Downloaded from orbit.dtu.dk on: Jun 23, 2021

DTU Library

Comparison of Prediction-Error-Modelling Criteria

Jørgensen, John Bagterp; Jørgensen, Sten Bay

Published in:
American Control Conference, 2007. ACC '07

Link to article, DOI:
10.1109/ACC.2007.4283020

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Jørgensen, J. B., & Jørgensen, S. B. (2007). Comparison of Prediction-Error-Modelling Criteria. In *American Control Conference, 2007. ACC '07* IEEE. https://doi.org/10.1109/ACC.2007.4283020

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Proceedings of the 2007 American Control Conference
Marriott Marquis Hotel at Times Square
New York City, USA, July 11-13, 2007

Comparison of Prediction-Error-Modelling Criteria

John Bagterp Jørgensen*
Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby, Denmark
jbj@imm.dtu.dk

Sten Bay Jørgensen
Department of Chemical Engineering
Technical University of Denmark
DK-2800 Kgs. Lyngby, Denmark
sbj@kt.dtu.dk

***Abstract*— Single and multi-step prediction-error-methods based on the maximum likelihood and least squares criteria are compared. The prediction-error methods studied are based on predictions using the Kalman filter and Kalman predictors for a linear discrete-time stochastic state space model, which is a realization of a continuous-discrete multivariate stochastic transfer function model. The proposed prediction error-methods are demonstrated for a SISO system parameterized by the transfer functions with time delays of a continuous-discrete-time linear stochastic system. The simulations for this case suggest to use the one-step-ahead prediction-error maximum-likelihood (or maximum a posteriori) estimator. It gives consistent estimates of all parameters and the parameter estimates are almost identical to the estimates obtained for long prediction horizons but with consumption of significantly less computational resources. The identification method is suitable for predictive control.**

I. Introduction

In this paper, we address construction of stochastic linear models using the *prediction-error-method* [1]–[3]. We parameterize the stochastic linear models as continuous-discrete-time transfer functions with delay and realize these models as discrete-time stochastic linear state space models. The Kalman filter and Kalman predictor for this system is used to generate the prediction errors and covariances need by the prediction-error identification criteria. We investigate multi-step prediction error identification and compare it to single-step predictor error identification. Shah and coworkers [4]–[6] apply a similar multi-step approach based on impulse response models and a least-squares criterion. The approach presented in this paper distinguishes itself by being general for linear systems, by applying least-squares as well as maximum likelihood criteria for the prediction errors in the estimator, and in particular by being directly applicable to state space model based predictive control in its modern implementation. predictive control. We propose a method to address the request for better identification methods tailored for predictive control in order to potentially improve the closed-loop performance of such control systems [7]–[11].

II. Prediction-Error-Methods

A. Standard Regression Problem

The essence of regression is to select some parameters, θ, such that the predicted outputs, $\hat{y}_k(\theta)$, match the measured outputs, y_k, as well as possible for all measurements $k = 0, 1, \ldots, N-1$. The estimation problem is often stated as the stochastic relation

$$\boldsymbol{y}_k = \hat{y}_k(\theta) + \boldsymbol{e}_k,\ \boldsymbol{e}_k \sim N(0, R_k),\ k = 0, 1, \ldots, N-1 \quad (1)$$

The predictor or estimator, $\hat{y}_k(\theta)$, is a function of the parameters, $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$. For the measured realization, $\{y_k\}_{k=0}^{N-1}$, of the outputs, $\{\boldsymbol{y}_k\}_{k=0}^{N-1}$, the parameters, θ, are computed such that some measure, e.g. the least squares measure, of the residuals, $\{e_k(\theta) = y_k - \hat{y}_k(\theta)\}_{k=0}^{N-1}$, is minimized. This is the standard nonlinear regression problem [12], [13], which can be stated as the optimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} V(\theta) \quad (2)$$

with the objective function $V(\theta) = V_{LS}(\theta)$ being

$$V_{LS}(\theta) = \frac{1}{2} \sum_{k=0}^{N-1} \|e_k(\theta)\|_2^2 \quad (3)$$

in the least squares case. The maximum-likelihood estimate corresponds to using negative log-likelihood function in (2), i.e. $V(\theta) = V_{ML}(\theta)$ with $V_{ML}(\theta)$ defined as

$$\begin{aligned} V_{ML}(\theta) = \frac{N n_y}{2} \ln(2\pi) + \frac{1}{2} \sum_{k=0}^{N-1} \ln\left(\det R_k(\theta)\right) \\ + \frac{1}{2} \sum_{k=0}^{N-1} e_k(\theta)' R_k(\theta)^{-1} e_k(\theta) \end{aligned} \quad (4)$$

The maximum *a posteriori* estimate assumes that a priori the parameters stem from the distribution $\boldsymbol{\theta} \sim N(\theta_0, P_{\theta_0})$ in which $\theta_0 \in \Theta \subset \mathbb{R}^{n_\theta}$. Then using Bayes rule the negative log-likelihood *a posteriori* function is

$$\begin{aligned} V_{MAP}(\theta) = V_{ML}(\theta) + \frac{n_\theta}{2} \ln(2\pi) + \frac{1}{2} \ln\left(\det P_{\theta_0}\right) \\ + \frac{1}{2} (\theta - \theta_0)' P_{\theta_0}^{-1} (\theta - \theta_0) \end{aligned} \quad (5)$$

Hence, the maximum *a posteriori* estimate is obtained by applying $V(\theta) = V_{MAP}(\theta)$ in (2).

B. Parametrization, Realization and Prediction

One way to represent multivariate stochastic distributed processes is through the input-output representation in the LaPlace domain

$$\boldsymbol{Z}(s) = G(s;\theta) U(s) + H(s;\theta) \boldsymbol{E}(s) \quad (6a)$$

$$\boldsymbol{y}(t_k) = \boldsymbol{z}(t_k) + \boldsymbol{v}(t_k) \quad (6b)$$

*Corresponding author

1-4244-0989-6/07/$25.00 ©2007 IEEE.

