

Dynamic system calibration: The low primary output sampling rate case

ROLF ERGON†

Keywords: *Product quality, estimation, system identification*

In many industrial cases it is not feasible to measure primary outputs, e.g. product quality, from production processes on-line. It is thus of interest to estimate such outputs from known process inputs and secondary process measurements. In an earlier paper it is shown that optimal estimators can be identified from data recorded during an informative experiment, with the primary outputs sampled at the same high rate as the inputs and secondary outputs. In the present paper it is shown that optimal estimators can also be found from data where the primary outputs are sampled at a low and possibly irregular rate.

1. Introduction

The theoretical basis for identification of the optimal estimator for primary outputs y_1 from a linear, time invariant and dynamic system, is presented in Ergon (1998). It is then assumed that the y_1 outputs are not measured at all or measured only at a low sampling rate. The optimal estimator utilizes all available information in known inputs u and measured secondary outputs y_2 , which are assumed to be available at a sufficiently high sampling rate. The basic insight behind this is that the y_2 measurements may carry valuable information about the process noise v , which it is possible to utilize when estimating y_1 , as illustrated in Fig. 1.

In Ergon (1998) it is shown how the optimal prediction (a priori) and current (a posteriori) estimators can be identified by use of an ordinary prediction error method with u and y_2 as inputs and y_1 as output. It is then assumed that also y_1 data at a high sampling rate is available from an informative experiment with sufficient excitation (e.g. Goodwin and Payne, 1977).

The present paper extends the theory in Ergon (1998) to cover the case where y_1 experimental data is available only at a low and possibly irregular sampling rate. A method for identification of the optimal estimators under this assumption is outlined in section 2, including a method for finding initial parameter values for the optimization. A simulation example in section 3 demonstrates the feasibility of the proposed methods, and conclusions are given in section 4.

2. Theory

2.1. Statement of problem

Consider the discrete system model

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

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

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

Received 8 February 1998.

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

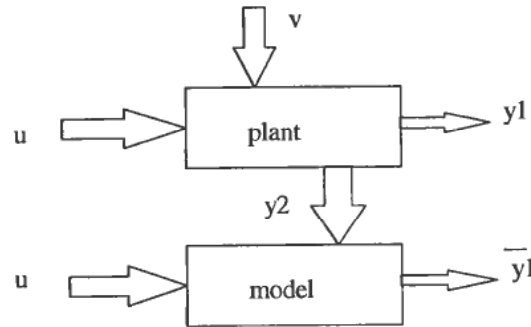


Figure 1. Basic principle for estimation of primary system outputs y_1 from known inputs u and measured secondary outputs y_2 in presence of process noise v .

where x_k is the state vector, while v_k and $w_k = [w_{1,k}^T \ w_{2,k}^T]^T$ are white and independent process and measurement noise vectors with covariances

$$R_v = E v_k v_k^T \text{ and } R_{w,k} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} E w_{1,k} w_{1,k}^T & E w_{1,k} w_{2,k}^T \\ E w_{2,k} w_{1,k}^T & E w_{2,k} w_{2,k}^T \end{bmatrix}.$$

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

Further assume that high sampling rate data records for u_k and $y_{2,k}$, with $k = 1, 2, \dots, N$, are at hand, with u_k persistently exciting of appropriate order. Also assume that $y_{1,j}$ data is available at a low and possibly irregular sampling rate, with $j = 1, 2, \dots, J$, where $J < N$ is a sufficiently high number and where each sampling of $y_{1,j}$ coincides in time with one of the u_k and $y_{2,k}$ samplings. The problem is now to identify the optimal one-step-ahead (a priori) $\bar{y}_{1,k|k-1}$ predictor based on past and present u_k and past $y_{2,k}$ values, and the optimal $\bar{y}_{1,k|k}$ current (a posteriori) estimator based also on present $y_{2,k}$ values.

