Noise Handling Capabilities of Multivariate Calibration Methods

ROLF ERGON*

Keywords: *PCR, PLSR, noise, prediction, spectra*

The noise handling capabilities of principal component regression (PCR) and partial least squares regression (PLSR) are somewhat disputed issues, especially regarding regressor noise. In an attempt to indicate an answer to the question, this article presents results from Monte Carlo simulations assuming a multivariate mixing problem with spectroscopic data. Comparisons with the best linear unbiased estimator (BLUE) based on Kalman filtering theory are included. The simulations indicate that both PCR and PLSR perform comparatively well even at a considerable regressor noise level. The results are also discussed in relation to estimation of pure spectra for the mixing constituents, i.e. to identification of the data generating system. In this respect solutions to well-posed least squares problems serve as references.

1. Introduction

The noise handling capabilities of principal component regression (PCR) and partial least squares regression (PLSR) are somewhat disputed issues, especially regarding regressor noise (**X**-noise). In an attempt to indicate an answer to the question, this article presents results from Monte Carlo simulations assuming a typical multivariate calibration problem, where several constituents with unknown spectroscopic properties are mixed.

The performances of PCR and PLSR are certainly noise dependent, but to which degree? A more specific question is how well these methods handle noise of different levels, as compared with theoretically best possible prediction results, which in the simulations are found by use of the best linear unbiased estimator (BLUE) based on Kalman filtering theory. The results are also discussed in relation to estimation of pure constituent spectra, i.e. to identification of the data generating system. In this respect solutions to well-posed least squares (LS) problems are used for comparisons.

The theoretical background based on latent variables (LV) modeling is summarized in Section 2, with references to more detailed treatments of PCR and PLSR. The simulated mixing problem and the simulation results are presented in Section 3, and conclusions are given in Section 4. Some details concerning PLSR modeling and constituent profile estimation are collected in Appendix A and B.

2. Theoretical background

Model assumptions and problem statement

Assume centered data generated according to the LV model

$$\mathbf{y}_k = \mathbf{C}_1 \mathbf{z}_k + \mathbf{f}_k \tag{1}$$

$$\mathbf{x}_k = \mathbf{C}_2 \mathbf{z}_k + \mathbf{e}_k, \tag{2}$$

where $\mathbf{z}_k \in \mathbb{R}^{A \times 1}$ is a random vector of latent variables, i.e. the expectation $E\mathbf{z}_j\mathbf{z}_k^T = \mathbf{0}$ for all $j \neq k$, and where $\mathbf{y}_k \in \mathbb{R}^{m \times 1}$ is a vector of response variables, while $\mathbf{x}_k \in \mathbb{R}^{p \times 1}$ is a vector of regressor variables. $\mathbf{C}_1 \in \mathbb{R}^{m \times A}$ and $\mathbf{C}_2 \in \mathbb{R}^{p \times A}$ are time-invariant matrices, while $\mathbf{f}_k$ and $\mathbf{e}_k$ are independent and random noise vectors of appropriate dimensions.

Also assume $m \leqslant A$ and independent components of $\mathbf{z}$ and $\mathbf{y}$, i.e. diagonal expectations $E\mathbf{z}_k\mathbf{z}_k^T$ and $E\mathbf{y}_k\mathbf{y}_k^T$. Without loss of generality we may then assume an LV representation such that

$$\mathbf{C}_1 = [\mathbf{I}_m \quad \mathbf{0}], \tag{3}$$

i.e. we assume that each response variable is a latent variable plus some random noise. Collection of data from N observations in matrices $\mathbf{Y} \in \mathbb{R}^{N \times m}$ and $\mathbf{X} \in \mathbb{R}^{N \times p}$ thus gives

$$\mathbf{Y} = \mathbf{Z}\mathbf{C}_1^T + \mathbf{F} = [\mathbf{Z}_Y \quad \mathbf{Z}_{OSC}] \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix} + \mathbf{F} = \mathbf{Z}_Y + \mathbf{F} \tag{4}$$

$$\mathbf{X} = \mathbf{Z}\mathbf{C}_2^T + \mathbf{E} = \mathbf{Z}_Y\mathbf{C}_Y^T + \mathbf{Z}_{OSC}\mathbf{C}_{OSC}^T + \mathbf{E} = \mathbf{Y}\mathbf{C}_Y^T - \mathbf{F}\mathbf{C}_Y^T + \mathbf{Z}_{OSC}\mathbf{C}_{OSC}^T + \mathbf{E}, \tag{5}$$

where it is a part of the assumptions that $A << N < p$. The OSC notation is borrowed from recent articles on orthogonal signal correction (Wold *et al.* (1998), Fearn (2000), Trygg and Wold (2001), Westerhuis *et al.* (2001), Trygg (2001)). The matrix $\mathbf{Z}_{OSC}\mathbf{C}_{OSC}^T$ thus contains the structured but **Y**-orthogonal information in **X**. With the assumptions given the columns of $\mathbf{C}_2$ may typically be scaled versions of pure constituent spectral profiles.

The assumption $A << N < p$ makes it natural to use PCR or PLSR for calibration purposes, and we will in the following focus on the PCR and PLSR noise sensitivity in relation to two problems:

- The multivariate calibration problem of finding an estimator $\hat{\mathbf{B}}$ for prediction of new responses from new regressor observations according to

$$\mathbf{y}_{new}^T = \mathbf{x}_{new}^T\hat{\mathbf{B}} \tag{6}$$

- The problem of estimating the pure constituent profiles in $\mathbf{C}_2$, i.e. the problem of identifying the data generating system.

$\hat{\boldsymbol{B}}$ from ordinary least squares regression

The ordinary LS solution for $\hat{\mathbf{B}}$ obtained from the data is (e.g. Johnson and Wichern, 1998)

$$\hat{\mathbf{B}}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{7}$$

Under the present assumptions with a large number p of **x** variables relative to the number N of observations, the underlying LS problem will be ill-posed. In this case there is a need for regularization, which can be based on LV modeling and PCR or PLSR as summarized below.

$\hat{C}_2$ from ordinary least squares regression

The columns $\mathbf{C}_Y$ of $\mathbf{C}_2$ that are directly related to the responses $\mathbf{Y}$ can be found from (5) using LS regression according to

$$\hat{\mathbf{C}}_Y = \mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}. \tag{8}$$

Under the assumptions given the underlying LS problem is well-posed. We will later discuss this result in relation to the PCR and PLSR methods.

Multivariate calibration model

Multivariate calibration using PCR or PLSR assumes a model

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \tag{9}$$

$$\mathbf{X} = \mathbf{T}\mathbf{L}^T + \mathbf{E}, \tag{10}$$

resulting from (4,5) through an unknown similarity transformation. Here, $\mathbf{T}$ is the matrix of scores, while $\mathbf{L}$ is the matrix of loadings. For the two problems under study we may note the following:

- PCR and PLSR use different factorizations of $\mathbf{Z}\mathbf{C}_1^T$ and $\mathbf{Z}\mathbf{C}_2^T$, as summarized below.
- The pure constituent profiles in $\mathbf{C}_2 = [\mathbf{C}_Y \quad \mathbf{C}_{OSC}]$ may be confounded and scaled in $\mathbf{L}$.