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

in which $U(s)$ is the process input vector, $\boldsymbol{E}(s)$ is a vector with white noise components, $\boldsymbol{Z}(s)$ is the process output vector. $\boldsymbol{v}(t_k) \sim N(0, R_{vv}(\theta))$ is the measurement noise vector and $\boldsymbol{y}(t_k)$ is the measured process output vector at time t_k. The elements, $\{g_{ij}(s)\}$ and $\{h_{ij}(s)\}$, of the transfer function matrices, $G(s)$ and $H(s)$, are rational transfer functions with time delays

$$g_{ij}(s) = \frac{b_{ij}(s;\theta)}{a_{ij}(s;\theta)} \exp(-\tau_{ij}(\theta)s) \tag{7a}$$

$$h_{ij}(s) = \frac{d_{ij}(s;\theta)}{c_{ij}(s;\theta)} \exp(-\lambda_{ij}(\theta)s) \tag{7b}$$

Assuming that $u(t)$ is a zero-order-hold input, (6) may be realized as a discrete-time stochastic linear state space model

$$\boldsymbol{x}_{k+1} = A(\theta)\boldsymbol{x}_k + B(\theta)u_k + \boldsymbol{w}_k \tag{8a}$$

$$\boldsymbol{y}_k = C(\theta)\boldsymbol{x}_k + \boldsymbol{v}_k \tag{8b}$$

in which

$$\begin{bmatrix}\boldsymbol{w}_k \\ \boldsymbol{v}_k\end{bmatrix} \sim N_{iid}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}R_{ww}(\theta) & R_{wv}(\theta)\\ R_{wv}(\theta)' & R_{vv}(\theta)\end{bmatrix}\right) \tag{8c}$$

and

$$\boldsymbol{x}_0 \sim N(\hat{x}_{0|-1}(\theta), P_{0|-1}(\theta)) \tag{8d}$$

The Kalman filter and Kalman predictor are optimal estimators for (8) and therefore also optimal estimators for (6) [14], [15]. The Kalman filter equations

$$e_k = y_k - C\hat{x}_{k|k-1} \tag{9a}$$

$$R_{e,k} = CP_{k|k-1}C' + R_{vv} \tag{9b}$$

$$K_{fx,k} = P_{k|k-1}C'R_{e,k}^{-1} \tag{9c}$$

$$K_{fw,k} = R_{wv}R_{e,k}^{-1} \tag{9d}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_{fx,k}e_k \tag{9e}$$

$$\hat{w}_{k|k} = K_{fw,k}e_k \tag{9f}$$

$$P_{k|k} = P_{k|k-1} - K_{fx,k}R_{e,k}K_{fx,k}' \tag{9g}$$

$$Q_{k|k} = R_{ww} - K_{fw,k}R_{e,k}K_{fw,k}' \tag{9h}$$

provide the mean and covariance for the conditional distributions $\boldsymbol{x}_k|\mathcal{I}_k \sim N(\hat{x}_{k|k}, P_{k|k})$ and $\boldsymbol{w}_k|\mathcal{I}_k \sim N(\hat{w}_{k|k}, Q_{k|k})$. $\mathcal{I}_k = \{(y_j, u_j)\}_{j=0}^{k}$. The Kalman one-step state predictor equations

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + B\hat{u}_{k|k} + \hat{w}_{k|k} \tag{10a}$$

$$P_{k+1|k} = AP_{k|k}A' + Q_{k|k} - AK_{fx,k}R_{wv}' - R_{wv}K_{fx,k}'A' \tag{10b}$$

and the j-step ($j > 1$) state predictor equations

$$\hat{x}_{k+j|k} = Ax_{k+j-1|k} + B\hat{u}_{k+j-1|k} \tag{11a}$$

$$P_{k+j|k} = AP_{k+j-1|k}A' + R_{ww} \tag{11b}$$

provide the mean and covariance for the conditional distribution $\boldsymbol{x}_{k+j}|\mathcal{I}_k \sim N(\hat{x}_{k+j|k}, P_{k+j|k})$ for $j \geq 1$. The output predictions

$$\hat{y}_{k+j|k} = C\hat{x}_{k+j|k} \tag{12a}$$

$$R_{k+j|k} = CP_{k+j|k}C' + R_{vv} \tag{12b}$$

provide the conditional mean and covariance for $\boldsymbol{y}_{k+j}|\mathcal{I}_k \sim N(\hat{y}_{k+j|k}, R_{k+j|k})$ for $j \geq 1$. The Kalman filter and Kalman predictor are implemented numerically robust using the array algorithm propagating the square root of the covariances rather than the covariances themselves [14].

C. The PE Method as a Regression Problem

The family of prediction-error-methods can be considered as solving a general regression problem similar to (1). The estimate of the prediction error estimates are obtained by solving an optimization problem like (2) for some criteria (LS, ML, MAP) and some predictors. For the case considered in this paper, the predictors in the prediction error method are the Kalman predictors, $\hat{y}_{k+j|k}(\theta)$. The prediction errors, $\varepsilon_{k+j|k} = y_{k+j} - \hat{y}_{k+j|k}(\theta)$, correspond to the residuals in the standard regression problem. Therefore, the prediction-error-method is a standard regression problem with a predictor generated by the Kalman filter and predictor. In the following, the statistical properties of the predictors and the prediction errors will be discussed and various criteria for estimating the parameters in the prediction error framework are presented.

