

Dynamic system multivariate calibration

Rolf Ergon *

Telemark Institute of Technology, N-3914 Porsgrunn, Norway

Abstract

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

Keywords: Estimation; System identification; Multivariate calibration

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 \mathbf{y}_1 from a linear, time invariant and stable dynamic system, utilizing all available information in known inputs \mathbf{u} and measured secondary outputs \mathbf{y}_2 . The basic insight behind this is that the \mathbf{y}_2 measurements may carry valuable information about the process noise \mathbf{v} , which it should be possible to utilize when estimating \mathbf{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.

The second aim is to study the relations between such methods for dynamic system identification (SI) and methods for static regression and multivariate calibration widely used in, e.g., chemometrics. In the static case we would find a model for estimation of primary dependent variables \mathbf{y}_1 from independent variables \mathbf{u} and secondary dependent variables \mathbf{y}_2 . The *final aim* is to indicate how these relations can be used to develop multivariate calibration methods that can also handle time series data generated by dynamic systems.

The paper is organized in the following way: Section 2 gives some background and preliminaries. In Section 3 the theory for use of secondary \mathbf{y}_2 measurements as inputs to SI procedures is established. This results in the central relation in the paper, giving the optimal \mathbf{y}_1 estimator (Eq. (19)) for linear, time invariant and stable dynamic systems. An analysis of

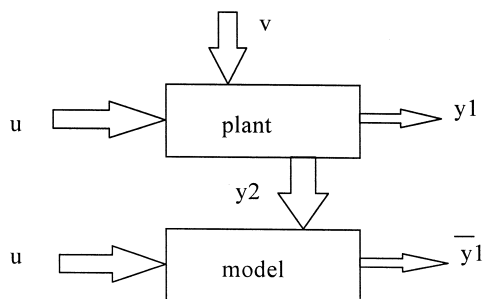


Fig. 1. Basic principle for estimation of primary system outputs y_1 from known inputs u and measured secondary outputs y_2 in presence of process noise v .

the covariance of the y_1 estimate is also given. In Section 4 it is shown that least squares estimation (LSE), principal component regression (PCR) and partial least squares regression (PLSR) can be seen as special static cases of this dynamical solution. 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 presented in Section 4. Section 5 gives some numerical examples and Monte Carlo simulations, and concluding remarks are given in Section 6.

2. Background and preliminaries

Linear regression and static calibration methods have roots in the classical least squares technique used by Gauss around 1800. When the number of estimator variables is large and the number of observations is limited, the ordinary solution to the least squares problem may have very large variance due to overfitting. This situation requires some form of regularization, e.g., PCR or PLSR [1,2]. 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 100 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. Such tools for multivariate data analysis are used in many scientific fields like biometrics, chemometrics, econometrics and psychometrics.

Linear regression can also be used to identify the parameters in dynamic system finite impulse response (FIR) models or autoregressive models with external inputs (ARX) [3]. Due to lack of noise modeling, this will normally result in biased parameter estimates, and the FIR truncation error comes in addition. Identification of FIR and ARX models by PCR and PLSR have also been investigated, see, e.g., Refs. [4,5].

In parallel with the development of the PCR and PLSR methods, the field of general dynamic SI has been developed into a sophisticated set of methods and practical tools. Classical SI methods are summarized in comprehensive books, e.g., Refs. [3,6]. At present, subspace identification methods attract a great deal of interest, see, e.g., Ref. [7] 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.

SI is also closely linked to the Kalman filtering theory [8]. 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 SI 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 Ref. [9], and the present paper includes a further and more detailed attempt to do so. 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 vari-

ables \mathbf{y} from both independent and dependent known variables \mathbf{x} . In SI terminology, this means methods for estimation of unknown system outputs \mathbf{y}_1 from both independent system inputs \mathbf{u} and dependent system outputs \mathbf{y}_2 . The basic observation here is that also dependent outputs \mathbf{y}_2 have to be used as inputs in the SI procedure.

(2) When the multivariate calibration models are used for estimation, the \mathbf{y}_1 outputs are not known, and this will also be the case for the corresponding dynamical models found by SI. We are, therefore, lead to consider output error (OE) models and not the qualitatively different ARMAX (autoregressive moving average with external inputs) type of models used for, e.g., control design based on known \mathbf{y}_1 outputs.

(3) In order to find the optimal \mathbf{y}_1 estimate, the underlying Kalman filter must be of the predictor-corrector form, which is normally not the case when innovation models are used in SI.

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. The use of both independent inputs \mathbf{u} and dependent \mathbf{y}_2 measurements as inputs in a SI procedure raises questions about identifiability and applications on deterministic and perfect measurement systems. A preliminary discussion of this is given in Ref. [10]. A detailed comparison of ARMAX and OE models for prediction of \mathbf{y}_1 based on \mathbf{u} and \mathbf{y}_2 is given in Ref. [11].

3. Secondary measurements as inputs in system identification

3.1. Statement of problem

Consider the discrete system model

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

where \mathbf{x} is the state vector, while \mathbf{v} and $\mathbf{w} = [\mathbf{w}_1^T \mathbf{w}_2^T]^T$ are white and independent process and measurement noise vectors. Also assume a stable system with $(\mathbf{A}, \mathbf{G}\sqrt{\mathbf{R}_v})$ reachable, where \mathbf{R}_v is given by the expecta-

tion $\mathbf{R}_v = \mathbf{E}\mathbf{v}_k\mathbf{v}_k^T$. Note that some or all of the secondary \mathbf{y}_2 measurements may be collinear with some or all of the primary \mathbf{y}_1 measurements.

Further assume that input-output data are available from an informative experiment [12], i.e., that data records for \mathbf{u}_k , $\mathbf{y}_{1,k}$ and $\mathbf{y}_{2,k}$ for $k = 1, 2, \dots, N$ are at hand, with \mathbf{u}_k persistently exciting of appropriate order and N sufficiently high. The problem is now to identify the optimal one-step-ahead $\mathbf{y}_{1,k}$ predictor based on past and present \mathbf{u}_k and past $\mathbf{y}_{2,k}$ values, and the optimal $\mathbf{y}_{1,k}$ current estimator based also on present $\mathbf{y}_{2,k}$ values.

Note that it is a part of the problem that $\mathbf{y}_{1,k}$ is not available as a basis for the prediction estimate $\bar{\mathbf{y}}_{1,k|k-1}$ or the current estimate $\bar{\mathbf{y}}_{1,k|k}$. 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 \mathbf{y}_1 is available

The model (Eq. (1)) can be expressed in the ordinary innovation form [6] given by the following equations, where $\mathbf{A}\mathbf{K} = \mathbf{A}[\mathbf{K}_1 \mathbf{K}_2]$ is the gain in a predictor type Kalman filter formulation with white innovations \mathbf{e}_1 and \mathbf{e}_2 :

$$\begin{aligned}\bar{\mathbf{x}}_{k+1} &= \mathbf{A}\bar{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k + \mathbf{A}[\mathbf{K}_1 \mathbf{K}_2] \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}_k \\ \mathbf{y}_{1,k} &= \mathbf{C}_1\bar{\mathbf{x}}_k + \mathbf{D}_1\mathbf{u}_k + \mathbf{e}_{1,k} \\ \mathbf{y}_{2,k} &= \mathbf{C}_2\bar{\mathbf{x}}_k + \mathbf{D}_2\mathbf{u}_k + \mathbf{e}_{2,k}.\end{aligned}\quad (2)$$

The optimal one-step-ahead \mathbf{y}_1 predictor with all measurements available and a known \mathbf{u}_k will then be

$$\begin{aligned}\bar{\mathbf{x}}_{k+1} &= \mathbf{A}(\mathbf{I} - \mathbf{K}_1\mathbf{C}_1 - \mathbf{K}_2\mathbf{C}_2)\bar{\mathbf{x}}_k \\ &\quad + (\mathbf{B} - \mathbf{A}\mathbf{K}_1\mathbf{D}_1 - \mathbf{A}\mathbf{K}_2\mathbf{D}_2)\mathbf{u}_k \\ &\quad + \mathbf{A}\mathbf{K}_1\mathbf{y}_{1,k} + \mathbf{A}\mathbf{K}_2\mathbf{y}_{2,k} \\ \bar{\mathbf{y}}_{1,k} &= \mathbf{C}_1\bar{\mathbf{x}}_k + \mathbf{D}_1\mathbf{u}_k.\end{aligned}\quad (3)$$

This will be the best linear one-step-ahead predictor if \mathbf{x}_0 , \mathbf{v}_k and \mathbf{w}_k have arbitrary statistics, and the optimal predictor assuming that \mathbf{x}_0 , \mathbf{v}_k and \mathbf{w}_k are

normally distributed [8]. This is also the predictor normally used in prediction error identification methods [3,6].