The Helland predictor

The PCR and PLSR regularizations are based on the latent variables model (9, 10) above. The LS solution of (10) is

$$\hat{\mathbf{T}} = \mathbf{X}\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}, \tag{11}$$

and from (9) and (11) we thus find the LS predictor related to the latent variables

$$\hat{\mathbf{Q}}^T = (\hat{\mathbf{T}}^T\hat{\mathbf{T}})^{-1}\hat{\mathbf{T}}^T\mathbf{Y} = ((\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{X}^T\mathbf{X}\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1})^{-1}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{X}^T\mathbf{Y}, \tag{12}$$

which after some simplifications results in fitted experimental responses according to (9)

$$\hat{\mathbf{Y}} = \hat{\mathbf{T}}\hat{\mathbf{Q}}^T = \mathbf{X}\mathbf{L}(\mathbf{L}^T\mathbf{X}^T\mathbf{X}\mathbf{L})^{-1}\mathbf{L}^T\mathbf{X}^T\mathbf{Y}. \tag{13}$$

The regularized LV predictor $\hat{\mathbf{B}}$ to be used in (6) thus becomes

$$\hat{\mathbf{B}}_{LV} = \mathbf{L}(\mathbf{L}^T\mathbf{X}^T\mathbf{X}\mathbf{L})^{-1}\mathbf{L}^T\mathbf{X}^T\mathbf{Y}. \tag{14}$$

This predictor was first presented by Helland (1988), although there not explicitly based on an LV model.

The problem now is to find $\mathbf{L}$, or more realistically good estimates $\hat{\mathbf{L}}$. A simple choice is $\hat{\mathbf{L}} = \mathbf{I}_p$, which brings us back to the LS solution (7). Other choices give the PCR and PLSR solutions.

The PCR predictor

In PCR the loading matrix is $\hat{\mathbf{L}} = \hat{\mathbf{P}}$, where $\hat{\mathbf{P}}$ is found from a principal component analysis (PCA) of $\mathbf{X}$ (e.g. Johnson and Wichern, 1998). We may also find $\hat{\mathbf{T}} = \mathbf{U}_1\mathbf{S}_1$ and $\hat{\mathbf{P}} = \mathbf{V}_1$ from the singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{USV}^T = [\mathbf{U}_1 \quad \mathbf{U}_2]\begin{bmatrix}\mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T \\ \mathbf{V}_2^T\end{bmatrix} \tag{15}$$

$$= \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^T + \mathbf{U}_2\mathbf{S}_2\mathbf{V}_2^T = \hat{\mathbf{T}}\hat{\mathbf{P}}^T + \mathbf{E}.$$

The LV represented by the score matrix $\hat{\mathbf{T}}$ are thus based on $\mathbf{X}$ information only.

The PLSR predictor

Regarding PLSR we will discuss two algorithms:

- The original method (Wold *et al.*, 1982) with an orthogonal score matrix $\hat{\mathbf{T}}_W$.
- The alternative method (e.g. Martens and Næs, 1989) with a non-orthogonal score matrix $\hat{\mathbf{T}}_M$.

Some of the latent variables represented in the PCR score matrix $\hat{\mathbf{T}}$ may often be very weakly correlated with the response variable in $\mathbf{y}$. The PLSR solution to this is to use both $\mathbf{X}$ and $\mathbf{y}$ information in order to find improved versions of $\hat{\mathbf{T}}$ and $\hat{\mathbf{L}}$. In the Wold and Martens algorithms this is done by step-wise computations (e.g. Martens and Næs, 1989), but a one-step procedure is also available (Di Ruscio, 2000). The original orthogonal PLSR algorithm of Wold is based on the factorization

$$\mathbf{X} = \hat{\mathbf{T}}_W\hat{\mathbf{P}}_W^T\hat{\mathbf{W}}\hat{\mathbf{W}}^T + \mathbf{E}, \tag{16}$$

where $\hat{\mathbf{P}}_W$ is a special non-orthogonal loading matrix. Both $\hat{\mathbf{T}}_W$ and the loading weight matrix $\hat{\mathbf{W}}$ are orthogonal, and $\hat{\mathbf{W}}^T\hat{\mathbf{W}} = \mathbf{I}_A$ (see Appendix A for a detailed discussion).

The non-orthogonal PLSR algorithm of Martens is based on the factorization

$$\mathbf{X} = \hat{\mathbf{T}}_M\hat{\mathbf{W}}^T + \mathbf{E}, \tag{17}$$

where $\hat{\mathbf{T}}_M$ is non-orthogonal, while $\hat{\mathbf{W}}$ is the same as in the Wold algorithm. Since $\hat{\mathbf{P}}_W^T\hat{\mathbf{W}}$ is a low dimensional and invertible matrix, application of (14) with $\hat{\mathbf{L}} = \hat{\mathbf{W}}\hat{\mathbf{W}}^T\hat{\mathbf{P}}_W$ and $\hat{\mathbf{L}} = \hat{\mathbf{W}}$ give the same result,

$$\hat{\mathbf{b}}_{PLSR} = \hat{\mathbf{W}}(\hat{\mathbf{W}}^T\mathbf{X}^T\mathbf{X}\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}^T\mathbf{X}^T\mathbf{y}. \tag{18}$$

See e.g. Martens and Næs (1989) for detailed descriptions of the algorithms, and Ergon and Esbensen (2001) for a new didactic version.

The optimal predictor

In order to obtain a basis for comparisons we need an optimal predictor formulation. The optimal predictor may be found by use of general Kalman filtering theory (e.g. Grewal and Andrews, 1993). We will, however, derive the optimal solution directly by introduction of the optimal state estimate related to the LV model (1,2),

$$\hat{\mathbf{z}}_k = \mathbf{K}\mathbf{x}_k, \tag{19}$$

where $\mathbf{K}$ is chosen such that the expectation

$$\begin{aligned}\mathbf{R} &= E(\mathbf{z}_k - \hat{\mathbf{z}}_k)(\mathbf{z}_k - \hat{\mathbf{z}}_k)^T = E[\mathbf{z}_k - \mathbf{K}(\mathbf{C}_2\mathbf{z}_k + \mathbf{e}_k)][\mathbf{z}_k - \mathbf{K}(\mathbf{C}_2\mathbf{z}_k + \mathbf{e}_k)]^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{C}_2)E(\mathbf{z}_k\mathbf{z}_k^T)(\mathbf{I} - \mathbf{K}\mathbf{C}_2)^T + \mathbf{K}E(\mathbf{e}_k\mathbf{e}_k^T)\mathbf{K}^T\end{aligned} \tag{20}$$

is minimized. Using $E\mathbf{z}_k\mathbf{z}_k^T = \mathbf{R}_z$ and $E\mathbf{e}_k\mathbf{e}_k^T = \mathbf{R}_e$ we find (e.g. Gelb, 1974)

$$\frac{\partial}{\partial \mathbf{K}} trace(\mathbf{R}) = -2(\mathbf{I} - \mathbf{K}\mathbf{C}_2)\mathbf{R}_z\mathbf{C}_2^T + 2\mathbf{K}\mathbf{R}_e, \tag{21}$$

i.e. $\partial(\text{trace}(\mathbf{R}))/\partial\mathbf{K} = \mathbf{0}$ gives the optimal solution

$$\mathbf{K} = \mathbf{R}_z\mathbf{C}_2^T(\mathbf{C}_2\mathbf{R}_z\mathbf{C}_2^T + \mathbf{R}_e)^{-1}. \tag{22}$$

This intermediate result, derived from general Kalman filtering theory, was first presented by Berntsen (1988).