If it is possible to know the true structure of the system, $\mathcal{S}$, and the model identified, $\mathcal{M}(\theta)$, is equal to the true system, $\mathcal{M}(\theta) = \mathcal{S}$, then this model will be optimal in a statistical sense no matter for what purpose it is to be used and what consistent estimator (criterion) used for determining the parameters. In any realistic situation, it is almost impossible to know the true model structure due to changing process conditions, changing disturbance properties and nonlinearities. Therefore, in practice the model should be suited and be identified for the purpose it is going to be used. In predictive control this corresponds to minimization of multi-step predictions compatible with the regulator objective function.

D. Single-Step j-Step-Ahead Prediction Error

Let $\{(u_k, y_k)\}_{k=0}^{N-1}$ denote the IO-data for identification and let N_p denote the prediction horizon. Let the time indices be $k = -1, 0, 1, \ldots, N-1-j$ and the prediction index be $1 \leq j \leq N_p$. This implies that $0 \leq k+j \leq N-1$. The conditional outputs, $\boldsymbol{y}_{k+j}|\mathcal{I}_k$, have the distribution

$$\boldsymbol{y}_{k+j}|\mathcal{I}_k \sim N(\hat{y}_{k+j|k}, R_{k+j|k}) \tag{13}$$

and their correlation may be computed by [14]

$$\begin{aligned} R_{(i,j)|k} &= \langle(\boldsymbol{y}_{k+i}|\mathcal{I}_k) - \hat{y}_{k+i|k}, (\boldsymbol{y}_{k+j}|\mathcal{I}_k) - \hat{y}_{k+j|k}\rangle \\ &= \begin{cases} CA^{i-j-1}N_{k+j|k} & i > j \\ CP_{k+i|k}C' + R_{vv} & i = j \\ N_{k+i|k}'(A^{j-i-1})'C' & i < j \end{cases} \end{aligned} \tag{14}$$

in which $1 \leq i \leq N_p$, $1 \leq j \leq N_p$, and

$$N_{k+i|k} = AP_{k+i|k}C' + R_{wv} \tag{15}$$

Hence, the single-step j-step-ahead prediction error problem may be stated as

$$\boldsymbol{y}_{k+j}|\mathcal{I}_k = \hat{y}_{k+j|k}(\theta) + \boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k \quad k = -1, 0, \ldots, N-1-j \tag{16}$$

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

with $\boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k \sim N(0, R_{k+j|k})$ in the ideal case when the system and the model on which the predictor is computed are identical. $\boldsymbol{\varepsilon}_{k+j,k}$ denotes the residual of the single-step j-step-ahead predictor at time k. This corresponds to a standard regression problem in which some measure of the j-step prediction error

$$\varepsilon_{k+j|k} = y_{k+j} - \hat{y}_{k+j|k}(\theta),\ k = -1, 0, \ldots, N-1-j \quad (17)$$

is minimized. $\varepsilon_{k+j|k}$ can be regarded as the realization of $\boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k \sim N(0, R_{k+j|k})$. When the structure of the model and the system are different, $\boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k$ may have a non-zero mean and a covariance different from $R_{k+j|k}$. Even in such cases it seems reasonable to minimize some measure of the prediction error, $\varepsilon_{k+j|k}$. However, as the distribution of $\boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k$ is unknown maximum likelihood based procedures can only be considered as approximation, i.e. quasi maximum likelihood.

As $\hat{y}_{k+j|k}(\theta)$ is not a simple function of θ, the analytical derivatives of $\varepsilon_{k+j|k} = \varepsilon_{k+j|k}(\theta)$ with respect to θ are generally not available. Hence, the optimization algorithms for solving the parameter estimation problem must compute the derivatives of the objective functions numerically, i.e. by finite difference.

The one-step prediction-error estimates may be regarded as special versions of the j-step prediction-error estimates. However, in that case no extra effort is needed for computing $\varepsilon_{k+1|k}$ and $R_{k+1|k}$ as they are already computed as part of the Kalman filter updates. In the j-step prediction case with $j > 1$, $\varepsilon_{k+j|k}$ and $R_{k+j|k}$ must be computed explicitly if needed in the parameter estimation objective function.

E. Multi-Step Maximum Likelihood Predictors

To deduce true multi-step prediction-error (quasi) maximum likelihood and maximum a posteriori estimators, the correlation between $\boldsymbol{\varepsilon}_{k+i,k}|\mathcal{I}_k$ and $\boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k$ for $i \neq j$ must be taken into account. This correlation is

$$\langle \boldsymbol{\varepsilon}_{k+i,k}|\mathcal{I}_k, \boldsymbol{\varepsilon}_{k+j,k}|\mathcal{I}_k \rangle = R_{(i,j)|k} \quad (18)$$

Define

$$\boldsymbol{Y}_k = \begin{bmatrix}\boldsymbol{y}'_{k+1} & \cdots & \boldsymbol{y}'_{k+N_p}\end{bmatrix}' \ k = -1, 0, \ldots, N-1-N_p$$
$$\boldsymbol{Y}_k = \begin{bmatrix}\boldsymbol{y}'_{k+1} & \cdots & \boldsymbol{y}'_{N-1}\end{bmatrix}' \ k = N-N_p, \ldots, N-2$$

and the corresponding multi-step predictions

$$\hat{Y}_k(\theta) = \begin{bmatrix}\hat{y}'_{k+1|k} & \cdots & \hat{y}'_{k+N_p|k}\end{bmatrix}' \ k = -1, 0, \ldots, N-1-N_p$$
$$\hat{Y}_k(\theta) = \begin{bmatrix}\hat{y}'_{k+1|k} & \cdots & \hat{y}'_{N-1|k}\end{bmatrix}' \ k = N-N_p, \ldots, N-2$$

Furthermore, define the conditional multi-step prediction error vector as

$$\boldsymbol{\epsilon}_k|\mathcal{I}_k = \begin{bmatrix}\boldsymbol{\varepsilon}_{k+1,k}|\mathcal{I}_k \\ \boldsymbol{\varepsilon}_{k+2,k}|\mathcal{I}_k \\ \vdots \\ \boldsymbol{\varepsilon}_{k+N_p,k}|\mathcal{I}_k\end{bmatrix} \quad \boldsymbol{\epsilon}_k|\mathcal{I}_k = \begin{bmatrix}\boldsymbol{\varepsilon}_{k+1,k}|\mathcal{I}_k \\ \boldsymbol{\varepsilon}_{k+2,k}|\mathcal{I}_k \\ \vdots \\ \boldsymbol{\varepsilon}_{N-1,k}|\mathcal{I}_k\end{bmatrix} \quad (20)$$

for $k = -1, 0, \ldots, N-1-N_p$ (the left vector) and $k = N-N_p, \ldots, N-2$ (the right vector), respectively.