3.3. Optimal one-step-ahead predictor when \mathbf{y}_1 is not available

When the \mathbf{y}_1 measurements are not available as a basis for prediction, the ARMAX predictor (Eq. (3)) is no longer optimal [11]. The obvious reason for this is that Eq. (3) is based on an underlying Kalman filter driven by \mathbf{y}_1 in addition to \mathbf{u} and \mathbf{y}_2 , and the information in the \mathbf{y}_2 measurements will then not be utilized in an optimal way when \mathbf{y}_1 is not available.

In a prediction error identification method, we must instead base the prediction on an underlying Kalman filter driven by \mathbf{u} and only the \mathbf{y}_2 measurements. With the assumption that $(\mathbf{C}_2, \mathbf{A})$ is detectable, the following innovation form can then be derived from Eq. (1):

$$\begin{aligned}\bar{\mathbf{x}}_{k+1}^{\text{OEP}} &= \mathbf{A}\bar{\mathbf{x}}_k^{\text{OEP}} + \mathbf{B}\mathbf{u}_k + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{e}_{2,k} \\ \mathbf{y}_{2,k} &= \mathbf{C}_2\bar{\mathbf{x}}_k^{\text{OEP}} + \mathbf{D}_2\mathbf{u}_k + \mathbf{e}_{2,k}.\end{aligned}\quad (4)$$

The \mathbf{y}_1 output is then given by

$$\mathbf{y}_{1,k} = \mathbf{C}_1\bar{\mathbf{x}}_k^{\text{OEP}} + \mathbf{D}_1\mathbf{u}_k + \mathfrak{d}_k^{\text{OEP}}, \quad (5)$$

where

$$\mathfrak{d}_k^{\text{OEP}} = \mathbf{C}_1(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEP}}) + \mathbf{w}_{1,k} \quad (6)$$

is colored noise.

The underlying Kalman filter is governed by the well known Kalman filter equations [8]. The Kalman gain is determined by

$$\mathbf{K}_2^{\text{OE}} = \mathbf{P}^{\text{OEP}}\mathbf{C}_2^T(\mathbf{C}_2\mathbf{P}^{\text{OEP}}\mathbf{C}_2^T + \mathbf{R}_{22})^{-1}, \quad (7)$$

where the prediction state estimation covariance $\mathbf{P}^{\text{OEP}} = \mathbf{E}(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEP}})(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEP}})^T$ is given by the Riccati equation

$$\mathbf{P}^{\text{OEP}} = \mathbf{A}\mathbf{P}^{\text{OEP}}\mathbf{A}^T + \mathbf{G}\mathbf{R}_v\mathbf{G}^T - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{C}_2\mathbf{P}^{\text{OEP}}\mathbf{A}^T, \quad (8)$$

and where $\mathbf{R}_v = \mathbf{E}\mathbf{v}_k\mathbf{v}_k^T$ and $\mathbf{R}_{22} = \mathbf{E}\mathbf{w}_{2,k}\mathbf{w}_{2,k}^T$.

Theoretically, it is possible to identify the system determined by Eqs. (4) and (5) using \mathbf{y}_1 and \mathbf{y}_2 as outputs, i.e., to identify Eq. (2) with a simplified noise model employing $\mathbf{K}_1 = 0$. With many secondary \mathbf{y}_2 measurements it is, however, a simpler task to use \mathbf{y}_2

as an input signal, and identify the OE prediction model (OEP model)

$$\begin{aligned}\bar{\mathbf{x}}_{k+1}^{\text{OEP}} &= \mathbf{A}(\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)\bar{\mathbf{x}}_k^{\text{OEP}} \\ &\quad + (\mathbf{B} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{D}_2)\mathbf{u}_k + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{y}_{2,k}\end{aligned}$$

$$\mathbf{y}_{1,k} = \mathbf{C}_1\bar{\mathbf{x}}_k^{\text{OEP}} + \mathbf{D}_1\mathbf{u}_k + \mathfrak{d}_k^{\text{OEP}}. \quad (9)$$

The corresponding input–output model is then

$$\begin{aligned}\mathbf{y}_{1,k} &= \mathbf{C}_1[\mathbf{q}\mathbf{I} - \mathbf{A} + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{C}_2]^{-1} \\ &\quad \cdot [(\mathbf{B} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{D}_2)\mathbf{u}_k + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{y}_{2,k}] \\ &\quad + \mathbf{D}_1\mathbf{u}_k + \mathfrak{d}_k^{\text{OEP}} \\ &= \bar{\mathbf{y}}_{1,k|k-1} + \mathfrak{d}_k^{\text{OEP}},\end{aligned}\quad (10)$$

where $\bar{\mathbf{y}}_{1,k|k-1}$ is the optimal prediction estimate.

This model can also be expressed as

$$\mathbf{y}_{1,k} = \mathbf{G}_1(q^{-1})\mathbf{u}_k + \mathbf{G}_2(q^{-1})\mathbf{y}_{2,k} + \mathfrak{d}_k^{\text{OEP}}, \quad (11)$$

where q^{-1} is the unit delay operator. The transfer functions are here

$$\mathbf{G}_1(q^{-1}) = \mathbf{C}_1(\mathbf{q}\mathbf{I} - \tilde{\mathbf{A}})^{-1}(\mathbf{B} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{D}_2) + \mathbf{D}_1 \quad (12)$$

and

$$\mathbf{G}_2(q^{-1}) = \mathbf{C}_1(\mathbf{q}\mathbf{I} - \tilde{\mathbf{A}})^{-1}\mathbf{A}\mathbf{K}_2^{\text{OE}}, \quad (13)$$

with $\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{C}_2$.

In order to identify the deterministic part of the system (10), i.e., \mathbf{G}_1 and \mathbf{G}_2 , we model $\mathfrak{d}_k^{\text{OEP}}$ by some unknown white noise sequence and use the prediction

$$\hat{\mathbf{y}}_{1,k} = \hat{\mathbf{G}}_1(q^{-1};\theta)\mathbf{u}_k + \hat{\mathbf{G}}_2(q^{-1};\theta)\mathbf{y}_{2,k}. \quad (14)$$

The prediction error is then

$$\begin{aligned}\boldsymbol{\epsilon}_{1,k} &= \mathbf{y}_{1,k} - \hat{\mathbf{y}}_{1,k} \\ &= [\mathbf{G}_1(q^{-1}) - \hat{\mathbf{G}}_1(q^{-1};\theta)]\mathbf{u}_k \\ &\quad + [\mathbf{G}_2(q^{-1}) - \hat{\mathbf{G}}_2(q^{-1};\theta)]\mathbf{y}_{2,k} + \mathfrak{d}_k^{\text{OEP}}.\end{aligned}\quad (15)$$

When evaluating the result of minimizing a criterion function $V_N(\theta) = \text{tr}((1/N)\sum_{k=1}^N \boldsymbol{\epsilon}_{1,k}\boldsymbol{\epsilon}_{1,k}^T)$, we must now consider the fact that $\mathbf{y}_{2,k}$ and $\mathfrak{d}_k^{\text{OEP}}$ are not independent. We then note that the predictor (Eq. (14))

has the form of an observer driven by \mathbf{u} and the \mathbf{y}_2 measurements, and that the criterion function determined by the prediction error (Eq. (15)) under the assumption of Gaussian noise therefore is minimized when and only when both

- the deterministic model is correct, and
- the observer gain is a Kalman gain, i.e., $\hat{\mathbf{K}}_2 \equiv \mathbf{K}_2^{\text{OE}}$.

