On Primary Output Estimation by Use of Secondary Measurements as Input Signals in System Identification

Rolf Ergon

Abstract— **In many cases, vital output variables in, e.g., industrial processes cannot be measured online. It is then of interest to estimate these primary variables from manipulated and measured inputs and the secondary output measurements that are available. In order to identify an optimal estimator from input–output data, a suitable model structure must be chosen. The paper compares use of ARMAX and output error (OE) structures in prediction error identification methods, theoretically and through simulations.**

Index Terms—**Estimation, product quality, system identification.**

I. Introduction

An important use of system identification methods is to find models for estimation of primary output variables y_1 that are not normally available online. In such cases all available information should be utilized, including secondary measurements y_2. A typical industrial application would be estimation of a product quality y_1 from manipulated inputs u_m, measured disturbances u_d, and available process measurements y_2. The practical use of the estimated y_1 output variables may be operator support, failure detection, and possibly closed-loop control.

From a system identification point of view, it is very natural to include the secondary measurements as input signals [1]. The basic idea in the present context is that for output estimation purposes, knowledge of the system model as such is not necessary. What is needed are the dynamical relations between the known input signals $u = [u_m^T \;\; u_d^T]^T$, the available secondary measurements y_2, and the primary output variables y_1, and these relations can often be identified with better accuracy than the relations between u and y_1 alone. The reason for this is that disturbances and noise entering early in the system will be indirectly measured by the secondary measurements later in the system. Here we assume, of course, that a representative data record of sufficient length and including also y_1 is available from an informative identification experiment.

The use of dependent y_2 variables as inputs to a system identification procedure raises several questions concerning identifiability, deterministic systems, and perfect measurement systems, and these topics are treated in [2]. In the present paper we assume a discrete-time system that is observable from the y_2 measurements. We then assume a prediction error identification method and compare identified Auto Regressive Moving Average with eXogenous inputs (ARMAX) and Output Error (OE) models using u and y_2 as inputs. It is shown that use of the OE structure asymptotically will result in optimal y_1 estimators giving minimized estimation covariance. The ARMAX structure will not give minimized estimation covariance due to the fact that past y_1 values are not available as a basis for the y_1 estimation, although such values are used in the system identification procedure. The result of this is that the y_2 information is not optimally utilized in the y_1 estimator.

A simulation example that supports the theoretical results is also presented.

Manuscript received December 2, 1996; revised October 1, 1997. Recommended by Associate Editor, J. C. Spall. The author is with the Telemark Institute of Technology, N-3914 Porsgrunn, Norway.

Publisher Item Identifier S 0018-9286(99)02102-9.

II. Theory

A. Statement of Problem

Consider the discrete-time system model

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gv_k \\ y_{1,k} &= C_1x_k + D_1u_k + w_{1,k} \\ y_{2,k} &= C_2x_k + D_2u_k + w_{2,k} \end{aligned} \tag{1}$$

where x_k is the state vector, while v_k and $w_k = [w_{1,k}^T \;\; w_{2,k}^T]^T$ are white, independent, and normal process and measurement noise vectors with covariance matrices $R_v = Ev_kv_k^T$ and $R_w = Ew_kw_k^T = \left[\begin{smallmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{smallmatrix}\right]$. Also assume that (C_2, A) is observable and that $(A, G\sqrt{R_v})$ is stabilizable. The assumptions of noise independence and state observability may be relaxed with appropriate theoretical modifications. This is, however, beyond the scope of the present paper.

Further assume that input–output data is available from an informative experiment [3], i.e., that data records for u_k, $y_{1,k}$, and $y_{2,k}$ for $k = 1, 2, \cdots, N$ are at hand, with u_k persistently exciting of appropriate order. The problem is now to identify the optimal one-step-ahead $\bar{y}_{1,k|k-1}$ prediction estimator based on past and present u_k and past $y_{2,k}$ values, and the optimal $\bar{y}_{1,k|k}$ current estimator based also on present $y_{2,k}$ values.

Note that it is a part of the problem that past $y_{1,k}$ values are not available as a basis for the estimates. 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.

B. Preliminary Discussion

In the following, three different estimation models will be discussed. Subsection II-C assumes identification of an ARMAX model using both y_1 and y_2 as outputs. The resulting one-step-ahead predictor is then clearly not optimal when past y_1 values are not available.

