

## Dynamic system multivariate calibration based on multirate sampling data

ROLF ERGON and MATHS HALSTENSEN

Keywords: *Dynamic system, latent variables, multirate sampling*

The statistical principal component regression (PCR) and chemometric partial least squares regression (PLSR) algorithms based on latent variables (LV) modeling are effective tools for handling ill-conditioned regression data. In many process related cases the data form time series, and it may then be possible to improve the prediction/estimation results by utilizing the autocorrelation in the observations. This can be done by use of estimators found from experimental data by use of a combination of statistical/chemometric and system identification methods. In important industrial cases, the response variables are product qualities which also in the experimental data are sampled at a low and possibly irregular rate, while the regressor variables are sampled at a higher rate. After a discussion of the options available, the paper shows how the autocorrelation of the regressor variables in such multirate sampling cases may be utilized by identification of latent variables based output error (LV + OE) estimators. An example using acoustic power spectrum regressor data is finally presented.

### 1. Introduction and problem statement

In many important practical cases there is a need to estimate primary system responses  $y_1$  from known system inputs  $u$  and secondary system outputs  $y_2$ . In typical industrial cases the  $y_1$  responses are product quality measurements that cannot be made available on-line, while  $u$  are manipulated process inputs and  $y_2$  are various on-line process measurements. Dynamic estimators for this purpose may be identified from experimental  $u$ ,  $y_1$  and  $y_2$  data (Ergon, 1999b).

The experimental data are often characterized by few observations and a large number of nearly collinear  $y_2$  variables. In the static case the statistical/chemometric methods PCR (Principal Component Regression) or PLSR (Partial Least Squares Regression) based on latent variables (LV) models provide powerful tools for handling such ill-conditioned problems. The experimental  $u$  and  $y_2$  data are then collected in an  $X$  matrix, while the  $y_1$  data are collected in a  $y$  vector (assuming a single response variable), and the  $X$  information is compressed into an estimated latent variables matrix  $\hat{T}$ . In many process related cases the  $X$  and  $y$  data in ordinary PCR or PLSR form time series, and it may then be possible to improve the prediction/estimation results by utilizing the autocorrelation in the observations. After compression of the  $X$  data into  $\hat{T}$ , this can basically be done by identification of FIR (Finite Impulse Response), ARMAX (AutoRegressive Moving Average with eXtra inputs), ARX (AutoRegressive with eXtra inputs) or OE (Output Error) type of models. The estimator may thus be found from experimental data by a combination of multivariate calibration methods used in statistics/chemometrics (e.g. Martens and Næs (1989) and Esbensen (2000)) and system identification methods (e.g. Ljung (1999), Söderström and Stoica (1989) and Van Overshee and De Moor (1996)).

In some cases of practical interest it is possible to obtain experimental  $X$  data at a high sampling rate, while the experimental  $y$  data are obtained at a lower and possibly also irregular rate. In typical cases the  $y$  data are product quality measurements that require chemical or other forms of analyses, such that the results for practical and economical reasons cannot be made available at a high rate. The problem of the present paper is to find an identification method that handles such multivariate and multirate sampling time series data.

After an introductory discussion of the options available, a solution to the problem using an LV + OE model structure is presented. An example using acoustic power spectrum  $X$  data demonstrates the feasibility of this method.

## 2. Static multivariate calibration

As a background for a comparison of static and dynamic estimators, we start with a short presentation of static multivariate calibration methods based on latent variables models. We then focus on the common situation where the  $X$  variables are secondary outputs from a static system, and where no manipulated or otherwise known system inputs are available. Since  $u = 0$ , the regressor data is then  $X = [U \ Y_2] = Y_2$ , and the following is based on this assumption.

### *The static multivariate calibration problem*

Assuming a static system with a scalar primary output or response variable  $y_1$  and multivariate secondary outputs  $y_2$ , the static calibration problem is to find an estimator  $\hat{b}$  from experimental data that may be used to predict a new response  $y_{1,0}$  from new observations  $y_{2,0}$  according to

$$\hat{y}_{1,0} = y_{2,0}^T \hat{b}.$$

A classical example is the estimation of protein content in whole wheat kernels based on near infrared (NIR) spectroscopy (Norris, 1993). Here, the protein content is the primary output  $y_1$ , while the NIR reflectance at a large number of frequencies gives rise to the  $y_2$  variables. A process related example is the estimation of distillation product composition from a number of temperature measurements along the distillation tower (Mejdell and Skogestad, 1989). The fundamental problem in such cases is that the number of  $y_2$  variables may be much larger than the number of observations in the experimental data. A comprehensive discussion of the multivariate calibration problem is given in Martens and Næs (1989), while Esbensen (2000) focuses on the interpretational and practical application aspects of latent variables methods.

### *Least squares estimation*

Assuming experimental data from  $N$  observations,  $y_1 = [y_{11} \ y_{12} \ \dots \ y_{1N}]^T$  and  $Y_2 = [y_{21} \ y_{22} \ \dots \ y_{2N}]^T$ , and independent observation errors, we find the least squares (LS) solution

$$\hat{b}_{LS} = (Y_2^T Y_2)^{-1} Y_2^T y_1. \quad (2)$$

With a large number  $p$  of  $y_2$  variables, this solution will be very noise and collinearity sensitive, and in practical applications the LS method will give satisfactory results only when  $p$  is well below  $N$ . A thorough discussion of this problem is given in Belsley (1991).

*Latent variables model*

In many practical situations, fortunately, the multivariate  $y_2$  variables are highly collinear, and the  $y_2$  information may then be compressed into small number  $a$  of estimated latent variables  $\tau = [\tau_1 \ \tau_2 \ \cdots \ \tau_a]^T$ . The model underlying such data compression is the latent variables (LV) model

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{w}_k \\ y_{1,k} &= \mathbf{C}_1 \mathbf{x}_k + v_{1,k} \\ y_{2,k} &= \mathbf{C}_2 \mathbf{x}_k + v_{2,k}, \end{aligned} \quad (3)$$

where  $\mathbf{w}_k$ ,  $v_{1,k}$  and  $v_{2,k}$  are white noise sequences. After an appropriate similarity transformation the output equations may be written

$$\begin{aligned} y_{1,k} &= \tilde{\mathbf{C}}_1 \tau_k + v_{1,k} \\ y_{2,k} &= \mathbf{W} \tau_k + v_{2,k}, \end{aligned} \quad (4)$$

where  $\mathbf{W}$  is orthonormal, e.g.  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ .

With  $N$  observations and  $\mathbf{T} = [\tau_1 \ \tau_2 \ \cdots \ \tau_N]^T$ ,  $\mathbf{v}_1 = [v_{11} \ v_{12} \ \cdots \ v_{1N}]^T$  and  $\mathbf{V}_2 = [v_{21} \ v_{22} \ \cdots \ v_{2N}]^T$  this gives

$$\mathbf{y}_1 = \mathbf{T} \tilde{\mathbf{C}}_1^T + \mathbf{v}_1 \quad (5)$$

$$\mathbf{Y}_2 = \mathbf{T} \mathbf{W}^T + \mathbf{V}_2. \quad (6)$$

*Regularized solutions*

In PCR and PLSR (the Martens algorithm) the  $\mathbf{Y}_2$  data matrix is compressed into a score matrix  $\hat{\mathbf{T}}$  by use of the factorization

$$\mathbf{Y}_2 = \hat{\mathbf{T}} \hat{\mathbf{W}}^T + \mathbf{E}, \quad (7)$$

where  $\mathbf{E}$  is a residual matrix. Here, the  $\hat{\mathbf{W}}$  weighting matrix is orthonormal, i.e.  $\hat{\mathbf{W}}^T \hat{\mathbf{W}} = \mathbf{I}$ , and the LS solution of (7) is thus

$$\hat{\mathbf{T}} = \mathbf{Y}_2 \hat{\mathbf{W}}. \quad (8)$$

From (5) we now find the ordinary LS estimator related to the  $\tau$  variables

$$\hat{\mathbf{C}}_1^T = (\hat{\mathbf{T}}^T \hat{\mathbf{T}})^{-1} \hat{\mathbf{T}}^T \mathbf{y}_1 = (\hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{Y}_2 \hat{\mathbf{W}})^{-1} \hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{y}_1, \quad (9)$$

which results in fitted experimental responses according to (5) and (8)

$$\hat{\mathbf{y}}_1 = \hat{\mathbf{T}} \hat{\mathbf{C}}_1^T = \mathbf{Y}_2 \hat{\mathbf{W}} \hat{\mathbf{C}}_1^T = \mathbf{Y}_2 \hat{\mathbf{W}} (\hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{Y}_2 \hat{\mathbf{W}})^{-1} \hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{y}_1. \quad (10)$$

The regularized latent variables estimator thus becomes

$$\hat{\mathbf{b}}_{LV} = \hat{\mathbf{W}} (\hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{Y}_2 \hat{\mathbf{W}})^{-1} \hat{\mathbf{W}}^T \mathbf{Y}_2^T \mathbf{y}_1. \quad (11)$$

This LV estimator formulation is earlier given by Helland (1988), although not directly based on a LV model. We will return to this expression in connection with PCR and PLSR estimators using  $\hat{\mathbf{W}} = \hat{\mathbf{W}}_{\text{PCR}} = \hat{\mathbf{P}}$  and  $\hat{\mathbf{W}} = \hat{\mathbf{W}}_{\text{PLS}}$ .

*Principal component regression*

In PCR the weighting matrix is  $\hat{\mathbf{W}} = \hat{\mathbf{P}}$ , where  $\hat{\mathbf{P}}$  is the loading matrix related to the principal components of  $\mathbf{Y}_2$ . We may find  $\hat{\mathbf{P}}$  from the singular value decomposition (SVD)

$$\begin{aligned}
 Y_2 &= USV^T = [U_1 \quad U_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} \hat{P}^T \\ V_2^T \end{bmatrix} \\
 &= U_1 S_1 \hat{P}^T + U_2 S_2 V_2^T = \hat{T} \hat{P}^T + E.
 \end{aligned} \tag{12}$$

The latent variables estimates represented by the score matrix  $\hat{T}$  are thus based on only  $Y_2$  information.

### Partial least squares regression

Some of the latent variables estimates used in a PCR solution may be very weakly correlated with the response variable  $y_1$ , and the PLSR solution to this is to make use of both  $Y_2$  and  $y_1$  information in order to find an improved version of the  $\hat{W}$  matrix. This is normally done step-wise by finding one column  $\hat{w}_a$  of  $\hat{W}_{PLS}$  at a time (e.g. Martens and Næs, 1989), although a one-step procedure is recently developed (Di Ruscio, 2000). Note that  $\hat{T}$  according to (8) is the non-orthogonal score matrix in the Martens PLSR algorithm, while the orthogonal score matrix in the Wold algorithm is somewhat differently defined (e.g. Martens and Næs, 1989).

The basic *ad hoc* idea in the step-wise PLSR algorithms is to find a given column vector  $\hat{w}_a$  by minimization of the sample covariance between the corresponding score vector  $\hat{t}_a = Y_2 \hat{w}_a$  according to (8) and the residual  $\varepsilon_{a-1}$  of  $y_1$  that is not explained by earlier columns  $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_{a-1}$  (in the original Wold and Martens algorithms also residuals of  $Y_2$  and not  $Y_2$  itself is used). A straightforward and simple step-wise PLSR algorithm is developed in Ergon and Esbensen (2001), where the relation between the LV estimator (11) and Kalman filtering theory is also discussed. The algorithm is as follows:

1. Set  $a = 1$  and  $\varepsilon_0 = y_1$ .
2. Compute

$$\hat{w}_a = Y_2^T \varepsilon_{a-1}, \tag{13}$$

$$\hat{w}_a \leftarrow \frac{\hat{w}_a}{\sqrt{\hat{w}_a^T \hat{w}_a}}, \tag{14}$$

$$\hat{W}_a = [\hat{w}_1 \quad \hat{w}_2 \quad \dots \quad \hat{w}_a], \tag{15}$$

$$\hat{b}_a = \hat{W}_a (\hat{W}_a^T Y_2^T Y_2 \hat{W}_a)^{-1} \hat{W}_a^T Y_2^T y \tag{16}$$

and

$$\varepsilon_a = y - Y_2 \hat{b}_a. \tag{17}$$

3. Let  $a \leftarrow a + 1$  and go to step 2.

This algorithm produces a sequence of predictors  $\hat{b}_1, \hat{b}_2$  etc. The appropriate number of column vectors  $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_A$  to use must then be decided after validation against an independent data set. The score and loading matrices that are necessary for the data evaluation and interpretation that is an essential part of chemometrics may also be computed in the algorithm, or separately for specific values of  $a$ .

### 3. Alternatives for dynamic estimation

Dynamic system extensions of the PCR and PLSR algorithms have been discussed by a number of authors, see Ergon and Halstensen (2000) for further references. In

relation to the present problem the alternatives available can be summarized as follows:

- The FIR model

$$y_k \approx \sum_{i=0}^L \mathbf{h}_i^T \mathbf{u}_{k-i} + \eta_k, \quad (18)$$

may be identified by use of LS regression, and this is possible also when most of the  $y_k$  values are missing, in which case we use only the  $y_j$  samples that are available, together with the corresponding past and present  $\mathbf{u}_k$  values. However, a large number of Markov parameters  $\mathbf{h}_i$  must normally be identified, and this requires a correspondingly large number of response observations. This is a serious drawback, especially since it often is a part of the problem that the number of response observations is very limited. In addition the FIR model is biased due to the lack of noise modeling.

- An ARMAX model of the type

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} - \dots + \mathbf{b}_0^T \mathbf{u}_k + \mathbf{b}_1^T \mathbf{u}_{k-1} + \mathbf{b}_2^T \mathbf{u}_{k-2} + \dots + e_k + c_1 e_{k-1} + c_2 e_{k-2} + \dots, \quad (19)$$

where  $e_k$  is a white noise sequence, may be identified by use of for example a prediction error method (e.g. Söderström and Stoica, 1989) or a subspace method (e.g. Van Overshee and De Moor (1996) and Di Ruscio (1997)). This is, however, not possible when  $y$  is sampled at a lower rate than  $\mathbf{u}$ , and an ARMAX estimator is in any case non-optimal when also secondary system outputs are used as inputs (Ergon, 1999a).

