

Dynamic system multivariate calibration by system identification methods

ROLF ERGON†

Keywords: *Estimation, system identification, multivariate calibration*

In the first part of the paper, the optimal estimator for normally nonmeasured primary outputs from a linear and time invariant dynamic system is developed. The estimator is based on an underlying Kalman filter, utilizing all available information in known inputs and measured secondary outputs. Assuming sufficient experimental data, the optimal estimator can be identified by specifying an output error model in a standard prediction error identification method. It is further shown that static estimators found by the ordinary least squares method or multivariate calibration by means of principal component regression (PCR) or partial least squares regression (PLSR) can be seen as special cases of the optimal dynamic estimator. Finally, it is shown that dynamic system PCR and PLSR solutions can be developed as special cases of the general estimator for dynamic systems.

1. Introduction

The first aim of the paper is to develop the theoretical basis for identification of the optimal estimator for normally nonmeasured primary outputs y_1 from a linear, time invariant and dynamic system, utilizing all available information in known inputs u and measured secondary outputs y_2 . The basic insight behind this is that the y_2 measurements may carry valuable information about the process noise v , which it should be possible to utilize when estimating y_1 , as illustrated in Fig. 1. The problem is then to find a method for identification of the optimal estimator model, and the practical use may be operator support, failure detection and feedback control.

Although the primary outputs y_1 are not measured directly or measured only at a slow sampling rate, information about y_1 may to a large extent be contained in the y_2 measurements. The plant itself may thus act as an instrument for measuring y_1 , and the process of finding the relation between y_2 and y_1 can therefore be seen as a form of calibration. Since the plant generally is dynamic in nature, and since the available data may be highly multivariate, it seems appropriate to use the terminology *Dynamic system multivariate calibration*. It should be emphasized, though, that the optimal estimator is found by use of well established system identification methods. As shown in the paper, this is done by specifying an output error (OE) model with u and y_2 used as inputs.

The second aim is to study the relations between the optimal estimator for dynamic systems and the static estimators obtained from linear regression and the multivariate calibration methods widely used in e.g. chemometrics. In the static case we will find models for estimation of primary dependent variables y_1 from independent variables u and secondary dependent variables y_2 . *The final aim* is to indicate how these relations can be used to develop multivariate calibration methods that can also handle collinear time series data generated by dynamic systems.

Received 28 January 1997.

†Telemark Institute of Technology, N-3914 Porsgrunn, Norway. E-mail: Rolf.Ergon@hit.no.

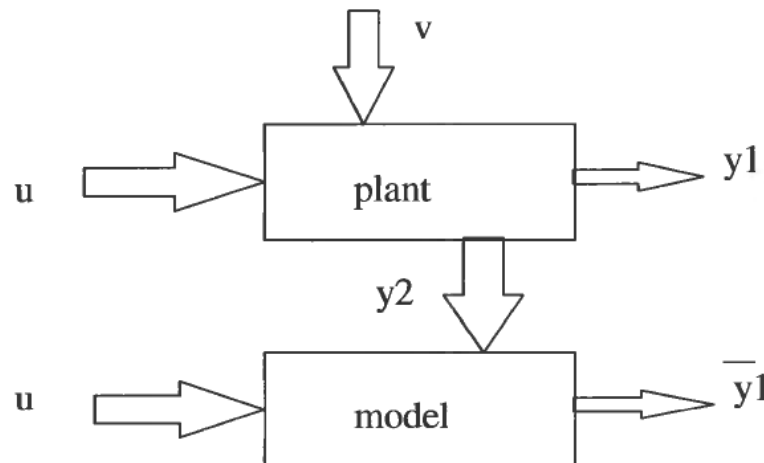


Figure 1. Basic principle for estimation of primary system outputs y_1 from known inputs u and measured secondary outputs y_2 in presence of process noise v .

The use of both independent inputs u and dependent measurements y_2 as inputs in a system identification procedure raises questions about identifiability and applications on deterministic and perfect measurement systems. A preliminary discussion of this is given in Ergon and Di Ruscio (1997). A comparison of ARMAX and OE models for prediction of y_1 based on u and y_2 is given in Ergon (1998a).

The present paper is organized in the following way: Section 2 gives some background and preliminaries. In section 3 the theory for use of secondary y_2 measurements as inputs to system identification procedures is established. This results in the central relations in the paper, giving the optimal y_1 prediction (a priori) and current (a posteriori) estimators (17) and (23) for a class of linear, time invariant and stable dynamic systems. It is further shown that identification by use of a standard prediction error method results in unbiased estimators. Asymptotic estimation covariance results are also given. In section 4 it is shown that least square estimation (LSE), principal component regression (PCR) and partial least squares regression (PLSR) give estimators that can be seen as special static cases of the optimal current estimator for dynamic systems. The relations between these data based estimators and theoretical estimators based on known or assumed static models and noise properties are also presented, and the relation between two different PLSR algorithms falls out as a neat result. Extensions of the PCR and PLSR methods to cover also dynamic systems with collinear measurements are discussed in section 5. Section 6 gives some numerical examples and Monte Carlo simulations, while conclusions and some remarks on further research are given in section 7.

2. Background and preliminaries

Linear regression and static calibration methods have roots in the classical least squares technique used by Gauss already in 1795 (Grewal and Andrews 1993), while the name regression goes back to an anthropological paper by Galton in 1885 (Johnson and Wichern 1992). The linear regression model arises in two different settings, that are both special cases of the model used in Chapter 4 of the paper. The classical model is concerned with the association between a random variable whose mean depends on

a collection of fixed and known predictor variables (known inputs u), while in the other setting all variables are random (secondary measurements y_2).

When the number of estimator variables is large and the number of observations is limited, the ordinary solution to the least squares problem gives very large prediction covariance due to overfitting. This situation requires some form of regularization, e.g. PCR or PLSR (Martens and Næs 1989). In many cases of great practical interest, the estimator variables far outnumber the observations at hand. An example is product quality characterization by use of near infrared spectroscopy, with several thousand estimator variables (frequencies) and often less than one hundred observations. In such cases, the estimator variables are often strongly collinear, and most of the information can then be compressed into a few latent variables within a subspace of the variable space. Basic tools for this data compression are singular value decomposition (SVD) and principal component analysis (PCA), and the regression method directly based on this is PCR, while PLSR combines data compression and regression in an iterative approach. These tools for static multivariate data analysis are used in many scientific fields like biometrics, chemometrics, econometrics and psychometrics. An industrial example is given in Mejdell and Andersson (1994).

In parallel with the development of the PCR and PLSR methods, the field of dynamic system identification (SI) has been developed into a sophisticated set of methods and practical tools. The field of classical SI is summarized in comprehensive books, e.g. Ljung (1987) and Söderström and Stoica (1989). At present, subspace identification methods attract a great deal of interest, see e.g. Di Ruscio (1996) with further references. In all forms of SI, one finds that LSE is used as a basic tool. It is, however, refined and in some cases replaced by e.g. prediction optimization methods in order to account for the noise influence in a proper way.

System identification is also closely linked to the Kalman filtering theory (Grewal and Andrews 1993). This is done by use of innovation models, where the different process and measurement noise sources are replaced by the white noise innovations in an underlying Kalman filter.

From a system identification and Kalman filtering point of view, it is intuitively evident that the classical linear regression and the modern multivariate calibration methods may be seen as special static cases of the more general parametric SI methods for dynamic systems. An early attempt to look into these similarities was made in Berntsen (1988), and the present paper includes a further and more detailed attempt to do so (further developed in Ergon (1998b)). When these similarities are to be investigated, three basic facts have to be acknowledged:

1. Methods of multivariate calibration are used to find models for estimation of unknown output variables y from both independent and dependent known variables x . In SI terminology this means methods for estimation of unknown systems outputs y_1 from both independent systems inputs u and dependent system outputs y_2 . The basic observation here is that also dependent outputs y_2 have to be used as inputs in the SI procedure.
2. When the multivariate calibration models are used for estimation, the y_1 outputs are not known, and this will also be the case for the corresponding dynamic system models found by SI. We are therefore lead to consider output error (OE) models and not the qualitatively different ARMAX type of models used for e.g. control design based on known y_1 outputs.
3. In order to find the optimal y_1 current estimate, the underlying Kalman filter must

be of the predictor-corrector form, which is normally not the case when innovation models are used in system identification.