2.2. Optimal one-step-ahead predictor

As past y_1 values are not available, the prediction must be based on an underlying Kalman filter driven by u and only the y_2 measurements (Ergon 1999). With the assumption that (C_2, A) is detectable, the following innovation form can then be derived from (1):

$$\bar{x}_{k+1}^{OEP} = A\bar{x}_k^{OEP} + Bu_k + AK_2^{OE}e_{2,k} \quad (2a)$$

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

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

$$K_2^{OE} = P^{OEP}C_2^T(C_2P^{OEP}C_2^T + R_{22})^{-1}, \quad (3)$$

where

$$P^{OEP} = E(x_k - \bar{x}_k^{OEP})(x_k - \bar{x}_k^{OEP})^T$$

is given by the Riccati equation

$$P^{OEP} = AP^{OEP}A^T + GR_vG^T - AP^{OEP}C_2^T(C_2P^{OEP}C_2^T + R_{22})^{-1}C_2P^{OEP}A^T. \quad (4)$$

The y_1 output is then given by

$$y_{1,k} = C_1\bar{x}_k^{OEP} + D_1u_k + v_k, \quad (5)$$

where

$$v_k = C_1(x_k - \bar{x}_k^{OEP}) + w_{1,k} \quad (6)$$

is colored noise.

With both u and y_2 used as inputs, the system determined by (2) and (5) is turned into the output error prediction model (OEP model)

$$\bar{x}_{k+1}^{OEP} = A(I - K_2^{OE}C_2)\bar{x}_k^{OEP} + (B - AK_2^{OE}D_2)u_k + AK_2^{OE}y_{2,k} \quad (7a)$$

$$y_{1,k} = C_1\bar{x}_k^{OEP} + D_1u_k + v_k. \quad (7b)$$

The corresponding input-output model is then

$$y_{1,k} = G_1(q^{-1}; \theta)u_k + G_2(q^{-1}; \theta)y_{2,k} + v_k, \quad (8)$$

where

$$G_1(q^{-1}; \theta) = C_1(qI - A + AK_2^{OE}C_2)^{-1}(B - AK_2^{OE}D_2) + D_1 \quad (9)$$

and

$$G_2(q^{-1}; \theta) = C_1(qI - A + AK_2^{OE}C_2)^{-1}AK_2^{OE}. \quad (10)$$

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

In order to identify the deterministic part of the system (8), i.e. $G_1(q^{-1}; \theta)$ and $G_2(q^{-1}; \theta)$, we model v_k by some unknown white noise sequence and use the prediction

$$\hat{y}_{1,k|k-1} = G_1(q^{-1}; \hat{\theta})u_k + G_2(q^{-1}; \hat{\theta})y_{2,k}, \quad (11)$$

where $\hat{\theta}$ is the parameter vector used in the prediction model. The prediction error

$$\varepsilon_{1,k} = y_{1,k} - \hat{y}_{1,k|k-1} \quad (12)$$

is then

$$\begin{aligned} \varepsilon_{1,k} &= [G_1(q^{-1}; \theta) - G_1(q^{-1}; \hat{\theta})]u_k \\ &\quad + [G_2(q^{-1}; \theta) - G_2(q^{-1}; \hat{\theta})]y_{2,k} + v_k. \end{aligned} \quad (13)$$

In a standard prediction error method, we should now minimize a scalar criterion function (Söderström and Stoica 1989), e.g.

$$V_N(\theta) = \det \left[\frac{1}{N} \sum_{k=1}^N \varepsilon_{1,k} \varepsilon_{1,k}^T \right]. \quad (14)$$

Since $y_{1,k}$ is not generally available in the low y_1 sampling rate case, we must give the prediction errors zero weight when $y_{1,k}$ does not exist. We thus minimize

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

where $\alpha_k = 1$ for those sampling instants where $y_{1,k}$ exists, while $\alpha_k = 0$ when $y_{1,k}$ does not exist. Since past $y_{1,k}$ values are not used in $\hat{y}_{1,k|k-1}$, we will asymptotically obtain the same result by use of (15) as by use of (14). In any case, however, we must consider the fact that $y_{2,j}$ and v_j are not independent. Note, though, that when $G_1(q^{-1}; \hat{\theta}) \equiv G_1(q^{-1}; \theta)$ and $G_2(q^{-1}; \hat{\theta}) \equiv G_2(q^{-1}; \theta)$, we will from (13), (6) and (4) simultaneously obtain

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

Since P^{OEP} is the minimized prediction state estimation covariance given the y_2 measurements, this represents a true minimum. We will thus be able to identify the optimal one-step-ahead prediction estimator given by (8) to (10):

