Technical report number 2007-03

Application of a New Data-based Covariance Estimation Technique to a Nonlinear Industrial Blending Drum*

Murali R. Rajamani† and James B. Rawlings‡
Department of Chemical and Biological Engineering
University of Wisconsin-Madison
Madison, WI 53706

Tyler A. Soderstrom§
ExxonMobil Chemical
Baytown Technology Engineering Complex - East #3319
4500 Bayway Drive, Baytown TX 77522-9728

24 September 2007

*A version of this report was submitted for publication in *IEEE Transactions on Control Systems Technology*

†rajamani@bevo.che.wisc.edu, Author to whom correspondence should be addressed
‡rawlings@engr.wisc.edu
§tyler.a.soderstrom@exxonmobil.com

Abstract

State estimation requires knowledge about the noise statistics affecting the states and the measurements. These noise statistics are usually unknown and need to be estimated from operating data. We present a time-varying Autocovariance Least-Squares (ALS) technique to estimate the noise covariances using autocorrelations of the data at different time lags. The ALS technique is formulated for nonlinear models and simulations are presented to illustrate the value of the technique. An industrial application of the ALS technique is then presented on a nonlinear blending drum model and industrial operating data. The estimator calculated using the ALS technique gives improved state estimates as compared to rough industrial estimates for the covariances.

Keywords
Autocovariance least-squares, state estimation, time-varying systems, nonlinear state estimation, model predictive control

1 Introduction

Most physical systems have nonlinear dynamics when modelled from first principles. Even a simple reaction such as $A + B \rightarrow C$ has a nonlinear rate of reaction $r = kC_AC_B$ and the rate constant k has a nonlinear dependence on the temperature. A linear time-invariant approximation to the nonlinear model is often not valid when the dynamics are significantly nonlinear or the plant is operated under batch mode eg. a co-polymerization reactor requiring batch transitions. Techniques that have been developed for linear time-invariant models then have to be extended to nonlinear models and their validity tested.

Advanced control schemes for nonlinear models such as model predictive control often have a state estimator that reconciles past inputs and measurements to make an estimate of the current state of the system. The regulator then uses the current state estimate and the model to optimize future control inputs. The state estimator requires knowledge of the noise statistics affecting the plant for optimal performance. If the noises are assumed to have Gaussian distributions with zero means, the covariances are required to specify the noise statistics. In practice, these covariances are typically chosen arbitrarily by the process engineer to get satisfactory closed-loop performance. If the state estimates are inaccurate, then these in turn negatively affect the prediction of the controller.

In this paper we present the Autocovariance Least-Squares (ALS) technique that uses correlations in the data to estimate the separate noise covariances affecting the measurements and the states. For linear models, pioneering work in correlation based techniques was done by Mehra (1970) continued by Carew and Bélanger (1973); Neethling and Young (1974) and more recently by Odelson, Rajamani, and Rawlings (2006b); Rajamani and Rawlings (2007). In Section 2, the Autocovariance least-squares technique is formulated for nonlinear and time-varying models. The formulation is then applied to real industrial data from a nonlinear blending drum used at *ExxonMobil Chemical Company* [1]. Section 3 presents the model and the results to illustrate the value of the nonlinear extension to the ALS technique. The conclusions are then given in Section 4.

2 Noise Covariance Estimation for Nonlinear Models

Consider the following nonlinear state-space model:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + g(x_k, u_k) w_k \\ y_k &= h(x_k) + v_k \end{aligned} \tag{1}$$

where, $x \in \mathbb{R}^n$ is the state, $y \in \mathbb{R}^p$ are the measurements, $u_k \in \mathbb{R}^m$ are the inputs w_k, v_k and the noises corrupting the state and the measurements respectively. The noises w_k, v_k are assumed to be from a Gaussian distribution with mean zero and covariances Q_w, R_v. Also, w_k, v_k are assumed to be uncorrelated with each other.