Subsection II-D discusses the use of ARMAX models of the form

$$Ay_{1,k} = B_1u_k + B_2y_{2,k} + Ce_{1,k} \tag{2}$$

where $A = A(q^{-1})$ etc. are matrix polynomials in the unit delay operator q^{-1}, and where $e_{1,k}$ is an innovation process in an underlying Kalman filter. Such a model can be constructed after identification of the model used in Section II-C, or alternatively directly identified by use of y_2 as an input signal as shown in Subsection II-D. The innovation $e_{1,k}$ will in general be correlated with $y_{2,k}$, and thus

$$\bar{y}_{1,k|k-1} = A^{-1}B_1u_k + A^{-1}B_2y_{2,k} \tag{3}$$

will not in general be the optimal predictor given only past and present inputs u_k and past secondary outputs $y_{2,k}$.

Subsection II-E discusses identification of an OE model

$$y_{1,k} = F^{-1}B_1u_k + F^{-1}B_2y_{2,k} + \vartheta_k \tag{4}$$

where ϑ_k is colored noise, and where $y_{2,k}$ is used as input signal. Although ϑ_k here is correlated with $y_{2,k}$, the result will still be an optimal predictor. The reason for this is that the expectation $E\vartheta_k\vartheta_k^T$ is minimized when and only when the correct parameters are found.

TABLE I
VALIDATION RMSE MEAN VALUES WITH STANDARD DEVIATIONS AND THEORETICAL MEAN VALUES FOR ARMAX2 MODEL (MULTIPLIED BY 10 000)

r_{11}	$ARMAX_2$	$ARMAX_{2,theor.}$
10^{-8}	352 ± 15	354
10^{-7}	341 ± 19	341
10^{-6}	312 ± 17	315
10^{-5}	278 ± 15	276
10^{-4}	256 ± 13	247
10^{-3}	372 ± 5	369

TABLE II
VALIDATION RMSE MEAN VALUES WITH STANDARD DEVIATIONS AND THEORETICAL MEAN VALUES FOR OE MODELS (MULTIPLIED BY 10 000)

r_{11}	OEP	$OEP_{th.}$	OEC	$OEC_{th.}$
10^{-8}	177 ± 5	177	173 ± 6	173
10^{-7}	177 ± 5	177	173 ± 6	173
10^{-6}	177 ± 5	177	173 ± 5	173
10^{-5}	181 ± 5	180	177 ± 5	176
10^{-4}	204 ± 6	203	200 ± 5	200
10^{-3}	363 ± 4	362	361 ± 3	360

C. ARMAX Model with y_2 As Output

System (1) can be expressed in the ordinary innovation form [4], based on an underlying Kalman filter driven by u and the y_1 and y_2 measurements. This form is given by the following equations, where $K = [K_1 \;\; K_2]$ is the predictor-corrector Kalman gain, and where $e_{1,k}$ and $e_{2,k}$ are white innovation processes

$$\begin{aligned} \bar{x}_{k+1} &= A\bar{x}_k + Bu_k + [AK_1 \quad AK_2]\begin{bmatrix} e_{1,k} \\ e_{2,k} \end{bmatrix} \\ y_{1,k} &= C_1\bar{x}_k + D_1u_k + e_{1,k} \\ y_{2,k} &= C_2\bar{x}_k + D_2u_k + e_{2,k}. \end{aligned} \tag{5}$$

In a prediction error identification method with u_k as input and $y_{1,k}$ and $y_{2,k}$ as outputs, the predictor would asymptotically ($N \rightarrow \infty$) and after minimization of an appropriate criterion function [4] become

$$\begin{aligned} \bar{x}_{k+1} &= \tilde{A}\bar{x}_k + \tilde{B}u_k + AK_1y_{1,k} + AK_2y_{2,k} \\ \bar{y}_{1,k} &= C_1\bar{x}_k + D_1u_k \\ \bar{y}_{2,k} &= C_2\bar{x}_k + D_2u_k \end{aligned} \tag{6}$$

where $\tilde{A} = A - AK_1C_1 - AK_2C_2$ and $\tilde{B} = B - AK_1D_1 - AK_2D_2$. This is 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 are normally distributed [5].