The multi-step prediction error problem can then be expressed as the stochastic model

$$\boldsymbol{Y}_k|\mathcal{I}_k = \hat{Y}_k(\theta) + \boldsymbol{\epsilon}_k|\mathcal{I}_k \quad k = -1, 0, \ldots, N-2 \quad (21)$$

with $\boldsymbol{\epsilon}_k|\mathcal{I}_k \sim N(0, R_k)$ and

$$R_k = \langle \boldsymbol{\epsilon}_k|\mathcal{I}_k, \boldsymbol{\epsilon}_k|\mathcal{I}_k \rangle = \begin{bmatrix} R_{(1,1)|k} & R_{(1,2)|k} & \cdots & R_{(1,N_p)|k} \\ R_{(2,1)|k} & R_{(2,2)|k} & \cdots & R_{(2,N_p)|k} \\ \vdots & \vdots & & \vdots \\ R_{(N_p,1)|k} & R_{(N_p,2)|k} & \cdots & R_{(N_p,N_p)|k}\end{bmatrix} \quad (22)$$

The realization of the multi-step prediction-error vector for $k = -1, 0, \ldots, N-1-N_p$ is

$$\epsilon_{k|k} = Y_k - \hat{Y}_k(\theta) = \begin{bmatrix} y_{k+1} - \hat{y}_{k+1|k} \\ y_{k+2} - \hat{y}_{k+2|k} \\ \vdots \\ y_{k+N_p} - \hat{y}_{k+N_p|k}\end{bmatrix} = \begin{bmatrix}\varepsilon_{k+1|k} \\ \varepsilon_{k+2|k} \\ \vdots \\ \varepsilon_{k+N_p|k}\end{bmatrix} \quad (23)$$

The negative log likelihood function for the multi-step prediction is

$$V_{1:N_p,ML}(\theta) = \frac{n_y f}{2}\ln(2\pi) + \frac{1}{2}\sum_{k=-1}^{N-2}\left(\ln(\det R_k) + \epsilon'_{k|k}R_k^{-1}\epsilon_{k|k}\right) \quad (24)$$

in which $f = N_p\left[N - \frac{1}{2}(N_p - 1)\right]$. In computation of the multi-step prediction-error maximum likelihood estimate, $\ln(\det R_k)$ and $\epsilon'_{k|k}R_k^{-1}\epsilon_{k|k}$ must be computed. $\epsilon_{k|k}$ is obtained by computing the j-step prediction errors. This is accomplished using (10)-(12) for $j = 1, 2, \ldots, N_p$ given $\mathcal{I}_k$ and $\hat{u}_{k+j|k} = u_{k+j}$. By construction the covariance matrix, R_k, has the special structure that arise from a state space model. This implies that [14]

$$R_k = L_k R_{\epsilon,k} L'_k \quad (25)$$

in which the factorization, L_k and $R_{\epsilon,k}$, is computed using the Kalman filter recursions (9)-(12). Using the one-step predictive Kalman gain

$$K_{p,k} = AK_{fx,k} + K_{fw,k} \quad (26)$$

the block lower triangular matrix, L_k, may be computed as

$$L_k = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK_{p,k+1} & I & \cdots & 0 \\ CAK_{p,k+1} & CK_{p,k+2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ CA^{N_p-2}K_{p,k+1} & CA^{N_p-3}K_{p,k+2} & \cdots & I\end{bmatrix} \quad (27)$$

and the block diagonal matrix, $R_{\epsilon,k}$, is

$$R_{\epsilon,k} = \begin{bmatrix} R_{e,k+1} & & & \\ & R_{e,k+2} & & \\ & & \ddots & \\ & & & R_{e,k+N_p}\end{bmatrix} \quad (28)$$

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

Hence, the determinant of R_k may be computed as

$$\det R_k = \det R_{\epsilon,k} = \prod_{j=1}^{N_p} \det R_{e,k+j} \tag{29}$$

which implies

$$\ln(\det R_k) = \ln\left(\prod_{j=1}^{N_p} \det R_{e,k+j}\right) = \sum_{j=1}^{N_p} \ln(\det R_{e,k+j}) \tag{30}$$

Consequently, the term $\sum_{k=-1}^{N-2} \ln(\det R_k)$ in (24) may be evaluated as

$$\begin{aligned}\sum_{k=-1}^{N-2} \ln(\det R_k) = &\sum_{k=0}^{N_p-2} (k+1)\ln(\det R_{e,k}) \\ &+ N_p \sum_{k=N_p-1}^{N-1} \ln(\det R_{e,k})\end{aligned} \tag{31}$$