Let the state estimates be obtained starting from an arbitrary initial value using a time-varying stable filter gain sequence L_k. An example of such a filter gain would be the sequence obtained by implementing the extended Kalman filter (EKF). The only condition on the time-varying gains L_k is that they are stable (see Assumption 1). The state estimation can then be described by the following equations:

$$\begin{aligned} \hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, u_k) \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} - L_k(y_k - \hat{y}_{k|k-1}) \\ \hat{y}_{k|k-1} &= h(\hat{x}_{k|k-1}) \end{aligned} \tag{2}$$

Here, $\hat{x}_{k|k-1}$ represents the predicted estimate of the state x_k given measurements and inputs up to time t_{k-1} and $\hat{x}_{k|k}$ the filtered estimate given measurements and inputs up to time t_k. Subtracting the predicted state estimates in

[1] *ExxonMobil Chemical Company* is a division of *ExxonMobil Corporation*

Equation 2 from the plant in Equation 1, we get an approximate time-varying linear model for the innovations:

$$\begin{aligned}\varepsilon_{k+1} &\approx \underbrace{(A_k - A_k L_k C_k)}_{\bar{A}_k} \varepsilon_k + \underbrace{\begin{bmatrix} G_k & -A_k L_k \end{bmatrix}}_{\bar{G}_k} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k} \\ \mathscr{Y}_k &\approx C_k \varepsilon_k + v_k\end{aligned} \tag{3}$$

where, $\varepsilon_k = (x_k - \hat{x}_{k|k-1})$ denotes the state estimate error and $\mathscr{Y}_k$ the innovations. The noises w_k, v_k driving the innovations sequence are assumed to drawn from time-invariant covariances Q_w, R_v.

The time-varying approximation of the nonlinear model is defined by the following linearization:

$$A_k = \left.\frac{\partial f(x_k, u_k)}{\partial x_k^T}\right|_{(\hat{x}_{k|k-1}, u_k)}, \quad G_k = g(\hat{x}_{k|k-1}, u_k), \quad C_k = \left.\frac{\partial h(x_k)}{\partial x_k^T}\right|_{\hat{x}_{k|k-1}} \tag{4}$$

A second possibility for the linear approximation is to evaluate the terms in Equation 4 at $\hat{x}_{k|k}$ instead of $\hat{x}_{k|k-1}$ (Anderson and Moore, 1979). The next section presents a technique for estimating the covariances Q_w, R_v using autocovariances of data at different time lags.

2.1 Time-varying Autocovariance Least-Squares Technique

The Autocovariance Least-Squares (ALS) covariance estimation technique described in Odelson et al. (2006b) and Rajamani and Rawlings (2007) was applied to a linear time-invariant model. When using nonlinear or time-varying models, a key difference is that the estimate error covariance $P_k = E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})]$ is the time-varying solution to the Riccati equation and does not reach a steady state value. No simple equation can then be written for P_k in terms of Q_w, R_v and the system matrices as in the linear time-invariant case. The following Assumption 1 however, allows extension of the ALS technique to time-varying and nonlinear systems.

Assumption 1. *The time-varying filter gain sequence L_k used in Equation 2 are such that when used in the approximate linearization given by Equation 3 they produce a sequence of $\bar{A}_k = (A_k - A_k L_k C_k)$ matrices such that the product $(\prod_{k=0}^{j} \bar{A}_i) \approx 0$ as j increases.*

Starting from an arbitrary initial condition ε_0 at t_0, consider the evolution of

Equation 3 up to time t_k to obtain the following:

$$\begin{aligned}
\mathscr{Y}_k &= C_k(\prod_{i=0}^{k-1} \bar{A}_i)\varepsilon_0 + C_k(\prod_{i=1}^{k-1} \bar{A}_i\bar{G}_0\bar{w}_0 + \prod_{i=2}^{k-1} \bar{A}_i\bar{G}_1\bar{w}_1 + \cdots + \bar{G}_{k-1}\bar{w}_{k-1}) + v_k \\
\mathscr{Y}_{k+j} &= C_{k+j}(\prod_{i=0}^{k+j-1} \bar{A}_i)\varepsilon_0 + C_{k+j}(\prod_{i=1}^{k+j-1} \bar{A}_i\bar{G}_0\bar{w}_0 + \prod_{i=2}^{k+j-1} \bar{A}_i\bar{G}_1\bar{w}_1 + \cdots + \bar{G}_{k+j-1}\bar{w}_{k+j-1}) + v_{k+j}
\end{aligned} \tag{5}$$