Once the model (5) is identified, and assuming normal statistics, the optimal one-step-ahead prediction of $y_{1,k}$ based on u_k and past $y_{1,k}$ and $y_{2,k}$ measurements could be constructed as

$$\bar{y}_{1,k|k-1} = C_1[qI - \tilde{A}]^{-1} \cdot [\tilde{B}u_k + AK_1y_{1,k} + AK_2y_{2,k}] + D_1u_k. \tag{7}$$

When past outputs $y_{1,k}$ are not available, i.e., with $y_{1,k} = 0$, the information in $y_{2,k}$ will not be utilized in an optimal way. A simple example occurs when $C_1 \equiv C_2$ and $D_1 \equiv D_2$, i.e., when the $y_{1,k}$ and $y_{2,k}$ outputs are identical except for the noise term. Then perfect y_1 measurements, i.e., $R_{11} \rightarrow 0$, would result in $K_2 \rightarrow 0$. With $y_{1,k} = 0$, the predictor (7) would thus be based almost entirely on the information in u_k, also if $y_{2,k}$ was obtained at a low measurement noise level.

D. ARMAX Model with y_2 as Input

A different choice when y_1 is not available as a basis for estimation would be to set also $K_1 = 0$, i.e., to assume an underlying observer driven only by u and y_2. The one-step-ahead predictor (7) would then be modified into

$$\begin{aligned} \bar{y}_{1,k|k-1}^{\text{ARMAX}_2} &= C_1[qI - A + AK_2C_2]^{-1} \\ &\quad \cdot [(B - AK_2D_2)u_k + AK_2y_{2,k}] + D_1u_k. \end{aligned} \tag{8}$$

This is a predictor of the form given in (3) and thus not optimal. The underlying ARMAX form (2) is here obtained by elimination of $e_{2,k}$ in the state equation in (5).

Assuming that (C_2, A) is observable, the state estimation error $\tilde{x}_k = x_k - \bar{x}_k^{\text{ARMAX}_2}$ in the underlying nonoptimal observer would be governed by

$$\tilde{x}_{k+1} = (A - AK_2C_2)\tilde{x}_k + Gv_k - AK_2w_{2,k} \tag{9}$$

resulting in the asymptotic prediction covariance

$$\begin{aligned} \text{Cov}\left(\bar{y}_{1,\text{theor.}}^{\text{ARMAX}_2}\right) &= E\left(y_{1,k} - \bar{y}_{1,k|k-1}^{\text{ARMAX}_2}\right) \cdot \left(y_{1,k} - \bar{y}_{1,k|k-1}^{\text{ARMAX}_2}\right)^T \\ &= C_1P^{\text{ARMAX}_2}C_1^T + R_{11} \end{aligned} \tag{10}$$

where $P^{\text{ARMAX}_2} = E\tilde{x}_k\tilde{x}_k^T$ is determined by (9) through the Lyapunov equation

$$\begin{aligned} P^{\text{ARMAX}_2} &= (A - AK_2C_2)P^{\text{ARMAX}_2} \cdot (A - AK_2C_2)^T \\ &\quad + GR_vG^T + AK_2R_{22}K_2^TA^T. \end{aligned} \tag{11}$$

Since (10) is a sum of nonnegative terms, it is evident that $\text{Cov}(\bar{y}_{1,\text{theor.}}^{\text{ARMAX}_2})$ is minimized only when P^{ARMAX_2} is minimized, which requires an optimal gain K_2. This will be obtained only when the prediction is based on an underlying Kalman filter driven only by u and y_2 and not also by y_1 (see also Subsection II-E).

The estimator (8) may be constructed after identification of (5). For complex systems with a number of secondary y_2 measurements, however, identification of (5) is a difficult task [1], involving minimization of, e.g., the criterion function $V_N(\hat{\theta}) = \text{tr}(\frac{1}{N}\sum \varepsilon_{1,k}\varepsilon_{1,k}^T) + \text{tr}(\frac{1}{N}\sum \varepsilon_{2,k}\varepsilon_{2,k}^T)$, where $\varepsilon_{1,k} = y_{1,k} - \hat{y}_{1,k}$ and $\varepsilon_{2,k} = y_{2,k} - \hat{y}_{2,k}$. Here, $\hat{y}_{1,k}$ and $\hat{y}_{2,k}$ are determined by (6) with A, AK, etc. replaced by estimates $\hat{A}$, $\widehat{AK}$ etc. Another and more appealing choice, especially with only one or a few primary y_1 measurements, would be to reorganize (5) into the partitioned innovation model

$$\begin{aligned} \bar{x}_{k+1} &= (A - AK_2C_2)\bar{x}_k + (B - AK_2D_2)u_k + AK_2y_{2,k} \\ &\quad + AK_1e_{1,k} \\ y_{1,k} &= C_1\bar{x}_k + D_1u_k + e_{1,k} \end{aligned} \tag{12}$$