The resulting optimal response estimate is

$$\hat{\mathbf{y}}_k = \mathbf{C}_1\mathbf{K}\mathbf{x}_k, \tag{23}$$

i.e. the optimal predictor is

$$\hat{\mathbf{B}}_{KF} = \mathbf{K}^T\mathbf{C}_1^T = (\mathbf{C}_2\mathbf{R}_z\mathbf{C}_2^T + \mathbf{R}_e)^{-1}\mathbf{C}_2\mathbf{R}_z\mathbf{C}_1^T. \tag{24}$$

Optimality here means that (24) gives the best linear unbiased estimate (BLUE), and the best possible estimate whatsoever assuming Gaussian noise distribution (e.g. Grewal and Andrews, 1993). This predictor will be used as a source of reference in the simulations in Section 3.

Pure spectra estimation from PLSR and PCR results

Pure spectra estimates may be found from the well-conditioned LS solution (8), and there is thus no need for use of the PCR and PLSR results for this purpose. It is, however, a central part of the PLSR algorithms that the first loading weight vector with a single response variable y_j is found as (e.g. Martens and Næs, 1989)

$$\hat{\mathbf{w}}_{j1} = \frac{\mathbf{X}^T\mathbf{y}_j}{\sqrt{\mathbf{y}_j^T\mathbf{X}\mathbf{X}^T\mathbf{y}_j}}. \tag{25}$$

From (8) thus follows that the column of $\mathbf{C}_2$ corresponding to $\mathbf{y}_j$ is estimated as

$$\hat{\mathbf{C}}_{2j} = \sqrt{\mathbf{y}_j^T\mathbf{X}\mathbf{X}^T\mathbf{y}_j}(\mathbf{y}_j^T\mathbf{y}_j)^{-1}\hat{\mathbf{w}}_{j1}. \tag{26}$$

With the representation used in (4), i.e. $\mathbf{Y} = \mathbf{Z}_Y + \mathbf{F}$, the first loading weight vector $\hat{\mathbf{w}}_{j1}$ thus gives a scaled LS estimate of $\mathbf{C}_{2j}$. In relation to the noise handling capabilities it is reassuring to know that a single response PLSR (PLS1) under the given assumptions results in a pure spectrum estimate that is identical with the result from a well-posed LS problem (see also simulation results in Section 3). For PCR the situation is more involved (see Appendix B).

Isolation of Y-orthogonal components

From (5) follows

$$\mathbf{X} - \mathbf{Y}\mathbf{C}_Y^T = \mathbf{Z}_{OSC}\mathbf{C}_{OSC}^T - \mathbf{F}\mathbf{C}_Y^T\mathbf{T} + \mathbf{E}. \tag{27}$$

Using $\hat{\mathbf{C}}_{\mathrm{Y}}^{T}$ from (8) we may compute $\mathbf{X}-\mathbf{Y}\hat{\mathbf{C}}_{\mathrm{Y}}^{T}$, and PCA/SVD of the result gives

$$\mathbf{X}-\mathbf{Y}\hat{\mathbf{C}}_{\mathrm{Y}}^{T}=\mathbf{USV}^{T}=[\mathbf{U}_{\mathrm{OSC}}\quad \mathbf{U}_{\mathrm{E}}]\begin{bmatrix}\mathbf{S}_{\mathrm{OSC}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{\mathrm{E}}\end{bmatrix}\begin{bmatrix}\mathbf{V}_{\mathrm{OSC}}^{T}\\ \mathbf{V}_{\mathrm{E}}^{T}\end{bmatrix}=\mathbf{U}_{\mathrm{OSC}}\mathbf{S}_{\mathrm{OSC}}\mathbf{V}_{\mathrm{OSC}}^{T}+\mathbf{E}. \quad (28)$$

Choosing $\hat{\mathbf{Z}}_{\mathrm{OSC}}=\mathbf{U}_{\mathrm{OSC}}$ we will thus find the confounded and scaled profiles of the **Y**-orthogonal interferants in

$$\hat{\mathbf{C}}_{\mathrm{OSC}}=\mathbf{V}_{\mathrm{OSC}}\mathbf{S}_{\mathrm{OSC}}^{T}. \quad (29)$$

Note that the scaled and sign indeterminate profile of a single unknown interferant will be found directly from (29). For correct scaling we would need additional information.

Consequences of data centering and standardization

Centering of the data, i.e. using $\mathbf{X}\leftarrow\mathbf{X}-\bar{\mathbf{X}}$ and $\mathbf{Y}\leftarrow\mathbf{Y}-\bar{\mathbf{Y}}$ where $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ are column mean values, has no effect on the $\hat{\mathbf{C}}_2$ estimate according to (8) and (29). However, standardization of the columns of **X** and **Y** to unit variance does affect $\hat{\mathbf{C}}_2$, and must thus be properly accounted for.

3. Monte Carlo simulation

The practical case behind the following simulation example could be a spectroscopic measurement of a solution with three different chemical constituents. A typical simulation result is shown in Fig. 1. Note the overlapping peaks and considerable **X**-noise.

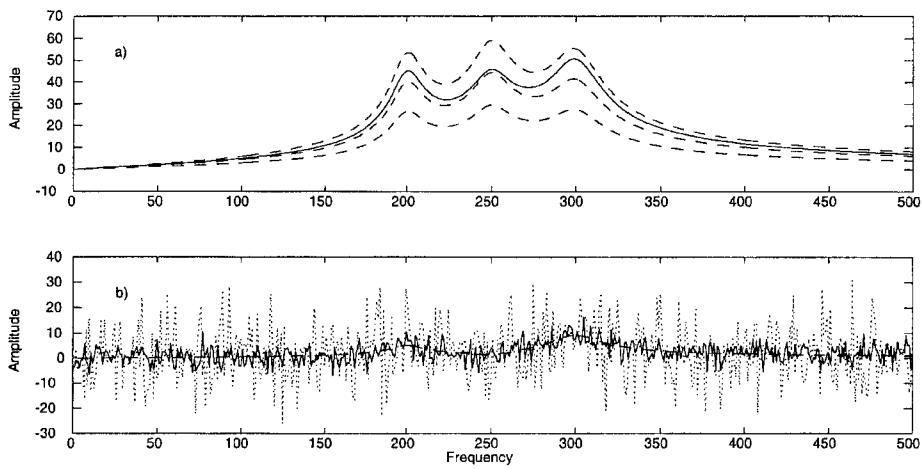

Figure 1. Mean spectrum and standard deviations (Fig. a – dashed lines) plus a typical realization of a noise free original spectrum (Fig. a – solid line), and corresponding centered and noise corrupted spectra (Fig. b) of a mixture of three chemical constituents. The **X**-noise covariances are here $r_{ee}=10$ (Fig. b – solid line) and $r_{ee}=100$ (Fig. b – dotted line) (see relation to signal-noise-ration below). The centred noise free spectrum is shown by dashed line in Fig. b.