The covariance of $\bar{w}_k = \begin{bmatrix} w_k \\ v_k \end{bmatrix}$ is given by $\mathbb{Q} = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$.

From Assumption 1, the index $k > 0$ is chosen to be large enough such that $\prod_{i=0}^{k-1} \bar{A}_i \approx 0$. The effect of the initial estimate error ε_0 is then negligible in $\mathscr{Y}_{k+j}$ for $j > 0$. The terms involving ε_0 in Equation 5 can then be neglected. From Equation 5 we get the following expressions for the expectation of the autocovariances at different lags:

$$\begin{aligned}
E[\mathscr{Y}_k\mathscr{Y}_k^T] &= C_k(\prod_{i=1}^{k-1} \bar{A}_i\bar{G}_0 + \cdots + \bar{G}_{k-1})\mathbb{Q}(\bar{G}_0^T \prod_{i=1}^{k-1} \bar{A}_i^T + \cdots + \bar{G}_{k-1}^T)C_k^T + R_v \\
E[\mathscr{Y}_{k+j}\mathscr{Y}_k^T] &= C_{k+j}(\prod_{i=1}^{k+j-1} \bar{A}_i\bar{G}_0 + \cdots + \prod_{i=k}^{k+j-1} \bar{A}_i\bar{G}_{k-1})\mathbb{Q}(\bar{G}_0^T \prod_{i=1}^{k-1} \bar{A}_i^T + \cdots + \bar{G}_{k-1}^T)C_k^T \\
&\quad - C_{k+j}(\prod_{i=k+1}^{k+j-1} \bar{A}_i)A_kL_kR_v
\end{aligned} \tag{6}$$

In the above Equation 6, we define $\prod_{i=k}^{j} \bar{A}_i = I$ for all $j \le k$.

Remark 1. *In general, instead of starting from the initial condition ε_0 at t_0, we can start from an initial condition ε_m at t_m and calculate the expectations as above, provided the indices k and m are such that the following condition holds as in Assumption 1:*

$$\prod_{i=m}^{m+k-1} \bar{A}_i \approx 0$$

Let the autocovariance matrix $\mathscr{R}_k(N)$ be defined as the expectation of the innovations data at different time lags over a user defined window N (Jenkins and Watts, 1968).

$$\mathscr{R}_k(N) = E\begin{bmatrix} \mathscr{Y}_k\mathscr{Y}_k^T \\ \vdots \\ \mathscr{Y}_{k+N-1}\mathscr{Y}_k^T \end{bmatrix} \tag{7}$$

Using Equations 5 and 7, we get:

$$\mathscr{R}_k(N) = \underbrace{\begin{bmatrix} I \\ -C_{k+1}A_kL_k \\ \vdots \\ -C_{k+N-1}(\prod_{i=k+1}^{k+N-2}\bar{A}_i)A_kL_k \end{bmatrix}}_{\Psi_1} R_v + \Gamma\Omega_1 \bigoplus_{i=1}^{k} Q_w \Omega_1^T \Gamma_1^T + \Gamma\Omega_2 \bigoplus_{i=1}^{k} R_v \Omega_2^T \Gamma_1^T \tag{8}$$

where the matrices are defined as follows and dimensioned appropriately:

$$\Gamma = \begin{bmatrix} C_k(\prod_{i=1}^{k-1}\bar{A}_i) & \cdots & C_k \\ \vdots & \ddots & \vdots \\ C_{k+N-1}(\prod_{i=1}^{k+N-2}\bar{A}_i) & \cdots & C_{k+N-1}(\prod_{i=k}^{k+N-2}\bar{A}_i) \end{bmatrix} \tag{9}$$

and

$$\Omega_1 = \begin{bmatrix} G_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & G_{k-1} \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} -A_0L_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -A_{k-1}L_{k-1} \end{bmatrix}, \quad \Gamma_1^T = \begin{bmatrix} (\prod_{i=1}^{k-1}\bar{A}_i^T)C_k^T \\ \vdots \\ C_k^T \end{bmatrix} \tag{10}$$