These basic facts must be reflected in the theoretical analysis of the relations between SI and LSE, PCR and PLSR, and this is quite independent of the specific SI methods considered.

3. Secondary measurements as inputs in system identification

3.1. Statement of problem

Consider the discrete system model

$$x_{k+1} = Ax_k + Bu_k + Gv_k \quad (1a)$$

$$y_{1,k} = C_1x_k + D_1u_k + w_{1,k} \quad (1b)$$

$$y_{2,k} = C_2x_k + D_2u_k + w_{2,k} \quad (1c)$$

where x is the state vector, while v and $w = [w_1^T \ w_2^T]^T$ are white and independent process and measurement noise vectors with covariances

$$R_v = Evv^T \text{ and } R_w = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} Ew_1w_1^T & Ew_1w_2^T \\ Ew_2w_1^T & Ew_2w_2^T \end{bmatrix}.$$

Also assume a stable system with (C_2, A) observable and $(A, G\sqrt{R_v})$ controllable. The assumptions of noise independence and state observability may be relaxed with appropriate theoretical modifications, which, however, is beyond the scope of the present paper. Note that some or all of the secondary y_2 measurements may be collinear with some or all of the primary y_1 measurements.

Further assume that input-output data are available from an informative experiment (e.g. Goodwin and Payne 1977), i.e. that data records for u_k , $y_{1,k}$ and $y_{2,k}$ for $k = 1, 2, \dots, N$ are at hand, with u_k persistently exciting of appropriate order and a sufficiently high number of samples. The problem is now to identify the optimal one-step-ahead (a priori) $\bar{y}_{1,k|k-1}$ predictor based on past and present u_k and past $y_{2,k}$ values, and the optimal $\bar{y}_{1,k|k}$ current (a posteriori) estimator based also on present $y_{2,k}$ values.

Note that it is a part of the problem that past values of y_1 is not available as a basis for the estimate of the present value of y_1 . This is a common situation in industrial applications, e.g. in polymer extruding, where product quality measurements involve costly laboratory analyses. Product samples are then collected at a rather low sampling rate, and product quality estimates at a higher rate may thus be valuable.

3.2. Optimal one-step-ahead predictor when y_1 is available

The model (1) can be expressed in the ordinary innovation form (Ljung 1987) given by the following equations, where $AK = A[K_1 \ K_2]$ is the gain in a predictor type Kalman filter formulation with white innovations e_1 and e_2 :

$$\bar{x}_{k+1} = A\bar{x}_k + Bu_k + AK \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}_k \quad (2a)$$

$$y_{1,k} = C_1\bar{x}_k + D_1u_k + e_{1,k} \quad (2b)$$

$$y_{2,k} = C_2\bar{x}_k + D_2u_k + e_{2,k}. \quad (2c)$$

The optimal one-step-ahead y_1 predictor with all measurements available and a known u_k will then be

$$\bar{x}_{k+1} = A(I - K_1 C_1 - K_2 C_2) \bar{x}_k \quad (3a)$$

$$+ (B - AK_1 D_1 - AK_2 D_2) u_k + AK_1 y_{1,k} + AK_2 y_{2,k} \quad (3b)$$

$$\bar{y}_{1,k|k-1} = C_1 \bar{x}_k + D_1 u_k. \quad (3c)$$

This will be the best linear one-step-ahead predictor if x_0 , v_k and w_k have arbitrary statistics, and the optimal predictor assuming that x_0 , v_k and w_k have Gaussian statistics (Lewis 1986). This is also the predictor normally used in predictor error identification methods (Söderstrom and Stoica 1989).

3.3. Optimal one-step-ahead predictor when y_1 is not available

As the predictor (3) is based on an underlying Kalman filter driven by both y_1 and y_2 , it will no longer be optimal when past y_1 values are not available, e.g. when used with $y_{1,k} = 0$. In a predictor error identification method, we must instead base the prediction on an underlying Kalman filter driven by u and only the y_2 measurements. With the assumption that (C_2, A) is observable, the following innovation form can then be derived from (1):

$$\bar{x}_{k+1}^{OEP} = A \bar{x}_k^{OEP} + B u_k + A K_2^{OE} e_{2,k} \quad (4a)$$

$$y_{2,k} = C_2 \bar{x}_k^{OEP} + D_2 u_k + e_{2,k}. \quad (4b)$$

Here, K_2^{OE} is determined by (Lewis 1986)

$$K_2^{OE} = P^{OEP} C_2^T (C_2 P^{OEP} C_2^T + R_{22})^{-1}, \quad (5)$$

where $P^{OEP} = E(x_k - \bar{x}_k^{OEP})(x_k - \bar{x}_k^{OEP})^T$ is determined by the Riccati equation

$$P^{OEP} = A P^{OEP} A^T + G R_v G^T - A P^{OEP} C_2^T (C_2 P^{OEP} C_2^T + R_{22})^{-1} C_2 P^{OEP} A^T. \quad (6)$$

The y_1 output is then given by

$$y_{1,k} = C_1 \bar{x}_k^{OEP} + D_1 u_k + \vartheta_k, \quad (7)$$

where

$$\vartheta_k = C_1(x_k - \bar{x}_k^{OEP}) + w_{1,k} \quad (8)$$

is colored noise.

Theoretically, it is possible to identify the system determined by (4) and (7) using y_1 and y_2 as outputs, i.e. to identify (2) with a simplified noise model employing $K_1 = 0$. With many secondary y_2 measurements it is, however, a simpler task to use y_2 as an input signal, and identify the output error prediction model (OEP model)

$$\bar{x}_{k+1}^{OEP} = A(I - K_2^{OE} C_2) \bar{x}_k^{OEP} + (B - AK_2^{OE} D_2) u_k + AK_2^{OE} y_{2,k} \quad (9a)$$

$$y_{1,k} = C_1 \bar{x}_k^{OEP} + D_1 u_k + \vartheta_k. \quad (9b)$$

The corresponding input-output model is then

$$y_{1,k} = G_1(q^{-1}; \theta) u_k + G_2(q^{-1}; \theta) y_{2,k} + \vartheta_k, \quad (10)$$

where

$$G_1(q^{-1}; \theta) = C_1(qI - A + AK_2^{OE}C_2)^{-1}(B - AK_2^{OE}D_2) + D_1 \quad (11)$$

and

$$G_2(q^{-1}; \theta) = C_1(qI - A + AK_2^{OE}C_2)^{-1}AK_2^{OE}. \quad (12)$$

Here, θ is the true parameter vector, while q^{-1} is the unit time delay operator.

In order to identify the deterministic part of the system (10), i.e. $G_1(q^{-1}; \theta)$ and $G_2(q^{-1}; \theta)$, we model ϑ_k by some unknown white noise sequence and use the prediction

$$\hat{y}_{1,k|k-1} = G_1(q^{-1}; \hat{\theta})u_k + G_2(q^{-1}; \hat{\theta})y_{2,k}, \quad (13)$$

where $\hat{\theta}$ is the parameter vector used in the prediction model. The prediction error $\varepsilon_{1,k} = y_{1,k} - \hat{y}_{1,k|k-1}$ is then

$$\begin{aligned} \varepsilon_{1,k} &= [G_1(q^{-1}; \theta) - G_1(q^{-1}; \hat{\theta})]u_k \\ &\quad + [G_2(q^{-1}; \theta) - G_2(q^{-1}; \hat{\theta})]y_{2,k} + \vartheta_k. \end{aligned} \quad (14)$$

When evaluating the result of minimizing a scalar criterion function (Söderström and Stoica 1989)

$$V_N(\hat{\theta}) = \det \left[\frac{1}{N} \sum_{k=1}^N \varepsilon_{1,k} \varepsilon_{1,k}^T \right], \quad (15)$$

we must now consider the fact that $y_{2,k}$ and ϑ_k are not independent. Note, however, that when $G_1(q^{-1}; \hat{\theta}) \equiv G_1(q^{-1}; \theta)$ and $G_2(q^{-1}; \hat{\theta}) \equiv G_2(q^{-1}; \theta)$, we will from (8) simultaneously obtain

$$E\varepsilon_{1,k} \varepsilon_{1,k}^T = E\vartheta_k \vartheta_k^T = C_1 P^{OEP} C_1^T + R_{11}. \quad (16)$$