before the identification. In this model $e_{1,k}$ is uncorrelated with u_s and $y_{2,s}$ for $s < k$, and we therefore have $y_{1,k} = z_k + e_{1,k}$ with z_k and $e_{1,k}$ uncorrelated. From this it follows that $z_k = \bar{y}_{1,k|k-1}$ according to (7) is the optimal predictor, just as when (5) is identified directly. The predictor in a prediction error identification method would also be the same as when identifying (5), with the optimal predictor given by (6). The difference would be that a simplified criterion function, e.g., $V_N^1(\hat{\theta}) = \text{tr}(\frac{1}{N}\sum \varepsilon_{1,k}\varepsilon_{1,k}^T)$, was used, and that $A - AK_2C_2$ and $B - AK_2D_2$ were treated as single matrices. Identification of (12) would therefore give the predictor (8) as the deterministic part of the solution, including the y_2 contribution. Regardless of the way we find it, however, the predictor (8) is not the optimal solution, since K_2 is found from the innovation forms (5) or (12) based on an underlying Kalman filter driven by u and both y_1 and y_2.

E. OE Model with y_2 as Input

Based on the assumption that (C_2, A) is observable and on an underlying Kalman filter driven by u and the y_2 measurements, the

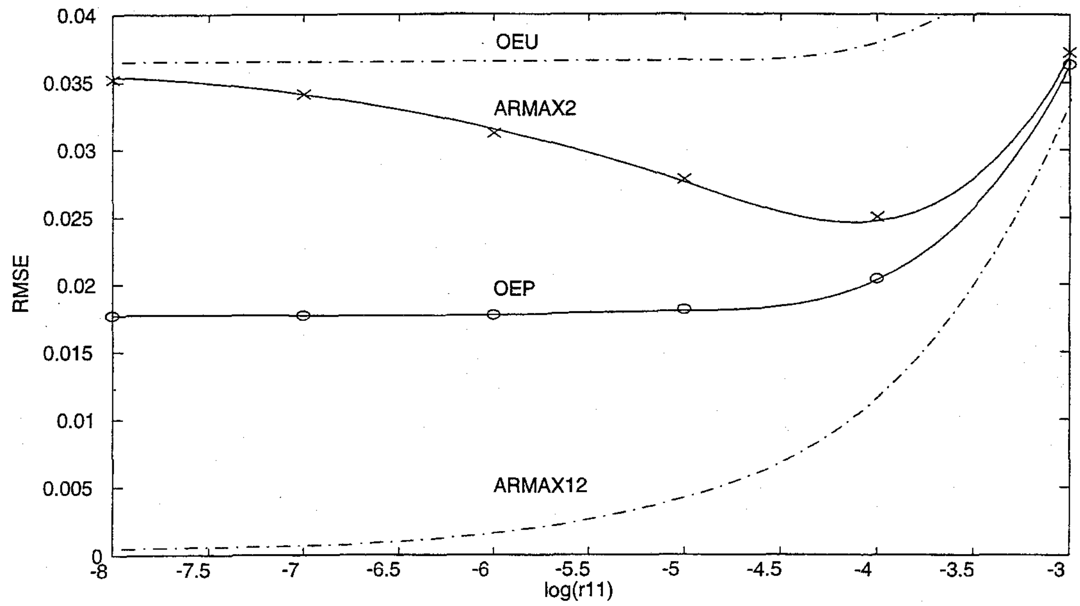

Fig. 1. Validation RMSE values for identified ARMAX$_2$ (x-markings) and OEP (o-markings) estimators as a function of $\log(r_{11})$ with $r_v = 0.1, r_{22} = 0.01$, and $N = 10\,000$ ($N = 50\,000$ for the ARMAX$_2$ model at $r_{11} = 10^{-4}$). These estimators utilize the information in both u and y_2. Theoretical values are shown as lines, including RMSE values for estimates based only on u (OEU) and on u and past y_1 as well as past y_2 values (ARMAX$_{12}$).