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

2.3. Initial parameter values

Minimization of the criterion function (15) requires reasonably good initial parameter values. In the ordinary high y_1 sampling rate case, i.e. when (14) is minimized, a least squares approximation (ARX model) can be used for this purpose (Ljung 1991). In the present low y_1 sampling rate case, however, this is not possible. A solution to this initial model problem is first to identify (2) by use of a prediction error or subspace identification method (see, e.g. Di Ruscio 1997), and then to find the parameters in (5) by ordinary least squares regression. An example given in section 3 demonstrates that this initial model is then improved by minimization of the criterion (15). Identification of (2) can also be done by use of some of the y_2 measurements as input signals, as shown in Ergon (1999).

2.4. Optimal current estimator

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

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

Here we introduce the colored noise

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

based on

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

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

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

with

$$P^{OEC} = E(x_j - \bar{x}_j^{OEC})(x_j - \bar{x}_j^{OEC})^T$$

given by

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

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

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

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

2.5. Theoretical y_1 estimation covariance

When the OEP model (7) or the OEC model (18) is identified using a large data set, i.e. when $J \rightarrow \infty$, the estimate $\bar{y}_{1,k|k-1}$ or $\bar{y}_{1,k|k}$ will be asymptotically unbiased when we use either only u or both u and y_2 as input signals. The asymptotic estimation covariance, however, will depend on the model and the quality of the data. In the following we assume perfect model and noise information, and derive theoretical asymptotic expressions for the prediction and current y_1 estimation covariances.

The underlying Kalman filter driven by u and the y_2 measurements is governed by the well known Kalman filter equations (3) and (4). As the prediction estimate $\bar{y}_{1,k|k-1}$ is directly based on \bar{x}_k^{OEP} , the theoretical asymptotic prediction estimate covariance becomes

$$\text{Cov}(\bar{y}_{1,k|k-1}) = E(y_{1,k} - \bar{y}_{1,k|k-1})(y_{1,k} - \bar{y}_{1,k|k-1})^T = C_1 P^{OEP} C_1^T + R_{11}, \quad (24)$$

with P^{OEP} given by (4). The theoretical asymptotic y_1 current estimation covariance is

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

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

Assume now for convenience a scalar y_1 measurement. When the estimators (17) and (23) are identified and validated by use of independent data sets with $N \rightarrow \infty$ and $J \rightarrow \infty$, we will then find the theoretical root mean square error

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_{1,k} - y_{1,k}^{est})^2} \rightarrow \sqrt{C_1 P C_1^T + R_{11}}, \quad (26)$$

where $y_{1,k}^{est}$ is either $\bar{y}_{1,k|k-1}$ or $\bar{y}_{1,k|k}$ according to (17) or (23), while $P = P^{OEP}$ or $P = P^{OEC}$ according to (4) or (22).

3. Simulation example

Simulation studies are undertaken by use of a prediction error method implemented in *Matlab* by use of a numerical Gauss-Newton method. With an appropriate OE model specified, the optimal prediction estimator (17) or the optimal current estimator (23) is then identified.

The simulation example is basically the same as an example presented in Ergon (1998), only modified with respect to the y_1 sampling rate. The intention is now primarily to verify that the optimal estimators will be found also in the low y_1 sampling rate case, and we choose to limit the simulations to the OEP estimator (17). As a starting point, the following continuous second-order process model with an additional first-order process noise model was used (e.g. interacting mixing tanks or thermal processes):

$$\dot{x} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} v \quad (27a)$$

$$y_1 = [1 \quad 0 \quad 0] x + w_1 \quad (27b)$$

$$y_2 = [0 \quad 1 \quad 0] x + w_2. \quad (27c)$$

The system was discretized assuming zero order hold elements on the u and v inputs and a sampling interval $T = 0.1$ (see Ergon (1998) for the discrete model). The system was then simulated with u_k as a filtered PRBS signal with autocovariance $r_u(p) = 0.8^{|p|}$ (Söderström and Stoica (1989), example 5.11 with $\alpha = 0.8$), i.e. an input that was persistently exciting of sufficient order. The scalar noise sources v_k , $w_{1,k}$ and $w_{2,k}$ were independent and normally distributed white noise sequences with zero mean and variances $r_v = 0.1$, $r_{11} = 0.0001$ and $r_{22} = 0.01$.