Since P^{OEP} is the minimized prediction state estimation covariance given the y_2 measurements, this represents a true minimum. From (10) to (12) we thus obtain the optimal one-step-ahead prediction estimator

$$\bar{y}_{1,k|k-1} = C_1(qI - A + AK_2^{OE}C_2)^{-1}[(B - AK_2^{OE}D_2)u_k + AK_2^{OE}y_{2,k}] + D_1u_k. \quad (17)$$

3.4. Optimal current estimator when y_1 is not available

In order to utilize also current y_2 values, the optimal estimator must be based on an underlying predictor-corrector Kalman filter (Lewis 1986). This results in the following output error model (OEC model):

$$\begin{aligned} y_{1,k} &= C_1(I - K_2^{OE}C_2)(qI - A + AK_2^{OE}C_2)^{-1}[(B - AK_2^{OE}D_2)u_k + AK_2^{OE}y_{2,k}] \\ &\quad + C_1K_2^{OE}(y_{2,k} - D_2u_k) + D_1u_k + \psi_k. \end{aligned} \quad (18)$$

Here we introduce the colored noise

$$\psi_k = C_1(x_k - \bar{x}_k^{OEC}) + w_{1,k}, \quad (19)$$

based on

$$\bar{x}_k^{OEC} = (I - K_2^{OE}C_2)\bar{x}_k^{OEP} + K_2^{OE}(y_{2,k} - D_2u_k). \quad (20)$$

Minimization of the criterion function (15) will now result in an optimal estimator only if

$$E\psi_k\psi_k^T = C_1 P^{OEC} C_1^T + R_{11} - C_1 K_2^{OE} R_{21} - R_{12} (K_2^{OE})^T C_1^T, \quad (21)$$

with $P^{OEC} = E(x_k - \bar{x}_k^{OEC})(x_k - \bar{x}_k^{OEC})^T$ given by

$$P^{OEC} = (I - K_2^{OE} C_2) P^{OEP} (I - K_2^{OE} C_2)^T + K_2^{OE} R_{22} (K_2^{OE})^T, \quad (22)$$

simultaneously is at a minimum. Since P^{OEC} is the minimized current state estimation covariance, this is true only when $R_{12} = R_{21}^T = 0$, and an optimal current estimator is therefore obtained only in this special case. It will then become

$$\begin{aligned} \bar{y}_{1,k|k} &= C_1 (I - K_2^{OE} C_2) (qI - A + AK_2^{OE} C_2)^{-1} [(B - AK_2^{OE} D_2)u_k + AK_2^{OE} y_{2,k}] \\ &\quad + C_1 K_2^{OE} (y_{2,k} - D_2 u_k) + D_1 u_k. \end{aligned} \quad (23)$$

If $R_{12} \neq 0$, the common part of the w_1 and w_2 noise may be modeled as delayed white noise, which requires an extension of the state vector. A special example of that is given in section 4.2 below. It is generally doubtful, however, if an OEC model with an extended state vector in a practical case will give better identification results than an OEP model.

3.5. Theoretical y_1 estimation covariance

When the OEP model (10) or the OEC model (18) are identified using a large data set, i.e. when $N \rightarrow \infty$, the estimates $\bar{y}_{1,k|k-1}$ and $\bar{y}_{1,k|k}$ will be asymptotically unbiased when we use either only u or both u and y_2 as input signals. The asymptotic estimation covariance, however, will depend on the model and the quality of the data. In the following we assume perfect model and noise information, and derive theoretical asymptotic expressions for the prediction and current y_1 estimation covariances.

The underlying Kalman filter driven by u and the y_2 measurements is governed by the well known Kalman filter equations (5) and (6). As the prediction estimate $\bar{y}_{1,k|k-1}$ is directly based on \bar{x}_k^{OEP} , the theoretical asymptotic prediction estimate covariance becomes

$$\text{Cov}(\bar{y}_{1,k|k-1}) = E(y_{1,k} - \bar{y}_{1,k|k-1})(y_{1,k} - \bar{y}_{1,k|k-1})^T = C_1 P^{OEP} C_1^T + R_{11}.$$

The theoretical asymptotic y_1 current estimation covariance is

$$\text{Cov}(\bar{y}_{1,k|k}) = E(y_{1,k} - \bar{y}_{1,k|k})(y_{1,k} - \bar{y}_{1,k|k})^T = C_1 P^{OEC} C_1^T + R_{11}, \quad (25)$$

with P^{OEC} given by (22).

For comparison purposes we also state the asymptotic covariance results for an estimator based on only u as input. The state estimation covariance $P^{OEU} = E(x_k - \bar{x}_k^{OEU})(x_k - \bar{x}_k^{OEU})^T$ is then determined by the Lyapunov equation

$$P^{OEU} = AP^{OEU}A^T + GR_iG^T, \quad (26)$$

and the resulting asymptotic y_1 estimation covariance becomes

$$\text{Cov}(\bar{y}_{1,k}) = E(y_{1,k} - \bar{y}_{1,k})(y_{1,k} - \bar{y}_{1,k})^T = C_1 P^{OEU} C_1^T + R_{11}. \quad (27)$$

Assume now for convenience a scalar y_1 measurement. When the estimators (17)

and (23) are identified and validated by use of independent data sets with $N \rightarrow \infty$, we will then find the theoretical root mean square error

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_{1,k} - y_{1,k}^{est})^2} \rightarrow C_1 P C_1^T + R_{11}, \quad (28)$$

where $y_{1,k}^{est}$ is either $\bar{y}_{1,k|k-1}$, $\bar{y}_{1,k|k}$ or $\bar{y}_{1,k}$, while $P = P^{OEP}$, $P = P^{OEC}$ or $P = P^{OEU}$ according to (6), (22) or (26).

4. Multivariate calibration as special cases

4.1. Assumptions according experimental setup and data

Consider again the system (1) with the optimal y_1 current estimator (23), and expand the input u with a vector d of unknown offsets or disturbances, i.e. use $u = [d^T u_m^T]^T$, where u_m is the known vector of manipulated or measured inputs. Let the input u_k be piecewise constant over periods that are much longer than both the time constants in the underlying continuous system and the discretization sampling time, and assume possibly collinear observations $y_{1,j}$ and $y_{2,j}$ at the end of each such period. Also assume that d_j is a white noise sequence, i.e. that the unknown offsets and disturbances are independent from one observation to the next. With a piecewise static input vector u_k and enough time for settlement, it follows from (1) that the observations will be given by

$$y_{1,j} = [C_1(I-A)^{-1}B + D_1] \begin{bmatrix} d \\ u_m \end{bmatrix}_j + \sum_{k=-\infty}^j v_k g_{1,j-k} + w_{1,j} \quad (29a)$$

$$y_{2,j} = [C_2(I-A)^{-1}B + D_2] \begin{bmatrix} d \\ u_m \end{bmatrix}_j + \sum_{k=-\infty}^j v_k g_{2,j-k} + w_{2,j}, \quad (29b)$$

where g_1 and g_2 stand for the impulse responses from v to y_1 and y_2 . All measurements are thus linear combinations of d and u_m plus noise, and since we assume a stable system with piecewise constant inputs and a settling time shorter than the data sampling time, this noise will be approximately white. Note, however, that since the noise terms in (29) are partly determined by the common process noise v_k , they will not be independent, as required for the optimal current estimator (23). For calibration purposes it is also a normal procedure to use mean values of the measurements over a certain period of time in order to reduce the noise, but this does not affect the theoretical analysis.

4.2. Least squares estimation

If both d and u_m are completely known, there is no need to utilize the information in the y_2 measurements, we can simply solve (29a) as an ordinary least squares problem. In our case, however, we consider d as unknown, and the y_2 measurements may then give valuable information about d and indirectly also about y_1 . In the following analysis we assume that $u_{m,j}$ is a persistently exciting stochastic signal, and that all data are centralized, i.e. that d_j , $u_{m,j}$, $y_{1,j}$ and $y_{2,j}$ are stochastic variables with zero mean. For details about centralization and the subsequent modification of the estimator, see e.g. Martens and Næs (1989). We also assume observations of $u_{m,j}$, $y_{1,j}$ and $y_{2,j}$ from an informative experiment with samples for $j = 1, 2, \dots, J$.