Γ_1 is the first row block of the Γ matrix. In the above equation and in the remainder of this paper we use standard properties and symbols of Kronecker products (Steeb, 1991; Graham, 1981). $\otimes$ is the standard symbol for the Kronecker product and $\bigoplus$ is the symbol for the Kronecker sum, satisfying the following property:

$$\bigoplus_{i=1}^{k} Q_w = (I_k \otimes Q_w)$$

We use the subscript 's' to denote the column-wise stacking of the elements of a matrix into a vector. Stacking both sides of Equation 8 we then get:

$$[\mathscr{R}_k(N)]_s = (\Gamma_1\Omega_1 \otimes \Gamma\Omega_1)\mathcal{I}_{g,k}(Q_w)_s + \left[(\Gamma_1\Omega_2 \otimes \Gamma\Omega_2)\mathcal{I}_{p,k} + I_p \otimes \Psi\right](R_v)_s \tag{11}$$

Here, $\mathcal{I}_{p,N}$ is a permutation matrix containing 1's and 0's and satisfying the relation:

$$\left(\bigoplus_{i=1}^{N} R_v\right)_s = \mathcal{I}_{p,N}(R_v)_s$$

If we have an estimate of the autocovariance matrix $\mathscr{R}_k(N)$ denoted by $\hat{\mathscr{R}}_k(N)$ and let $\hat{b}_k = [\hat{\mathscr{R}}_k(N)]_s$, then from Equation 11 we can formulate a positive semidefinite constrained least-squares problem in the unknown covariances Q_w, R_v (Odelson et al., 2006b). The optimization to be solved is given by:

$$\begin{aligned} \Phi_k = \min_{Q_w, R_v} \left\| \mathscr{A}_k \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \hat{b}_k \right\|_W^2 \\ \text{subject to,} \quad Q_w, R_v \geq 0, \quad Q_w = Q_w^T, R_v = R_v^T \end{aligned} \tag{12}$$

where,

$$\begin{aligned} \mathscr{A}_k &= \begin{bmatrix} \mathscr{A}_{k1} & \mathscr{A}_{k2} \end{bmatrix} \\ \mathscr{A}_{k1} &= (\Gamma_1 \Omega_1 \otimes \Gamma \Omega_1) \mathcal{I}_{g,k} \\ \mathscr{A}_{k2} &= \left[(\Gamma_1 \Omega_2 \otimes \Gamma \Omega_2) \mathcal{I}_{p,k} + I_p \otimes \Psi \right] \end{aligned} \tag{13}$$

We will refer to the optimization in Equation 12 as the Autocovariance Least-Squares (ALS) technique. Necessary and sufficient conditions for the uniqueness of the ALS optimization in Equation 12 are given in (Rajamani and Rawlings, 2007). For the estimates of Q_w, R_v in the ALS optimization to have minimum variance the weighting matrix W in the ALS objective is given by $W = T_k^{-1}$ where T_k is the covariance of $\hat{b}_k$ (Aitken, 1935).

The matrices $\mathscr{A}_k$ and the vector $\hat{b}_k$ in the Equation 12 have the time subscript 'k' to emphasize that these quantities are time-varying and based on the time-varying approximation given in Equation 3.

