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 $d$, 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 $u = [u_{m1}^T, u_{m2}^T]^T$, 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.

II. THEORY

A. Statement of Problem

Consider the discrete-time system model

$$
\begin{align*}
x_{k+1} &= Ax_k + Bu_k + Gv_k \\
y_{1,k} &= C_1 x_k + D_1 u_k + w_{1,k} \\
y_{2,k} &= C_2 x_k + D_2 u_k + w_{2,k}
\end{align*}
$$

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 $R_p = E v_k v_k^T$ and $R_w = E w_k w_k^T = [R_{w1}^T, R_{w2}^T]$. Also assume that $(C_2, A)$ is observable and that $(A, G, R_w)$ 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 $k = 1, 2, \ldots, N$ are at hand, with $u_k$ persistently exciting of appropriate order. The problem is now to identify the optimal one-step-ahead $\hat{y}_{1,k|k-1}$ prediction estimator based on past and present $u_k$ and past $y_{2,k}$ values, and the optimal $\hat{y}_{1,k|k}$ 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 $y_{1,k}$ values are not available.

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

$$
A y_{1,k} = B_1 u_k + B_2 y_{2,k} + C_{1,k}
$$

where $A = A(q^{-1})$ etc. are matrix polynomials in the unit delay operator $q^{-1}$, and $C_{1,k}$ 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

$$
\hat{y}_{1,k|k-1} = A^{-1} B_1 u_k + A^{-1} B_2 y_{2,k}
$$

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

$$
y_{1,k} = F^{-1} B_1 u_k + F^{-1} B_2 y_{2,k} + \theta_k
$$

where $\theta_k$ is colored noise, and where $y_{2,k}$ is used as input signal. Although $\theta_k$ here is correlated with $y_{2,k}$, the result will still be an optimal predictor. The reason for this is that the expectation $E \hat{y}_{1,k|k}^2$ is minimized when and only when the correct parameters are found.
### Table I

<table>
<thead>
<tr>
<th>$t_{11}$</th>
<th>ARMAX</th>
<th>ARMAX, theor.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{-8}</td>
<td>352±15</td>
<td>354</td>
</tr>
<tr>
<td>10^{-7}</td>
<td>341±19</td>
<td>341</td>
</tr>
<tr>
<td>10^{-6}</td>
<td>312±17</td>
<td>315</td>
</tr>
<tr>
<td>10^{-5}</td>
<td>278±15</td>
<td>276</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>265±13</td>
<td>247</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>372±5</td>
<td>369</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>$t_{11}$</th>
<th>OEP</th>
<th>OEP,th</th>
<th>OEC</th>
<th>OEC,th</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{-8}</td>
<td>177±5</td>
<td>177</td>
<td>173±6</td>
<td>173</td>
</tr>
<tr>
<td>10^{-7}</td>
<td>177±5</td>
<td>177</td>
<td>173±6</td>
<td>173</td>
</tr>
<tr>
<td>10^{-6}</td>
<td>177±5</td>
<td>177</td>
<td>173±5</td>
<td>173</td>
</tr>
<tr>
<td>10^{-5}</td>
<td>181±5</td>
<td>180</td>
<td>177±5</td>
<td>176</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>204±6</td>
<td>203</td>
<td>200±0</td>
<td>200</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>363±5</td>
<td>362</td>
<td>361±0</td>
<td>360</td>
</tr>
</tbody>
</table>

### C. ARMAX Model with y2 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.

$$\bar{x}_{k+1} = A\bar{x}_k + Bu_k + [AK_1 A K_2]\begin{bmatrix} e_{1,k} \\ e_{2,k} \end{bmatrix}$$

$$y_{1,k} = C_1\bar{x}_k + D_1u_k + e_{1,k}$$

$$y_{2,k} = C_2\bar{x}_k + D_2u_k + e_{2,k}.$$  \hspace{1cm} (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 ($N \to \infty$) and after minimization of an appropriate criterion function [4] become:

$$\tilde{x}_{k+1} = \bar{x}_k + B u_k + AK_1 y_{1,k} + AK_2 y_{2,k}.$$  \hspace{1cm} (6)

$$\tilde{y}_{1,k} = C_1\tilde{x}_k + D_1 u_k$$

$$\tilde{y}_{2,k} = C_2\tilde{x}_k + D_2 u_k$$

where $A = A - AK_1 C_1 - AK_2 C_2$ and $B = B - AK_1 D_1 - AK_2 D_2$. 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 $u_k$ 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:

