	0.51
STY	0.19	-0.15	-0.14	-0.14	-0.07	0.68
INI	0.35	-0.28	-0.33	-0.18	-0.20	-0.74
TOL	-0.19	0.14	0.18	0.07	0.06	0.26
TEMP	-0.98	0.99	0.99	0.99	1.00	0.33

Table 3.9: Correlations between the $\boldsymbol{x}$ variables and the second latent variables for different DRMs

COR	PLS	MOR	WMOR4	RRR	CCR	PCR
MMA	0.86	0.73	-0.70	-0.92	0.91	0.53
STY	0.66	0.75	-0.76	-0.57	0.58	0.38
INI	-0.38	-0.56	0.58	0.19	-0.25	0.22
TOL	-0.03	0.07	-0.09	0.03	-0.22	-0.52
TEMP	0.11	0.07	-0.04	-0.08	-0.03	-0.70

All the first latent variables but the first principal component (Table 3.8) have similar correlation pattern with the $\boldsymbol{x}$ variables, and are dominated by TEMP. With the exception of the second principal component, all other second latent variables are uncorrelated with temperature. Only the second principal component and the second CC variate have correlation greater than 0.1 (in absolute value) with the solvent, TOL (Table 3.9). All second latent variates, but the second principal component, have high correlation with the two monomers. Note that MOR and WMOR_4 latent variables are similar to those of the other predictive methods, thus different from principal components.

Tables 3.10-3.15 give R^2 and the Redundancy Indices (RI) for the $\boldsymbol{y}$ variables employing up to five latent components for the different methods.

Table 3.10: R^2 and Redundancy Index for PLS

PLS	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.001	0.813	0.023	0.819	0.061	0.453
2 comps	0.034	0.814	0.034	0.833	0.221	0.489
3 comps	0.034	0.861	0.034	0.892	0.221	0.507
4 comps	0.036	0.872	0.043	0.892	0.221	0.511
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

Table 3.11: R^2 and Redundancy Index for MOR

MOR	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.001	0.855	0.024	0.862	0.056	0.466
2 comps	0.028	0.855	0.032	0.865	0.199	0.497
3 comps	0.031	0.857	0.034	0.885	0.207	0.502
4 comps	0.037	0.861	0.039	0.889	0.222	0.508
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

Table 3.12: R^2 and Redundancy Index for WMOR4.

WMOR4	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.001	0.861	0.024	0.867	0.056	0.468
2 comps	0.029	0.861	0.033	0.873	0.202	0.499
3 comps	0.031	0.861	0.034	0.888	0.209	0.504
4 comps	0.036	0.864	0.039	0.891	0.222	0.509
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

Table 3.13: R^2 and Redundancy Index for RRR

RRR	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.001	0.872	0.024	0.879	0.052	0.471
2 comps	0.036	0.872	0.039	0.890	0.221	0.510
3 comps	0.037	0.873	0.044	0.891	0.223	0.511
4 comps	0.037	0.874	0.045	0.892	0.223	0.512
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

Table 3.14: R^2 and Redundancy Index for CCR

CCR	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.003	0.864	0.021	0.890	0.034	0.471
2 comps	0.033	0.873	0.037	0.890	0.215	0.508
3 comps	0.033	0.874	0.038	0.892	0.220	0.509
4 comps	0.037	0.874	0.045	0.892	0.223	0.510
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

Table 3.15: R^2 and Redundancy Index for PCR

PCR R^2	CPC	CR	RMW	RP	$\boldsymbol{Z}$	RI
1 comp	0.020	0.055	0.001	0.081	0.060	0.203
2 comps	0.023	0.497	0.026	0.447	0.201	0.366
3 comps	0.025	0.660	0.028	0.692	0.206	0.435
4 comps	0.037	0.692	0.034	0.742	0.222	0.455
5 comps	0.037	0.874	0.045	0.892	0.223	0.512

For all DRMs but PCR the first latent component explains most of the linear relationships between the $\boldsymbol{x}$ and the $\boldsymbol{y}$ variables. Also, contributions of the second components are similar but their addition in the predictive model gives only modest increases in the overall RI. The different behaviour of PCR is easily explained by the different nature of the PCA

decomposition, which does not depend on the $\boldsymbol{y}$ variables. The performance of PCR in this example is quite poor, compared with the other methods. The increase in RI follows a similar pattern for all methods but PLS. The value of RI for PLS prediction of the $\boldsymbol{y}$ variables is slightly lower than the others for the first two components, but after the third component is added to the model the RI becomes almost the same as that of the others. For the $\boldsymbol{x}$ variables we note that the RI's of PCR are highest for all number of components, while those of RRR are the lowest for all components.

3.3 Alternative Analysis

Ordinary Regression

We perform regression of each $\boldsymbol{y}$ variable on the $\boldsymbol{x}$-variables standardized to unit length. The regression coefficients and R^2 indices are given in Table 3.16.

Table 3.16: Regression coefficients and R^2 for the $\boldsymbol{x}$ variables standardized to unit length

Coeff	CPC	CR	RMW	RP	$\boldsymbol{Z}$
MMA	-0.171	-0.089	0.160	0.026	-0.333
STY	-0.040	0.041	-0.027	0.004	-0.207
INI	0.005	0.071	0.000	0.036	-0.015
TOL	-0.020	-0.041	0.036	-0.088	0.000
TEMP	-0.053	0.953	-0.148	0.960	0.168
R^2	0.037	0.874	0.044	0892	0.222
R^2 (Temp)	0.026	0.861	0.020	0.882	0.038

The regression coefficients and the R^2 indices confirm that only CR and RP are well predicted by the inputs and that temperature alone is significantly linearly correlated with these outputs. This is also confirmed by a regression of each y variable on temperature alone. Corresponding R^2 (Temp) indices are also shown in Table 3.16.