The term $\epsilon'_{k|k} R_k^{-1} \epsilon_{k|k}$ can be evaluated as

$$\begin{aligned}\epsilon'_{k|k} R_k^{-1} \epsilon_{k|k} &= \epsilon'_{k|k} (L_k R_{\epsilon,k} L_k')^{-1} \epsilon_{k|k} \\ &= (L_k^{-1}\epsilon_{k|k})' R_{\epsilon,k}^{-1} (L_k^{-1}\epsilon_{k|k}) \\ &= \sum_{j=1}^{N_p} \bar{e}'_{k+j|k} R_{e,k+j}^{-1} \bar{e}_{k+j|k}\end{aligned} \tag{32}$$

in which $\begin{bmatrix}\bar{e}'_{k+1|k} & \bar{e}'_{k+2|k} & \cdots & \bar{e}'_{k+N_p|k}\end{bmatrix}' = L_k^{-1}\epsilon_{k|k}$. $\{\bar{e}_{k+j|k}\}_{j=1}^{N_p}$ is efficiently computed using the Kalman filter recursions for $j = 1, 2, \ldots, N_p$

$$\bar{e}_{k+j|k} = \varepsilon_{k+j|k} - C\bar{x}_{k+j|k} \tag{33a}$$
$$\bar{x}_f = \bar{x}_{k+j|k} + K_{fx,k+j}\bar{e}_{k+j|k} \tag{33b}$$
$$\bar{w}_f = K_{fw,k+j}\,\bar{e}_{k+j|k} \tag{33c}$$
$$\bar{x}_{k+j+1|k} = A\bar{x}_f + \bar{w}_f \tag{33d}$$

with $\bar{x}_{k+1|k} = 0$. Note that (33b)-(33d) may be expressed as

$$\bar{x}_{k+j+1|k} = A\bar{x}_{k+j|k} + K_{p,k+j}\bar{e}_{k+j|k} \tag{34}$$

which implies that (33) can be expressed as

$$\bar{x}_{k+j+1|k} = (A - K_{p,k+j}C)\bar{x}_{k+j|k} + K_{p,k+j}\varepsilon_{k+j|k} \tag{35a}$$
$$\bar{e}_{k+j|k} = -C\bar{x}_{k+j|k} + \varepsilon_{k+j|k} \tag{35b}$$

Consequently, the term $\sum_{k=-1}^{N-2} \epsilon'_{k|k} R_k^{-1} \epsilon_{k|k}$ in (24) may be efficiently evaluated using

$$\begin{aligned}\sum_{k=-1}^{N-2} \epsilon'_{k|k} R_k^{-1} \epsilon_{k|k} = &\sum_{k=0}^{N_p-2} \sum_{j=1}^{k+1} \bar{e}'_{k|k-j} R_{e,k}^{-1} \bar{e}_{k|k-j} \\ &+ \sum_{k=N_p-1}^{N-1} \sum_{j=1}^{N_p} \bar{e}'_{k|k-j} R_{e,k}^{-1} \bar{e}_{k|k-j}\end{aligned} \tag{36}$$

and a bank of Kalman filter recursion (33) for computing $\bar{e}_{k|k-j}$ and $\bar{x}_{k+1|k-j}$ for $j = 1, 2, \ldots, N_p$. Hence, at each time instant k the multi-step prediction error $\epsilon_{k|k}$ is computed using the Kalman predictions. This vector is stored in memory for N_p iterations such that $\varepsilon_{k|k-j}$ can be used in computation of $\bar{e}_{k|k-j}$ and subsequent evaluation of the terms in (36). The advantage of this method compared to a naive implementation is that gains and covariances in the Kalman recursions need to be evaluated only once at each time step.

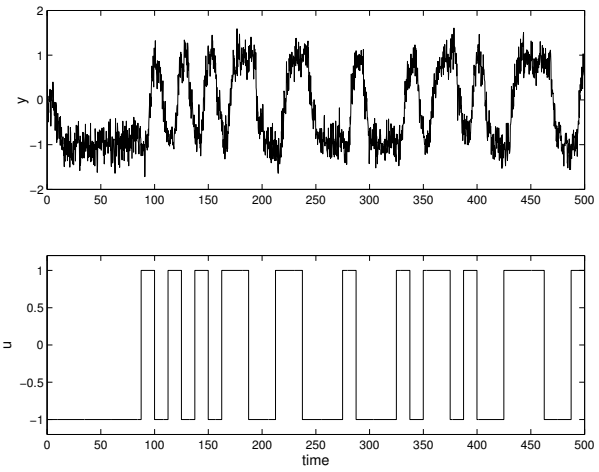

Fig. 1. IO-data for the SISO system, $\mathcal{S}$, defined by (37)-(38). The inputs, $\{u(t)\}$, are PRBS with bandwidth [0 0.02] and levels [−1 1].

III. SISO Example

To illustrate the identification criteria discussed in this paper, we consider the SISO system, $\mathcal{S} = \{g(s), h(s)\}$, defined as

$$\boldsymbol{Z}(s) = g(s)U(s) + h(s)\boldsymbol{E}(s) \tag{37a}$$
$$\boldsymbol{y}(t_k) = \boldsymbol{z}(t_k) + \boldsymbol{v}(t_k) \tag{37b}$$

in which $\boldsymbol{E}(s)$ is standard white noise and $\boldsymbol{v}(t_k) \sim N_{iid}(0, r^2)$. The transfer function, $g(s)$, from the process inputs, $U(s)$, to the process output, $\boldsymbol{Y}(s)$, and the disturbance transfer function, $h(s)$, are

$$g(s) = \frac{K}{(\alpha_1 s + 1)(\alpha_2 s + 1)}\, e^{-\tau s} \tag{38a}$$
$$h(s) = \frac{\sigma}{\gamma s + 1} \tag{38b}$$

The parameters defining the system $\mathcal{S}$ and used for generating the data are: $K = 1.0$, $\alpha_1 = 1.0$, $\alpha_2 = 3.0$, $\tau = 5.2$, $\sigma = 0.2$, $\gamma = 1.0$ and $r = 0.2$. The system is sampled with a sampling time of $T_s = 0.25$. The deterministic input, $U(s)$, is assumed to be implemented using a zero-order-hold circuit. The IO-data used for estimation of this system are illustrated in Figure 1.