$$\tilde{y}_{1,k} = C_1 [qI - A]^{-1} \cdot [B u_k + AK_1 y_{1,k} + AK_2 y_{2,k}] + D_1 u_k.$$  \hspace{1cm} (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 $C_1 = C_2$ and $D_1 = D_2$, i.e., when the $y_{1,k}$ and $y_{2,k}$ outputs are identical except for the noise term. Then perfect measurements, i.e., $R_{11} \to 0$, would result in $K_2 \to 0$. 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 y2 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:

$$\tilde{y}_{1,k|k-1} = C_1 [qI - A + AK_2 C_2]^{-1} \cdot [(B - AK_2 D_2) u_k + AK_2 y_{2,k}] + D_1 u_k.$$  \hspace{1cm} (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 $\tilde{x}_k = x_k - \tilde{x}_k$ in the underlying nonoptimal observer would be governed by:

$$\tilde{x}_{k+1} = (A - AK_2 C_2) \tilde{x}_k + G v_k - AK_2 w_{2,k}.$$  \hspace{1cm} (9)

resulting in the asymptotic prediction covariance:

$$\text{Cov} (\tilde{y}_{1,k|k-1}) = \text{E} ([\tilde{y}_{1,k|k-1}] - \tilde{y}_{1,k|k-1})^T = C_1 \sigma^2 \text{ARMAX}_2 C_1^T + R_{11}.$$  \hspace{1cm} (10)

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

$$\text{PARMAX}_2 = (A - AK_2 C_2) \text{PARMAX}_2 + (A - AK_2 C_2)^T + G R e G^T + AK_2 R_{22} K_2^T A^T.$$  \hspace{1cm} (11)

Since (10) is a sum of nonnegative terms, it is evident that $\text{Cov}(\tilde{y}_{1,k|k-1})$ is minimized only when $\text{PARMAX}_2$ 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 $V(y) = \text{tr}(Y^{-1}) \sum \epsilon_i^2 + \text{tr}(Y^{-1}) \sum e_{2,k}^2 + \text{tr}(Y^{-1}) \sum e_{2,k}^2$ for a model of the form:

$$\text{PARMAX}_2 = (A - AK_2 C_2) \text{PARMAX}_2 + (A - AK_2 C_2)^T + G R e G^T + AK_2 R_{22} K_2^T A^T.$$  \hspace{1cm} (12)

E. OE Model with y2 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 ARMAX2 (x-markings) and OEP (o-markings) estimators as a function of log($r_{11}$) with $r_e = 0.1$, $r_{22} = 0.01$, and $N = 10000$ ($N = 50000$ for the ARMAX2 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 (ARMAX12).

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 ($\mu_{\text{OEU}} = [0.3, 0.0, 0.3, 1]$, dotted, RMSE = 0.1078). The experimental conditions are given by $r_e = 1$, $r_{11} = 0.0001$, $r_{22} = 0.01$, and $N = 200$, and the ideal validation response is shown by a solid line.

The following innovation form can be derived from (1):

$$
\hat{x}_{k+1}^\text{OEP} = A\hat{x}_k^\text{OEP} + Bu_k + AK_2^\text{OE}e_{2,k}
$$

$$
y_{2,k} = C_2\hat{x}_k^\text{OEP} + D_2u_k + e_{2,k}.
$$

The $y_1$ output is then given by

$$
y_{1,k} = C_1\hat{x}_k^\text{OEP} + D_1u_k + \theta_k
$$

where

$$
\theta_k = C_1(x_k - \hat{x}_k^\text{OEP}) + w_{1,k}
$$

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

$$
\hat{x}_{k+1}^\text{OEP} = (A - AK_2^\text{OE}C_2)\hat{x}_k^\text{OEP} + (B - AK_2^\text{OE}D_2)u_k + AK_2^\text{OE}y_{2,k}
$$

$$
y_{1,k} = C_1\hat{x}_k^\text{OEP} + D_1u_k + \theta_k.
$$

The corresponding input-output model is then

$$
y_{1,k} = C_1[yI - A + AK_2^\text{OE}C_2]^{-1}[B - AK_2^\text{OE}D_2]u_k + [AK_2^\text{OE}y_{2,k}] + D_1u_k + \theta_k
$$

$$
= \hat{y}_{1,k|k-1} + \theta_k
$$
Minimization of the criterion function $V_0(\theta)$ will now result in an optimal estimator only if