In order to use the Kalman filter formalism, we model d_j as generated by white noise through a pure delay system. In the same way we model the common noise part $e_{c,j}$ in $y_{1,j}$ and $y_{2,j}$ as generated by a delayed white noise sequence. Expressing y_1 and y_2 as linear combinations of $z = [d^T \ e_c^T]^T$ and u_m , we then arrive at the dynamic system

$$z_{j+1} = \begin{bmatrix} d \\ e_c \end{bmatrix}_{j+1} = e_{z,j} \quad (30a)$$

$$y_{1,j} = L_{11}z_j + L_{12}u_{m,j} + e_{1,j} \quad (30b)$$

$$y_{2,j} = L_{21}z_j + L_{22}u_{m,j} + e_{2,j}, \quad (30c)$$

where the detailed expressions for the L matrices follow from (29), and where $e_{z,j}$, $e_{1,j}$ and $e_{2,j}$ are white and independent noise sequences. This is a dynamic system as given in (1) with $A = 0$, $B = 0$ and $G = I$, and the algebraic Riccati equation (6) then results in

$$P^{OEP} = P_z = E z_j z_j^T. \quad (31)$$

From (5) follows that the Kalman gain related to the y_2 measurements is

$$K_2^{OE} = P_z L_{21}^T (L_{21} P_z L_{21}^T + R_{22})^{-1}, \quad (32)$$

where $R_{22} = E e_{2,j} e_{2,j}^T$. With $A = B = 0$ and an appropriate change of notation according to (30), the OEC model (18) now gives

$$\begin{aligned} y_{1,j} &= L_{11} K_2^{OE} (y_{2,j} - L_{22} u_{m,j}) + L_{12} u_{m,j} + \psi_j \\ &= (L_{12} - L_{11} K_2^{OE} L_{22}) u_{m,j} + L_{11} K_2^{OE} y_{2,j} + \psi_j. \end{aligned} \quad (33)$$

With

$$y_{1,j}^T = [u_{m,j}^T \ y_{2,j}^T] \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} + \psi_j^T \quad (34)$$

this gives

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} (L_{12} - L_{11} K_2^{OE} L_{22})^T \\ (L_{11} K_2^{OE})^T \end{bmatrix}. \quad (35)$$

In the same way as with the parameters in (18), we can find B_1 and B_2 in (34) by identification of an OE model. In this special static case, however, we find from (9) that $\bar{x}^{OEP} = 0$, and from (19) and (20) with appropriate change of notation we then find

$$\psi_k = L_{11} z_j - L_{11} K_2^{OE} L_{21} z_j - L_{11} K_2^{OE} e_{2,j} + e_{1,j}, \quad (36)$$

which shows that ψ_k is a white noise sequence. We can thus find unbiased parameter estimates directly as the solution to a least squares (LS) problem. In order to show that this is the case, we collect $u_{m,j}^T$, z_j^T , $y_{1,j}^T$, $y_{2,j}^T$, $e_{1,j}^T$ and $e_{2,j}^T$ for $j = 1, 2, \dots, J$ in data matrices. From (30) we then find

$$Y_1 = Z L_{11}^T + U_m L_{12}^T + E_1 \quad (37a)$$

$$Y_2 = Z L_{21}^T + U_m L_{22}^T + E_2, \quad (37b)$$

and from (34) the ordinary LS solution

$$\begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} = \left(\begin{bmatrix} U_m^T \\ Y_2^T \end{bmatrix} [U_m \ Y_2] \right)^{-1} \begin{bmatrix} U_m^T \\ Y_2^T \end{bmatrix} Y_1. \quad (38)$$

Utilizing that z_j , $e_{1,j}$ and $e_{2,j}$ are independent white noise sequences, and introducing $R_u = E u_{m,j} u_{m,j}^T$, we further find

$$E \frac{1}{N} \begin{bmatrix} U_m^T \\ Y_2^T \end{bmatrix} [U_m \ Y_2] = \begin{bmatrix} R_u & R_u L_{22}^T \\ L_{22} R_u & L_{21} P_z L_{21}^T + L_{22} R_u L_{22}^T + R_{22} \end{bmatrix} \quad (39)$$

and

$$E \frac{1}{N} \begin{bmatrix} U_m^T \\ Y_2^T \end{bmatrix} Y_1 = \begin{bmatrix} R_u L_{12}^T \\ L_{21} P_z L_{11}^T + L_{22} R_u L_{12}^T \end{bmatrix}. \quad (40)$$

By use of the matrix inversion lemma (Kailath 1980), we now readily find

$$E \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} = \begin{bmatrix} (L_{12} - L_{11} K_2^{OE} L_{22})^T \\ (L_{11} K_2^{OE})^T \end{bmatrix}, \quad (41)$$

with K_2^{OE} according to (32). This shows that the LS solution asymptotically is equal to the theoretical result (35) based on Kalman filtering formalism.

Without known manipulated inputs, i.e. with $u_{m,j} = 0$, the model (30) is simplified to

$$z_{j+1} = e_{z,j} \quad (42a)$$

$$y_{1,j} = L_1 z_j + e_{1,j} \quad (42b)$$

$$y_{2,j} = L_2 z_j + e_{2,j}, \quad (42c)$$

resulting in the simplified theoretical estimate

$$\bar{y}_{1,j} = L_1 K_2^{OE} y_{2,j} = L_1 P_z L_2^T (L_2 P_z L_2^T + R_{22})^{-1} y_{2,j}, \quad (43)$$

or with $\bar{y}_{1,j}^T = y_{2,j}^T B$

$$B = (L_1 K_2^{OE})^T = (L_2 P_z L_2^T + R_{22})^{-1} L_2 P_z L_1^T. \quad (44)$$

The expectation (41) of the LS solution is in this case simplified to

$$E \hat{B} = E (Y_2^T Y_2)^{-1} Y_2^T Y_1 = (L_2 P_z L_2^T + R_{22})^{-1} L_2 P_z L_1^T. \quad (45)$$

This simplified connection between ordinary LS estimation and Kalman filtering without dynamics was found also in Berntsen (1988), but then without the general dynamic model (18) as a basis, and also limited to the case where $L_1 = I$ (or at least invertible) and $e_1 = 0$, i.e. the case where y_1 are noise free measurements of all states in the system (possibly after a similarity transformation).

We end this subsection on LS estimation by an analysis of the asymptotic covariance of the y_1 estimates. The models (30) and (42) imply a state estimation covariance matrix $P^{OEP} = P_z$ as given in (31). Application of the general asymptotic covariance expression (25) on the model (30) thus results in

$$\text{Cov}(\bar{y}_{1,k|k}) = L_{11}(I - K_2^{OE} L_{21}) P_z (I - K_2^{OE} L_{21})^T L_{11}^T + L_{11} K_2^{OE} R_{22} (K_2^{OE})^T L_{11}^T + R_{11}, \quad (46)$$

where K_2^{OE} is given by (32). If the simplified model (42) is used, L_{11} is here replaced by L_1 while L_{21} is replaced by L_2 .

4.3. Principal component regression

With a large number of y_2 variables and a limited number of observations, the estimator (38) may give very large estimation covariance due to overfitting. In the common case with collinear y_2 variables, we can then make use of the fact that the information can be compressed into a smaller number of latent variables determined by the total number of independent variables in u and z . We then first collect all input data in either $X = [U_m \ Y_2]$ or $X = Y_2$, dependent on the problem formulation. By use of an appropriate number of principal components A (Martens and Næs 1989), the data is then expressed as

$$X \approx TP^T = [t_1 \ t_2 \ \dots \ t_A]P^T = [\tau_1 \ \tau_2 \ \dots \ \tau_j]^T P^T, \quad (47)$$

where T is the score matrix and P is the loading matrix.