A. Identical Model and System Structure

Consider the situation in which the model and the system has the same structure. In this case the structure of the model,

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

$\hat{g}(s)$, and the disturbance model, $\hat{h}(s)$, are

$$\hat{g}(s) = \frac{\hat{K}}{(\hat{\alpha}_1 s + 1)(\hat{\alpha}_2 s + 1)} e^{-\hat{\tau} s} \tag{39a}$$

$$\hat{h}(s) = \frac{\hat{\sigma}}{\hat{\gamma} s + 1} \tag{39b}$$

Let $\mathcal{M} = \left\{ \hat{g}(s), \hat{h}(s) \right\}$. This implies that the true system, $\mathcal{S}$, is within the class of models, $\mathcal{M}$, estimated, i.e. $\mathcal{S} \in \mathcal{M}$.

The estimates for the single-step and multi-step least squares criteria and various prediction horizons are shown in Tables I-II.[1] From these results, it is apparent that the LS method cannot be used to uniquely estimate σ and r. However, their ratio seems to be constant for different starting guesses and decreases with increasing horizon. This implies that the identified model approaches an output error model for long prediction horizons.

The estimates for the single-step and multi-step maximum likelihood criteria and various prediction horizons are shown in Tables III-IV. σ and r are estimated consistently for various initial guesses. For long-range single-step maximum likelihood estimation, the estimated model is essentially an output error model. The step response for the true model and the models estimated by the multi-step maximum-likelihood criterion with prediction horizons $N_p = 1$ and $N_p = 200$ are shown in Figure 2. There is not much difference between the two estimated models, but a little steady difference compared to the true model. However, as can be read off from Table IV the main difference between the estimated models for prediction horizon $N_p = 1$ and prediction horizon $N_p = 200$ is not the deterministic transfer function, $\hat{g}(s)$, but the disturbance model, $\hat{h}(s)$, and the covariance of the measurement noise, $\hat{r}^2$.

B. Simplified Model with Output Integrator

In this subsection we will illustrate the methodology when the model structure, $\mathcal{M}$, is different from the system model, $\mathcal{S}$, used to generate the data, i.e. $\mathcal{S} \notin \mathcal{M}$. To do this consider the model

$$\hat{g}(s) = \frac{\hat{K}}{\hat{\alpha} s + 1} e^{-\hat{\tau} s} \tag{40a}$$

$$\hat{h}(s) = \frac{\hat{\sigma}}{s} \tag{40b}$$

In the process industries most stable models can be approximated quite well by delayed first-order transfer functions, $\hat{g}(s)$. The disturbance model, $\hat{h}(s)$, is chosen as an integrator to ensure off-set free control for step-type disturbances and model-plant mismatch in the resulting predictive control system for which the estimated model is applied. For internal model control (IMC) which can be considered as a restricted class of predictive control this modelling approach is commonplace [16].

In contrast to the least-squares prediction-error-methods, the maximum-likelihood prediction-error-methods yield unique estimates for the covariance matrices. Hence, only the maximum-likelihood prediction-error-estimates for the system (40) will be reported here. The single-step maximum-likelihood estimates for various prediction horizons are shown in Table V. As the prediction horizon increases, $\hat{\sigma}$ is decreased and the estimated model becomes essentially an output error model. For the case considered, the estimated process noise vanishes already at a prediction horizon of $j = 8$. The measurement noise is increased slightly as the prediction horizon increases to accommodate the output noise that is not caught by the process noise. The multi-step maximum-likelihood estimates are shown in Table VI. Compared to the single-step maximum likelihood estimates the multi-step maximum-likelihood estimates are much less sensitive to the chosen prediction horizon. In fact there is

[1] All computations are conducted using a 3.20 GHz Pentium IV processor. The CPU time is reported to indicate the order of magnitude of computing time needed to calculate the various estimates.

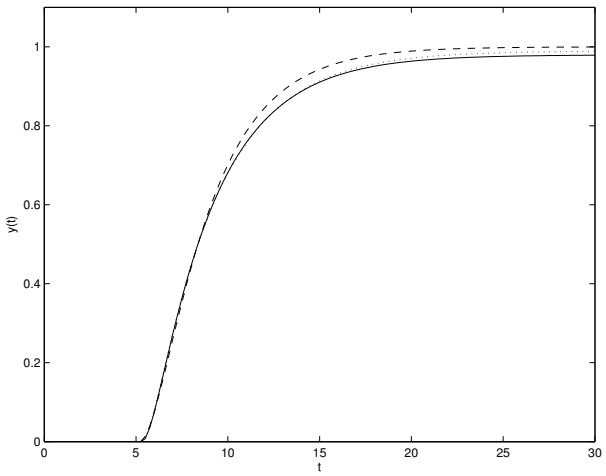

Fig. 2. Step response for the deterministic part of the SISO model estimated using a simplified model with an output integrator. Estimated model (39a) using the multi-step maximum likelihood criterion with $N_p = 1$ (solid line) and $N_p = 200$ (dotted line). Dashed line: True model (38a).

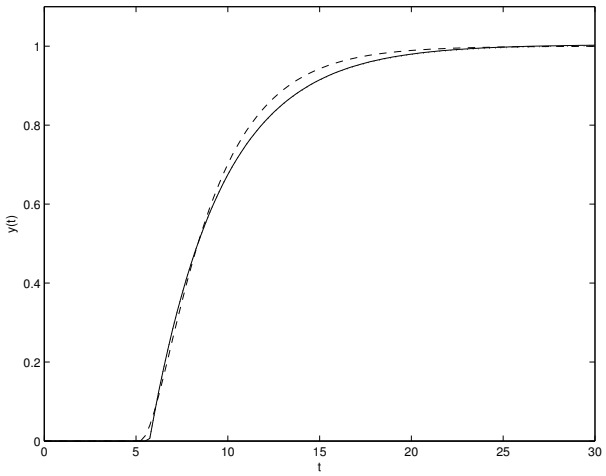

Fig. 3. Step response for the deterministic part of the SISO model estimated using a simplified model with an output integrator. Solid line: Estimated model (40a) using the multi-step maximum likelihood criterion with $N_p = 1$ and $N_p = 200$. Dashed line: True model (38a).