$$E\psi_k^T = C_1 P_{OEC} C_1^T + R_{11} - C_1 K_{2OEC}^T R_{21} - R_{12} (C_1 K_{2OEC})^T$$

with $P_{OEC} = E(x_k - \hat{x}_{k,OEC}) (x_k - \hat{x}_{k,OEC})^T$ given by

$$P_{OEC} = (I - K_{2OEC} C_2) P_{OEC} (I - K_{2OEC} C_2) + K_{2OEC} R_{22} (K_{2OEC})^T$$

simultaneously is at a minimum. Since $P_{OEC}$ is the minimized current state estimation covariance, this is true only when $R_{12} = R_{21} = 0$, and the asymptotic current estimation covariance then becomes

$$\text{Cov}(\hat{x}_{k,OEC}) = E\psi_k \psi_k^T = C_1 P_{OEC} C_1^T + R_{11}. \quad (31)$$

### 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):

$$\dot{x} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \end{bmatrix} v$$

$$y_1 = [1 \ 0 \ 0] x + w_1$$

$$y_2 = [0 \ 1 \ 0] x + w_2.$$

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

$$x_{k+1} = \begin{bmatrix} 0.0902 & 0.0863 & 0.0044 \\ 0.0863 & 0.8230 & 0.0063 \\ 0 & 0 & 0.9086 \end{bmatrix} x_k + \begin{bmatrix} 0.0045 \\ 0.0002 \\ 0.0562 \end{bmatrix} u_k$$

$$y_{1,k} = [1 \ 0 \ 0] x_k + w_{1,k}$$

$$y_{2,k} = [0 \ 1 \ 0] x_k + w_{2,k}.$$

The system was then simulated with $u_k$ as a filtered pseudorandom binary sequence (PRBS) with autocovariance $\tau_{uu}(p) = 0.8^{|p|}$ ([6, example 5.11] with $\alpha = 0.8$), i.e., an input that was persistently exciting of sufficient order. The noise sources $v_{1,k}$, $w_{1,k}$, 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 ARMAX2, OEP, and OEC models with $u_k$ and $y_{2,k}$ as input signals and $y_{1,k}$ as output signal, using $N = 10000$ samples.
The ARMAX_2 model (12) was specified as (see [8] for definition of \( n_n \))

\[
n_{ARMAX_2} = \begin{bmatrix} 3 & 3 & 3 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}
\]

i.e., a model

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

with

\[
A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3}
\]

\[
B_1(q^{-1}) = b_{11}q^{-1} + b_{12}q^{-2} + b_{13}q^{-3}
\]

\[
B_2(q^{-1}) = b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}
\]

\[
C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + c_3q^{-3}
\]

The OEP model (17) was specified as

\[
n_{OEP} = \begin{bmatrix} 0 & 3 & 3 & 0 & 0 & 3 & 3 & 1 & 1 \end{bmatrix}
\]

i.e., a model

\[
y_{1,k} = \frac{B_1(q^{-1})}{F_1(q^{-1})}u_k + \frac{B_2(q^{-1})}{F_2(q^{-1})}y_{2,k} + \theta_k
\]

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

\[
F_1(q^{-1}) = 1 + f_{11}q^{-1} + f_{12}q^{-2} + f_{13}q^{-3}
\]

\[
F_2(q^{-1}) = 1 + f_{21}q^{-1} + f_{22}q^{-2} + f_{23}q^{-3}
\]

The OEC model (26) was specified as

\[
n_{OEC} = \begin{bmatrix} 0 & 3 & 4 & 0 & 0 & 3 & 3 & 1 & 0 \end{bmatrix}
\]

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

\[
B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}
\]

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

\[
B_2(q^{-1}) = b_{20} + b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}
\]

The tables show an obvious agreement between results based on simulation and theory. The only exception is the ARMAX_2 result for \( r_{11} = 10^{-1} \), where repeated simulations show a mean deviation of approximately \( 10 \cdot 10^{-4} \). When the number of samples was increased to \( N = 50000 \), this specific result was altered to \( RMSE = (250 \pm 6) \cdot 10^{-4} \). 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 OEP based only on the independent input \( u \) and for the one-step-ahead predictor ARMAX_2 based on (7), i.e., utilizing also past \( y_1 \) values.

The results in the tables and Fig. 1 were obtained from \( N = 10000 \) samples (one exception with \( N = 50000 \)). 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 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.

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

REFERENCES