For convenience we now limit the treatment to the case where $u_{m,j} = 0$, i.e. to the case where $X = [y_{2,1} \ y_{2,2} \ \dots \ y_{2,N}]^T$. By making use of the fact that $P^T P = I$, i.e. $T = XP$ and $\tau_j \approx P^T y_{2,j}$, the system (42) is now replaced by

$$z_{j+1} = e_{z,j} \quad (48a)$$

$$y_{1,j} = L_1 z_j + e_{1,j} \quad (48b)$$

$$\tau_j \approx P^T L_2 z_j + P^T e_{2,j}. \quad (48c)$$

With $y_{2,j}$ replaced by τ_j and using $\bar{y}_{1,j}^T \approx \tau_j^T B_T \approx \bar{y}_2^T P B_T$, i.e. $B \approx P B_T$, we now find the theoretical estimator (44) replaced by

$$B \approx P(P^T L_2 P_z L_2^T P + P^T R_{22} P)^{-1} P^T L_2 P_z L_1^T, \quad (49)$$

while the LS solution $\hat{B} = (Y_2^T Y_2)^{-1} Y_2^T Y_1$ is replaced by the data based PCR estimator

$$\hat{B} = P(T^T T)^{-1} T^T Y_1 = P(P^T X^T X P)^{-1} P^T X^T Y_1. \quad (50)$$

4.4. Partial least squares regression

The aim of partial least squares regression (PLSR) is to improve PCR by finding latent variables that explain both the X and the Y_1 data, and there exist at least two slightly different PLSR algorithms (Martens and Næs 1989). Also here we limit the treatment to the case where $u_{m,j} = 0$, and it is convenient to start with the PLSR method of Martens that makes use of linear combinations $\tau_M = W^T y_2$ (where W is found iteratively and where $T_M^T T_M$ in (53) below is non-diagonal). The result of this is that (42) is replaced by the PLSR_M model

$$z_{j+1} = e_{z,j} \quad (51a)$$

$$y_{1,j} = L_1 z_j + e_{1,j} \quad (51b)$$

$$\tau_{M,j} \approx W^T L_2 z_j + W^T e_{2,j}. \quad (51c)$$

The theoretical PCR estimator (49) is then replaced by the theoretical PLSR estimator

$$B \approx W(W^T L_2 P_z L_2^T W + W^T R_{22} W)^{-1} W^T L_2 P_z L_1^T, \quad (52)$$

while (50) is replaced by the data based PLSR estimator

$$\hat{B} = W(T_M^T T_M)^{-1} T_M^T Y_1 = W(W^T X^T X W)^{-1} W^T X^T Y_1. \quad (53)$$

The original PLSR method of Wold uses linear combinations $\tau_w = (W^T P_w)^{-1} W^T y_2$,

with the same W matrix as Martens and with a special loading matrix P_w and a diagonal matrix $T_w^T T_w$. The model (42) is then replaced by the PLSR_w model

$$z_{j+1} = e_{z,j} \quad (54a)$$

$$y_{1,j} = L_1 z_j + e_{1,j} \quad (54b)$$

$$\tau_{w,j} \approx (W^T P_w)^{-1} W^T L_2 z_j + (W^T P_w)^{-1} W^T e_{2,j}. \quad (54c)$$

With $W(W^T P_w)^{-1}$ instead of W , the theoretical PLSR estimator (52) becomes

$$\begin{aligned} B &\approx W(\cdot)^{-T} \left((\cdot)^{-1} W^T L_2 P_z L_2^T W(\cdot)^{-T} \right)^{-1} (\cdot)^{-1} W^T L_2 P_z L_1^T \\ &\quad + (\cdot)^{-1} W^T R_{22} W(\cdot)^{-T} \Big)^{-1} (\cdot)^{-1} W^T L_2 P_z L_1^T \\ &= W(W^T L_2 P_z L_2^T W + W^T R_{22} W)^{-1} W^T L_2 P_z L_1^T, \end{aligned} \quad (55)$$

while the data based PLSR estimator (53) becomes

$$\begin{aligned} \hat{B} &= W(\cdot)^{-T} [(\cdot)^{-1} W^T X^T X W(\cdot)^{-T}]^{-1} (\cdot)^{-1} W^T X^T Y_1 \\ &= W(P_w^T W)^{-1} (T_w^T T_w)^{-1} T_w^T y = W(W^T X^T X W)^{-1} W^T X^T Y_1. \end{aligned} \quad (56)$$

We see from this that P_w disappears from the estimator expressions, and that the final theoretical as well as data based estimators are the same for the Wold and Martens algorithms. This equivalence is of course well known (Martens and Næs 1989), although the treatment by use of Kalman filtering formalism is new.

5. Dynamic system PCR and PLSR solutions

Assuming independent measurement noise, the optimal y_1 current estimate for dynamic systems given in (23), may also form a basis for dynamic system solutions using PCR or PLSR (DPCR or DPLSR). It is then natural to split the secondary measurements into $y_{2,k} = [y_{21,k}^T y_{22,k}^T]^T$, where $y_{21,k}$ are the secondary measurements that are linked to y_1 only through a static system. When some or all of the $y_{21,k}$ measurements are internally collinear, they can be replaced by latent variables as in (48), (51) and (54), i.e. both PCR and PLSR may be used. Using e.g. the score definition in the PLSR method of Martens, i.e. $\tau = W^T y_{21}$, the OEC model (23) will then be replaced by

$$\begin{aligned} \bar{y}_{1,k|k} &= C_1(I - K_{\tau}^{OE} W^T C_{21} - K_{22}^{OE} C_{22})[qI - A(I - K_{\tau}^{OE} W^T C_{21} - K_{22}^{OE} C_{22})]^{-1} \\ &\quad [(B - AK_{\tau}^{OE} W^T D_{21} - AK_{22}^{OE} D_{22})u_k + AK_{\tau}^{OE} \tau_k + AK_{22}^{OE} y_{22,k}] \\ &\quad + C_1 K_{\tau}^{OE} (\tau_k - W^T D_{21} u_k) + C_1 K_{22}^{OE} (y_{22,k} - D_{22} u_k) + D_1 u_k. \end{aligned} \quad (57)$$

The Kalman gains are here determined as the solution to the Kalman filter equations (5) and (6) with $C_2 = [(W^T C_{21})^T C_{22}^T]^T$ and

$$R_{22} = \begin{bmatrix} EW^T w_{21} w_{21}^T W & EW^T w_{21} w_{22}^T W \\ EW_{22} w_{21}^T W & EW_{22} w_{22}^T W \end{bmatrix}.$$

If we find the τ variables by use of the PLSR method of Wold, we have to replace W^T with $(W^T P_w)^{-1} W^T$, while the PCR method uses P^T instead of W^T .

When the current estimator (57) is identified by use of a prediction error method, also past τ_k values will be used as a basis for determining $\bar{y}_{1,k|k}$, with reduced estimation covariance as the expected results, and we can in fact look and treat the latent variables as ordinary measurement signals. An essential assumption is here that the linear

relations between $y_{21,k}$ and τ_k given by P^T , W^T or $(W^T P_w)^{-1} W^T$ are time invariant and determined as in the static case either by PCA or by the iterative PLSR algorithms. Note, however, that time invariance is an essential assumption also in the general model (23).

If all or some of the y_{22} measurements are internally collinear, these measurements may also be replaced by latent variables in order to reduce the variance in the solution. However, since y_{22} is linked to y_1 through a dynamic system, the iterative PLSR method cannot be expected to work, and we must be content with using SVD or PCA to find these latent variables. They may also then be combined with known inputs and other measurements, and also with other latent variables found by PCA or PLSR.

With $u_k = 0$ and $y_{22,k} = 0$, equation (57) is simplified to

$$\bar{y}_{1,k|k} = C_1(I - K_\tau^{OE} W^T C_{21})[qI - A(I - K_\tau^{OE} W^T C_{21})]^{-1} A K_\tau^{OE} \tau_k + C_1 K_\tau^{OE} \tau_k, \quad (58)$$

showing the dynamic relation between the collinear time series y_{21} represented by τ and the time series y_1 .

6. Simulation examples

Simulation studies are undertaken by use of *Matlab*, primarily the *dlsim.m* function in the control system toolbox (Grace *et al.* 1992), the prediction error method implemented in the *pem.m* function in the system identification toolbox (Ljung 1991), the *least squares operator* and the *eig.m* function. With an appropriate OE model specified, the *pem.m* function identifies the optimal prediction estimator (17) or the optimal current estimator (23), where u and the secondary measurements y_2 are used as input signals. The corresponding OE model using only u as an input, may also be identified. For validation comparisons, the RMSE criterion in (28) was used.