Minimization will therefore asymptotically ($N \rightarrow \infty$) result in $\hat{\mathbf{G}}_1 \equiv \mathbf{G}_1$ and $\hat{\mathbf{G}}_2 \equiv \mathbf{G}_2$, with \mathbf{G}_1 and \mathbf{G}_2 given by Eqs. (12) and (13). The prediction estimate $\bar{\mathbf{y}}_{1,k|k-1}$ will thus be asymptotically unbiased.

3.4. Optimal current estimator when \mathbf{y}_1 is not available

Utilizing also current \mathbf{y}_2 values, the optimal estimator considering that \mathbf{y}_1 is not available will be found by identifying the following OE model based on an underlying predictor–corrector Kalman filter [8] utilizing also current data (OEC model):

$$\begin{aligned} \mathbf{y}_{1,k} = & \mathbf{C}_1(\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)[q\mathbf{I} - \mathbf{A} + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{C}_2]^{-1} \\ & \cdot [(\mathbf{B} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{D}_2)\mathbf{u}_k + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{y}_{2,k}] \\ & + \mathbf{C}_1\mathbf{K}_2^{\text{OE}}(\mathbf{y}_{2,k} - \mathbf{D}_2\mathbf{u}_k) + \mathbf{D}_1\mathbf{u}_k + \boldsymbol{\vartheta}_k^{\text{OEC}}. \end{aligned} \quad (16)$$

Here we introduce the colored noise

$$\boldsymbol{\vartheta}_k^{\text{OEC}} = \mathbf{C}_1(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEC}}) + \mathbf{w}_{1,k}, \quad (17)$$

based on

$$\bar{\mathbf{x}}_k^{\text{OEC}} = (\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)\bar{\mathbf{x}}_k^{\text{OEP}} + \mathbf{K}_2^{\text{OE}}(\mathbf{y}_{2,k} - \mathbf{D}_2\mathbf{u}_k). \quad (18)$$

From Eq. (16), we find the asymptotically unbiased and optimal \mathbf{y}_1 current estimator

$$\begin{aligned} \bar{\mathbf{y}}_{1,k|k} = & \mathbf{C}_1(\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)[q\mathbf{I} - \mathbf{A} + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{C}_2]^{-1} \\ & \cdot [(\mathbf{B} - \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{D}_2)\mathbf{u}_k + \mathbf{A}\mathbf{K}_2^{\text{OE}}\mathbf{y}_{2,k}] \\ & + \mathbf{C}_1\mathbf{K}_2^{\text{OE}}(\mathbf{y}_{2,k} - \mathbf{D}_2\mathbf{u}_k) + \mathbf{D}_1\mathbf{u}_k. \end{aligned} \quad (19)$$

This is the central relation in the paper, showing how past and present \mathbf{u} and \mathbf{y}_2 values can be utilized in an optimal way to find the current estimate $\bar{\mathbf{y}}_{1,k|k}$. It is straightforward to show, however, that identification of Eq. (19) by use of a prediction error method will

result in a correct result only when $\mathbf{w}_{1,k}$ and $\mathbf{w}_{2,k}$ are uncorrelated [11].

The optimal estimator (Eq. (19)) is also the basis for Section 4, where LSE, PCR and PLSR are found as special static cases, and for the dynamic PCR and PLS solutions presented in Section 5.

3.5. Theoretical \mathbf{y}_1 current estimation covariance

When the OEC model (Eq. (16)) is identified using a large data set, i.e., $N \rightarrow \infty$, the estimate $\bar{\mathbf{y}}_{1,k|k}$ will be asymptotically unbiased when we use either only \mathbf{u} or both \mathbf{u} and \mathbf{y}_2 as input signals. The asymptotic covariance will, however, 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 \mathbf{y}_1 current estimation covariance.

The underlying Kalman filter driven by \mathbf{u} and the \mathbf{y}_2 measurements is governed by Eqs. (7) and (8). The current state estimate is given by Eq. (18), and the current state estimation covariance $\mathbf{P}^{\text{OEC}} = \mathbf{E}(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEC}})(\mathbf{x}_k - \bar{\mathbf{x}}_k^{\text{OEC}})^T$ is thus

$$\begin{aligned} \mathbf{P}^{\text{OEC}} = & (\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)\mathbf{P}^{\text{OEP}}(\mathbf{I} - \mathbf{K}_2^{\text{OE}}\mathbf{C}_2)^T \\ & + \mathbf{K}_2^{\text{OE}}\mathbf{R}_{22}(\mathbf{K}_2^{\text{OE}})^T. \end{aligned} \quad (20)$$

As the current estimate $\bar{\mathbf{y}}_{1,k|k}$ is directly based on $\bar{\mathbf{x}}_k^{\text{OEC}}$, the theoretical asymptotic \mathbf{y}_1 current estimation covariance becomes

$$\begin{aligned} \text{Cov}(\bar{\mathbf{y}}_{1,k|k}) & = \mathbf{E}(\mathbf{y}_{1,k} - \bar{\mathbf{y}}_{1,k|k})(\mathbf{y}_{1,k} - \bar{\mathbf{y}}_{1,k|k})^T \\ & = \mathbf{C}_1\mathbf{P}^{\text{OEC}}\mathbf{C}_1^T + \mathbf{R}_{11}, \end{aligned} \quad (21)$$

with \mathbf{P}^{OEC} given by Eq. (20) and $\mathbf{R}_{11} = \mathbf{E}\mathbf{w}_{1,k}\mathbf{w}_{1,k}^T$.

Assume now for convenience a scalar \mathbf{y}_1 measurement. When the model is identified and validated by use of independent data sets with $N \rightarrow \infty$, we will then find the theoretical root mean square error (RMSE)

$$\begin{aligned} \text{RMSE}|_{\mathbf{u},\mathbf{y}_2} & = \sqrt{\frac{1}{N} \sum_{k=1}^N (\mathbf{y}_{1,k} - \bar{\mathbf{y}}_{1,k|k})^2} \\ & \rightarrow \sqrt{\mathbf{C}_1\mathbf{P}^{\text{OEC}}\mathbf{C}_1^T + \mathbf{R}_{11}}. \end{aligned} \quad (22)$$