Since the only data set available for estimating the time-varying quantity $\mathscr{R}_k(N)$ defined in Equation 7 is $\{\mathscr{Y}_k, \cdots, \mathscr{Y}_{k+N-1}\}$, the only calculable estimate of $\hat{b}_k$ is given by:

$$\hat{b}_k = [\hat{\mathscr{R}}_k(N)]_s = \begin{bmatrix} \mathscr{Y}_k \mathscr{Y}_k^T \\ \vdots \\ \mathscr{Y}_{k+N-1} \mathscr{Y}_k^T \end{bmatrix}_s \tag{14}$$

At every time instant t_k, we compute the quantities $\mathscr{A}_k$ and $\hat{b}_k$ from Equations 13 and 14. To simplify the computation, the matrices $\Gamma, \Omega_1, \Omega_2, \Gamma_1$ defined in Equations 9 and 10 and used in the calculation of $\mathscr{A}_k$, can be computed starting from an initial condition at time t_m rather than t_0 as given in Remark 1. We then use a sliding window strategy to compute the time-varying matrices $\mathscr{A}_k$ and $\hat{b}_k$. Figure 1 illustrates the calculation procedure.

Using the computed time-varying matrices and the ALS formulation in Equation 12, we can then solve the following optimization for a set of data of length

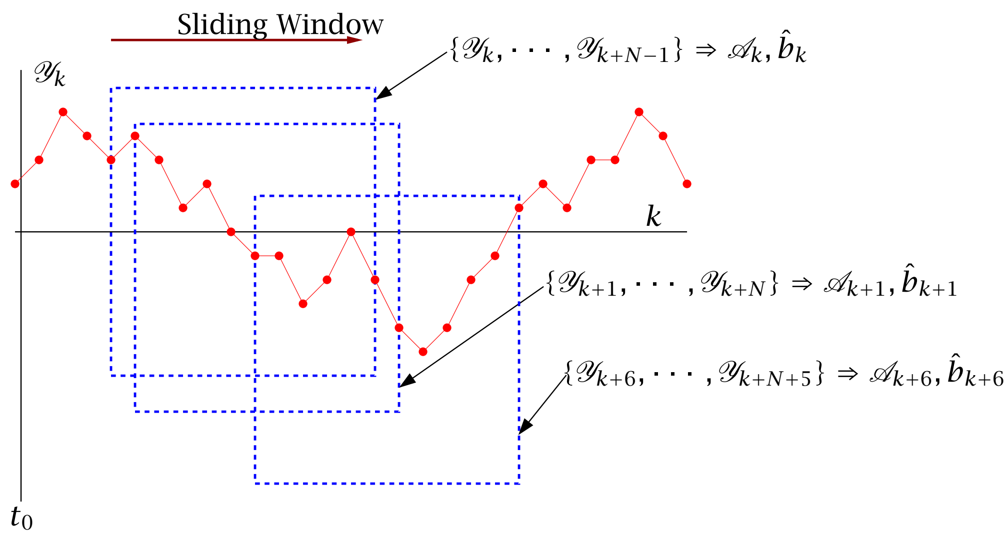

Figure 1: Strategy for calculating the time-varying $\mathscr{A}_k$ matrices in Equation 13

N_d to estimate Q_w, R_v:

$$\Phi = \min_{Q_w, R_v} \left\| \begin{bmatrix} \mathscr{A}_k \\ \vdots \\ \mathscr{A}_{N_d-N+1} \end{bmatrix} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \begin{bmatrix} \hat{b}_k \\ \vdots \\ \hat{b}_{N_d-N+1} \end{bmatrix} \right\|_{W_f}^2 \tag{15}$$
$$\text{subject to,} \quad Q_w, R_v \geq 0, \quad Q_w = Q_w^T, R_v = R_v^T$$

Since $\hat{b}_k, \hat{b}_{k+1}, \cdots$ are not independent, the weighting matrix W_f is not block diagonal. The formula for W_f is a complicated function of the unknown covariances Q_w, R_v and an iterative procedure is required for its calculation (Rajamani and Rawlings, 2007). We use $W_f = I$ to avoid the computationally infeasible and nonlinear calculation.