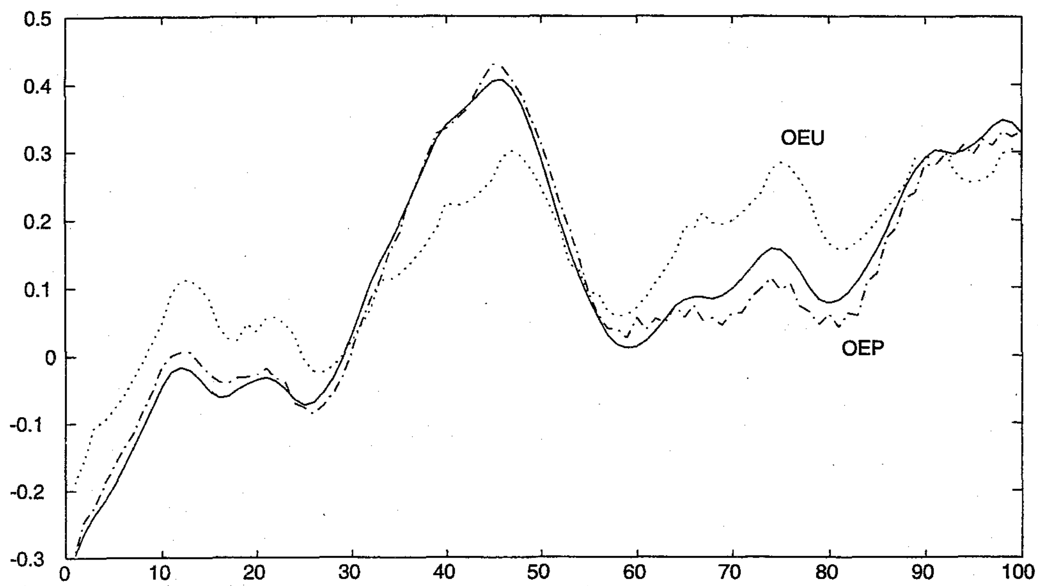

Fig. 2. Segment of validation responses for the OEP model (41) using both u and y_2 as inputs (dashed, RMSE = 0.0239) and an OE model using only u as input ($nn_{\mathrm{OEU}} = [0, 3, 0, 0, 3, 1]$, dotted, RMSE = 0.1078). The experimental conditions are given by $r_v = 1$, $r_{11} = 0.0001$, $r_{22} = 0.01$, and $N = 200$, and the ideal validation response is shown by a solid line.

following innovation form can be derived from (1):

$$\begin{aligned}\bar{x}_{k+1}^{\mathrm{OEP}} &= A\bar{x}_k^{\mathrm{OEP}} + Bu_k + AK_2^{\mathrm{OE}} e_{2,k}\\ y_{2,k} &= C_2\bar{x}_k^{\mathrm{OEP}} + D_2 u_k + e_{2,k}.\end{aligned} \tag{13}$$

The y_1 output is then given by

$$y_{1,k} = C_1\bar{x}_k^{\mathrm{OEP}} + D_1 u_k + \vartheta_k \tag{14}$$

where

$$\vartheta_k = C_1\left(x_k - \bar{x}_k^{\mathrm{OEP}}\right) + w_{1,k} \tag{15}$$

is colored noise.

The system determined by (13) and (14) can be identified by use of y_2 as an input signal in the output error prediction (OEP) model

$$\begin{aligned}\bar{x}_{k+1}^{\mathrm{OEP}} &= \left(A - AK_2^{\mathrm{OE}} C_2\right)\bar{x}_k^{\mathrm{OEP}} + \left(B - AK_2^{\mathrm{OE}} D_2\right)u_k\\ &\quad + AK_2^{\mathrm{OE}} y_{2,k}\\ y_{1,k} &= C_1\bar{x}_k^{\mathrm{OEP}} + D_1 u_k + \vartheta_k.\end{aligned} \tag{16}$$

The corresponding input–output model is then

$$\begin{aligned}y_{1,k} &= C_1\left[qI - A + AK_2^{\mathrm{OE}} C_2\right]^{-1}\\ &\quad \cdot \left[\left(B - AK_2^{\mathrm{OE}} D_2\right)u_k + AK_2^{\mathrm{OE}} y_{2,k}\right] + D_1 u_k + \vartheta_k\\ &= \bar{y}_{1,k|k-1}^{\mathrm{OEP}} + \vartheta_k\end{aligned} \tag{17}$$

or

$$y_{1,k} = G_1(q^{-1})u_k + G_2(q^{-1})y_{2,k} + \vartheta_k \tag{18}$$

where

$$G_1(q^{-1}) = C_1\left[qI - A + AK_2^{\text{OE}}C_2\right]^{-1} \cdot \left[B - AK_2^{\text{OE}}D_2\right] + D_1 \tag{19}$$

and

$$G_2(q^{-1}) = C_1\left[qI - A + AK_2^{\text{OE}}C_2\right]^{-1}AK_2^{\text{OE}}. \tag{20}$$