The simulations are based on assumed discrete frequency spectra in the range $0 < f \leqslant 500$ frequency units (f.u.),

$$x_k(f) = \frac{f_1 f}{\sqrt{(f_1^2 - f^2)^2 + (2\zeta f_1 f)^2}}(3 + z_{1,k}) + \frac{f_2 f}{\sqrt{(f_2^2 - f^2)^2 + (2\zeta f_2 f)^2}}(3 + z_{2,k})$$

$$+ \frac{f_3 f}{\sqrt{(f_3^2 - f^2)^2 + (2\zeta f_3 f)^2}}(3 + z_{3,k}) + e_k(f) \tag{30}$$

$$= 3\mathbf{C}_2(f)[1 \quad 1 \quad 1]^T + \mathbf{C}_2(f)\mathbf{z}_k + e_k(f),$$

with resonance frequencies $f_1 = 200$ f.u., $f_2 = 250$ f.u., $f_3 = 300$ f.u. and relative dampings $\zeta_1 = \zeta_2 = \zeta_3 = 0.05$, and with $\mathbf{C}_2(f) \in \mathbb{R}^{1 \times 3}$. It is also assumed that the variations in the concentration of Constituent 1, Constituent 2 and Constituent 3, denoted $z_{1,k}$, $z_{2,k}$ and $z_{3,k}$, are independent and randomly generated zero mean numbers with normal distributions and variances $r_{zz} = Ez_{1,k}^2 = Ez_{2,k}^2 = Ez_{3,k}^2 = 1$. The noise terms $e_k(f)$ are independent and randomly generated zero mean numbers with normal distribution and equal variances $r_{ee} = Ee_k^2(f)$. Several r_{ee} values were used in the simulations.

Signal-noise-ratio for the **X**-data

The total signal-noise-ratio (total *SNR*) for the **X**-data used in the simulations can be defined as the ratio between the total variances in the centered matrices $\mathbf{ZC}_2^T$ and **E** in (5). This gives the expectation (e.g. Johnson and Wichern, 1998)

$$E\{\text{total } SNR\} = E\left\{\frac{trace(\mathbf{C}_2\mathbf{Z}^T\mathbf{Z}\mathbf{C}_2^T)}{trace(\mathbf{E}^T\mathbf{E})}\right\} = \frac{trace(\mathbf{C}_2\mathbf{C}_2^T)}{pr_{ee}}. \tag{31}$$

The expected total *SNR* for the different values of r_{ee} used in the simulations are given in Table 1. Note the very low total SNR for $r_{ee} = 100$. Also note, however, that the signal-to-noise ratio in the central part of the spectrum is better than that. The highest expected column *SNR* value is found at the frequency $f = 250$ as (in Matlab notation)

$$E\{\text{max. column } SNR\} = \frac{trace(\mathbf{C}_2(250,:)\mathbf{C}_2^T(250,:))}{r_{ee}}, \tag{32}$$

and is also included in Table 1.

Case with a single response variable

It was initially assumed a single response variable

$$y_k = z_{2,k} = [0 \quad 1 \quad 0]\mathbf{z}_k + f_k. \tag{33}$$

Table 1. Expected total and maximum column SNR for different values of **X**-noise variance r_{ee}.

r_{ee}	1	3.2	10	32	100
total *SNR*	22.6	7.06	2.26	0.71	0.23
max. column *SNR*	111.6	34.8	11.16	3.49	1.12

In a practical case this would mean that the primary response of interest would be the concentration of one of the three chemical constituents, while the others would be treated as interferants.

The total model with centered data is then

$$y_k = [0 \quad 1 \quad 0]\mathbf{z}_k + f_k \tag{34}$$

$$\mathbf{x}_k = \mathbf{C}_2\mathbf{z}_k + \mathbf{e}_k, \tag{35}$$

where $k = 1,2,\ldots,N$ indicates sequences of y and $\mathbf{x}$ observations corresponding to different concentrations of the three constituents, and where $\mathbf{C}_2 \in \mathbb{R}^{500 \times 3}$, $E\mathbf{z}_k\mathbf{z}_k^T = \mathbf{I}_3$, $r_{ff} = Ef_k^2 = 0.0001$ and $\mathbf{R}_e = E\mathbf{e}_k\mathbf{e}_k^T = r_{ee}\mathbf{I}_{500}$.

Prediction ability. Based on a modeling set with $N = 100$ and a validation set with $N_{\text{val}} = 1000$ centered observations, $M = 100$ Monte Carlo runs gave the mean root mean square errors of prediction (RMSEP) as shown in Fig. 2. Mean RMSEP values based on the theoretical Kalman predictor (24) are also plotted.

PLSR results at different noise levels and the corresponding results for PCR and the theoretical Kalman predictor (24) are shown in Table 2.

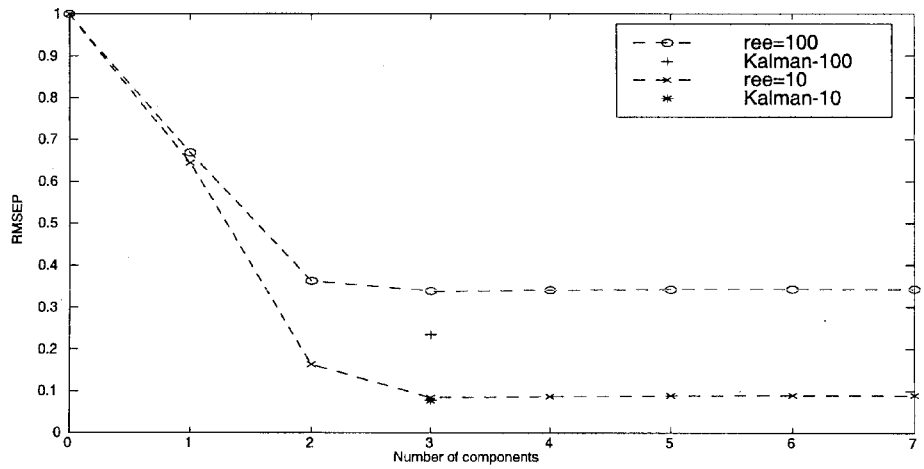

Figure 2. Mean validation RMSEP values for different numbers of PLSR components, based on $M = 100$ Monte Carlo runs using $N = 100$ observations in the modeling set. Two different **X**-noise levels, $r_{ee} = 10$ and $r_{ee} = 100$ are used. The mean validation RMSEP values based on the theoretical Kalman predictor (24) for $A = 3$ components are included.