The results presented so far provide overwhelming evidence to the fact that the only significant regressions are those of CR and RP on temperature. It should also be clear that a linear multivariate approach for this kind of data is redundant. It is evident that variables CPC, RMW and Z are not predicted well at all.

Time Series Analysis

The poor performance of the linear models for prediction may be explained by the serial correlation in the responses. In analyzing the time structure of the $\boldsymbol{y}$ variables it is evident that CPC, RMW and Z have strong correlation with their past values. The analysis of the sample autocorrelation and partial autocorrelation functions of each $\boldsymbol{y}$ variable individually leads to the following Autoregressive Integrated Moving Average (ARIMA) models (Box and Jenkins (1976)).

- CPC: ARIMA(2,1,0)
  Non-stationary model: $U_{1t} - 1.04U_{1t-1} + 0.47U_{1t-2} = a_{1t}$
  where $U_{1t} = y_{1t} - y_{1t-1}$ and a_{1t} is white noise
- CR, White noise
- RMW, ARIMA(1,0,1): $y_{3t} - 0.85y_{3t-1} = a_{3t} + 0.54a_{3t-1}$
- RP, White noise
- Z: ARIMA(1,0,1), $y_{5t} - 0.77y_{5t-1} = a_{5t} + 0.78a_{5t-1}$

This indicates that for prediction of CPC, RMW and Z we can use their past values while CR and RP can be predicted with temperature. It should also be noted that the correlations as well as regressions of the filtered $\boldsymbol{y}$ variables (residuals from the ARIMA fit) with the $\boldsymbol{x}$ variables remain almost the same as those of the non-filtered $\boldsymbol{y}$'s. Also, the canonical correlations between the filtered $\boldsymbol{y}$ and $\boldsymbol{x}$ are almost the same as those from the non-filtered $\boldsymbol{y}$:

$$\rho_1^2 = 0.958,\ \rho_2^2 = 0.410,\ \rho_3^2 = 0.326,\ \rho_4^2 = 0.105,\ \rho_5^2 = 0.012$$

However, predictions of CPC, RMW and Z can be substantially improved by the use of their past values. This is indicated by the R^2 (Filter) obtained by regressing each of CPC, RMW and Z on their past values. (see Table 3.17). We also show the R^2 values from Table 3.16.

Table 3.17 R^2 (Filter) and R^2 (OLS)

	CPC	RMW	Z
R^2 (Filter)	0.952	0.878	0.852
R^2 (OLS)	0.037	0.044	0.222

It is evident that R^2 (Filter) is substantially larger than R^2 (OLS) for these variables. In summary

- Responses CR and RP are strongly correlated with temperature and not at all correlated with the other input variables. The R^2 coefficients shown in Table 3.16 indicate that almost 90% of the variability in the two variables is explained by temperature alone. Addition of the other variables to the linear model improves the R^2 only slightly.

- The other responses, CPC, RMW and Z, show a strong time dependence with their previous values and are not predicted well by the $\boldsymbol{x}$ variables. Instead, these variables are well explained by their past values. After filtering the time correlation in the responses through an ARIMA model, the residuals are still not explained by the input variables. However, these series are well fitted by ARIMA models.

The conclusion we draw from the analysis is that once the copolymer process has reached its steady state, it is very sensitive to changes in temperature and not in changes of the other variables. Of course, this conclusion is only valid within the specified tolerance region shown in Table 3.1.

4. Concluding Remarks

In this paper we briefly reviewed some common DRMs such as PCR, RRR, CCR and PLS. We also discussed the relatively new techniques MOR and WMOR. We note here that there have been attempts to look for a common framework for these procedures (see Burnham et al. (1995)). Merola (1998) and Merola and Abraham (1998) have discussed a common objective function from which the objective functions of the various DRMs can be obtained by changing some parameters.

Acknowledgements

B. Abraham was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors would also like to thank Materials Manufacturing Ontario for partial support. Thanks are also extended to Jun Gao for his help with the copolymer data.

References

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series analysis forecasting and control.* Holden Day, San Francisco.

Burnham, A.J., Viveros, R., and MacGregor, J.F. (1996). Frameworks for latent variable multivariate regression. *J. of Chemometrics*, 10, 31-45.

de Jong, S. (1993). Simpls: an alternative approach to partial least squares regression. *Chemom. and Intell. Lab. Systems*, 18:251-263.

Gelaldi, P. and Kowalski, B.R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185:1-17.

Helland, I.S. (1988). On the structure of partial least squares. *Comm. Stat.-sim*, 17(2):581-607.

Hoskuldsson, P. (1988). PLS regression methods. *J. of Chemometrics*, 2:211-228.

Hotelling, H. (1936). Relation between two sets of variates. *Biometrica*, 28:321-377.

Izenman, A.J. (1975). Reduced-rank regression for the multivariate bilinear model. *J. of Multivariate Analysis*, 5:248-264.

Merola, G.M. (1998). Dimensionality reduction methods in multivariate prediction. Unpublished Ph.D. Thesis. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada.

Merola, G.M. and Abraham, B. (1998). An objective function approach for dimensionality reduction methods in prediction. Research Report, IIQP, University of Waterloo, Waterloo, Canada.

Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley

Nomikos, P. and MacGregor, J.F. (1994). Monitoring of batch processes using multi-way principal component analysis. *A.I.Ch.E. J.*, 40:1361-1375.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A*, 26:329-358.

Schmidli, H. (1995). *Reduced Rank Regression.* Contributions to Statistics. Physica-Verlag.

Wold, H. (1982). Soft modelling, the basic design and some extensions. In Joresorg, K. and Wold, H., editors, *Systems Under Indirect Observation*, volume II, pages 589-591. Wiley and Sons.