4. Multivariate calibration as special cases

4.1. Assumptions according experimental setup and data

Consider again the system (Eq. (1)) with the optimal \mathbf{y}_1 current estimator (Eq. (19)), and expand the input \mathbf{u} with a vector \mathbf{d} of unknown offsets or disturbances, i.e., use $\mathbf{u} = [\mathbf{d}^T \mathbf{u}_m^T]^T$, where \mathbf{u}_m is the known vector of manipulated or measured inputs. Let the input \mathbf{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 $\mathbf{y}_{1,j}$ and $\mathbf{y}_{2,j}$ at the of each such period. Also assume that \mathbf{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 \mathbf{u}_k and enough time for settlement, it follows from (Eq. (1)) that the observations will be given by

$$\mathbf{y}_{1,j} = [\mathbf{C}_1(\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}_1] \begin{bmatrix} \mathbf{d} \\ \mathbf{u}_m \end{bmatrix}_j + \sum_{k=-\infty}^j \mathbf{v}_k \mathbf{g}_{1,j-k} + \mathbf{w}_{1,j} \quad (23a)$$

$$\mathbf{y}_{2,j} = [\mathbf{C}_2(\mathbf{I} - \mathbf{A})^{-1} \mathbf{B} + \mathbf{D}_2] \begin{bmatrix} \mathbf{d} \\ \mathbf{u}_m \end{bmatrix}_j + \sum_{k=-\infty}^j \mathbf{v}_k \mathbf{g}_{2,j-k} + \mathbf{w}_{2,j}, \quad (23b)$$

where \mathbf{g}_1 and \mathbf{g}_2 stand for the impulse responses from \mathbf{v} to \mathbf{y}_1 and \mathbf{y}_2 . All measurements are thus linear combinations of \mathbf{d} and \mathbf{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 Eqs. (23a) and (23b) are partly determined by the common process noise \mathbf{v}_k , they will not be independent, as required for the optimal current estimator (Eq. (19)). 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 \mathbf{d} and \mathbf{u}_m are completely known, there is no need to utilize the information in the \mathbf{y}_2 measurements, we can simply solve Eq. (23a) as an ordinary least squares problem. In our case, however, we consider \mathbf{d} as unknown, and the \mathbf{y}_2 measurements may then give valuable information about \mathbf{d} and indirectly also about \mathbf{y}_1 . In the following analysis we assume that $\mathbf{u}_{m,j}$ is a persistently exciting stochastic signal, and that all data are centralized, i.e., that \mathbf{d}_j , $\mathbf{u}_{m,j}$, $\mathbf{y}_{1,j}$ and $\mathbf{y}_{2,j}$ are stochastic variables with zero mean. For details about centralization and the subsequent modification of the estimator, see, e.g., Ref. [1]. We also assume observations of $\mathbf{u}_{m,j}$, $\mathbf{y}_{1,j}$ and $\mathbf{y}_{2,j}$ from an informative experiment with samples for $j = 1, 2, \dots, J$.

In order to use the Kalman filter formalism, we model \mathbf{d}_j as generated by a white noise sequence $\mathbf{e}_{1,j}$ through a pure delay system. In the same way we model the common noise part $\boldsymbol{\eta}_{c,j}$ in $\mathbf{y}_{1,j}$ and $\mathbf{y}_{2,j}$ as generated by a white noise sequence $\mathbf{e}_{2,j}$. Expressing \mathbf{y}_1 and \mathbf{y}_2 as linear combinations of $\mathbf{z} = [\mathbf{d}^T \boldsymbol{\eta}_c^T]^T$ and \mathbf{u}_m , we then arrive at the following dynamic system

$$\begin{aligned} \mathbf{z}_{j+1} &= \begin{bmatrix} \mathbf{d} \\ \boldsymbol{\eta}_c \end{bmatrix}_{j+1} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}_j = \mathbf{e}_j \\ \mathbf{y}_{1,j} &= \mathbf{L}_{11} \mathbf{z}_j + \mathbf{L}_{12} \mathbf{u}_{m,j} + \boldsymbol{\eta}_{1,j} \\ \mathbf{y}_{2,j} &= \mathbf{L}_{21} \mathbf{z}_j + \mathbf{L}_{22} \mathbf{u}_{m,j} + \boldsymbol{\eta}_{2,j}, \end{aligned} \quad (24)$$

where the detailed expressions for the \mathbf{L} matrices follow from Eqs. (23a) and (23b), and where $\boldsymbol{\eta}_{1,j}$ and $\boldsymbol{\eta}_{2,j}$ are white and independent noise sequences. This is a dynamic system as given in Eq. (1) with $\mathbf{A} = 0$, $\mathbf{B} = 0$ and $\mathbf{G} = \mathbf{I}$, and the algebraic Riccati (8) then results in

$$\mathbf{P} = \mathbf{P}_z = \mathbf{R}_e = \mathbf{E} \mathbf{e}_j \mathbf{e}_j^T. \quad (25)$$

From Eq. (7) follows that the Kalman gain related to the \mathbf{y}_2 measurements is

$$\mathbf{K}_2^{\text{OE}} = \mathbf{R}_e \mathbf{L}_{21}^T (\mathbf{L}_{21} \mathbf{R}_e \mathbf{L}_{21}^T + \mathbf{R}_{22})^{-1}, \quad (26)$$

where $\mathbf{R}_{22} = \mathbf{E} \boldsymbol{\eta}_{2,j} \boldsymbol{\eta}_{2,j}^T$. With $\mathbf{A} = \mathbf{B} = 0$ and appropriate change of notation according to Eq. (26), the

optimal current estimator (19) now gives the theoretical static estimator

$$\begin{aligned}\bar{\mathbf{y}}_{1,j/j} &= \mathbf{L}_{11}\mathbf{K}_2^{\text{OE}}(\mathbf{y}_{2,j} - \mathbf{L}_{22}\mathbf{u}_{m,j}) + \mathbf{L}_{12}\mathbf{u}_{m,j} \\ &= (\mathbf{L}_{12} - \mathbf{L}_{11}\mathbf{K}_2^{\text{OE}}\mathbf{L}_{22})\mathbf{u}_{m,j} + \mathbf{L}_{11}\mathbf{K}_2^{\text{OE}}\mathbf{y}_{2,j},\end{aligned}\quad (27)$$

or with

$$\begin{aligned}(\bar{\mathbf{y}}_{1,j/j})^T &= [\mathbf{u}_{m,j}^T \mathbf{y}_{2,j}^T] \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} \\ \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} &= \begin{bmatrix} (\mathbf{L}_{12} - \mathbf{L}_{11}\mathbf{K}_2^{\text{OE}}\mathbf{L}_{22})^T \\ (\mathbf{L}_{11}\mathbf{K}_2^{\text{OE}})^T \end{bmatrix}.\end{aligned}\quad (28)$$

Without known manipulated inputs, i.e., with $\mathbf{u}_{m,j} = 0$, we have a simplified model

$$\begin{aligned}\mathbf{z}_{j+1} &= \mathbf{e}_j \\ \mathbf{y}_{1,j} &= \mathbf{L}_1\mathbf{z}_j + \boldsymbol{\eta}_{1,j} \\ \mathbf{y}_{2,j} &= \mathbf{L}_2\mathbf{z}_j + \boldsymbol{\eta}_{2,j},\end{aligned}\quad (29)$$

resulting in the simplified theoretical estimate

$$\bar{\mathbf{y}}_{1,j/j} = \mathbf{L}_1\mathbf{R}_e\mathbf{L}_2^T(\mathbf{L}_2\mathbf{R}_e\mathbf{L}_2^T + \mathbf{R}_{22})^{-1}\mathbf{y}_{2,j},\quad (30)$$

or with $(\bar{\mathbf{y}}_{1,j/j})^T = \mathbf{y}_{2,j}^T\mathbf{B}$

$$\mathbf{B} = (\mathbf{L}_2\mathbf{R}_e\mathbf{L}_2^T + \mathbf{R}_{22})^{-1}\mathbf{L}_2\mathbf{R}_e\mathbf{L}_1^T.\quad (31)$$

In the same way as with the parameters in Eq. (19), we can find \mathbf{B}_1 and \mathbf{B}_2 in Eq. (28) or \mathbf{B} in Eq. (31) by identification of an OE model. In this special static case, however, we can also find the parameters directly as the solution to a least squares problem. Let us start by comparing the theoretical estimator Eq. (31) with a data based least squares solution. By stacking the collected data $\mathbf{y}_{1,1}^T, \mathbf{y}_{1,2}^T, \dots, \mathbf{y}_{1,N}^T$ in a data matrix \mathbf{Y}_1 and $\mathbf{y}_{2,1}^T, \mathbf{y}_{2,2}^T, \dots, \mathbf{y}_{2,N}^T$ in a data matrix \mathbf{Y}_2 , we can express the relation between \mathbf{Y}_2 and \mathbf{Y}_1 as

$$\mathbf{Y}_1 = \mathbf{Y}_2\mathbf{B} + \mathbf{E},\quad (32)$$

resulting in the least squares estimator

$$\hat{\mathbf{B}} = (\mathbf{Y}_2^T\mathbf{Y}_2)^{-1}\mathbf{Y}_2^T\mathbf{Y}_1 = \left(\frac{1}{N}\mathbf{Y}_2^T\mathbf{Y}_2\right)^{-1} \cdot \frac{1}{N}\mathbf{Y}_2^T\mathbf{Y}_1.\quad (33)$$