Table 2. Mean validation PCR, PLSR and Kalman predictor results from $M = 100$ Monte Carlo runs using $N = 100$ observations in the modeling set, $A = 3$ components and different values of r_{ee}.

r_{ee}	$\text{RMSEP}_{\text{PCR}}$	$\text{RMSEP}_{\text{PLSR}}$	RMSEP_{KF}
1	0.0251	0.0251	0.0244
3.2	0.0460	0.0461	0.0436
10	0.0866	0.0862	0.0770
32	0.1799	0.1751	0.1357
100	0.3718	0.3393	0.2343

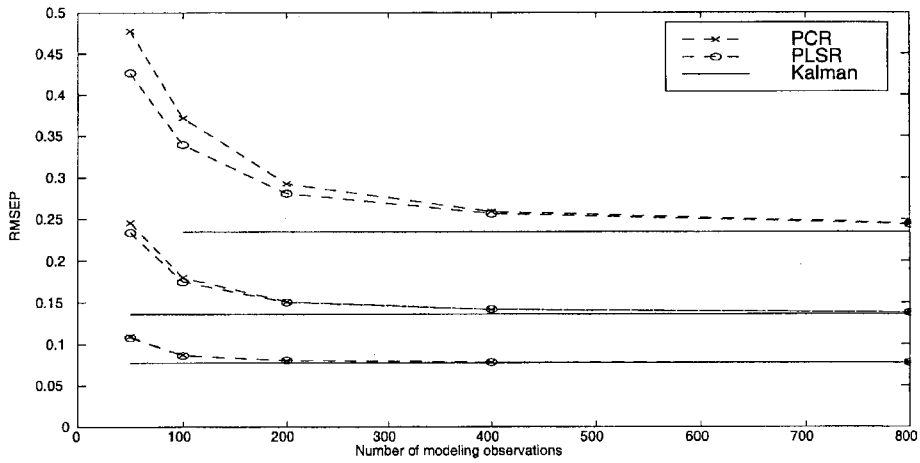

Figure 3. Mean validation PCR and PLSR results from $M = 100$ Monte Carlo runs using $r_{ee} = 10$, 32 and 100 (expected total $SNR = 2.26$, 0.71 and 0.23), $A = 3$ and different numbers N of modeling observations. The Kalman predictor results are shown by solid lines.

The corresponding PCR and PLSR results at different noise levels r_{ee} and with different numbers N of modeling observations are shown in Fig. 3. Not surprisingly, the predictors deteriorate for small values of N, especially at high noise levels. Note that the difference between PCR and PLSR is more pronounced at high noise levels, and that for large values of N the predictions seems to approach the theoretical Kalman predictions.

From Table 2 and Fig. 3 it may be concluded that both PCR and PLSR in this case handles **X**-noise well, as compared with the theoretical Kalman predictor, especially at noise levels up to $r_{ee} = 10$ (total $SNR = 2.26$), where for both PLSR and PCR the relative RMSEP increase due to noise is 12% for $N = 100$. For $N = 400$ the relative RMSEP increase due to noise is 9 to 10% at $r_{ee} = 100$ (total $SNR = 0.23$). See also Fig. 1 for an illustration of the noise levels.

Spectra estimation. An LS estimation according to (8) resulted in a typical case with $A = 3$, $N = 200$ and $r_{ee} = 10$ in the estimated spectral profile for Constituent 2 shown in Fig. 4a, while the profile estimates according to (29) for the unknown constituents 1 and 3 are shown in Fig. 4b and 4c (sign indeterminate and assuming the same scaling factor $(y^T y)^{-1}$ as for Constituent 2). As can be seen, the known constituent profile is estimated fairly well, while the profiles of the two unknown constituents are confounded.

Case with two response variables

The output model (33) was in this case replaced by

$$\begin{bmatrix} y_{1,k} \\ y_{2,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_{1,k} \\ z_{2,k} \\ z_{3,k} \end{bmatrix} + \begin{bmatrix} f_{1,k} \\ f_{2,k} \end{bmatrix}. \tag{36}$$

Two separate PLSR models with $\mathbf{C}_1 = [1 \quad 0 \quad 0]$ and $\mathbf{C}_1 = [0 \quad 1 \quad 0]$, and with $A = 3$

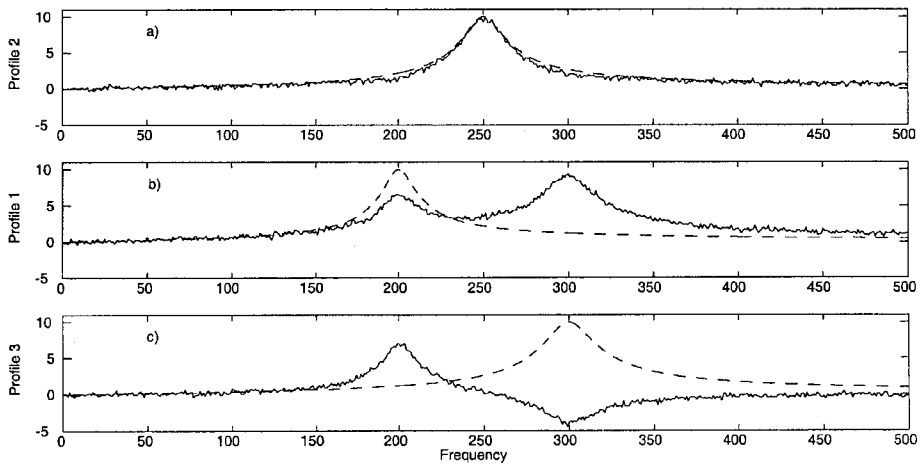

Figure 4. Estimated spectral profiles for a single known constituent (Fig. a) and for two unknown constituents (Fig. b and c), using $A = 3$ components, noise variances $r_{ee} = 10$ and $N = 200$ observations in the modeling set. The known reference profiles are shown by dashed lines.

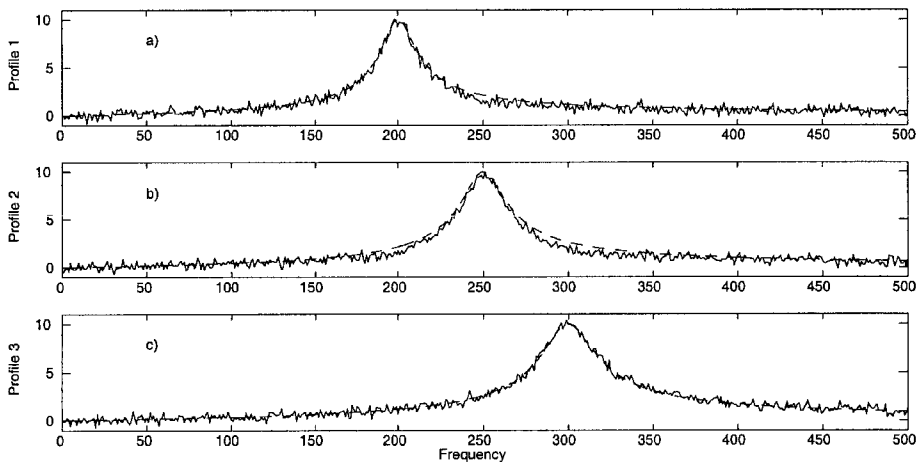

Figure 5. Estimated profiles for two known constituents (Fig. a and b) and for a third unknown constituent (Fig. c), using noise variances $r_{ee} = 10$ and $N = 200$ observations in the modeling set. The known reference profiles are shown by dashed lines.

components, gave in a typical case with $N = 200$ and $r_{ee} = 10$ the individually estimated profiles shown in Fig. 5a and b, while the profile for the unknown Constituent 3 is shown in Fig. 5c (assuming the same scaling factor $(y^T y)^{-1}$ as for Constituent 1). In this case all constituent profiles are estimated fairly well, including the unknown interferant profile.

Conclusions

The noise handling capabilities of PCR and PLSR have been tested by simulations of a typical multivariate mixing problem, using spectra with $p = 500$ discrete frequen-

cies and different **X**-noise levels. Comparisons with optimal Kalman predictors show that both PCR and PLSR perform well even at a considerable noise level (ca. 12% relative increase in RMSEP for $N = 100$ observations at a total signal-to-noise ratio $SNR = 2.26\%$, and ca. 10% relative increase in RMSEP for $N = 400$ observations at a total $SNR = 0.23$). Prediction errors due to **X**-noise as functions of total SNR and N are presented in Fig. 3. Corresponding tests on constituent profile LS + PCA/SVD estimation show similar good noise handling capabilities.