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

TABLE I

SINGLE-STEP LS ESTIMATION IN MODEL (39).

j	K	α_1	α_2	τ	σ	γ	r	σ/r	V	CPU sec.
1	0.9797	0.5641	3.4216	5.3171	0.8705	1.3757	0.8198	1.0618	109.9	113
4	0.9792	0.5798	3.4053	5.3119	0.4802	1.1548	0.3756	1.2785	116.9	154
8	0.9790	0.7239	3.3496	5.2037	0.8301	1.1885	0.5636	1.4730	120.0	227
20	0.9832	0.7086	3.3815	5.2019	1.9202	9.7771	2.0543	0.9347	119.2	351
40	0.9786	0.8639	3.2871	5.1016	0.1824	0.9056	0.1776	1.0268	122.5	325
80	0.9719	0.7612	3.3374	5.1578	0.1800	0.9000	0.1800	1.0000	129.5	394
100	1.0087	0.9820	3.3471	4.8656	0.1800	0.9000	0.1800	1.0000	200.3	376
200	0.9428	1.1445	2.8801	5.0634	0.1800	0.9000	0.1800	1.0000	130.4	532

TABLE II

MULTI-STEP LS ESTIMATION IN MODEL (39).

N_p	K	α_1	α_2	τ	σ	γ	r	σ/r	V	CPU sec.
1	0.9797	0.5632	3.4219	5.3179	0.3377	1.3754	0.3180	1.0620	110.0	87
4	0.9796	0.5657	3.4194	5.3170	0.0080	1.3861	0.0075	1.0603	449.1	421
8	0.9794	0.6136	3.3981	5.2827	0.3805	1.3641	0.3595	1.0585	924.2	251
20	0.9792	0.7116	3.3563	5.2107	0.3053	1.4301	0.2897	1.0539	2370	393
40	0.9823	0.7394	3.3641	5.1836	0.7805	8.0826	0.8318	0.9383	4763	644
80	0.9804	0.7597	3.3357	5.1767	0.4734	7.1230	0.4975	0.9514	9481	1101
100	0.9796	0.7586	3.3305	5.1782	0.6825	6.5664	0.7271	0.9386	11804	1426
200	0.9760	0.7739	3.2966	5.1802	0.5728	6.1969	0.6151	0.9314	23023	2382

TABLE III

SINGLE-STEP ML ESTIMATION IN MODEL (39).

j	K	α_1	α_2	τ	σ	γ	r	σ/r	V	CPU sec.
1	0.9797	0.5651	3.4211	5.3164	0.2204	1.3762	0.2077	1.0613	-63.27	115
4	0.9793	0.5798	3.4048	5.3120	0.2432	1.1321	0.1868	1.3022	-1.987	371
8	0.9789	0.7184	3.3500	5.2088	0.2853	1.0881	0.1526	1.8694	23.57	725
20	0.9832	0.7081	3.3829	5.2007	0.2243	9.0000	0.2391	0.9384	17.62	2038
40	0.9786	0.8639	3.2871	5.1060	0.0002	0.1380	0.2475	0.0007	45.37	3297
80	0.9719	0.7608	3.3376	5.1580	0.0002	0.1985	0.2545	0.0007	100.7	6082
100	1.0088	0.9817	3.3474	4.8656	0.0002	0.2748	0.3165	0.0006	536.9	6696
200	0.9426	1.1232	2.8931	5.0718	0.0002	0.0291	0.2553	0.0007	107.4	13227

TABLE IV

MULTI-STEP ML ESTIMATION IN MODEL (39).

j	K	α_1	α_2	τ	σ	γ	r	σ/r	V	CPU sec.
1	0.9797	0.5651	3.4211	5.3164	0.2204	1.3762	0.2077	1.0613	-63.27	111
4	0.9798	0.5607	3.4251	5.3186	0.2182	1.4745	0.2102	1.0383	-248.7	160
8	0.9798	0.5307	3.4369	5.3425	0.2051	1.5453	0.2140	0.9581	-468.5	237
20	0.9801	0.5262	3.4700	5.3392	0.1860	1.5200	0.2189	0.8498	-829.0	490
40	0.9835	0.5302	3.5242	5.3124	0.1807	1.3614	0.2189	0.8254	-1204	878
80	0.9872	0.5294	3.5495	5.3043	0.1840	1.3603	0.2190	0.8403	-1977	1616
100	0.9879	0.5295	3.5535	5.3027	0.1843	1.3535	0.2189	0.8421	-2397	2508
200	0.9898	0.5298	3.5640	5.2987	0.1840	1.3310	0.2185	0.8420	-4715	6730

not much difference between the estimated parameters for the one-step ahead maximum likelihood estimate and the multi-step maximum likelihood estimate with a very long prediction horizon, i.e. $N_p = 200$. The step responses for the estimated multi-step maximum likelihood estimate with a prediction horizon of $N_p = 1$, i.e. the one-step maximum likelihood estimate, and a prediction horizon of $N_p = 200$ are shown in Figure 3. They can hardly be distinguished. Hence, for all practical purposes they can be considered identical. This suggests that the one-step ahead prediction maximum-likelihood estimate should be applied in practice as the computing time for the one-step ahead prediction maximum-likelihood estimate is considerably lower than the computing time for the multi-step maximum-likelihood prediction with a long prediction horizon ($N_p = 200$). Figure 3 also depicts the step response of the true system. It is evident that the step responses of the estimated models approximate the true step response quite well.