For a theoretical analysis we also stack $(\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_N^T)$, $(\boldsymbol{\eta}_{1,1}^T, \boldsymbol{\eta}_{1,2}^T, \dots, \boldsymbol{\eta}_{1,N}^T)$ and $(\boldsymbol{\eta}_{2,1}^T, \boldsymbol{\eta}_{2,2}^T, \dots, \boldsymbol{\eta}_{2,N}^T)$ in data matrices \mathbf{Z} , \mathbf{E}_1 and \mathbf{E}_2 , and by use of Eq. (29) and for $N \rightarrow \infty$ we will then find

$$\begin{aligned}\frac{1}{N}\mathbf{Y}_2^T\mathbf{Y}_2 &= \frac{1}{N}(\mathbf{Z}\mathbf{L}_2^T + \mathbf{E}_2)^T(\mathbf{Z}\mathbf{L}_2^T\mathbf{E}_2) \\ &\rightarrow \mathbf{L}_2\mathbf{R}_e\mathbf{L}_2^T + \mathbf{R}_{22},\end{aligned}\quad (34)$$

and

$$\begin{aligned}\frac{1}{N}\mathbf{Y}_2^T\mathbf{Y}_1 &= \frac{1}{N}(\mathbf{Z}\mathbf{L}_2^T + \mathbf{E}_2)^T(\mathbf{Z}\mathbf{L}_1^T + \mathbf{E}_1) \\ &\rightarrow \mathbf{L}_2\mathbf{R}_e\mathbf{L}_1^T.\end{aligned}\quad (35)$$

Here, we make use of the fact that $(1/N)\mathbf{Z}^T\mathbf{Z} \rightarrow \mathbf{R}_e$ and that $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are independent white noise sequences, which means that $(1/N)\mathbf{L}_2\mathbf{Z}^T\mathbf{E}_2 \rightarrow 0$, $(1/N)\mathbf{L}_2\mathbf{Z}^T\mathbf{E}_1 \rightarrow 0$, $(1/N)\mathbf{E}_2^T\mathbf{Z}\mathbf{L}_1^T \rightarrow 0$ and $(1/N)\mathbf{E}_2^T\mathbf{E}_1 \rightarrow 0$ when $N \rightarrow \infty$. By inserting Eqs. (34) and (35) into Eq. (33), we now find the estimator (31), showing that the estimators (31, 33) are asymptotically equivalent. This connection between Kalman filtering without dynamics and ordinary LSE was found also in Ref. [9], but then without the general dynamic estimator (19) as a basis, and also limited to the case were $\mathbf{L}_1 = \mathbf{I}$ (or at least invertible) and $\boldsymbol{\eta}_1 = 0$, i.e., the case were \mathbf{y}_1 are noise-free measurements of all states in the system (possibly after a similarity transformation).

In a similar although somewhat more involved way we can also show that the estimator

$$\hat{\mathbf{B}} = \left(\frac{1}{N}\begin{bmatrix} \mathbf{U}_m^T \\ \mathbf{Y}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{U}_m \\ \mathbf{Y}_2 \end{bmatrix}\right)^{-1} \frac{1}{N}\begin{bmatrix} \mathbf{U}_m^T \\ \mathbf{Y}_2^T \end{bmatrix} \mathbf{Y}_1\quad (36)$$

is asymptotically equivalent with Eq. (28).

4.3. Principal component regression

With a large number of \mathbf{y}_2 variables and a limited number of observations, the estimators (33, 36) may have very large variance. In the common case with collinear \mathbf{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 \mathbf{u}_m and \mathbf{z} . We then collect all input data in either $\mathbf{X} = [\mathbf{U}_m \mathbf{Y}_2]$ as in Eq. (36) or $\mathbf{X} = \mathbf{Y}_2$ as in Eq. (33), dependent

on the problem formulation. By use of an appropriate number of principal components [1,2], the data is then expressed as

$$\mathbf{X} \approx \mathbf{T}\mathbf{P}^T, \quad (37)$$

where \mathbf{T} is the score matrix and \mathbf{P} is the loading matrix.

For convenience and due to space limitations we now limit the treatment to the case where $\mathbf{u}_{m,j} = 0$, i.e., to the case where $\mathbf{X} = [\mathbf{y}_{2,1}, \mathbf{y}_{2,2}, \dots, \mathbf{y}_{2,N}]^T$. We then replace the measured variables $\mathbf{y}_{2,j}$ with latent variables $\boldsymbol{\tau}_j = \mathbf{P}^T \mathbf{y}_{2,j}$, and make use of the fact that $\mathbf{P}^T \mathbf{P} = \mathbf{I}$, and the system (29) is thus replaced by

$$\begin{aligned} \mathbf{z}_{j+1} &= \mathbf{e}_j \\ \mathbf{y}_{1,j} &= \mathbf{L}_1 \mathbf{z}_j + \boldsymbol{\eta}_{1,j} \\ \boldsymbol{\tau}_j &\approx \mathbf{P}^T \mathbf{L}_2 \mathbf{z}_j + \mathbf{P}^T \boldsymbol{\eta}_{2,j}. \end{aligned} \quad (38)$$

The theoretical estimator (31) is then replaced by

$$\mathbf{B} = \mathbf{P}(\mathbf{P}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_2^T \mathbf{P} + \mathbf{P}^T \mathbf{R}_{22} \mathbf{P})^{-1} \mathbf{P}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_1^T. \quad (39)$$

By replacing \mathbf{y}_2 with \mathbf{T} and inserting Eqs. (34) and (35) into Eq. (39), we find the corresponding data based PCR estimator

$$\hat{\mathbf{B}} = \mathbf{P}(\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y}_1 = \mathbf{P}(\mathbf{P}^T \mathbf{X}^T \mathbf{X} \mathbf{P})^{-1} \mathbf{P}^T \mathbf{X}^T \mathbf{Y}_1. \quad (40)$$

4.4. Partial least squares regression

The aim of PLSR is to improve PCR by finding $\boldsymbol{\tau}$ variables that explain both the \mathbf{X} and the \mathbf{Y}_1 data, and there exist at least two slightly different PLSR algorithms [1]. Also here we limit the treatment to the case where $\mathbf{u}_{m,j} = 0$, and it is convenient to start with the PLSR method of Martens that makes use of linear combinations $\boldsymbol{\tau}_M = \mathbf{W}^T \mathbf{y}_2$ (where the weight matrix \mathbf{W} is found iteratively). The result of this is that Eq. (29) is replaced by the PLSR_M model

$$\begin{aligned} \mathbf{z}_{j+1} &= \mathbf{e}_j \\ \mathbf{y}_{1,j} &= \mathbf{L}_1 \mathbf{z}_j + \boldsymbol{\eta}_{1,j} \\ \boldsymbol{\tau}_{M,j} &\approx \mathbf{W}^T \mathbf{L}_2 \mathbf{z}_j + \mathbf{W}^T \boldsymbol{\eta}_{2,j}. \end{aligned} \quad (41)$$