In order to identify the deterministic part of the system (17), i.e., G_1 and G_2, we model ϑ_k by some unknown white noise sequence and use the prediction

$$\hat{y}_{1,k} = \hat{G}_1(q^{-1};\hat{\theta})u_k + \hat{G}_2(q^{-1};\hat{\theta})y_{2,k} \tag{21}$$

where θ is the parameter vector. The prediction error is then

$$\begin{aligned}\varepsilon_{1,k} = y_{1,k} - \hat{y}_{1,k} = [G_1(q^{-1}) - \hat{G}_1(q^{-1};\hat{\theta})]u_k + [G_2(q^{-1})\\ - \hat{G}_2(q^{-1};\hat{\theta})]y_{2,k} + \vartheta_k.\end{aligned} \tag{22}$$

When evaluating the result of minimizing a scalar criterion function, e.g., $V_N^1(\hat{\theta}) = \text{tr}(\frac{1}{N}\sum_{k=1}^{N}\varepsilon_{1,k}\varepsilon_{1,k}^T)$, we must now consider the fact that $y_{2,k}$ and ϑ_k are not independent. Note, however, that when $\hat{G}_1(q^{-1};\hat{\theta}) \equiv G_1(q^{-1})$ and $\hat{G}_2(q^{-1};\hat{\theta}) \equiv G_2(q^{-1})$, we will in the asymptotic case ($N \to \infty$) simultaneously obtain

$$\begin{aligned}\text{Cov}\left(\bar{y}_{1,\text{theor.}}^{\text{OEP}}\right) &= E\varepsilon_{1,k}\varepsilon_{1,k}^T = E\vartheta_k\vartheta_k^T\\ &= C_1P^{\text{OEP}}C_1^T + R_{11}\end{aligned} \tag{23}$$

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

$$P^{\text{OEP}} = AP^{\text{OEP}}A^T + GR_vG^T - AK_2^{\text{OE}}C_2P^{\text{OEP}}A^T \tag{24}$$

with

$$K_2^{\text{OE}} = P^{\text{OEP}}C_2^T\left[C_2P^{\text{OEP}}C_2^T + R_{22}\right]^{-1}. \tag{25}$$

Since P^{OEP} is the minimized prediction state estimation covariance given the y_2 measurements, this represents a true minimum, resulting in consistent parameter estimates.

Note that the prediction covariance (23) is derived in the same way as the prediction covariance (10) for the ARMAX_2 case (with $K_1 = 0$), only that we now have a minimized P^{OEP} covariance matrix due to the use of $K_2 = K_2^{\text{OE}}$.

Utilizing also current y_2 values, the optimal estimator considering that y_1 is not available will be found by identifying the following output error model based also on current data (OEC model):

$$\begin{aligned}y_{1,k} &= C_1\left(I - K_2^{\text{OE}}C_2\right)\left[qI - A + AK_2^{\text{OE}}C_2\right]^{-1}\\ &\quad \cdot \left[(B - AK_2^{\text{OE}}D_2)u_k + AK_2^{\text{OE}}y_{2,k}\right]\\ &\quad + C_1K_2^{\text{OE}}(y_{2,k} - D_2u_k) + D_1u_k + \psi_k\\ &= \bar{y}_{1,k|k}^{\text{OEC}} + \psi_k.\end{aligned} \tag{26}$$

Here we introduce the colored noise

$$\psi_k = C_1\left(x_k - \bar{x}_k^{\text{OEC}}\right) + w_{1,k} \tag{27}$$

based on

$$\bar{x}_k^{\text{OEC}} = \left(I - K_2^{\text{OE}}C_2\right)\bar{x}_k^{\text{OEP}} + K_2^{\text{OE}}(y_{2,k} - D_2u_k). \tag{28}$$

Minimization of the criterion function $V_N^1(\hat{\theta})$ will now result in an optimal estimator only if

$$E\psi_k\psi_k^T = C_1P^{\text{OEC}}C_1^T + R_{11} - C_1K_2^{\text{OE}}R_{21} - R_{12}\left(C_1K_2^{\text{OE}}\right)^T \tag{29}$$

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

$$P^{\text{OEC}} = \left(I - K_2^{\text{OE}}C_2\right)P^{\text{OEP}}\left(I - K_2^{\text{OE}}C_2\right) + K_2^{\text{OE}}R_{22}\left(K_2^{\text{OE}}\right)^T \tag{30}$$

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 the asymptotic current estimation covariance then becomes

$$\text{Cov}\left(\bar{y}_{1,\text{theor.}}^{\text{OEC}}\right) = E\psi_k\psi_k^T = C_1P^{\text{OEC}}C_1^T + R_{11}. \tag{31}$$

III. Simulation Results

Simulation studies are undertaken, using *dlsim.m* in the Control system toolbox for use with Matlab [7], and the prediction error method implemented in *pem.m* in the System identification toolbox for use with Matlab [8]. The *pem.m* function identifies the system matrices and the Kalman gain, based on the general innovation model (5), or the partitioned innovation model (12) when the y_2 measurements are also used as input signals. Provided a proper parameterization, it also identifies the OEP model (17) and the OEC model (26).