6.1. Example 1. A second order system with a first order process noise model

The intention of this simulation example is primarily to support the theoretical development of the optimal prediction and current estimators (17) and (23). As a starting point, the following continuous second-order process model with an additional first-order process noise model was used (e.g. interacting mixing tanks or thermal processes):

$$\dot{x} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} v \quad (59a)$$

$$y_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x + w_1 \quad (59b)$$

$$y_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} x + w_2. \quad (59c)$$

The system was discretized assuming zero order hold elements on the u and v inputs and a sampling interval $T = 0.1$, resulting in the discrete model

$$x_{k+1} = \begin{bmatrix} 0.9092 & 0.0863 & 0.0044 \\ 0.0863 & 0.8230 & 0.0863 \\ 0 & 0 & 0.9048 \end{bmatrix} x_k + \begin{bmatrix} 0.0045 \\ 0.0908 \\ 0 \end{bmatrix} u_k + \begin{bmatrix} 0.0002 \\ 0.0045 \\ 0.0952 \end{bmatrix} v_k \quad (60a)$$

$$y_{1,k} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x_k + w_{1,k} \quad (60b)$$

$$y_{2,k} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} x_k + w_{2,k}. \quad (60c)$$

The system was then simulated with u_k as a filtered PRBS signal with autocovariance $r_u(p) = 0.5^{|p|}$ (Söderström and Stoica 1989, example 5.11 with $\alpha = 0.5$), i.e. an input that was persistently exciting of sufficient order. The scalar noise sources v_k , $w_{1,k}$ and $w_{2,k}$ were independent and normally distributed white noise sequences with zero mean and variances given below.

The simulated system was identified using OEP and OEC models with u_k and $y_{2,k}$ as input signals and $y_{1,k}$ as output signal, using $N = 10\,000$ samples.

The OEP model (10) was specified as

$$nn_{OEP} = [0, [3\ 3], 0, 0, [3\ 3], [1\ 1]], \quad (61)$$

i.e. a model

$$y_{1,k} = \frac{B_1(q^{-1})}{F_1(q^{-1})} u_k + \frac{B_2(q^{-1})}{F_2(q^{-1})} y_{2,k} + e_k^{OEP} \quad (62)$$

with

$$B_1(q^{-1}) = b_{11}q^{-1} + b_{12}q^{-2} + b_{13}q^{-3} \quad (63)$$

$$B_2(q^{-1}) = b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3} \quad (64)$$

$$F_1(q^{-1}) = 1 + f_{11}q^{-1} + f_{12}q^{-2} + f_{13}q^{-3} \quad (65)$$

$$F_2(q^{-1}) = 1 + f_{21}q^{-1} + f_{22}q^{-2} + f_{23}q^{-3} \quad (66)$$

The OEC model (18) was specified as

$$nn_{OEC} = [0, [3\ 4], 0, 0, [3\ 3], [1\ 0]], \quad (67)$$

i.e. the same model as (62), but with $B_2(q^{-1})$ altered to

$$B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}. \quad (68)$$

As the main purpose of the simulations was to support the theory, no attempt was made to find the model order and model structure from the data. The model order can, however, be found by ordinary use of one of the several available subspace identification methods, e.g. Di Ruscio (1996), and a systematic method for finding the structure is presented in Ergon and Di Ruscio (1997). No attempt was made to force $F_1(q^{-1})$ and $F_2(q^{-1})$ to be identical, which they theoretically should be.

Each identified model was validated against an independent data set with the same number of samples and the same noise variances as used for identification. Validation comparisons between the different identified models were based on the root mean square error criterion (28).

As a basis for comparisons given a specific experimental condition, each model was identified and validated in $M = 100$ Monte Carlo runs using independent data sets. In order to limit the influence of local minima problems, each identification and validation given a specific data set was repeated $R = 5$ times with randomized initial B parameters ($b_{ij,r+1} = b_{ij,r}(1 + 0.5e)$, with e as a zero mean and normal random variable with variance 1).

The mean RMSE values and RMSE standard deviations for $N = 10\,000$, $r_v = 0.1$, $r_{22} = 0.01$ and varying r_{11} values are given in Table 1. The table also includes theoretical RMSE values $\sqrt{\text{Cov}(\bar{y}_{1,k|k-1})}$ and $\sqrt{\text{Cov}(\bar{y}_{1,k|k})}$ computed according to (24) and (25). The agreement between results based on simulation and theory is obvious.

Table 1. Validation RMSE mean values with standard deviations and theoretical mean values for OE models with varying primary measurement noise variance. The number of samples was $N = 10\,000$

r_{11}	OEP	OEP _{th}	OEC	OEC _{th}
10^{-8}	0.0177 ± 0.0005	0.0177	0.0173 ± 0.0006	0.0173
10^{-7}	0.0177 ± 0.0005	0.0177	0.0173 ± 0.0006	0.0173
10^{-6}	0.0177 ± 0.0005	0.0177	0.0173 ± 0.0005	0.0173
10^{-5}	0.0181 ± 0.0005	0.0180	0.0177 ± 0.0005	0.0176
10^{-4}	0.0204 ± 0.0006	0.0203	0.0200 ± 0.0005	0.0200
10^{-3}	0.0363 ± 0.0004	0.0362	0.0361 ± 0.0003	0.0360

The results in Table 1 were obtained by use of $N = 10\,000$ samples in each data set. To indicate expected results for a more realistic number of samples, Monte Carlo validation results for models based on also $N = 1000$ and $N = 200$ samples are shown in Table 2. Here the y_1 variance is kept constant, $r_{11} = 0.0001$, otherwise the conditions are the same as above.

Table 2. Validation RMSE mean values with standard deviations and theoretical mean values for different numbers of samples. Here, $r_{11} = 0.0001$, and the RMSE values are multiplied by 10 000

N	OEU	OEU _{th}	OEP	OEP _{th}	OEC	OEC _{th}
200	350 ± 115	379	233 ± 66	203	223 ± 54	200
1000	379 ± 59	379	209 ± 18	203	204 ± 18	200
10 000	379 ± 17	379	203 ± 5	203	200 ± 5	200

Table 2 also includes results for an OE model with only u used as input. The model was then specified as

$$nn_{OEU} = [0, 3, 0, 0, 3, 1], \quad (69)$$

i.e.

$$y_{1,k} = \frac{B(q^{-1})}{F(q^{-1})} u_k + e_k \quad (70)$$

with

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3} \quad (71)$$

$$F(q^{-1}) = 1 + f_1 q^{-1} + f_2 q^{-2} + f_3 q^{-3}. \quad (72)$$

As expected, Table 2 shows increased estimation error when the number of samples is reduced from $N = 10\,000$ to more realistic values.

In order to visualize the degree of model misfit behind the RMSE values in the tables, specific validation responses for models based on $N = 200$ samples are shown in Fig. 2. This figure also gives a representative picture of the improvement achieved by including y_2 as an input signal.

6.2. Example 2. Least squares estimation

For an application of the least squares estimation in subsection 4.2, assume a process stream with varying but known concentrations $u_{m,1}$, $u_{m,2}$ and $u_{m,3}$ of three substances A, B and C, and varying concentrations d_1 , d_2 and d_3 of three other substances D, E and

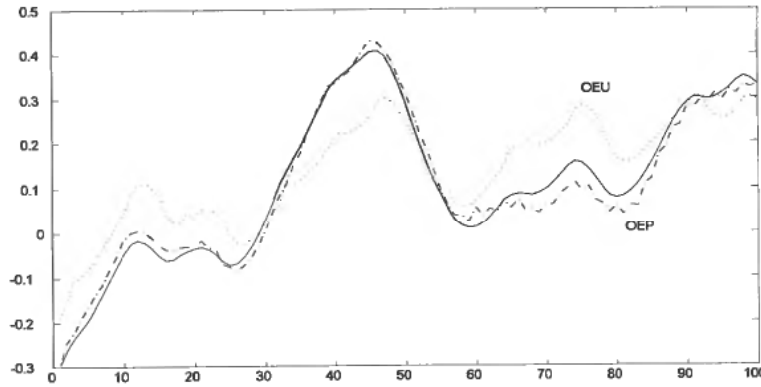


Figure 2. Segment of validation responses for the OEP model (10) using both u and y_2 as inputs (dashed, $RMSE = 0.0239$) and the OEU model (70) using only u as input (dotted, $RMSE = 0.1078$). The experimental conditions are given by $r_v = 1$, $r_{11} = 0.0001$, $r_{22} = 0.01$ and $N = 200$, and the ideal validation response is shown by solid line.