- ARX models are ARMAX models without noise modeling, e.g. as given in (19) but with  $c_1 = c_2 = \dots = 0$ . In the present case they have the same drawbacks as ARMAX models, and bias due to the lack of noise modeling in addition.
- OE estimators as described below can be identified also when most of the ordinary  $y$  observations are missing and are theoretically optimal in the sense that they are unbiased and have minimized variances when secondary system outputs are used as estimator inputs (Ergon, 1999a).

#### 4. Theoretical OE estimator

Assume the known discrete-time state-space model

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{G}\mathbf{w}_k \\ y_{1,k} &= \mathbf{C}_1\mathbf{x}_k + \mathbf{D}_1\mathbf{u}_k + v_{1,k} \\ y_{2,k} &= \mathbf{C}_2\mathbf{x}_k + \mathbf{D}_2\mathbf{u}_k + v_{2,k}, \end{aligned} \quad (20)$$

where  $\mathbf{x}_k$  is the state vector. Here  $y_{1,k}$  for  $k=1, 2, \dots, N$  are the primary response observations that in ordinary static PCR/PLSR with a single response variable would be collected in the  $\mathbf{y}$  vector, while the known inputs  $\mathbf{u}_k$  and the secondary outputs  $y_{2,k}$  would be collected in the  $\mathbf{X}$  matrix. The process and measurement noise sequences  $\mathbf{w}_k$ ,  $v_{1,k}$  and  $v_{2,k}$  are assumed to be independent and white. Note that the LV model (3) is a special case of this dynamic model.

Assuming the system matrices known and past and present  $y_{1,k}$  values not available, the optimal  $y_1$  estimator is obtained by a Kalman filter driven by  $u_k$  and  $y_{2,k}$ . The result is

$$\hat{y}_{1,k|k}^{\text{OE}} = C_1 \hat{x}_{k|k} + D_1 u_k = C_1 [\hat{x}_{k|k-1} + K_2^{\text{OE}} (y_{2,k} - D_2 u_k - C_2 \hat{x}_{k|k-1})] + D_1 u_k, \quad (21)$$

where

$$\hat{x}_{k|k-1} = (qI - A + AK_2^{\text{OE}}C_2)^{-1}[(B - AK_2^{\text{OE}}D_2)u_k + AK_2^{\text{OE}}y_{2,k}]. \quad (22)$$

Here,  $K_2^{\text{OE}}$  is the Kalman gain found from the algebraic Riccati equation

$$P^{\text{OE}} = AP^{\text{OE}}A^T + GR_eG^T - AP^{\text{OE}}C_2^T(C_2P^{\text{OE}}C_2^T + R_{22})^{-1}C_2P^{\text{OE}}A^T \quad (23)$$

and

$$K_2^{\text{OE}} = P^{\text{OE}}C_2^T(C_2P^{\text{OE}}C_2^T + R_{22})^{-1}, \quad (24)$$

where  $P^{\text{OE}} = E(x_k - \hat{x}_{k|k-1}^{\text{OE}})(x_k - \hat{x}_{k|k-1}^{\text{OE}})^T$  is the minimized prediction state estimate covariance (see e.g. Grewal and Andrews, 1993).

The estimator equation (21) corresponds to the OE model

$$y_{1,k} = \frac{B_1(q^{-1})u_k + B_2(q^{-1})y_{2,k}}{A(q^{-1})} + \vartheta_k, \quad (25)$$

where  $A(q^{-1})$ ,  $B_1(q^{-1})$  and  $B_2(q^{-1})$  are polynomials in the unit time delay operator  $q^{-1}$ , while  $\vartheta_k$  is a non-white stochastic process.

Note that for a first-order system with  $C_1 = 1$  and  $u_k = 0$ , equation (21) can be simplified to

$$\hat{y}_{1,k|k}^{\text{OE}} = [1 - (A - AK_2^{\text{OE}}C_2)q^{-1}]^{-1}K_2^{\text{OE}}y_{2,k}. \quad (26)$$

## 5. Estimator identification

### Block diagram

The block diagram for the estimator (21) is shown as a part of Figure 1, where the notation  $(\cdot)^{\text{pred}}$  indicates parameter values that are not necessarily correct (although some useful initial values are needed). The figure also illustrates the prediction error method (PEM) for system identification, adapted to minimize the error in the current estimate  $y_{1,j}^{\text{pred}}$  while considering both  $u$  and  $y_2$  as input signals. A parsimonious canonical form with  $C = [1 \ 0 \ \dots \ 0]$  is assumed, and the output matrix is therefore not adjusted in the minimization procedure. It is further assumed that  $y_1$  samples are available only at some of the  $u_k$  and  $y_{2,k}$  time instances  $k$ , resulting in a sequence of low and possibly irregular sampling rate observations  $y_{1,j}$ .

### Estimation error minimization

As indicated in Figure 1, the parameter values are successively modified until a criterion function

$$V_{N_1}(\theta) = \frac{1}{N_1} \sum_{j=1}^{N_1} \varepsilon_{1,j}^2(\theta) \quad (27)$$

is minimized. Here

$$\varepsilon_{1,j}(\theta) = y_{1,j} - y_{1,j}^{\text{pred}} \quad (28)$$

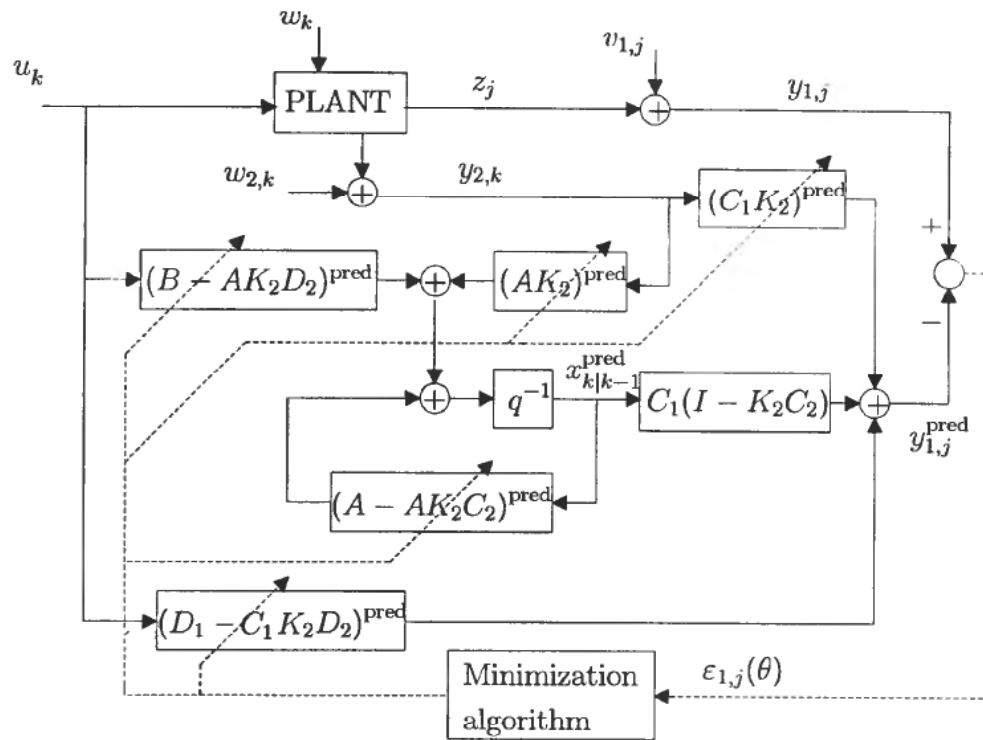


Figure 1. Identification of OE estimator by estimation error minimization.

is the estimation error, while  $N_1$  is the number of  $y_1$  observations in the experimental data (see e.g. Söderström and Stoica (1989) for a general discussion of this prediction error method, and Ergon (1999a), for the case with also  $y_2$  used as input).