The main aim of the simulations is to support the theoretical asymptotic covariance expressions (10), (23), and (31), using a simple system and a high number of samples. Note, however, that the theoretical expressions are based on perfect model information, which would not be available in a practical situation (see [9] for a general discussion of practical cases).

As a starting point, the following continuous-time second-order process model with an additional first-order process noise model was used (e.g., interacting mixing tanks or thermal processes):

$$\begin{aligned}\dot{x} &= \begin{bmatrix}-1 & 1 & 0\\ 1 & -2 & 1\\ 0 & 0 & -1\end{bmatrix}x + \begin{bmatrix}0\\1\\0\end{bmatrix}u + \begin{bmatrix}0\\0\\1\end{bmatrix}v\\ y_1 &= [1 \quad 0 \quad 0]x + w_1\\ y_2 &= [0 \quad 1 \quad 0]x + w_2.\end{aligned} \tag{32}$$

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

$$\begin{aligned}x_{k+1} &= \begin{bmatrix}0.9092 & 0.0863 & 0.0044\\ 0.0863 & 0.8230 & 0.0863\\ 0 & 0 & 0.9048\end{bmatrix}x_k + \begin{bmatrix}0.0045\\0.0908\\0\end{bmatrix}u_k\\ &\quad + \begin{bmatrix}0.0002\\0.0045\\0.0952\end{bmatrix}v_k\\ y_{1,k} &= [1 \quad 0 \quad 0]x_k + w_{1,k}\\ y_{2,k} &= [0 \quad 1 \quad 0]x_k + w_{2,k}.\end{aligned} \tag{33}$$

The system was then simulated with u_k as a filtered pseudorandom binary sequence (PRBS) with autocovariance $r_{uu}(p) = 0.8^{|p|}$ ([6, example 5.11] with $\alpha = 0.8$), i.e., an input that was persistently exciting of sufficient order. The 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 ARMAX_2, 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 ARMAX_2 model (12) was specified as (see [8] for definition of nn)

$$nn_{\mathrm{ARMAX}_2} = [3, [3 \quad 3], 3, 0, [0 \quad 0], [1 \quad 1]] \tag{34}$$

i.e., a model

$$A(q^{-1})y_{1,k} = B_1(q^{-1})u_k + B_2(q^{-1})y_{2,k} + C(q^{-1})e_{1,k} \tag{35}$$

with

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3} \tag{36}$$
$$B_1(q^{-1}) = b_{11}q^{-1} + b_{12}q^{-2} + b_{13}q^{-3} \tag{37}$$
$$B_2(q^{-1}) = b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3} \tag{38}$$
$$C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + c_3q^{-3}. \tag{39}$$

The OEP model (17) was specified as

$$nn_{\mathrm{OEP}} = [0, [3 \quad 3], 0, 0, [3 \quad 3], [1 \quad 1]] \tag{40}$$

i.e., a model

$$y_{1,k} = \frac{B_1(q^{-1})}{F_1(q^{-1})}u_k + \frac{B_2(q^{-1})}{F_2(q^{-1})}y_{2,k} + \vartheta_k \tag{41}$$

with $B_1(q^{-1})$ and $B_2(q^{-1})$ as in (37) and (38), and

$$F_1(q^{-1}) = 1 + f_{11}q^{-1} + f_{12}q^{-2} + f_{13}q^{-3} \tag{42}$$
$$F_2(q^{-1}) = 1 + f_{21}q^{-1} + f_{22}q^{-2} + f_{23}q^{-3}. \tag{43}$$

The OEC model (26) was specified as

$$nn_{\mathrm{OEC}} = [0, [3 \quad 4], 0, 0, [3 \quad 3], [1 \quad 0]] \tag{44}$$

i.e., the same model as (41), but with $B_2(q^{-1})$ altered to

$$B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}. \tag{45}$$

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, e.g., [10], and a systematic method for finding the structure is presented in [2]. For the OEP and OEC models, no attempt was made to force $F_1(q^{-1})$ and $F_2(q^{-1})$ to be identical, as 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 (RMSE) criterion

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(y_{1,k} - y_{1,k}^{\mathrm{est.}}\right)^2} \tag{46}$$

where $y_{1,k}^{\mathrm{est.}} = \bar{y}_{1,k|k-1}^{\mathrm{ARMAX}_2}$ according to (8) for the ARMAX model (35) $y_{1,k}^{\mathrm{est.}} = \bar{y}_{1,k|k-1}^{\mathrm{OEP}}$ according to (17) for the OEP model (41) and $y_{1,k}^{\mathrm{est.}} = \bar{y}_{1,k|k}^{\mathrm{OEC}}$ according to (26) for the OEC model specified by (44).