The theoretical PCR estimator (39) is then replaced by the theoretical PLSR estimator

$$\begin{aligned} \mathbf{B} &= \mathbf{W}(\mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_2^T \mathbf{W} + \mathbf{W}^T \mathbf{R}_{22} \mathbf{W})^{-1} \\ &\quad \times \mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_1^T, \end{aligned} \quad (42)$$

while Eq. (40) is replaced by the data based PLSR estimator

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{W}(\mathbf{T}_M^T \mathbf{T}_M)^{-1} \mathbf{T}_M^T \mathbf{Y}_1 \\ &= \mathbf{W}(\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{Y}_1. \end{aligned} \quad (43)$$

The original PLSR method of Wold uses linear combinations $\boldsymbol{\tau}_W = (\mathbf{W}^T \mathbf{P}_W)^{-1} \mathbf{W}^T \mathbf{y}_2$, with the same \mathbf{W} matrix as Martens and with a special loading matrix \mathbf{P}_W . The model (29) is then replaced by the PLSR_W model

$$\begin{aligned} \mathbf{z}_{j+1} &= \mathbf{e}_j \\ \mathbf{y}_{1,j} &= \mathbf{L}_1 \mathbf{z}_j + \boldsymbol{\eta}_{1,j} \\ \boldsymbol{\tau}_{W,j} &\approx (\mathbf{W}^T \mathbf{P}_W)^{-1} \mathbf{W}^T \mathbf{L}_2 \mathbf{z}_j + (\mathbf{W}^T \mathbf{P}_W)^{-1} \mathbf{W}^T \boldsymbol{\eta}_{2,j}. \end{aligned} \quad (44)$$

With $\mathbf{W}(\mathbf{W}^T \mathbf{P}_W)^{-1}$ instead of \mathbf{W} , the theoretical PLSR estimator (42) becomes

$$\begin{aligned} \mathbf{B} &= \mathbf{W}(\cdot)^{-T} \left((\cdot)^{-1} \mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_2^T \mathbf{W}(\cdot)^{-T} + (\cdot)^{-1} \mathbf{W}^T \mathbf{R}_{22} \mathbf{W}(\cdot)^{-T} \right)^{-1} \\ &\quad \cdot (\cdot)^{-1} \mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_1^T \\ &= \mathbf{W}(\mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_2^T \mathbf{W} + \mathbf{W}^T \mathbf{R}_{22} \mathbf{W})^{-1} \\ &\quad \cdot \mathbf{W}^T \mathbf{L}_2 \mathbf{R}_e \mathbf{L}_1^T, \end{aligned} \quad (45)$$

while the data based PLSR estimator (43) becomes

$$\begin{aligned} \hat{\mathbf{B}} &= \mathbf{W}(\cdot)^{-T} \left((\cdot)^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W}(\cdot)^{-T} \right)^{-1} \\ &\quad \cdot (\cdot)^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{Y}_1 \\ &= \mathbf{W}(\mathbf{P}_W^T \mathbf{W})^{-1} (\mathbf{T}_W^T \mathbf{T}_W)^{-1} \mathbf{T}_W^T \mathbf{y} \\ &= \mathbf{W}(\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{Y}_1. \end{aligned} \quad (46)$$

We see from this that \mathbf{P}_W disappears from the estimator expressions, and that the final theoretical as well as the data based estimators are the same for the Wold and Martens algorithms. This is, of course, well known [1], although the relation to the underlying Kalman filter is new.

4.5. Dynamic system PCR and PLSR solutions

The optimal \mathbf{y}_1 estimator for dynamic systems given in Eq. (19) may also form a basis for dynamic system solutions using PCR or PLSR. It is then natural to split the secondary measurements into $\mathbf{y}_{2,k} = [\mathbf{y}_{21,k}^T \mathbf{y}_{22,k}^T]^T$, where $\mathbf{y}_{21,k}$ are the secondary measurements that are linked to \mathbf{y}_1 only through a static system. The $\mathbf{y}_{21,k}$ measurements are then internally collinear, and they can thus be replaced by latent $\boldsymbol{\tau}$ variables as in Eqs. (38), (41) and (44), i.e., both PCR and PLSR may be used. Using, e.g., the score definition in the PLSR method of Martens, i.e., $\boldsymbol{\tau} = \mathbf{W}^T \mathbf{y}_2$, the OEC estimator (19) will be replaced by

$$\begin{aligned} \bar{\mathbf{y}}_{1,k|k} = & \mathbf{C}_1 (\mathbf{I} - \mathbf{K}_\tau^{\text{OE}} \mathbf{W}^T \mathbf{C}_{21} - \mathbf{K}_{22}^{\text{OE}} \mathbf{C}_{22}) \\ & \cdot (q\mathbf{I} - \mathbf{A}_{\text{est}})^{-1} \cdot (\mathbf{B}_{\text{est}} \mathbf{u}_k + \mathbf{A} \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} \boldsymbol{\tau}_k \\ & + \mathbf{A} \mathbf{K}_{22}^{\text{OE}} \mathbf{y}_{22,k}) + \mathbf{C}_1 \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} (\boldsymbol{\tau}_k - \mathbf{W}^T \mathbf{D}_{21} \mathbf{u}_k) \\ & + \mathbf{C}_1 \mathbf{K}_{22}^{\text{OE}} (\mathbf{y}_{22,k} - \mathbf{D}_{22} \mathbf{u}_k) + \mathbf{D}_1 \mathbf{u}_k, \end{aligned} \quad (47)$$

where $\boldsymbol{\tau}_k = \mathbf{W}^T \mathbf{y}_{21,k}$, $\mathbf{A}_{\text{est}} = \mathbf{A} (\mathbf{I} - \mathbf{K}_\tau^{\text{OE}} \mathbf{W}^T \mathbf{C}_{21} - \mathbf{K}_{22}^{\text{OE}} \mathbf{C}_{22})$ and $\mathbf{B}_{\text{est}} = \mathbf{B} - \mathbf{A} \mathbf{K}_\tau^{\text{OE}} \mathbf{W}^T \mathbf{D}_{21} - \mathbf{A} \mathbf{K}_{22}^{\text{OE}} \mathbf{D}_{22}$. The Kalman gains are here determined as the solution to the Kalman filter (7, 8) with

$$\begin{aligned} \mathbf{C}_2 = & \left[(\mathbf{W}^T \mathbf{C}_{21})^T \mathbf{C}_{22}^T \right]^T \text{ and} \\ \mathbf{R}_{22} = & \begin{bmatrix} \mathbf{E} \mathbf{W}^T \mathbf{w}_{21} \mathbf{w}_{21}^T \mathbf{W} & \mathbf{E} \mathbf{W}^T \mathbf{w}_{21} \mathbf{w}_{22}^T \\ \mathbf{E} \mathbf{w}_{22} \mathbf{w}_{21}^T \mathbf{W} & \mathbf{E} \mathbf{w}_{22} \mathbf{w}_{22}^T \end{bmatrix}. \end{aligned}$$

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

When the current estimator (47) is identified by use of, e.g., a prediction error method, also past $\boldsymbol{\tau}_k$ values will be used as a basis for determining $\bar{\mathbf{y}}_{1,k|k}$, with reduced variance as the expected result, and we can in fact look at and treat the latent variables as ordinary measurement signals. An essential assumption is here that the linear relations between $\mathbf{y}_{21,k}$ and $\boldsymbol{\tau}_k$ given by \mathbf{P}^T , \mathbf{W}^T or $(\mathbf{W}^T \mathbf{P}_w)^{-1} \mathbf{W}^T$ 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 estimator (19).

If all or some of the \mathbf{y}_{22} measurements follow the same dynamic response except for noise, and thus are

internally collinear, such measurements may also be replaced by latent variables in order to reduce the variance in the solution. However, since \mathbf{y}_{22} is linked to \mathbf{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 or other measurements, and also with other latent variables found by PCA or PLSR.