Note that when not all  $y_{1,k}$  values are available, the standard initial value procedure based on a least squares approach (Ljung, 1995) cannot be used. The initial value problem may therefore be the most difficult part of the problem, although several *ad hoc* approaches are available (Ergon, 1998, 1999b).

Minimization of equation (27) can be performed only after a specification of the model order, and the normal procedure is then to select a chain of canonical structures with increasing model order, starting with a low-order model (Ljung, 1999). An alternative would be to find the model order by use of a subspace identification method (e.g. Di Ruscio, 1997), which would give an ARMAX type of model. However, this is not a feasible choice in cases with low primary output sampling rate.

#### Numerical algorithm

The practical identification can be performed by a modified version of the *pem* function in the System Identification Toolbox for use with Matlab, which uses an iterative Gauss-Newton algorithm (Ljung, 1995). The modifications are necessary in order to handle the multirate sampling case with only some of the  $y_{1,k}$  observations available.

## 6. Latent variables solutions

With a large number of secondary  $y_2$  variables, the number of parameters to identify will be correspondingly large. However, assuming collinear  $y_2$  variables the  $Y_2$  data may be compressed into latent variables by use of principal component analysis (PCA) or PLSR.

### PCA solution

The factorization (7) with  $\hat{W} = \hat{P}$  as defined by (12) results in the estimated latent variables

$$\hat{\epsilon}_k = \hat{P}^T y_{2,k} = \hat{P}^T C_2 x_k + \hat{P}^T D_2 u_k + \hat{P}^T v_{2,k}. \quad (29)$$

With  $\hat{\epsilon}_k$  as a substitute for  $y_{2,k}$ , the estimator (21) is replaced by

$$\begin{aligned} \hat{y}_{1,k|k}^{PCA+OE} = & C_1(I - K_P \hat{P}^T C_2)(qI - A + AK_P \hat{P}^T C_2)^{-1} \\ & \times [(B - AK_P \hat{P}^T D_2)u_k + AK_P \hat{\epsilon}_k] + C_1 K_P (\hat{\epsilon}_k - \hat{P}^T D_2 u_k) + D_1 u_k, \end{aligned} \quad (30)$$

where the Kalman gain is given by appropriately modified versions of equations (23) and (24), i.e.

$$K_P = P_P C_2^T \hat{P} (\hat{P}^T C_2 P_P C_2^T \hat{P} + \hat{P}^T R_{22} \hat{P})^{-1}, \quad (31)$$

with  $P_P$  determined by the algebraic Riccati equation

$$P_P = AP_P A^T + GR_v G^T - AP_P C_2^T \hat{P} (\hat{P}^T C_2 P_P C_2^T \hat{P} + \hat{P}^T R_{22} \hat{P})^{-1} \hat{P}^T C_2 P_P A^T. \quad (32)$$

The identification of this estimator is performed as indicated in Figure 1, only that the  $y_{2,k}$  data is replaced by  $\hat{\epsilon}_k$  as given by equation (29). For the special case of a first-order estimator, (26) is then modified into

$$\hat{y}_{1,k|k}^{PCA+OE} = [1 - (A - AK_P \hat{P}^T C_2)q^{-1}]^{-1} K_P \hat{\epsilon}_k. \quad (33)$$

### PLSR solution

As an alternative to the PCA + OE estimator (30) we may use a PLSR + OE estimator with

$$\hat{\epsilon}_k = \hat{W}_{PLS}^T y_{2,k}, \quad (34)$$

where  $\hat{W}_{PLS}$  is found by use of low sampling rate  $y_{2,k}$  as well as  $y_{1,k}$  data. The first-order estimator (33) is then replaced by

$$\hat{y}_{1,k|k}^{PLSR+OE} = [1 - (A - AK_{PLS} \hat{W}_{PLS}^T C_2)q^{-1}]^{-1} K_{PLS} \hat{\epsilon}_k. \quad (35)$$

## 7. Initial value determination

Identification of the OE estimator (16), the PCA + OE estimator (33) or the PLSR + OE estimator (35) requires useful initial parameter values. As already mentioned, the standard procedure based on a least squares solution (ARX model) is not feasible due to the lack of high sampling rate  $y_1$  information, and one has to resort to *ad hoc* methods (Ergon, 1999b). In the present case it is natural to look for a solution using a static PCR or PLSR estimator.



Due to the difficulties involved we focus on a simple although general case. A first-order state space realization

$$\begin{aligned}x_{k+1} &= f x_k + \mathbf{h}^T \hat{\epsilon}_k \\ y_{1,k} &= x_k + \mathbf{m}^T \hat{\epsilon}_k\end{aligned}\quad (36)$$

corresponding to the OE estimator

$$\hat{y}_{1,k|k} = \frac{\mathbf{m}^T + (\mathbf{h}^T - f \mathbf{m}^T) q^{-1}}{1 - f q^{-1}} \tau_k \quad (37)$$

may be chosen such that the structure of the estimators (33) and (35) is obtained. We may use this relation to find useful initial parameter values, and then let the values be free, or we may force the parameter values to comply to the relation during the entire minimization.

#### PCA + OE estimator

A comparison of (33) and (37) shows that  $f = A - A \mathbf{K}_p \hat{\mathbf{P}}^T \mathbf{C}_2$ ,  $\mathbf{m}^T = \mathbf{K}_p$  and  $\mathbf{h}^T = f \mathbf{m}^T$ , resulting in the static estimator (with  $q^{-1} = 1$ )

$$\hat{y}_{1,k|k} = \frac{\mathbf{m}^T}{1-f} \tau_k = \frac{\mathbf{m}^T}{1-f} \hat{\mathbf{P}}^T \mathbf{y}_{2,k} = \mathbf{y}_{2,k}^T \hat{\mathbf{P}} \frac{\mathbf{m}}{1-f} \quad (38)$$

It is natural to compare this with the PCR estimator according to (11) and (12)

$$\hat{\mathbf{b}}_{\text{PCR}}^T = \hat{\mathbf{P}} (\hat{\mathbf{P}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{Y}_{2,\text{low}} \hat{\mathbf{P}})^{-1} \hat{\mathbf{P}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{y}_1 = \hat{\mathbf{P}} \hat{\mathbf{b}}_{\text{T}}^{\text{PCR}}, \quad (39)$$

where  $\mathbf{Y}_{2,\text{low}}$  in the same way as  $\mathbf{y}_1$  contains only low sampling rate data. This comparison indicates the initial guesses

$$\mathbf{m}_0 = \pm (1 - f_0) \hat{\mathbf{b}}_{\text{T}}^{\text{PCR}}, \quad (40)$$

and thus

$$\mathbf{h}_0 = f_0 \mathbf{m}_0 = \pm f_0 (1 - f_0) \hat{\mathbf{b}}_{\text{T}}^{\text{PCR}}. \quad (41)$$

Here, the sign indeterminacy is due to the sign indeterminacy of the PCA involved. It now only remains to select  $f_0$  and the sign, and this can be done after a systematic search for good validation results.