F. Also assume a scalar primary property y_1 and $p = 3$ secondary noisy y_2 measurements that all are linear combinations of the six concentrations. In order to find the relation between the estimator variables u_m and y_2 , and the y_1 property, a calibration experiment is performed on the system, with a data sampling interval such that the unknown concentrations are independent white noise sequences. We thus have a system as given in (30) with $e_c = 0$.

Calibration experiments were performed in $M = 100$ Monte Carlo simulations. In each experiment, d , u_m , e_1 and e_2 were generated as normally distributed random numbers, with $r_{11} = 0.0001$ and with diagonal covariance matrices $R_d = P_z$, R_u and R_{22} . The parameters in P_z , R_u , L_{11} , L_{12} , L_{21} and L_{22} were uniformly distributed random numbers in the interval $(0, 1)$, while the parameters in R_{22} were uniformly distributed random numbers in the interval $(0, 0.01)$. The resulting mean parameter values for the theoretical estimator (35) were

$$b = 10^{-4} \cdot [1675 \ 3048 \ -1347 \ -4712 \ 13191 \ 1021]^T \quad (73)$$

Least squares estimation according to (38) with $N = 10000$ samples gave the corresponding mean parameter values

$$\hat{b}_{N=10000} = 10^{-4} \cdot [1692 \ 3049 \ -1350 \ -4720 \ 13186 \ 1030]^T, \quad (74)$$

while $N = 200$ samples gave the mean parameter values

$$\hat{b}_{N=200} = 10^{-4} \cdot [1863 \ 2658 \ -880 \ -4035 \ 12346 \ 840]^T. \quad (75)$$

The mean theoretical RMSE value $RMSE_{theor.} = \sqrt{Cov(\bar{y}_{1,k|k})}$ with $Cov(\bar{y}_{1,k|k})$ determined according to (46) was

$$RMSE_{theor.} = \sqrt{Cov(\bar{y}_{1,k|k})} = 0.1136. \quad (76)$$

Least squares estimation according to (38) with $N = 10000$ samples and validation against an independent data set gave the mean value

$$RMSE_{N=10000} = 0.1134, \quad (77)$$

while estimation and validation with $N = 200$ samples gave the mean value

$$RMSE_{N=200} = 0.1182. \quad (78)$$

We notice here that the reduction from $N = 10000$ to $N = 200$ gave a significant impairment in all entries of the estimated parameter vector, while the overall RMSE values are very much the same. This shows that the optimum represented by the theoretical solution is not a very distinct one, which explains why regularization methods like PCR and PLSR can give good y_1 estimates even though the estimated parameters might have significant errors.

6.3. Example 3. Multivariate calibration

Assume the same system as in Example 2, but now with $p = 200$ noisy y_2 measurements. As in Example 2, the mean RMSE results from Monte Carlo runs (reduced to $M = 17$) were determined using the theoretical least squares estimator (35). In addition, the data based least squares estimator (38), and the corresponding PCR and PLSR estimators with both u and y_2 as inputs (generalizations of (50) and (53) with $X = [U_m \ Y_2]$) were determined. The number of principal components used in the PCR and PLSR algorithms were $A = 6$, which was easily found from the eigenvalues of the $X^T X$ data matrix (for a typical simulation with $N = 200$ observations, these were in descending order $\hat{\lambda}_i = 31410, 2870, 2065, 1369, 862, 356, 4.6, 4.4, 4.2, \dots$).

The mean validation RMSE results are shown in Table 3, where $RMSE_{LS, theor.} = \sqrt{Cov(\bar{y}_{1,k})}$, with $Cov(\bar{y}_{1,k})$ determined according to (46).

Table 3. Theoretical and validation RMSE mean values for LS, PCR and PLSR models based on different number of samples.

N	$RMSE_{LS, theor.}$	$RMSE_{LS}$	$RMSE_{PCR}$	$RMSE_{PLSR}$
200	0.0112	0.2288	0.0169	0.0169
400	0.0112	0.0171	0.0171	0.0170
800	0.0112	0.0128	0.0157	0.0157

As can be seen from Table 3, the LS method gives very poor results for $N = 200$, which is as expected since 203 estimator parameters are determined. For $N = 800$, the least squares estimates have a mean variance that approaches the theoretical mean value. The PCR and PLSR methods give far better results for $N = 200$, but show little improvement for an increased number of samples. There are no obvious differences between the PCR and PLSR methods, which is as expected with the randomly generated data that is used.

6.4. Example 4. Dynamic system PCR and PLSR solutions

For an application of the dynamical DPCR and DPLSR solutions in section 5, three independent filtered white noise sequences $v = [v_1^T \ v_2^T \ v_3^T]^T$ were generated. The following continuous system of three independent second order systems was then used as a starting point:

$$\dot{x} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} v \quad (79a)$$

$$y_1 = [1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0] x + w_1 \quad (79b)$$

$$y_2 = [L_{21} \quad 0] x + w_2 \quad (79c)$$

Here, L_{21} was a 200×3 matrix with uniformly distributed random parameters in the interval (0, 1). The system was discretized assuming zero order hold elements on the inputs and a sampling interval $T = 0.1$. The system was then simulated with v , w_1 and w_2 as independent and normally distributed white noise sequences with zero mean. The R_v and R_{22} covariance matrices were diagonal, with uniformly distributed random parameters in the intervals (0, 1) and (0, r_{22}) respectively, while w_1 was normally distributed with variance $r_{11} = 0.0001$. Different values of r_{22} were used as described below.

The simulations started with $r_{22} = 0.01$. In order to find the appropriate number of components, the PCR and PLSR estimators (50) and (53) based on $N = 200$ samples were first determined for different numbers of components A . In addition the dynamical DPCR and DPLSR models according to (57) were identified using the OE model (see Ljung, 1991, for definition of nn)

$$nn = [0, [2, \dots, 2], 0, 0, [2, \dots, 2], [0, \dots, 0]] \quad (80)$$

Each model was identified in $M = 10$ Monte Carlo runs, with validation against independent data sets with $N = 200$ samples. The resulting mean RMSE values are plotted in Fig. 3.

From Fig. 3 we find the optimal number of components $A = 3$, which is not surprising since the system has three independent noise sources. The figure also

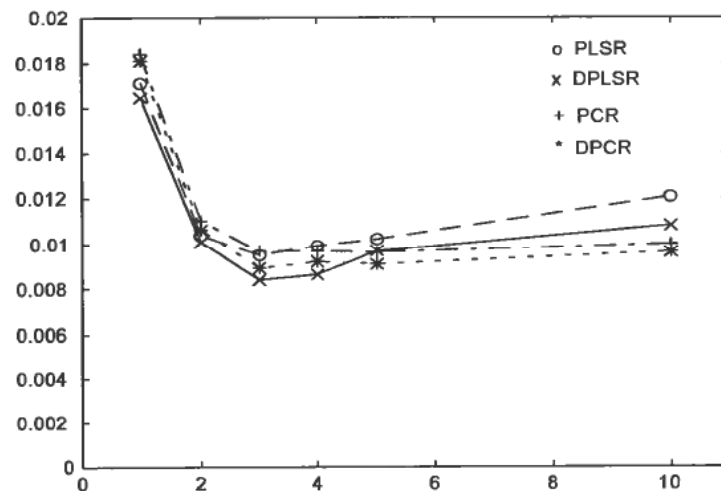


Figure 3. RMSE mean values as function of number of components used in PCR, PLSR, DPCR and DPLSR models for $r_{22} = 0.01$, based on 10 Monte Carlo runs with $N = 200$ samples.

indicates that PLSR is slightly better than PCR, and that the dynamic solutions are better than the static ones.

The RMSE values for different numbers of components are related to the explained proportion of the sample variance in a principal component analysis. This is a function of the sample eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$, given by (see, e.g. Johnson and Wichern 1992)

$$\text{explained proportion of sample variance} = \frac{\sum_{i=1}^A \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i}, \quad (81)$$

where $p = 200$ is the number of predictor variables.

For a simulation using $M = 100$ Monte Carlo runs with $r_{22} = 0.01$, the explained proportion of the sample variance for $A = 1, 2, 3, 4, 5$ and 10 were $0.594, 0.666, 0.689, 0.696, 0.703$ and 0.736 .

The models for $A = 2, 3$ and 4 were also determined using $M = 100$ Monte Carlo runs, with mean validation results and standard deviations as shown in Table 4.