Appendix A

Latent variables PLSR models

In Section 2 a new latent variables representation of the orthogonal PLSR factorization was presented. This calls for further discussion and argumentation.

The two PLSR algorithms of Wold and Martens may use LV models as starting points. The Martens algorithm is in this respect quite straightforward. Using the model

$$\mathbf{Y} = \mathbf{T}_{\mathrm{M}}\mathbf{Q}_{\mathrm{M}}^{T} + \mathbf{F} \tag{37}$$

$$\mathbf{X} = \mathbf{T}_{\mathrm{M}}\mathbf{W}^{T} + \mathbf{E}, \tag{38}$$

and following the derivation of the Helland predictor (14), the predictor is

$$\hat{\mathbf{B}}_{\mathrm{PLSR}} = \hat{\mathbf{W}}(\hat{\mathbf{W}}^{T}\mathbf{X}^{T}\mathbf{X}\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}^{T}\mathbf{X}^{T}\mathbf{Y}, \tag{39}$$

where $\hat{\mathbf{W}}$ is found from the modeling data through a step-wise procedure (e.g. Martens and Næs, 1989). Since $\hat{\mathbf{W}}^{T}\hat{\mathbf{W}} = \mathbf{I}_{A}$, (38) gives the LS estimate $\hat{\mathbf{T}}_{\mathrm{M}} = \mathbf{X}\hat{\mathbf{W}}$, and the estimate

$$\hat{\mathbf{X}}_{\mathrm{M}} = \hat{\mathbf{T}}_{\mathrm{M}}\hat{\mathbf{W}}^{T} = \mathbf{X}\hat{\mathbf{W}}\hat{\mathbf{W}}^{T}. \tag{40}$$

The Wold algorithm is normally associated with the model (e.g. Trygg, 2001)

$$\mathbf{Y} = \mathbf{T}_{\mathrm{W}}\mathbf{Q}_{\mathrm{W}}^{T} + \mathbf{F} \tag{41}$$

$$\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}_{\mathrm{W}}^{T} + \mathbf{E}, \tag{42}$$

although the step-wise algorithm also finds the same loading weight matrix $\hat{\mathbf{W}}$ as in the Martens algorithm, and thus an identical predictor (Ergon, 1998). This model is unfortunate in that the resulting $\hat{\mathbf{X}}_{\mathrm{W}} = \hat{\mathbf{T}}_{\mathrm{W}}\hat{\mathbf{P}}_{\mathrm{W}}^{T}$ is different from $\hat{\mathbf{X}}_{\mathrm{M}}$ according to (40), in spite of the fact that the predictors are identical. This is corrected by use of the model

$$\mathbf{Y} = \mathbf{T}_{\mathrm{W}}\mathbf{Q}_{\mathrm{W}}^{T} + \mathbf{F} \tag{43}$$

$$\mathbf{X} = \mathbf{T}_{\mathrm{W}}\mathbf{P}_{\mathrm{W}}^{T}\mathbf{W}\mathbf{W}^{T} + \mathbf{E}, \tag{44}$$

from which follows the LS estimate

$$\hat{\mathbf{T}}_{\mathrm{W}} = \mathbf{X}\hat{\mathbf{W}}(\hat{\mathbf{P}}_{\mathrm{W}}^{T}\hat{\mathbf{W}})^{-1}, \tag{45}$$

and

$$\hat{\mathbf{X}}_{\mathrm{W}} = \hat{\mathbf{T}}_{\mathrm{W}}\hat{\mathbf{P}}_{\mathrm{W}}^{T}\hat{\mathbf{W}}\hat{\mathbf{W}}^{T} = \mathbf{X}\hat{\mathbf{W}}(\hat{\mathbf{P}}_{\mathrm{W}}^{T}\hat{\mathbf{W}})^{-1}\hat{\mathbf{P}}_{\mathrm{W}}^{T}\hat{\mathbf{W}}\hat{\mathbf{W}}^{T} = \hat{\mathbf{X}}\mathbf{W}\hat{\mathbf{W}}^{T} = \hat{\mathbf{X}}_{\mathrm{M}}. \tag{46}$$

Note that (45) is used also in connection with the model (41, 42), although this is

less obvious (e.g. Helland, 1988). Another argument for use of (43,44) follows from the equations for profile estimation derived in Appendix B. It is there shown that the models (37,38) and (43,44) give the same $\hat{\mathbf{C}}_Y$ estimate, while (41,42), on the other hand, gives a different result. Considering that the Wold and Martens algorithms give the same predictor, a different $\hat{\mathbf{C}}_Y$ seems quite illogical. A closer look at the two alternative Wold models reveals that (42) gives

$$\mathbf{X} = \hat{\mathbf{T}}_W \hat{\mathbf{P}}_W^T + \mathbf{E} = \hat{\mathbf{t}}_1 \hat{\mathbf{p}}_1^T + \hat{\mathbf{t}}_2 \hat{\mathbf{p}}_2^T + \ldots + \hat{\mathbf{t}}_{A-1} \hat{\mathbf{p}}_{A-1}^T + \hat{\mathbf{t}}_A \hat{\mathbf{p}}_A^T + \mathbf{E}_1, \tag{47}$$

while (44) gives

$$\mathbf{X} = \hat{\mathbf{T}}_W \hat{\mathbf{P}}_W^T \hat{\mathbf{W}} \hat{\mathbf{W}}^T + \mathbf{E} = \hat{\mathbf{t}}_1 \hat{\mathbf{p}}_1^T + \hat{\mathbf{t}}_2 \hat{\mathbf{p}}_2^T + \ldots + \hat{\mathbf{t}}_{A-1} \hat{\mathbf{p}}_{A-1}^T + \hat{\mathbf{t}}_A \hat{\mathbf{w}}_A^T + \mathbf{E}_2, \tag{48}$$

i.e. there is a difference in the last component only. This is due to the bi-diagonal structure of $\mathbf{P}_W^T \hat{\mathbf{W}}$ (Manne, 1987). It is also a result of the step-wise PLSR algorithm (e.g. Martens and Næs, 1989), where $\hat{\mathbf{t}}_a$ for each component is found from the local model $\mathbf{X}_{a-1} = \mathbf{t}_a \hat{\mathbf{w}}_a^T + \mathbf{E}$, while $\hat{\mathbf{p}}_a$ is used only to find $\mathbf{X}_a$ used for computation of $\hat{\mathbf{t}}_{a+1}$ and $\hat{\mathbf{w}}_{a+1}$ etc.

Appendix B

Constituent profile estimation from PCR and PLSR results

As shown in (25, 26), the first loading weight vector $\mathbf{w}_{j1}$ related to a specific single response y_j gives an LS optimal estimate (possibly scaled) of the corresponding pure constituent spectrum $\mathbf{C}_{2j}$. As shown below, this result may also be found by reconstruction of the similarity transformation from the model (4, 5) to (9, 10). This may also be applied to PCR, although the result is then not LS optimal.