IV. CONCLUSION

A constructive method for estimation of parameters in continuous-discrete-time stochastic systems described by

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.

TABLE V
SINGLE-STEP ML ESTIMATION IN MODEL (40).

j	K	α	τ	σ	r	σ/r	V	CPU sec.
1	1.0043	3.8386	5.7243	0.0658	0.2226	0.2959	-19.78	123
4	0.9911	3.6390	5.7547	0.0124	0.2424	0.0511	30.24	399
8	0.9811	3.5585	5.7792	0.0006	0.2490	0.0025	34.02	857
20	0.9812	3.5568	5.7802	0.0004	0.2455	0.0018	29.20	2245
40	0.9822	3.5750	5.7697	0.0002	0.2479	0.0009	48.26	4192
80	0.9747	3.5664	5.7618	0.0002	0.2547	0.0008	102.8	7448
100	1.0107	3.6458	5.6487	0.0002	0.3169	0.0006	539.5	8212
200	0.9465	3.3713	5.8331	0.0006	0.2556	0.0023	110.5	17885

TABLE VI
MULTI-STEP ML ESTIMATION IN MODEL (40).

j	K	α	τ	σ	r	σ/r	V	CPU sec.
1	1.0043	3.8386	5.7243	0.0658	0.2226	0.2959	-19.78	120
4	1.0043	3.8387	5.7244	0.0659	0.2424	0.2962	-79.72	160
8	1.0043	3.8386	5.7243	0.0658	0.2490	0.2956	-161.4	257
20	1.0043	3.8389	5.7244	0.0660	0.2455	0.2968	-398.4	382
40	1.0044	3.8394	5.7245	0.0666	0.2479	0.2995	-780.4	722
80	1.0039	3.8319	5.7256	0.0669	0.2547	0.3011	-1541	1550
100	1.0033	3.8277	5.7261	0.0670	0.3169	0.3018	-1954	2082
200	1.0024	3.8209	5.7269	0.0672	0.2556	0.3027	-4234	4268

transfer functions with time delays has been described and demonstrated. The method applies prediction-error criteria and the predictions are generated using the Kalman filter and predictor for a stochastic linear discrete-time state space model equivalent to the continuous-discrete-time stochastic transfer function model with time delays. In particular, an efficient computing scheme for the multi-step maximum likelihood prediction-error estimator is developed. The multi-step prediction-error criteria may be selected such that they are compatible with the optimization criterion applied by the predictive controller that uses the identified model. Compared to the single-step least-squares and the single-step maximum likelihood estimators, the multi-step maximum likelihood estimator produces parameter estimates that are less sensitive to the prediction horizon applied. In contrast to the single-step and multi-step least squares estimators, the multi-step maximum likelihood estimator computes unique parameters for the process and measurement noise. Hence, the multi-step maximum likelihood estimators are recommended for predictive control. Depending on the prediction horizon, the multi-step maximum likelihood estimator requires much more computer resources than the single-step one-step ahead least-squares predictor.

Consequently, based on the SISO simulation example, *we recommend the maximum likelihood (or maximum a posteriori) estimator based on the one-step-ahead prediction-error*. The models obtained using the multi-step maximum-likelihood prediction-error method with a prediction horizon of one and a very long prediction horizon are essentially identical. However, the long prediction horizon demands much more computational resources than the criterion based on the one-step-ahead prediction.

REFERENCES

[1] K. J. Åström, "Maximum likelihood and prediction error methods," *Automatica*, vol. 16, pp. 551–574, 1980.

[2] L. Ljung, *System Identification. Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

[3] N. R. Kristensen, H. Madsen, and S. B. Jørgensen, "Parameter estimation in stochastic grey-box models," *Automatica*, vol. 40, pp. 225–237, 2004.

[4] D. S. Shook, C. Mohtadi, and S. L. Shah, "A control-relevant identification strategy for GPC," *IEEE Transactions on Automatic Control*, vol. 37, pp. 975–980, 1992.

[5] R. B. Gopaluni, R. S. Patwardhan, and S. L. Shah, "The nature of data pre-filters in MPC relevant identification - open- and closed-loop issues," *Automatica*, vol. 39, pp. 1617–1626, 2003.

[6] ——, "MPC relevant identification - tuning the noise model," *Journal of Process Control*, vol. 14, pp. 699–714, 2004.

[7] M. Morari and J. H. Lee, "Model predictive control: Past, present and future," *Computers and Chemical Engineering*, vol. 23, pp. 667–682, 1999.

[8] S. B. Jørgensen and J. H. Lee, "Recent advances and challenges in process identification," in *Chemical Process Control - 6: Assessment and New Directions for Research*, J. B. Rawlings and B. A. Ogunnaike, Eds., January 2001.

[9] H. Hjalmarsson, "From experiments to closed loop control," in *13th IFAC Symposium on System Identification, 27-29 August 2003, Rotterdam, The Netherlands*. IFAC, 2003.

[10] M. Gevers, "A personal view on the development of system identification," in *13th IFAC Symposium on System Identification, 27-29 August 2003, Rotterdam, The Netherlands*. IFAC, 2003.

[11] ——, "Identification for control: Achievements and open problems," in *DYCOPS 7, Cambridge, Massachussetts USA, July 5-7*. IFAC, 2004.

[12] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York: Wiley, 1989.

[13] J. D. Hamilton, *Time Series Analysis*. Princeton, New Jersey: Princeton University Press, 1994.

[14] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[15] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[16] M. Morari and E. Zafiriou, *Robust Process Control*. Englewood Cliffs, New Jersey: Prentice-Hall, 1989.

Authorized licensed use limited to: Danmarks Tekniske Informationscenter. Downloaded on November 18, 2009 at 07:05 from IEEE Xplore. Restrictions apply.