#### PLSR + OE estimator

In the same way as above, a comparison of (35) and (37) results in the initial values

$$\mathbf{m}_0 = \pm (1 - f_0) \hat{\mathbf{b}}_{\text{T}}^{\text{PLSR}} \quad (42)$$

and

$$\mathbf{h}_0 = f_0 \mathbf{m}_0 = \pm f_0 (1 - f_0) \hat{\mathbf{b}}_{\text{T}}^{\text{PLSR}}, \quad (43)$$

where  $\hat{\mathbf{b}}_{\text{T}}^{\text{PLSR}} = (\hat{\mathbf{W}}_{\text{PLS}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{Y}_{2,\text{low}} \hat{\mathbf{W}}_{\text{PLS}})^{-1} \hat{\mathbf{W}}_{\text{PLS}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{y}_1$ .

## 8. Acoustic data example

### Introduction

Acoustic chemometrics is based on signals from an acoustic sensor (accelerometer) placed for example on, or slightly downstream of, a standard orifice plate. Observa-

tions of the power spectrum of the sensor signal are collected in the  $\mathbf{X} = \mathbf{Y}_2$  matrix, and calibrated against physical  $y = y_1$  primary quantities like multi-component mixture concentrations, density etc., using for example a standard PLSR method (Esbensen *et al.*, 1999).

### Experiment

In an experiment on a test rig at Telemark University College, the flow rate of ordinary drinking water was measured by use of an ultrasonic flowmeter (Fuji Portaflow X) and used as the response variable  $y_1$ , while the acoustic power spectral densities at 1024 frequencies originating from an accelerometer placed on a standard orifice plate were used as  $y_2$  variables.

The flow rate was varied by means of a control valve with a control signal generated as filtered white noise. The resulting  $y_{1,k}$  and  $y_{2,k}$  signals were recorded for 230 observations with a sampling interval  $T_2 = 5.3$  sec. No known inputs  $\mathbf{u}$  were recorded. Each power spectrum representing an  $y_2$  observation was formed as the mean value of 50 consecutive power spectra in the sampling interval. This standard procedure in static multivariate calibration based on acoustic data is necessary in order to obtain a reasonable noise level. Only every 5th  $y_1$  sample was used in the estimator identification procedure, and the number of  $y_1$  observations in the modeling set were accordingly  $N_1 = 46$ . We thus consider a multirate system with  $y_2$  mean sampling with intervals  $T_2 = 5.3$  sec, while  $y_1$  is sampled with intervals  $T_1 = 26.5$  sec. We also consider a difficult case with only  $N_1 = 46$  samples of  $y_1$ . The  $y_1$  signal in the modeling set is shown in Figure 2.

The estimator model was validated against a separate data set with 170  $y_1$  and  $y_2$  observations. The modeling and validation data sets were chosen in time intervals where no obvious outliers in the acoustic power spectrum were recorded.

**Remark 8.1** *The use of every 5th observation from high sampling rate data is of course not optimal. However, if the response variable had been for example a concentration of a certain component in a multi-component mixture, and the measurements had to be done through laboratory analyses, the response sampling rate would very likely for practical and economical reasons have been lower than the obtainable acoustic data sampling rate. Also the validation would in such a case have to be done against low sampling rate data.*

### Multirate identification

In order to test the method presented above, static PLSR and dynamic PCA + OE and PLSR + OE estimators were identified using the multirate modeling data described above.

Both an ordinary PLSR estimator

$$\hat{\mathbf{b}}_{\text{PLSR}} = \hat{\mathbf{W}}_{\text{PLS}} (\hat{\mathbf{W}}_{\text{PLS}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{Y}_{2,\text{low}} \hat{\mathbf{W}}_{\text{PLS}})^{-1} \hat{\mathbf{W}}_{\text{PLS}}^T \mathbf{Y}_{2,\text{low}}^T \mathbf{y}_1 = \hat{\mathbf{W}}_{\text{PLS}} \hat{\mathbf{b}}_{\text{PLSR}}^{\text{PLSR}}, \quad (44)$$

and first-order PCA + OE and PLSR + OE estimators according to (36) with  $\tau$  given by (29) and (34) were determined, using only the  $N_1 = 46$  response observations considered available. Here,  $\mathbf{Y}_{2,\text{low}}$  contains only low sampling rate  $y_{2,k}$  data, while the PCA + OE and PLSR + OE dynamic estimators were identified by use of the high sampling rate  $y_{2,k}$  data. The best PCA + OE and PLSR results were obtained when

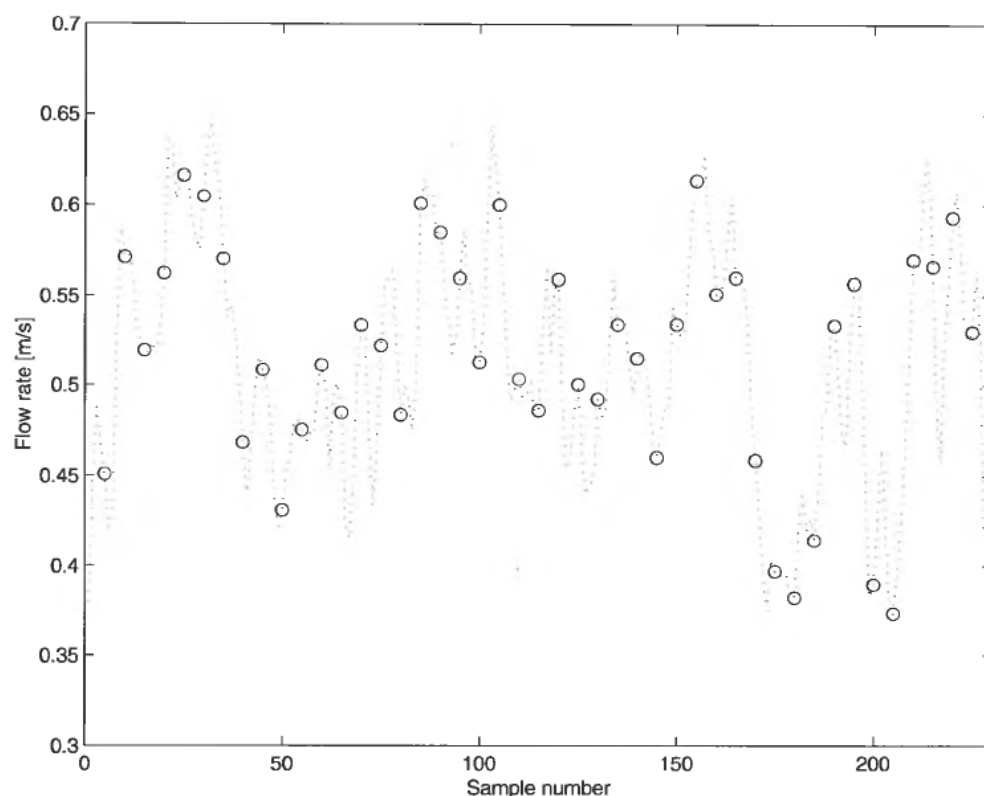


Figure 2. Measured flow rate (dotted) in the modeling set, with assumed low sampling rate measurements (o-marked).

the model structures (33) and (35) were used only for the initial parameter values, i.e. when the parameters in (36) were left to move freely during the minimization.