Reconstruction of similarity transformation

Assume that (9, 10) are obtained from (4, 5) through the similarity transformation

$$\mathbf{Y} = \mathbf{Z}\mathbf{S}^{-T}\mathbf{S}^T\mathbf{C}_1^T + \mathbf{F} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \tag{49}$$

$$\mathbf{X} = \mathbf{Z}\mathbf{S}^{-T}\mathbf{S}^T\mathbf{C}_2^T + \mathbf{E} = \mathbf{T}\mathbf{L}^T + \mathbf{E}, \tag{50}$$

i.e. $\mathbf{Z}\mathbf{S}^{-T} = \mathbf{T}$, $\mathbf{C}_1\mathbf{S} = \mathbf{Q}$ and $\mathbf{C}_2\mathbf{S} = \mathbf{L}$. From (3) thus follows that

$$\mathbf{Q} = \mathbf{C}_1\mathbf{S} = [\mathbf{I}_m \quad 0] \begin{bmatrix} \mathbf{S}_Y \\ \mathbf{S}_{OSC} \end{bmatrix} = \mathbf{S}_Y, \tag{51}$$

i.e. $\mathbf{S}_Y = \mathbf{Q}$. The first m rows in $\mathbf{S}$ are thus given by the m rows in $\mathbf{Q}$. From (49) follows further that an LS estimate of $\mathbf{S}_Y^T = \mathbf{Q}^T$ is found from

$$\hat{\mathbf{S}}_Y^T = \hat{\mathbf{Q}}^T = (\hat{\mathbf{T}}^T\hat{\mathbf{T}})^{-1}\hat{\mathbf{T}}^T\mathbf{Y}. \tag{52}$$

Pure constituent profile estimation

It further follows from (49, 50) that

$$\mathbf{S}\mathbf{T}^T\mathbf{T}\mathbf{S}^T(\mathbf{Z}^T\mathbf{Z})^{-1} = \mathbf{Z}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1} = \mathbf{I}_A, \tag{53}$$

and thus by use of (4)

$$\hat{\mathbf{C}}_2 = [\hat{\mathbf{C}}_Y \quad \hat{\mathbf{C}}_{\text{OSC}}] \approx \hat{\mathbf{C}}_2 \hat{\mathbf{S}} \hat{\mathbf{T}}^T \hat{\mathbf{T}} \hat{\mathbf{S}}^T (\hat{\mathbf{Z}}^T \hat{\mathbf{Z}})^{-1} \approx \hat{\mathbf{L}} \hat{\mathbf{T}}^T \hat{\mathbf{T}} \hat{\mathbf{S}}^T \begin{bmatrix} \mathbf{Y}^T\mathbf{Y} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{\text{OSC}}^T \mathbf{Z}_{\text{OSC}} \end{bmatrix}^{-1} \tag{54}$$

$$= \hat{\mathbf{L}} \hat{\mathbf{T}}^T \hat{\mathbf{T}} [\hat{\mathbf{S}}_Y^T \quad \mathbf{S}_{\text{OSC}}^T] \begin{bmatrix} (\mathbf{Y}^T\mathbf{Y})^{-1} & \mathbf{0} \\ \mathbf{0} & (\hat{\mathbf{Z}}_{\text{OSC}}^T \hat{\mathbf{Z}}_{\text{OSC}})^{-1} \end{bmatrix}.$$

The pure constituent profile estimate is thus by use of (52) found as

$$\hat{\mathbf{C}}_Y = \hat{\mathbf{L}} \hat{\mathbf{T}}^T \hat{\mathbf{T}} \hat{\mathbf{S}}_Y^T (\mathbf{Y}^T\mathbf{Y})^{-1} = \hat{\mathbf{L}} \hat{\mathbf{T}}^T \hat{\mathbf{T}} \hat{\mathbf{Q}}^T (\mathbf{Y}^T\mathbf{Y})^{-1} = \hat{\mathbf{L}} \hat{\mathbf{T}}^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1}, \tag{55}$$

where the columns of $\hat{\mathbf{C}}_Y$ are scaled estimates of the pure constituent profiles corresponding to $y_1, y_2, \ldots, y_m$.

The matrix $\hat{\mathbf{L}}$ will depend on the specific multivariate calibration method used. A PCR model uses $\hat{\mathbf{L}} = \hat{\mathbf{P}}$ and the LS estimate $\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{P}}$ from (15), and (55) then results in

$$\hat{\mathbf{C}}_Y^{\text{PCR}} = \hat{\mathbf{P}} \hat{\mathbf{T}}^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1} = \hat{\mathbf{P}} \hat{\mathbf{P}}^T \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1}. \tag{56}$$

The Martens PLSR algorithm uses $\hat{\mathbf{L}} = \hat{\mathbf{W}}$ and the LS estimate $\hat{\mathbf{T}}_M = \mathbf{X}\hat{\mathbf{W}}$ from (17), where $\hat{\mathbf{W}}$ is the loading weight matrix, and (55) then results in

$$\hat{\mathbf{C}}_Y^{\text{PLSR}} = \hat{\mathbf{W}} \hat{\mathbf{T}}_M^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1} = \hat{\mathbf{W}} \hat{\mathbf{W}}^T \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1}. \tag{57}$$

The Wold PLSR algorithm uses $\hat{\mathbf{L}} = \hat{\mathbf{W}} \hat{\mathbf{W}}^T \hat{\mathbf{P}}_W$ and $\hat{\mathbf{T}}_W = \mathbf{X}\hat{\mathbf{W}}(\mathbf{P}_W^T \hat{\mathbf{W}})^{-1}$ (see Appendix A), which results in

$$\hat{\mathbf{C}}_Y^{\text{PLSR}} = \hat{\mathbf{W}} \hat{\mathbf{W}}^T \hat{\mathbf{P}}_W \hat{\mathbf{T}}_W^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1} = \hat{\mathbf{W}} \hat{\mathbf{W}}^T \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T\mathbf{Y})^{-1}, \tag{58}$$

i.e. the same estimate as for the Martens algorithm. This is natural, since the two PLSR algorithms give the same predictor $\hat{\mathbf{B}}$ (see a detailed discussion of the two PLSR methods in Appendix A). For both PCR and PLSR the estimates of the columns of $\hat{\mathbf{C}}_Y$ may be found jointly (PCR or PLS2), or separately (PCR or PLS1). This will give identical results for PCR, while the PLS1 and PLS2 results normally are different.