With $\mathbf{u}_k = 0$ and $\mathbf{y}_{22,k} = 0$, Eq. (47) is simplified to

$$\begin{aligned} \bar{\mathbf{y}}_{1,k|k} = & \mathbf{C}_1 (\mathbf{I} - \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} \boldsymbol{\tau}^T \mathbf{C}_{21}) \\ & \cdot (q\mathbf{I} - \mathbf{A} (\mathbf{I} - \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} \boldsymbol{\tau}^T \mathbf{C}_{21}))^{-1} \\ & \cdot \mathbf{A} \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} \boldsymbol{\tau}_k + \mathbf{C}_1 \mathbf{K}_\tau^{\text{OE}} \boldsymbol{\tau} \boldsymbol{\tau}_k, \end{aligned} \quad (48)$$

showing the dynamic relation between the collinear time series \mathbf{y}_{21} represented by $\boldsymbol{\tau}$ and the time series \mathbf{y}_1 .

We end this subsection with a general discussion on dynamic system multivariate calibration methods. The proposed estimator (47) is based on the asymptotically optimal estimator (19). This is in contrast with PCR and PLSR methods for identification of FIR models [4], where also the asymptotical least squares solution is biased due to truncation as well as lack of noise modeling [3]. It is also in contrast with PLSR methods for identification of ARX models [5], where again the asymptotical least squares solution is biased when the observation error is colored [3,6].

In addition we must consider the fact that an ARX estimator would make use of past \mathbf{y}_1 values that are not available as the present problem is formulated in Section 3.1. As for the optimal ARMAX estimator (3), an ARX estimator would then not utilize secondary \mathbf{y}_2 information in an optimal way. One obvious effect of this would be that noisy \mathbf{y}_{21} measurements collinear with \mathbf{y}_1 would be effectively ignored when the identification experiment gives low noise \mathbf{y}_1 information.

The fact that FIR and ARX least squares solutions are not asymptotically optimal does not mean, of course, that the PCR and PLSR solutions presented in, e.g., Refs. [4,5] may not give good results in some realistic cases with a limited number of observations. An in-depth comparison between such known solu-

tions and the proposed estimator (47) is, however, beyond the scope of the present paper.

5. Simulation examples

Simulation studies are undertaken by use of MATLAB, primarily the *dlsim.m* function in the control system toolbox [13], and the prediction error method implemented in the *pem.m* function in the SI toolbox [14]. With an appropriate OE model specified, the *pem.m* function identifies the optimal current estimator (19), where the secondary measurements \mathbf{y}_2 are also used as input signals. For validation comparisons, the RMSE criterion in Eq. (22) was used, with $\bar{\mathbf{y}}_{1,k|k}$ replaced by the appropriate estimate.

5.1. Example 1: a second-order system with a first-order process noise model

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):

$$\mathbf{x} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mathbf{u} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mathbf{v} \quad (49)$$

$$\mathbf{y}_1 = [1 \ 0 \ 0] \mathbf{x} + \mathbf{w}_1$$

$$\mathbf{y}_2 = [0 \ 1 \ 0] \mathbf{x} + \mathbf{w}_2.$$

The system was discretized assuming zero-order hold elements on the \mathbf{u} and \mathbf{v} inputs and a sampling interval $\mathbf{T} = 0.1$, and then simulated with \mathbf{u}_k as a filtered PRBS signal with autocovariance $r_{\mathbf{u}\mathbf{u}}(\mathbf{p}) = 0.8^{|\mathbf{p}|}$ (in Ref. [3], example 5.11 with $\alpha = 0.8$), e.g., an input that was persistently exciting of sufficient order. The noise sources \mathbf{v}_k , $\mathbf{w}_{1,k}$ and $\mathbf{w}_{2,k}$ were independent and normally distributed white noise sequences with zero mean and variances given below.

The simulated system was identified using the OEP and OEC models (10, 16) with \mathbf{u}_k and $\mathbf{y}_{2,k}$ as input signals and $\mathbf{y}_{1,k}$ as output signal, using $N = 10\,000$ samples.

The OEP model (10) was specified as

$$\mathbf{nn}_{\text{OEP}} = [0, [3, 3], 0, 0, [3, 3], [1, 1]], \quad (50)$$

i.e., a model

$$\mathbf{y}_{1,k} = \frac{B_1(q^{-1})}{F_1(q^{-1})} \mathbf{u}_k + \frac{B_2(q^{-1})}{F_2(q^{-1})} \mathbf{y}_{2,k} + \mathbf{e}_k^{\text{OEP}} \quad (51)$$

with

$$B_1(q^{-1}) = b_{11}q^{-1} + b_{12}q^{-2} + b_{13}q^{-3} \quad (52)$$

$$B_2(q^{-1}) = b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3} \quad (53)$$

$$F_1(q^{-1}) = 1 + f_{11}q^{-1} + f_{12}q^{-2} + f_{13}q^{-3} \quad (54)$$

$$F_2(q^{-1}) = 1 + f_{21}q^{-1} + f_{22}q^{-2} + f_{23}q^{-3}. \quad (55)$$

The OEC model (Eq. (16)) was specified as

$$\mathbf{nn}_{\text{OEC}} = [0, [3, 4], 0, 0, [3, 3], [1, 0]], \quad (56)$$

i.e., the same model as Eq. (51), but with $B_2(q^{-1})$ altered to

$$B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}. \quad (57)$$

As the main purpose of the simulations was to verify 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 [7], and a systematic method for finding the structure is presented in Ref. [10]. For the OEP and OEC models, no attempt was made to force $F_1(q^{-1})$ and $F_2(q^{-1})$ to be identical, which they theoretically should be.

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 $J = 5$ times with randomized initial B parameters ($\mathbf{b}_{i,j+1} = \mathbf{b}_{i,j} \times (1 + 0.5\mathbf{e})$, with \mathbf{e} as a normal random variable with zero mean and variance 1).

The mean RMSE values and RMSE standard deviations for $N = 10\,000$, $\mathbf{r}_v = 0.1$, $\mathbf{r}_{22} = 0.01$ and varying \mathbf{r}_{11} values are given in Table 1, showing an obvious agreement between results based on simulation and theory. Table 1 also includes theoretical RMSE values $\sqrt{\text{Var}(\bar{\mathbf{y}}_{1,k|k-1})} = \sqrt{\mathbf{C}_1 \mathbf{P}^{\text{OEP}} \mathbf{C}_1^T + \mathbf{r}_{11}}$

Table 1

Validation RMSE mean values and standard deviations and theoretical mean values for OE models (multiplied with 10000)

\mathbf{r}_{11}	OEP	OEP _{th.}	OEC	OEC _{th.}
10^{-8}	177 ± 5	177	173 ± 6	173
10^{-6}	177 ± 5	177	173 ± 5	173
10^{-4}	204 ± 6	203	200 ± 5	200

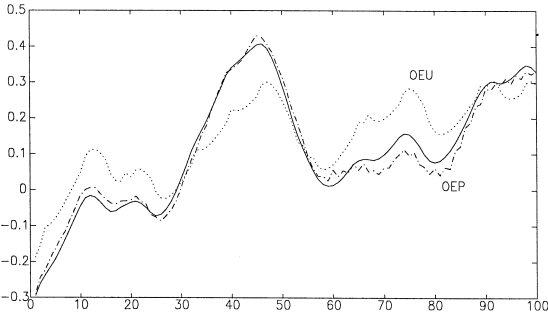


Fig. 2. Segment of validation responses for the OEP model (Eq. (51)) using both \mathbf{u} and \mathbf{y}_2 as inputs (dashed, RMSE = 0.0273) and an OE model using only \mathbf{u} as input ($\mathbf{nn}_{\text{OEU}} = [0,3,0,0,3,1]$, dotted, RMSE = 0.0906). The experimental conditions are given by $\mathbf{r}_v = 1$, $\mathbf{r}_{11} = 0.0001$, $\mathbf{r}_{22} = 0.01$ and $N = 200$, and the ideal validation response is shown by solid line.

and $\sqrt{\text{Var}(\bar{\mathbf{y}}_{1,k|k})} = \sqrt{\mathbf{C}_1 \mathbf{P}^{\text{OEC}} \mathbf{C}_1^T + \mathbf{r}_{11}}$ with \mathbf{P}^{OEP} and \mathbf{P}^{OEC} computed according to Eqs. (8) and (20).