#### Number of parameters

The number of independent parameters in the PLSR estimator is equal to the number of components  $a$  chosen (the number of latent variables), while the PCA + OE and PLSR + OE estimators has  $n + (n + 1)a$  parameters, where  $n$  is the model order (i.e. 5 parameters for  $a = 2$  and  $n = 1$ ).

#### Results

The best static PLSR estimator was obtained using the means of the last two  $y_2$  observations, i.e. by computing the mean of 100 power spectra as  $y_{2,k} \leftarrow (y_{2,k} + y_{2,k-1})/2$ . The best PCA + OE and PLSR + OE estimators were obtained using the original means of 50 power spectra.

The identified estimators were validated by use of the independent validation data set and computation of

$$RMSEVAL = \sqrt{\frac{1}{N_2} \sum_{k=1}^{N_2} (y_{1,k} - \hat{y}_{1,k})^2}, \quad (45)$$

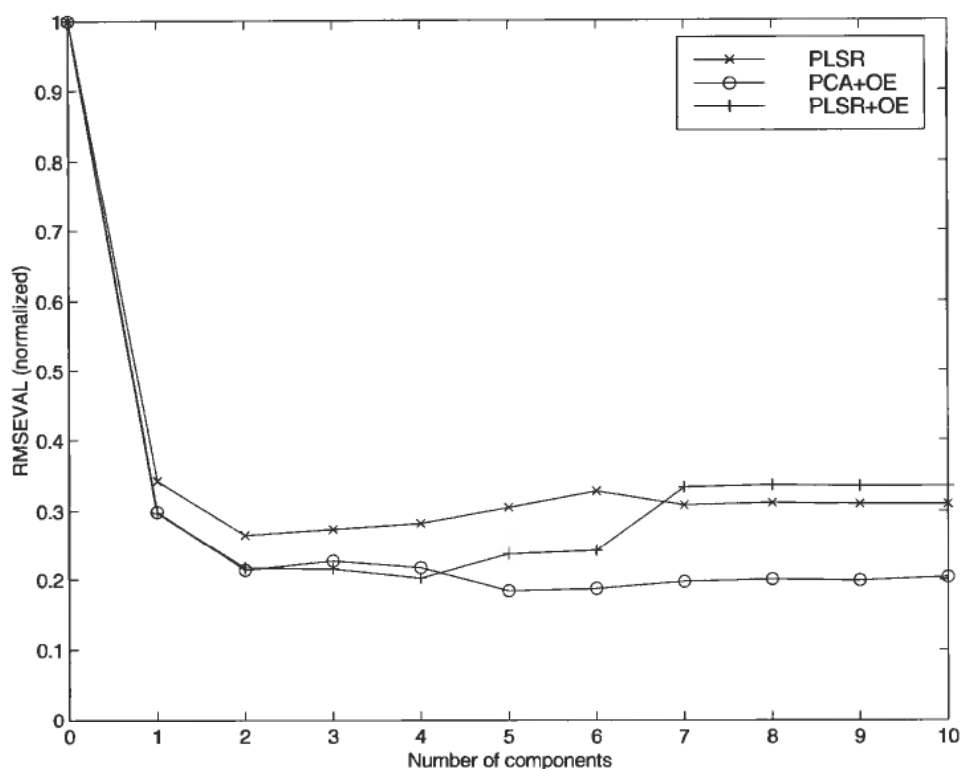


Figure 3. Validation RMSEVAL results for static PLSR and dynamic PCA+OE and PLSR+OE estimators based on low sampling rate  $y_1$  data.

i.e. all available validation data was used, including all validation data  $y_{1,k}$  observations.

The models for different numbers of components found in this way gave the validation RMSEVAL results shown in Figure 3.

As can be seen in Figure 3, the PCA+OE estimator gave an approximate 30% reduction of the validation RMSEVAL value at the optimal number of components  $a=5$ , as compared with the PLSR result at  $a=2$ . However, in a practical application we would most likely be content with 20% reduction, as given by the parsimonious PCA+OE and PLSR+OE estimators using only  $a=2$  components. Note that these estimators gave very similar results for  $a \leq 4$ .

The PCA+OE validation response for  $a=2$  components is shown in Figure 4. Note that the estimated response follows the (in this case) measured response well also between the sampled values. This is also the case for the low sampling rate PLSR estimator, although the dynamic estimator gives a considerably lower validation RMSEVAL value.

#### Further results and discussion

The fact that we need two PCA or PLSR components to explain the main part of the variance in  $y_1$  indicates that there is a second and independent source of the power spectrum variations in addition to the varying flow rate. This is most likely a temperature drift, or possibly air bubbles or particles in the water. Note that we use centered

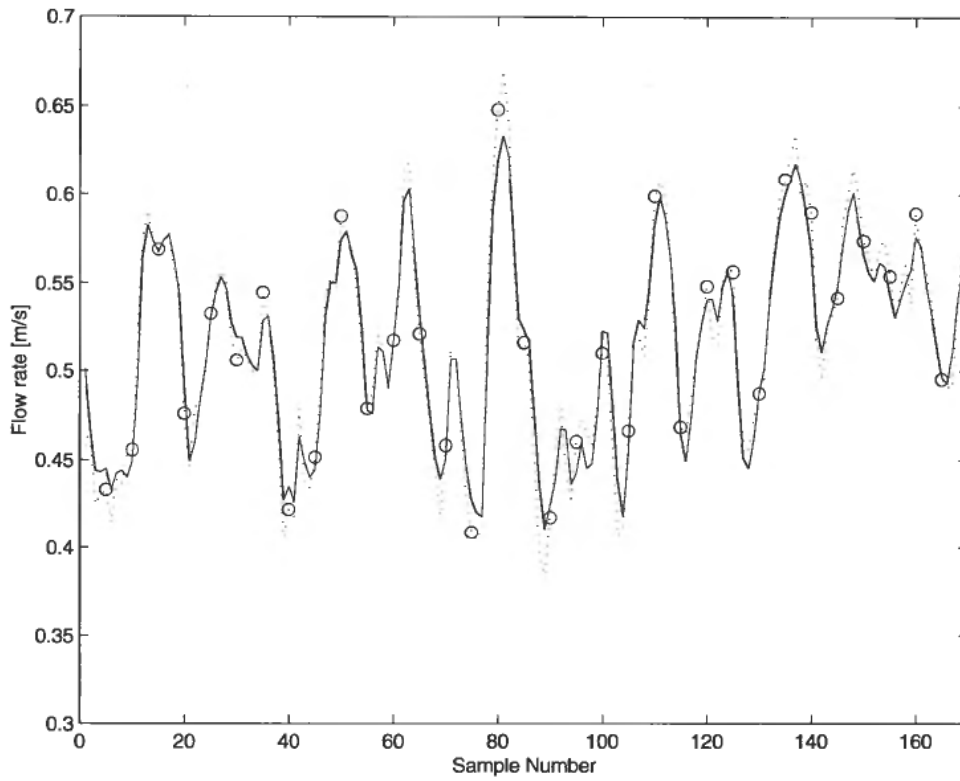


Figure 4. Validation response for dynamic PCA + OE estimator based on  $a=2$  PCA components and low sampling rate  $y_1$  data. The low sampling rate validation data is o-marked, while the (in this case) known intermediate validation values are shown by dotted line. The estimated values are shown by solid line.

data, with flow rate variations around a fixed mean value, i.e. power spectrum variations caused by such a second source may be present also without flow variations.