Remark 2. *Notice that if the system is time-invariant we have $\mathscr{A}_k = \cdots = \mathscr{A}_{N_d-N+1}$ and we can then recover the time-invariant ALS optimization presented in Odelson et al. (2006b) and Rajamani and Rawlings (June, 2007).*

Remark 3. *Assumption 1 is a simple practical requirement that the time-varying linear approximation of the full nonlinear model has a gain sequence that makes the estimate error asymptotically zero. This requirement is satisfied in most industrial applications that use a linear approximation to design the state estimator.*

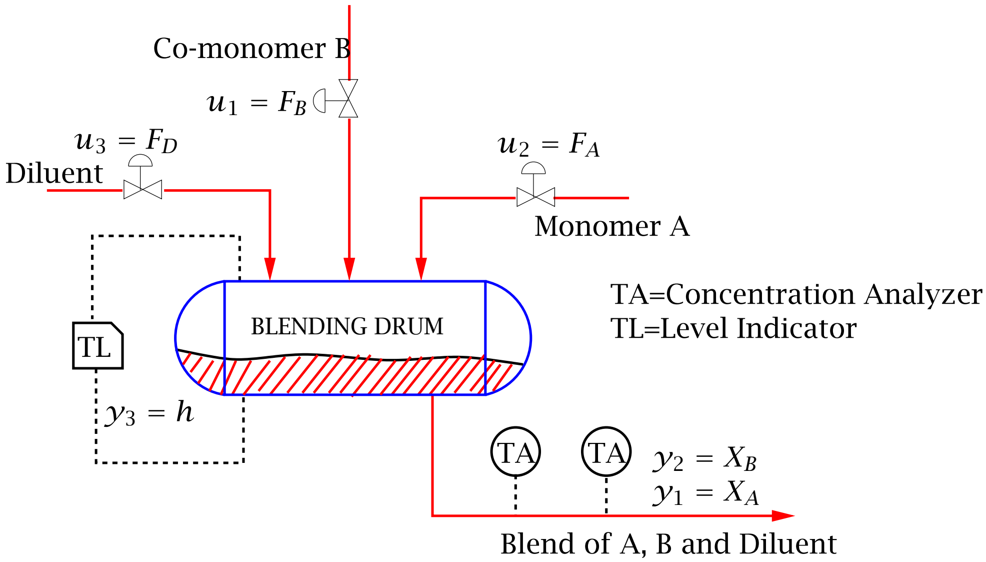

Figure 2: Schematic diagram of the blending drum: the level and mass fractions h, X_A, X_B are controlled by flows F_A, F_B, F_D

3 Industrial Blending Drum Example

Figure 2 shows a schematic diagram of a blending drum where two components, monomer A and co-monomer B are mixed with a diluent to give the correct proportions for the blends used as the feed for the different grades of polymers made in the reactors. The diagram is a representation of a blending process unit at *ExxonMobil Chemical Company*. The mass fractions of A and B inside the blending drum are represented by X_A and X_B. The blending tank is elliptical at the bottom and the dynamics of the blending process are nonlinear.

The states of the system are the mass fractions X_A, X_B and the level in the drum h. The nonlinearity in the drum is given by the following equations where V is the volume of the blend in the drum and C_1, C_2, C_3, C_4 are constants:

$$V = C_1 h^3 + C_2 h^2 + C_3 h + C_4$$
$$\frac{dV}{dh} = 3C_1 h^2 + 2C_2 h + C_3$$

A diluent with mass flow rate F_D is added to the blending tank to maintain the required monomer to co-monomer ratio. If the mass flow rates of A and B into the drum are given by F_A and F_B respectively, we can write simple mass balances for the dynamics of the states X_A, X_B, h.

$$\begin{aligned}
\frac{dX_A}{dt} &= \frac{F_A - X_A(F_A + F_B + F_D)}{\rho(C_1h^3 + C_2h^2 + C_3h + C_4)} \\
\frac{dX_B}{dt} &= \frac{F_B + X_{BD}F_D - X_B(F_A + F_B + F_D)}{\rho(C_1h^3 + C_2h^2 + C_3h + C_4)} \\
\frac{dh}{dt} &= \frac{(F_A + F_B + F_D) - F_{out}}{\rho(C_1h^2 + C_2h + C_3)}
\end{aligned} \tag{16}$$

where, X_{BD} is the mass fraction of B in the diluent feed, F_{out} is the measured out flow rate of the blend and ρ is the recorded density of the blend.