The results in Table 1 were obtained from $N = 10\,000$ samples. To indicate expected results for a more realistic number of samples, and at the same time visualize the degree of model misfit behind the RMSE values in Table 1, specific validation responses for models based on $N = 200$ samples are shown in Fig. 2. Fig. 2 also gives a representative picture of the improvement achieved by including \mathbf{y}_2 as an input signal in addition to \mathbf{u} .

5.2. Example 2: dynamic system PCR and PLSR solutions

For simulations of the dynamical DPCR and DPLSR solutions in Section 5, three independent filtered white noise sequences were generated. The following continuous system of three independent second order systems was used as a starting point (with $a = -1$):

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} a & 0 & 0 & 1 & 0 & 0 \\ 0 & a & 0 & 0 & 1 & 0 \\ 0 & 0 & a & 0 & 0 & 1 \\ 0 & 0 & 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & a \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \\ \mathbf{y}_1 &= [1 \ 1 \ 1 \ 0 \ 0 \ 0] \mathbf{x} + \mathbf{w}_1 \\ \mathbf{y}_2 &= [\mathbf{L}_{21} \ 0] \mathbf{x} + \mathbf{w}_2. \end{aligned} \quad (58)$$

Here, \mathbf{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 \mathbf{v} inputs and a sampling interval $\mathbf{T} = 0.1$. The system was then simulated with \mathbf{v} , \mathbf{w}_1 and \mathbf{w}_2 as independent and normally distributed white noise sequences with zero mean. The \mathbf{R}_v and \mathbf{R}_{22} covariance matrices were diagonal, with uniformly distributed random parameters in the intervals (0,1) and (0, \mathbf{r}_{22}), respectively, while the \mathbf{y}_1 variance was $\mathbf{r}_{11} = 0.0001$. Different values of \mathbf{r}_{22} were used as described below.

The simulations started with $\mathbf{r}_{22} = 0.01$ and $N = 200$. In order to find the appropriate number of components, the static PCR and PLSR estimators (40, 46) were first determined for different numbers of components \mathbf{A} . In addition the dynamical DPCR and DPLSR estimates according to Eq. (48) were identified using the OE model (see Ref. [14] for definition of \mathbf{nn})

$$\mathbf{nn} = [0, [2, \dots, 2], 0, 0, [2, \dots, 2], [0, \dots, 0]]. \quad (59)$$

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, and there we find the optimal number of components $\mathbf{A} = 3$. This is not surprising since the system has three independent noise sources. Fig. 3 also indicates that PLSR is slightly better than PCR, and that the dynamic solutions are better than the static ones.

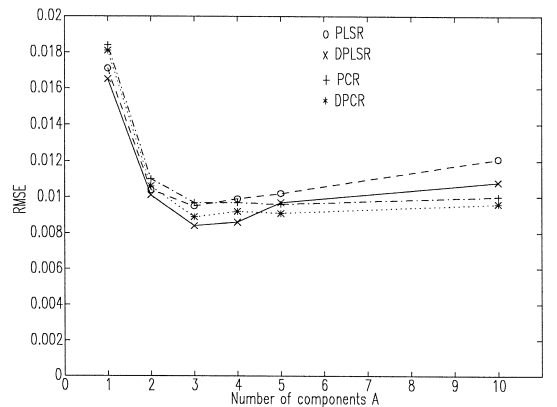


Fig. 3. RMSE mean values as function of number of components used in PCR, PLSR, DPCR and DPLSR models for $\mathbf{r}_{22} = 0.01$, based on 10 Monte Carlo runs with $N = 200$ samples.

Table 2
Validation RMSE mean values and standard deviations, based on $N = 200$ samples and $\mathbf{A} = 3$ components

\mathbf{r}_{22}	$\overline{\text{RMSE}}_{\text{PLSR}}$	$\overline{\text{RMSE}}_{\text{DPLSR}}$
0.001	0.00301 ± 0.00028	0.00305 ± 0.00020
0.01	0.0097 ± 0.0006	0.0086 ± 0.0007
0.1	0.0318 ± 0.0035	0.0257 ± 0.0054

The simulations were repeated using also $\mathbf{r}_{22} = 0.001$ and $\mathbf{r}_{22} = 0.1$. Mean validation results and standard deviations based on $M = 100$ Monte Carlo runs with $\mathbf{A} = 3$ are given in Table 2, indicating that the improvement obtained by using the dynamical model increases with increasing \mathbf{y}_2 noise level. At the same time the explained proportion of sample variance decreases (from 0.95 with $\mathbf{r}_{22} = 0.001$ to 0.19 with $\mathbf{r}_{22} = 0.1$). For $\mathbf{r}_{22} = 0.1$ the optimal number of PLSR components are in fact $\mathbf{A} = 2$, with slightly reduced RMSE values as compared with use of $\mathbf{A} = 3$.

6. Concluding remarks

An optimal estimator for nonmeasured primary output variables \mathbf{y}_1 from, e.g., an industrial plant is developed. The estimator utilizes all available information, also in secondary output measurements \mathbf{y}_2 , and the theoretical results are verified through simulation of a simple system. Industrial applications in, e.g., polymer extruding are now investigated, with promising preliminary results.

It is further shown that the general dynamic optimal estimator in the special static case results in a least squares estimator valid for problems where the estimator variables includes noisy measurements. This is also extended to cover PCR and 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.

Finally, it is indicated how dynamic system PCR (DPCR) and PLSR (DPLSR) solutions can be

achieved as special cases of the general dynamic system estimator. As demonstrated in a simple simulation example, this may result in a considerable reduction of the \mathbf{y}_1 estimation covariance, compared with ordinary PCR and PLSR. The full potential of this, theoretically and in practical applications, are open questions. A comparison with known PCR and PLSR methods for identification of dynamical FIR and ARX models is left for further research.

References

- [1] H. Martens, T. Næs, Multivariate Calibration, Wiley, Chichester, 1989.
- [2] A. Höskuldsson, Prediction Methods in Science and Technology, Thor Publishing, Denmark, 1996.
- [3] T. Söderström, P. Stoica, System Identification, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [4] B.M. Wise, N.L. Ricker, Identification of finite impulse response models with continuum regression, Journal of Chemometrics 7 (1) (1993) .
- [5] B.S. Dayal, J.F. MacGregor, Recursive exponentially weighted PLS and its applications to adaptive control and prediction, Journal of Process Control 7 (3) (1997) .
- [6] L. Ljung, System Identification, Theory for the User, Prentice-Hall, Englewood Cliffs, NJ, 1987.
- [7] D. Di Ruscio, Combined deterministic and stochastic system identification and realization: DSR-a subspace approach based on observations, Modeling, Identification and Control 17 (1996) 193–230.
- [8] F.L. Lewis, Optimal Estimation, With An Introduction To Stochastic Control Theory, Wiley, Chichester, 1986.
- [9] H. Berntsen, Utvidet Kalmanfilter og multivariabel kalibrering, Report STF48 A88019, SINTEF, Trondheim, Norway, 1988.
- [10] R. Ergon, D. Di Ruscio, Dynamic system calibration by system identification methods, The European Control Conference, Brussels, 1997.
- [11] R. Ergon, On the use of secondary measurements as input signals in system identification, scheduled for IEEE Transactions on Automatic Control, April 1999.
- [12] G.C. Goodwin, R.L. Payne, Dynamic System Identification, Academic Press, New York, 1977.
- [13] A. Grace et al., Control System Toolbox For Use With MATLAB, The MathWorks, 1992.
- [14] L. Ljung, System Identification Toolbox For Use With MATLAB, The MathWorks, 1991.