Assuming first-order dynamics for the two major and independent acoustic sources, the model (20) will in such a case result in

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{k+1} &= \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_k + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_k \\ y_{1,k} &= [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_k + v_{1,k} \\ y_{2,k} &= [C_{21} \quad C_{22}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_k + v_{2,k}, \end{aligned} \quad (46)$$

where a temperature drift can be modeled as integrated white noise by setting  $a_{22} = 1$ . After the data compression

$$\tau_k = \hat{W}^T y_{2,k} = \hat{W}^T [C_{21} \quad C_{22}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_k + \hat{W}^T v_{2,k}, \quad (47)$$

i.e. by using

$$\hat{\epsilon}_k = \hat{W}^T y_{2,k} = \hat{W}^T [C_{21} \quad C_{22}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_k \quad (48)$$

with  $\hat{W} = \hat{P}$  or  $\hat{W} = \hat{W}_{PLS}$ , this results in a theoretical primary response estimate on the form

$$\hat{y}_{1,k|k} = \frac{[k_{11} \quad k_{12}] + [b_{11} \quad b_{12}]q^{-1}}{1 + a_1q^{-1} + a_2q^{-2}} \begin{bmatrix} \hat{\epsilon}_{1,k} \\ \hat{\epsilon}_{2,k} \end{bmatrix}, \quad (49)$$

where  $k_{11}$  and  $k_{12}$  are Kalman gains. In the case that the second acoustic source does not result in a time series, i.e. when  $a_{22} = 0$ , this is simplified to the form given by equation (26)

$$\hat{y}_{1,k|k} = \frac{[k_{11} \quad k_{12}]}{1 + a_1q^{-1}} \begin{bmatrix} \hat{\epsilon}_{1,k} \\ \hat{\epsilon}_{2,k} \end{bmatrix}. \quad (50)$$

The theoretical estimators (49) and (50) give a background for a discussion of the identified dynamic estimators, although this is somewhat complicated by the fact that the  $y_2$  observations were obtained as mean values of various numbers of consecutive power spectra. The dynamic PLSR + OE estimator obtained with  $a = 2$  components and based on 50 consecutive power spectra was

$$\hat{y}_{1,k|k} = \frac{[0.0652 \quad 0.1315] + [0.2345 \quad 0.2228]q^{-1}}{1 - 0.0837q^{-1}} \begin{bmatrix} \hat{\epsilon}_{1,k} \\ \hat{\epsilon}_{2,k} \end{bmatrix} \cdot 10^{-3}, \quad (51)$$

which resulted in 20% reduction of the RMSEVAL value as compared with the static PLSR estimator based on the means of 100 consecutive power spectra (see Figure 3). For this estimator the initial values were chosen according to equations (40) and (41) with  $f_0 = 0.5$  (the  $f_0$  value was in no way critical), but the parameters were then allowed to move freely during the minimization. A comparison with the model in equation (49) shows that a  $q^{-2}$  term is missing in the denominator, i.e. we should have identified a second-order estimator. The reason for not doing so is the initial value problem discussed in Section 7.

When the parameter values in the PLSR + OE estimator were forced to comply with the theoretical estimators (35) and (50) also during the minimization the result was

$$\hat{y}_{1,k|k} = \frac{[0.1895 \quad 0.2018]}{1 - 0.4713q^{-1}} \begin{bmatrix} \hat{\epsilon}_{1,k} \\ \hat{\epsilon}_{2,k} \end{bmatrix} \cdot 10^{-3}. \quad (52)$$

However, this estimator gave no improvement in the RMSEVAL value as compared to the static PLSR estimator, which indicates that the second acoustic source gave time series data, i.e. that  $a_{22} \neq 0$ .

The dynamic estimators may also be compared to the static estimators. With  $q^{-1} = 1$  the PLSR + OE estimator (51) gives the corresponding static estimator

$$\hat{y}_1 = (0.3271\hat{\epsilon}_1 + 0.3867\hat{\epsilon}_2) \cdot 10^{-3}, \quad (53)$$

while the static PLSR estimator (based on the means of 100 consecutive power spectra) for  $a = 2$  components gave

$$\hat{y}_1 = (0.3684\hat{\epsilon}_1 + 0.3009\hat{\epsilon}_2) \cdot 10^{-3}. \quad (54)$$

Let us finally find the spectral properties of the underlying state variables  $x_1$  and  $x_2$  in equation (46). Based on collected data the second output equation in (46) gives

$$\mathbf{Y}_2 = [\mathbf{x}_1 \quad \mathbf{x}_2] \begin{bmatrix} \mathbf{C}_{21}^T \\ \mathbf{C}_{22}^T \end{bmatrix} + \mathbf{V}_2, \quad (55)$$

where  $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2]$  is orthogonal. A PLSR factorization using the Martens algorithm would result in  $\hat{\mathbf{Y}}_2 = \hat{\mathbf{T}}\hat{\mathbf{W}}_{\text{PLS}}^T$ , where the  $\hat{\mathbf{T}}$  matrix is non-orthogonal, and it is therefore more appropriate to use the orthogonal score matrix  $\hat{\mathbf{T}}_w$  in the Wold algorithm. Following Helland (1988) we have  $\mathbf{Y}_2\hat{\mathbf{W}}_{\text{PLS}} = \hat{\mathbf{T}}_w\hat{\mathbf{P}}_w^T\hat{\mathbf{W}}_{\text{PLS}}$ , where  $\hat{\mathbf{P}}_w$  is the Wold algorithm loading matrix. The corresponding expression (8) for the Martins algorithm is  $\mathbf{Y}_2\hat{\mathbf{W}}_{\text{PLS}} = \hat{\mathbf{T}}$ , and we may thus perform the factorization

$$\hat{\mathbf{Y}}_2 = \hat{\mathbf{T}}\hat{\mathbf{W}}_{\text{PLS}}^T = \hat{\mathbf{T}}_w\hat{\mathbf{P}}_w^T\hat{\mathbf{W}}_{\text{PLS}}\hat{\mathbf{W}}_{\text{PLS}}^T = \hat{\mathbf{T}}_w\hat{\mathbf{S}}^T = \hat{\mathbf{T}}_w \begin{bmatrix} \hat{\mathbf{s}}_1^T \\ \hat{\mathbf{s}}_2^T \end{bmatrix},$$

where  $\hat{\mathbf{s}}_1$  and  $\hat{\mathbf{s}}_2$  are scaled estimates of the spectra  $\mathbf{C}_{21}$  and  $\mathbf{C}_{22}$  in (55). The estimated spectra for the PLSR estimator (54) are plotted in Figure 5. Here, spectrum 1 shows the relative weight of the different frequency components in the power spectrum