The initial parameter values were found by identification of an ARMAX model using N samples with u_k as input and $y_{2,k}$ as output, followed by a least squares estimation (LSE) of the static relation between the state vector \bar{x}_k^{OEP} and $y_{1,k}$. The ARMAX model was specified as

$$A(q^{-1})y_{2,k} = B(q^{-1})u_k + C(q^{-1})e_k, \quad (28)$$

with

$$A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3}, \quad (29)$$

$$B(q^{-1}) = b_1q^{-1} + b_2q^{-2} + b_3q^{-3} \quad (30)$$

and

$$C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + c_3q^{-3}. \quad (31)$$

The OEP model (7) was then identified with u_k and $y_{2,k}$ as input signals and $y_{1,j}$ as output signal. It was assumed that $y_{1,j}$ was recorded at every tenth u_k and $y_{2,k}$ sampling, resulting in $J = \frac{N}{10}$ samples. The OEP model was specified as

$$y_{1,k} = \frac{B_1(q^{-1})}{F_1(q^{-1})}u_k + \frac{B_2(q^{-1})}{F_2(q^{-1})}y_{2,k} + e_k^{OEP} \quad (32)$$

with

$$B_1(q^{-1}) = b_{11}q^{-1} + b_{12}q^{-2} + b_{13}q^{-3}, \quad (33)$$

$$B_2(q^{-1}) = b_{21}q^{-1} + b_{22}q^{-2} + b_{23}q^{-3}, \quad (34)$$

$$F_1(q^{-1}) = 1 + f_{11}q^{-1} + f_{12}q^{-2} + f_{13}q^{-3} \quad (35)$$

and

$$F_2(q^{-1}) = 1 + f_{21}q^{-1} + f_{22}q^{-2} + f_{23}q^{-3}. \quad (36)$$

Table 1. Validation RMSE mean values and standard deviations and theoretical mean values for different numbers of samples. Here, $r_v = 0.1$, $r_{11} = 0.0001$ and $r_{22} = 0.01$, and the RMSE values are multiplied by 10 000

N	J	ARMAX + LSE	OEP	OEP _{theor.}
400	40	557 ± 298	353 ± 154	203
2000	200	354 ± 80	234 ± 40	203
10 000	1000	332 ± 35	216 ± 17	203

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

Each identified model was validated against an independent data set with the same number of samples and the same noise variances as used for identification. Validation comparisons between the different identified models were based on the root mean square error criterion (26).

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

The mean RMSE values and RMSE standard deviations for different numbers of samples are given in Table 1. The table also includes theoretical RMSE values $\sqrt{\text{Cov}(\bar{y}_{1,k|k-1})}$ computed according to (24).

As expected, Table 1 shows reduced estimation error when the number of samples is increased from realistically low values to $N = 10\,000$. In all cases, however, the OEP model is considerably better than the ARMAX + LSE model. For a high number of samples, the OEP estimation variance approaches the theoretical value. The results are quite similar to the corresponding results in the high y_1 sampling rate case with $J = N$ given in Ergon (1998).

In order to visualize the degree of model misfit behind the RMSE values in Table 1, specific validation responses for models based on $N = 400$ samples, are shown in Fig. 2. The figure also gives a representative picture of the improvement from the initial ARMAX + LSE to the final OEP solution. Note that the number of y_1 measurements behind each model is only $J = 40$.

4. Conclusions

Known optimal estimators for non-measured primary outputs y_1 from e.g. an industrial plant, assumes experimental identification data with y_1 sampled at the same rate as known inputs u and secondary measurements y_2 (Ergon 1998). In the present paper it is shown that these asymptotically optimal estimators can be found also from experimental data where y_1 is sampled at a lower and possibly also irregular rate. As in the previously known case, the optimal estimators are obtained by identification of