The controlled variables are the level of inventory in the drum h, the ratio of the mass fractions of A to B, $\frac{X_A}{X_B}$ in the outlet stream and the fraction X_A of A leaving the drum. Measurements are made for the states X_A, X_B, h at intervals of one minute.

Five sets of industrial operating data were provided by *ExxonMobil Chemical Company* with the characteristics described below:

- **Data Set I:** Steady state operating data
- **Data Set II:** Set point change in the liquid level h
- **Data Set III:** Step change in the in flow rate F_B
- **Data Set IV:** Set point change in the liquid level and the ratio of the monomer to co-monomer mass fractions
- **Data Set V:** Same as Data Set IV

The state-space model in discrete time for the above equations can then be represented as:

$$\begin{aligned}
x_{k+1} &= f(x_k, u_k) + Gw_k \\
y_k &= x_k + v_k
\end{aligned} \tag{17}$$

in which,

$$x_k = \begin{bmatrix} X_A \\ X_B \\ h \end{bmatrix}, \quad u_k = \begin{bmatrix} F_A \\ F_B \\ F_D \end{bmatrix}$$

The G matrix is chosen as the constant $G = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ in the model, which implies the state noise in h is more significant than the noise in the states X_A, X_B. The state x_{k+1} is calculated by integrating the nonlinear model given in Equation 16 from x_k over the discretization time of one minute while keeping the input u_k constant. The states and the measurements are corrupted by noises w_k, v_k with unknown covariances Q_w, R_v respectively.

3.1 Simulation Results

To illustrate the value of the ALS technique applied to a nonlinear model, the data were generated by drawing noises w_k and v_k from known covariances Q_w, R_v and simulating the nonlinear model from Equations 16 as given in Equation 17. Since a steady state simulation is equivalent to using a linear time-invariant approximation to the nonlinear model, the data are simulated by making multiple set point changes to the level in the drum and implementing a simple PI controller to maintain the level. The set point changes introduced are such that a simple linear time-invariant approximation of the full nonlinear model is not valid.

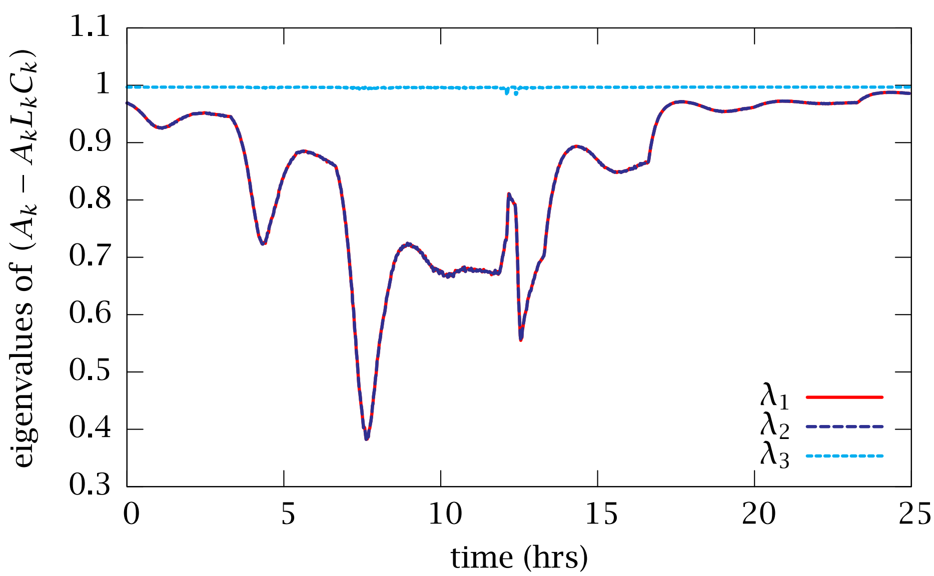

Figure 3: Eigenvalues of $(A_k - A_k L_k C_k)$ plotted against time for the simulated data