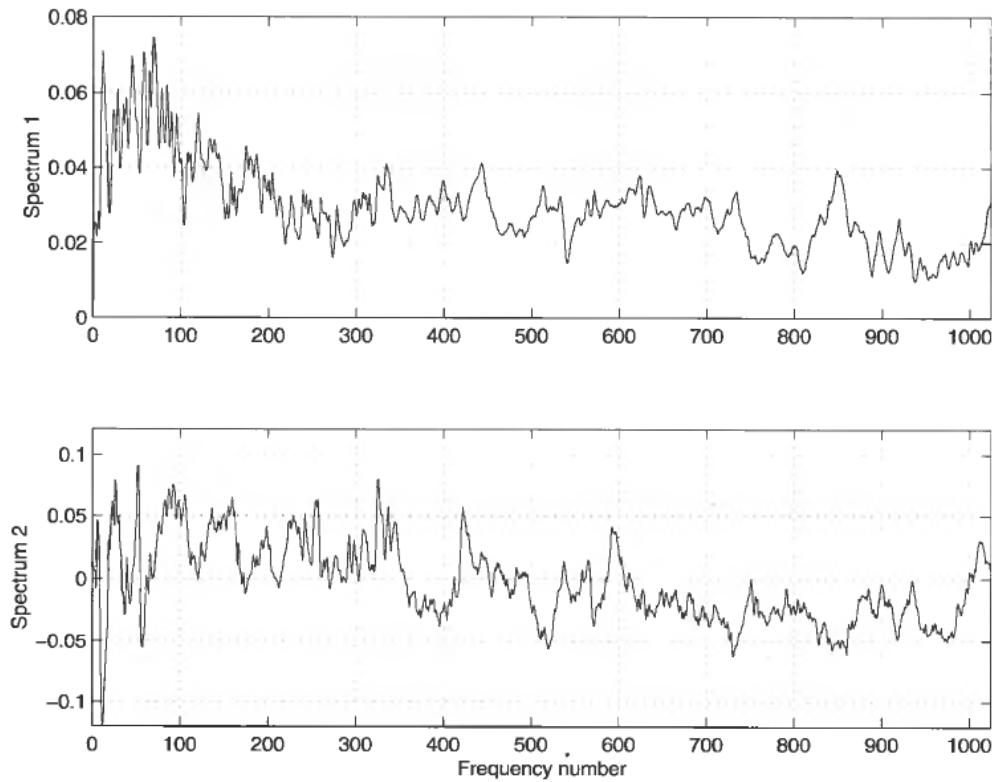


Figure 5. Spectra for PLSR estimator with two components.

variations caused by flow variations, while spectrum 2 reflects the influence of the second underlying source of variation. Since  $\hat{\mathbf{T}} = \mathbf{Y}_2 \hat{\mathbf{W}}_{\text{PLS}}$  is nearly orthogonal, these spectra are almost identical to the loading weight spectra based on  $\mathbf{W}_{\text{PLS}} = [\mathbf{w}_1 \quad \mathbf{w}_2]$ .

## 9. Conclusion

High sampling rate  $\mathbf{X}$  and low sampling rate  $\mathbf{y}$  time series data may be used to identify an output error (OE) response estimator based on an underlying Kalman filter. The  $\mathbf{X}$  variables may be both known system inputs  $\mathbf{u}_k$  and secondary system outputs  $\mathbf{y}_{2,k}$ , while the  $\mathbf{y}$  variables are the primary responses  $\mathbf{y}_{1,j}$ .

A high number of collinear  $\mathbf{X} = [\mathbf{U} \quad \mathbf{Y}_2]$  variables may be compressed by principal component analysis (PCA) or partial least squares regression (PLSR), resulting in PCA + OE and PLSR + OE methods.

A multirate sampling example with acoustic power spectrum  $\mathbf{Y}_2$  data and flow meter  $\mathbf{y}_1$  data demonstrates the feasibility of these methods, although the interpretation of the results is complicated by the fact that the  $\mathbf{Y}_2$  data are obtained as means of several consecutive power spectra.

## REFERENCES

- BELSLEY, A. D. (1991). *Conditioning Diagnosis: Collinearity and Weak Data in Regression*, John Wiley & Sons.
- DI RUSCIO, D. (1997). A method for identification of combined deterministic-stochastic systems, *Applications of Computer Aided Time Series Modeling*, M. Aoki and A. M. Havenner, Eds., Springer-Verlag, New York.
- ERGON, R. (1998). Dynamic system calibration: The low primary output sampling rate case. *Modeling, Identification and Control*, **19**:99–107.
- ERGON, R. (1999a). On primary output estimation by use of secondary measurements as input signals in system identification. *IEEE Trans. Autom. Control*, **44**:821–825.
- ERGON, R. (1999b). *Dynamic System Multivariate Calibration for Optimal Primary Output Estimation*. PhD thesis at the Norwegian University of Science and Technology, Trondheim, Norway.
- ERGON, R. and HALSTENSEN, M. (2000). Dynamic system multivariate calibration with low-sampling-rate  $\mathbf{y}$  data, *Journal of Chemometrics*, **14**:617–628.
- ERGON, R. and ESBENSEN, K. (2001). A didactically motivated PLS prediction algorithm, submitted to *Modeling, Identification and Control*, Jan. 2001.
- ESBENSEN, K., HOPE, B., LIED, T. T., HALSTENSEN, M., GRAVERMOEN, T. and SUNDBERG, K. (1999). Acoustic chemometrics for fluid flow quantifications-II: A small constriction will go a long way. *Journal of Chemometrics*, **13**:209–236.
- ESBENSEN, K. (2000). *Multivariate Data Analysis—in practice*, Camo ASA, Trondheim, Norway.
- 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**:581–607.
- LJUNG, L. (1999). *System Identification: Theory for the user*, Prentice Hall: New Jersey.
- LJUNG, L. (1995). *System Identification Toolbox for use with Matlab*, The MathWorks Inc., Mass.
- MARTENS, H. and NÆS, T. (1989). *Multivariate Calibration*, Wiley: Chichester, UK.
- MEJDELL, T. and SKOGSTAD, S. (1989). Estimate of process output from multiple secondary measurements, *Proc. American Control Conference*, 2112–2121.
- NORRIS, K. H. (1993). Extracting information from spectrophotometric curves. Predicting chemical composition from visible and near-infrared spectra, *Proc. IUFost Symp. Food Research and Data Analysis*, Sept. 1982, Oslo, Norway (Martens and Russworm, eds.), Applied Science Publ., 95–113.
- SÖDERSTRÖM, T. and STOICA, P. (1989). *System Identification*, Prentice Hall: Cambridge, UK.