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 ($b_{il,j+1} = b_{il,j} \cdot (1 + 0.5e)$, with e as a zero mean and normal random variable with variance 1).

The mean RMSE values and RMSE standard deviations for $N = 10000$, $r_v = 0.1$, $r_{22} = 0.01$, and varying r_{11} values are given in Tables I and II. The tables also include theoretical RMSE values $\sqrt{\mathrm{Var}(\bar{y}_{1,\mathrm{theor.}}^{\mathrm{ARMAX}_2})}$, $\sqrt{\mathrm{Var}(\bar{y}_{1,\mathrm{theor.}}^{\mathrm{OEP}})}$ and $\sqrt{\mathrm{Var}(\bar{y}_{1,\mathrm{theor.}}^{\mathrm{OEC}})}$ computed according to (10), (23), and (31).

The tables show an obvious agreement between results based on simulation and theory. The only exception is the ARMAX_2 result for $r_{11} = 10^{-4}$, where repeated simulations show a mean deviation of approximately $10 \cdot 10^{-4}$. When the number of samples was increased to $N = 50\,000$, this specific result was altered to $\mathrm{RMSE} = (250 \pm 6) \cdot 10^{-4}$. The reason for this extraordinary demand for a high number of samples is not investigated further.

The RMSE results for the ARMAX_2 and OEP models in Tables I and II are also shown in Fig. 1, together with the theoretical results for a one-step-ahead predictor OEU based only on the independent input u and for the one-step-ahead predictor ARMAX_{12} based on (7), i.e., utilizing also past y_1 values.

The results in the tables and Fig. 1 were obtained from $N = 10\,000$ samples (one exception with $N = 50\,000$). 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 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.

IV. Concluding Remarks

Through a theoretical development with established system identification theory as a basis, it is shown how one-step-ahead prediction and current estimation of nonmeasured primary output variables y_1 can be done in asymptotically optimal ways by use of identified models. The solution is to employ OE models with both the independent inputs u and secondary output variables y_2 as input signals. This can be achieved by use of a prediction error identification method. ARMAX models may utilize the y_2 information in a far from optimal way, due to the fact that past y_1 values are used in the identification stage, while such values are later not available as a basis for estimation. In both the OE and ARMAX cases, Kalman gains in underlying optimal observers will be part of the deterministic models for prediction and estimation of y_1.

The theoretical estimation covariance results are supported by Monte Carlo simulations of a third-order system.

References

[1] L. Ljung, "System identification," Dept. Electrical Engineering, Linköping Univ., Sweden, Technical Rep. LiTH-ISY-R-1763, 1995.

[2] R. Ergon and D. Di Ruscio, "Dynamic system calibration by system identification methods," in *Proc. Fourth European Control Conf. (ECC'97)*, Brussels, Belgium.

[3] G. C. Goodwin and R. L. Payne, *Dynamic System Identification*. New York: Academic, 1977.

[4] L. Ljung, *System Identification, Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[5] F. L. Lewis, *Optimal Estimation, with an Introduction to Stochastic Control Theory*. New York: Wiley, 1986.

[6] T. Söderström and P. Stoica, *System Identification*. New York: Prentice-Hall, 1989.

[7] A. Grace *et al.*, *Control System Toolbox for Use with Matlab*. Natick, MA: The MathWorks, 1992.

[8] L. Ljung, *System Identification Toolbox for Use with Matlab*. Natick, MA: The MathWorks, 1991.

[9] C. F. Ansley and R. Kohn, "Prediction mean squared error for state space models with estimated parameters," *Biometrika*, vol. 73, pp. 467–473, 1986.

[10] D. Di Ruscio, "A method for identification of combined deterministic stochastic systems," in *Applications of Computer Aided Time Series Modeling*, M. Aoki and A. M. Havenner, Eds. New York: Springer-Verlag, 1997, pp. 181–235.