Table 4. Validation RMSE mean values and standard deviations for PCR, DPCR, PLSR and DPLSR models for $r_{22} = 0.01$, based on $N = 200$ samples and with different numbers of components

A	$RMSE_{PCR}$	$RMSE_{DPCR}$	$RMSE_{PLSR}$	$RMSE_{DPLSR}$
2	0.0124 ± 0.0044	0.0121 ± 0.0059	0.0103 ± 0.0015	0.0095 ± 0.0020
3	0.0100 ± 0.0009	0.0091 ± 0.0017	0.0097 ± 0.0006	0.0086 ± 0.0007
4	0.0100 ± 0.0009	0.0091 ± 0.0011	0.0099 ± 0.0008	0.0089 ± 0.0008

The results clearly indicate the improvement obtained by use of the dynamical DPCR and DPLSR solutions. A statistical analysis is, however, beyond the scope of the present paper.

The PLSR and DPLSR simulations were finally repeated using also $r_{22} = 0.001$ and $r_{22} = 0.1$. Mean results based on $M = 100$ Monte Carlo runs with $A = 3$ are given in Table 5, indicating that the improvement obtained by use of the dynamical model increases with increasing y_2 noise level. At the same time the explained proportion of sample variance decreases. For $r_{22} = 0.1$ the optimal number of components is in fact $A = 2$, with slightly reduced RMSE values as compared with use of $A = 3$.

Table 5. Validation RMSE mean values and standard deviations for PCR, PLSR, DPCR and DPLSR models, based on $N = 200$ samples, $A = 3$ components and with different y_2 measurement noise. The explained proportion of sample variance with $A = 3$ is also given

r_{22}	$RMSE_{PLSR}$	$RMSE_{DPLSR}$	$\frac{RMSE_{DPLSR}}{RMSE_{PLSR}}$	expl. var.
0.001	0.00301 ± 0.00028	0.00305 ± 0.00020	0.99	0.95
0.01	0.0097 ± 0.0006	0.0086 ± 0.0007	0.89	0.69
0.1	0.0318 ± 0.0035	0.0257 ± 0.0054	0.81	0.19

7. Concluding remarks and further research

Optimal prediction (a priori) and current (a posteriori) estimators for non-measured primary output variables y_1 from e.g. an industrial plant, are developed. These estimators utilize all available information, also in secondary output measurements y_2 . The theoretical results are supported by results from a Monte Carlo simulation. Although the estimators at present are restricted to systems that are fully observable through the

y_2 outputs, preliminary results show that an extension to a wider class of systems is possible. Industrial applications in e.g. polymer extruding are now investigated, with promising preliminary results.

It is further shown that the general dynamic optimal current estimator in the special static case results in a least squares estimator valid for problems where the estimator variables include noisy measurements. This is also extended to cover principal component regression (PCR) and partial least squares regression (PLSR), and as a result it is shown how the PLSR methods of Wold and Martens are related to each other. The practical usefulness of this is not further investigated. The theoretical developments concerning static estimators are also supported by Monte Carlo simulations.

Finally it is indicated how dynamic system PCR (DPCR) and PLSR (DPLSR) solutions can be achieved as special cases of the general dynamic system current estimator. As demonstrated in a simple simulation example, this may result in some reduction of the y_1 estimation covariance, compared with ordinary PCR and PLSR. The full potential of this, theoretically and in practical applications, are open questions.

As the basic problem in the paper is formulated, the goal is to identify optimal estimators for y_1 in (1), i.e. for

$$y_{1,k} = C_1 x_k + D_1 u_k + w_{1,k}. \quad (82)$$

More precisely, though, we want to find estimators for the noise free part of y_1 , i.e. for

$$z_{1,k} = C_1 x_k + D_1 u_k, \quad (83)$$

and this is in fact done in the optimal estimators (17) and (23). We may, therefore, replace $\bar{y}_{1,k|k-1}$ and $\bar{y}_{1,k|k}$ in these estimators with $\bar{z}_{k|k-1}$ and $\bar{z}_{k|k}$. The theoretical asymptotic covariances (25) and (26) will then be altered to

$$\text{Cov}(\bar{z}_{k|k-1}) = E(z_k - \bar{z}_{k|k-1})(z_k - \bar{z}_{k|k-1})^T = C_1 P^{OEC} C_1^T \quad (84)$$

and

$$\text{Cov}(\bar{z}_{k|k}) = E(z_k - \bar{z}_{k|k})(z_k - \bar{z}_{k|k})^T = C_1 P^{OEC} C_1^T. \quad (85)$$

The methods presented in the paper have the obvious drawback that they require experimental data where also the primary y_1 measurements are sampled at a fast rate. In many cases, this may be very impractical and costly. It is, therefore, a need for methods that can identify the estimators from the y_1 data that are normally available from product quality analyses at a slow sampling rate. Since the optimal estimators are based on output error (OE) models, the solution to this important problem is in theory surprisingly simple. The OE models do not use past y_1 values, and it is therefore not quite necessary to minimize the scalar criterion function (15), $V_N(\hat{\theta}) = \det[\frac{1}{N} \sum_{k=1}^N \varepsilon_{1,k} \varepsilon_{1,k}^T]$, based on fast sampling rate y_1 data. Asymptotically, the same results will be obtained by minimization of

$$V_J(\hat{\theta}) = \det \left[\frac{1}{J} \sum_{j=1}^J \varepsilon_{1,j} \varepsilon_{1,j}^T \right], \quad (86)$$

based on y_1 values at a slow and possibly also irregular sampling rate. The standard prediction error identification algorithms must then be modified accordingly, and a preliminary analysis and simulation results show that this is quite feasible. The main problem is to find good initial parameter values for the minimization, and methods for this is now under investigation. To obtain useful results, the number J of sampled y_1 values must be sufficiently high, which may mean that data have to be collected over

a long period of time. This is no major problem with modern control and operator support systems.

The insight reflected in (86) also opens for the development of algorithms that update the estimators recursively, as the plant parameters vary slowly with time. This is a common situation in industrial plants, and recursive algorithms are therefore in many cases necessary.

REFERENCES

- BERNTSEN, H. (1988). Utvidet Kalmanfilter og Multivariabel kalibrering. *Report STF48 A88019*, SINTEF, Trondheim, Norway.
- DI RUSCIO, D. (1996). Combined Deterministic and Stochastic System Identification and Realization: DSR—a subspace approach based on observations. *Modeling, Identification and Control*, Vol. 17, No. 3, pp. 193–230.
- ERGON, R. and DI RUSCIO, D. (1997). Dynamic system calibration by system identification methods. *The European Control Conference, ECC97*, Brussels, CD file ECC 416.pfd.
- ERGON, R. (1998a). On Primary Output Estimation by use of Secondary Measurements as Input Signals in System Identification, accepted for publication in *IEEE Transactions on Automatic Control*.
- ERGON, R. (1998b). Static linear regression analysis by use of Kalman filtering formalism, submitted to *IEEE Transactions on Automatic Control*.
- GRACE, A., LAUB, A. J., LITTLE, J. N. and THOMPSON, C. M. (1992). *Control System Toolbox for use with Matlab*, The MathWorks, Inc.
- GOODWIN, G. C. and PAYNE, R. L. (1977). *Dynamic System Identification*, Academic Press, Inc.
- GREWAL, M. S. and ANDREWS, A. P. (1993). *Kalman filtering, Theory and Practice*, Prentice-Hall, Inc.
- JOHNSON, R. A. and WICHERN, D. W. (1992). *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc.
- KAILATH, T. (1980). *Linear Systems*, Prentice-Hall, Inc.
- LEWIS, F. L. (1986). *Optimal Estimation, with an Introduction to Stochastic Control Theory*, John Wiley & Sons.
- LJUNG, L. (1987). *System Identification, Theory for the User*, Prentice-Hall, Inc.
- LJUNG, L. (1991). *System Identification Toolbox for use with Matlab*, The MathWorks, Inc.
- MARTENS, H. and NÆS, T. (1989). *Multivariate Calibration*, John Wiley & Sons.
- MEJDELL, T. and ANDERSSON, B. O. (1994). Using temperature profile for product quality estimation on a distillation column, *ISA Transactions*, Vol. 33, No. 1, pp. 27–34.
- SÖDERSTRÖM, T. and STOICA, P. (1989). *System Identification*, Prentice-Hall, Inc.