Comparison of PCR and PLSR results

The pure constituent profile estimates (56) and (57, 58) should be compared with the estimate (8), which is optimal in the LS sense. Starting with PLSR we find for a single response variable y_j (PLS1)

$$\hat{\mathbf{C}}_{2j}^{\text{PLSR}} = \hat{\mathbf{W}}_j \hat{\mathbf{W}}_j^T \mathbf{X}^T \mathbf{y}_j (\mathbf{y}_j^T \mathbf{y}_j)^{-1} = [\hat{\mathbf{w}}_{j1} \quad \hat{\mathbf{w}}_{j2} \quad \ldots \quad \hat{\mathbf{w}}_{jA}] \begin{bmatrix} \hat{\mathbf{w}}_{j1}^T \\ \hat{\mathbf{w}}_{j2}^T \\ \vdots \\ \hat{\mathbf{w}}_{jA}^T \end{bmatrix} \mathbf{X}^T \mathbf{y}_j (\mathbf{y}_j^T \mathbf{y}_j)^{-1}. \tag{59}$$

However, it is a part of the PLSR algorithms that $\mathbf{X}^T\mathbf{y}_j = \sqrt{\mathbf{y}_j^T \mathbf{X}\mathbf{X}^T \mathbf{y}_j}\, \hat{\mathbf{w}}_{j1}$ and that $\hat{\mathbf{W}}_j^T \mathbf{W}_j = \mathbf{I}$ (e.g. Martens and Næs, 1989), and we thus find

$$\hat{\mathbf{C}}_{2j}^{\text{PLSR}} = \sqrt{\mathbf{y}_j^T \mathbf{X}\mathbf{X}^T \mathbf{y}_j} [\hat{\mathbf{w}}_{j1} \quad \hat{\mathbf{w}}_{j2} \quad \ldots \quad \hat{\mathbf{w}}_{jA}] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (\mathbf{y}_j^T \mathbf{y}_j)^{-1} = \mathbf{X}^T \mathbf{y}_j (\mathbf{y}_j^T \mathbf{y}_j)^{-1}. \tag{60}$$

Using this for all the columns of $\mathbf{C}_Y$ we will find the total estimate $\hat{\mathbf{C}}_Y$ given by the LS solution (8). The PSLR algorithms are thus optimal in the sense that the first loading weight vectors for the different single responses provide LS estimates of the corresponding pure constituent profiles. The PCR estimate (56), on the other hand, results in

$$\hat{\mathbf{C}}_{2j}^{\mathrm{PCR}} = \hat{\mathbf{P}}_j \hat{\mathbf{P}}_j^T \mathbf{X}^T \mathbf{y}_j (\mathbf{y}_j^T \mathbf{y}_j)^{-1} = [\hat{\mathbf{p}}_{j1} \quad \hat{\mathbf{p}}_{j2} \quad \cdots \quad \hat{\mathbf{p}}_{jA}] \begin{bmatrix} \hat{\mathbf{p}}_{j1}^T \\ \hat{\mathbf{p}}_{j2}^T \\ \vdots \\ \hat{\mathbf{p}}_{jA}^T \end{bmatrix} \mathbf{X}^T \mathbf{y}_j (\mathbf{y}_j^T \mathbf{y}_j)^{-1}, \qquad (61)$$

from which follows that the LS solution (8) is obtained for $A = p$ only, i.e. when $\hat{\mathbf{P}}_j \hat{\mathbf{P}}_j^T = \mathbf{I}_p$. When it comes to estimation of $\mathbf{C}_Y$, the PCR estimate is thus generally not optimal in the LS sense.

The difference between PLSR and PCR based constituent profile estimation may be small. As a test the PCR based results corresponding to Fig. 5 were computed by use of (61), and the differences from the PLSR results were hardly visible.

There is, however, a difference when it comes to the number of components necessary for constituent profile estimation. As shown in (26), the first PLSR component only is actually used, and the rest are thus unnecessary in this respect. When using PCR for this purpose, however, the optimal or a larger number of components must be used. As an example, Fig. 6 shows results corresponding to Fig. 5, but now with use of $A = 2$ PLSR and PCR components only.

The PCR results for $A = 2$ show confounded spectra for the two known constituents, while the PCR and PLSR estimates for the unknown constituent are quite similar. As pointed out above, the LS solution (8) is obtained for $A = p$, and simulations using the present mixing example with three constituents show very good results for $A \geqslant 3$.

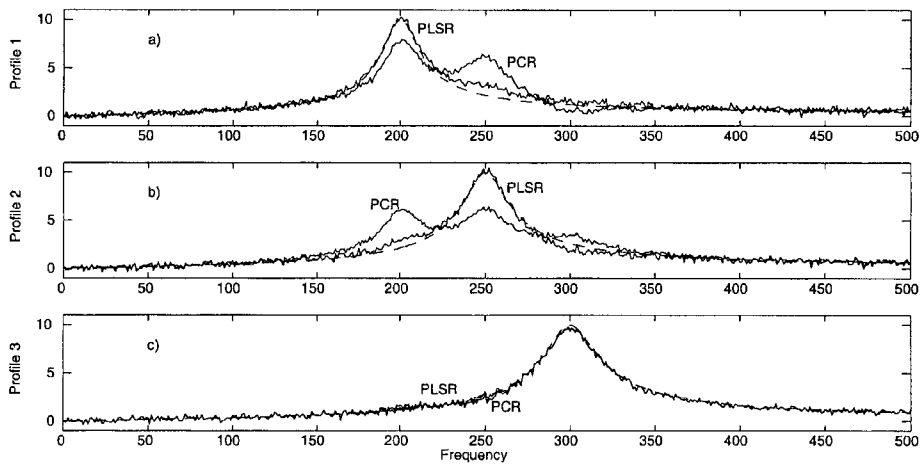

Figure 6. Estimated profiles for two known constituents (Fig. a and b) and for a third unknown constituent (Fig. c), using $N = 200$ modeling observations and $A = 2$ PLSR and PCR components. The **X**-noise variances are $r_{ee} = 10$.

References

BERNTSEN, H. (1988). *Utvidet Kalmanfilter og multivariabel kalibrering*, Report STF48 A88019, SINTEF, Trondheim, Norway.

DI RUSCIO, D. (2000), A weighted view on the partial least squares algorithm, *Automatica*, **36,** pp. 831–850.

ERGON, R. (1998). Dynamic system multivariate calibration by system identification methods, *Modeling, Identification and Control*, **19**, No. 2, pp. 77–97.

ERGON, R. and ESBENSEN, K. H. (2001). A didactically motivated PLS prediction algorithm, *Modeling, Identification and Control*, **22,** No. 3, pp. 131–139.

FEARN, T. (2000). On orthogonal signal correction, *Chemometrics Intell. Lab. Syst.*, **44,** pp. 229–244.

GELB, A. (1974). *Applied Optimal Estimation*, MIT Press, Mass.

GREWAL, M. S. and ANDREWS, A. P. (1993). *Kalman Filtering: Theory and Practice*, Prentice Hall, New Jersey.

HELLAND, I. S. (1988). On the structure of partial least squares regression, *Communications in statistics*, **17,** pp. 581–607.

JOHNSON, R. A. and WICHERN, D. W. (1998). *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey.

MANNE, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration, *Chemometrics Intell. Lab. Syst.*, **2,** pp. 187–97.

MARTENS, H. and NÆS, T. (1989). *Multivariate Calibration*, Wiley, New York.

TRYGG, J. and WOLD, S. (2001). Orthogonal projections to latent structures (O-PLS), *J. Chemometrics*, **15,** pp. 1–18.

TRYGG, J. (2001). *Parsimonious Multivariate Models*, Ph.D. thesis, Umeå University, Department of Chemistry, Sweden.

WESTERHUIS, J. A., DE JONG, S., SMILDE, A. K. (2001). Direct orthogonal signal correction, *Chemometrics Intell. Lab. Syst.*, **56,** pp. 13–25.

WOLD, S., MARTENS, H. and WOLD, H. (1982). The multivariate calibration problem in chemistry solved by the PLS method, *Proc. Conf. Matrix pencils.*

WOLD, S., ANTTI, H., LINDGREN, F. and OHMAN, J. (1988). Orthogonal signal correction of near-infrared spectra, *Chemometrics Intell. Lab. Syst.*, **44,** pp. 175–185.