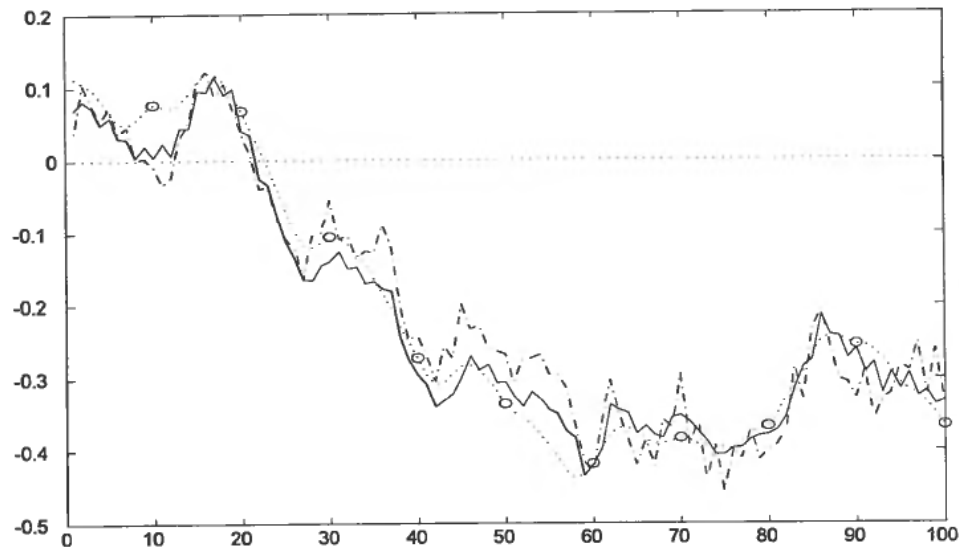


Figure 2. Segment of validation responses for an initial ARMAX + LSE model (dashdot, $RMSE = 0.0551$) and the OEP model (7) (solid, $RMSE = 0.0313$). The experimental conditions are given by $r_v = 0.1$, $r_{11} = 0.0001$, $r_{22} = 0.01$, $N = 400$ and $J = 40$, and the ideal validation response is shown by dotted line with o-markings at the j sampling instants.

output error (OE) models. In this case, however, initial parameter values must be found by a novel method, which is also outlined in the paper.

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

$$y_{1,k} = C_1 x_k + D_1 u_k + w_{1,k}. \quad (37)$$

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

$$z_{1,k} = C_1 x_k + D_1 u_k, \quad (38)$$

and this is in fact done in the optimal estimators (17) and (23). We may, therefore, replace $\bar{y}_{1,k|k-1}$ and $\bar{y}_{1,k|k}$ in these estimators with $\bar{z}_{k|k-1}$ and $\bar{z}_{k|k}$. The theoretical asymptotic covariances (24) and (25) will then be altered to

$$\text{Cov}(\bar{z}_{k|k-1}) = E(z_k - \bar{z}_{k|k-1})(z_k - \bar{z}_{k|k-1})^T = C_1 P^{OEP} C_1^T \quad (39)$$

and

$$\text{Cov}(\bar{z}_{k|k}) = E(z_k - \bar{z}_{k|k})(z_k - \bar{z}_{k|k})^T = C_1 P^{OEC} C_1^T. \quad (40)$$

The feasibility of the proposed method is demonstrated by a simulation example, using the proposed method for determination of initial parameter values followed by the proposed prediction error method implemented in Matlab.

REFERENCES

- DI RUSCIO, D. (1997). A method for identification of combined deterministic and stochastic systems. *Applications of Computer Aided Time Series Modeling*, edited by M. Aoki and A. M. Havenner, New York; Springer-Verlag, pp. 181–235.
- ERGON, R. and DI RUSCIO, D. (1997). Dynamic system calibration by system identification methods. *The European Control Conference, ECC97*, Brussels, CD file ECC416.pdf.

- ERGON, R. (1998). Dynamic system multivariate calibration, *Modeling, Identification and Control*, Vol. 19, No. 2.
- ERGON, R. (1999). On Primary Output Estimation by use of Secondary Measurements as Input Signals in System Identification, scheduled for publication in *IEEE Transactions on Automatic Control*, April 1999.
- GOODWIN, G. C. and PAYNE, R. L. (1977). *Dynamic System Identification, Experiment Design and Data Analysis*, Academic Press, Inc.
- LEWIS, F. L. (1986). *Optimal Estimation, with an Introduction to Stochastic Control Theory*, John Wiley & Sons.
- LJUNG, L. (1991). *System Identification Toolbox for use with Matlab*, The MathWorks, Inc.
- SÖDERSTRÖM, T. and STOICA, P. (1989). *System Identification*, Prentice Hall Inc.