Figure 3 shows a plot of the eigenvalues of the time-varying $(A_k - A_k L_k C_k)$ matrices from the simulation as a function of time. From the plot in Figure 3 it is clear that a linear time-invariant approximation of the nonlinear model is not applicable.

The covariances in the simulated data are chosen as:

$$Q_w = 2 \times 10^{-5}, \qquad R_v = \begin{bmatrix} 2 \times 10^{-9} & 0 & 0 \\ 0 & 3.2 \times 10^{-7} & 0 \\ 0 & 0 & 3 \times 10^{-3} \end{bmatrix}$$

The ALS technique with the time-varying strategy described in Section 2 is then applied to the simulated data to recover the noise covariances.

Figure 4 shows a scatter plot of the covariances estimated using the ALS technique repeated 200 times with $N_d = 200$ and $N = 15$. The mean of the

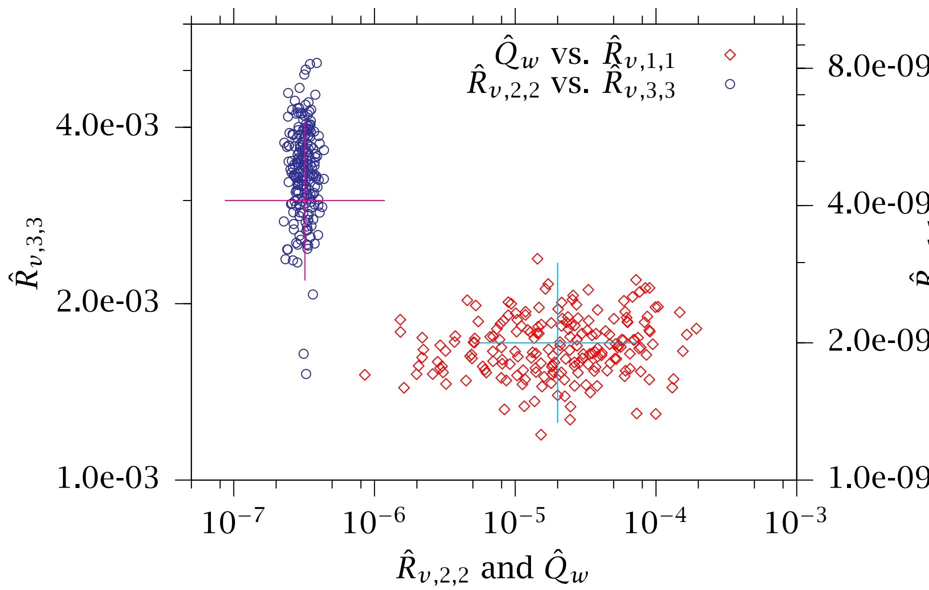

Figure 4: Scatter in the covariance estimates using the ALS technique on simulated data. The + sign shows the actual value of the covariances

estimates is calculated as:

$$\hat{Q}_w = 3.23 \times 10^{-5}, \qquad \hat{R}_v = \begin{bmatrix} 1.98 \times 10^{-9} & -2.71 \times 10^{-11} & 6.01 \times 10^{-9} \\ -2.71 \times 10^{-11} & 3.17 \times 10^{-7} & -1.12 \times 10^{-7} \\ 6.01 \times 10^{-9} & -1.12 \times 10^{-7} & 3.42 \times 10^{-3} \end{bmatrix}$$

The variance in the estimates seen in the Figure 4 decreases as N_d is increased. The covariance estimates using the ALS technique are thus close to the actual covariances used in the simulation.

3.2 Using Real Industrial Data

The ALS technique is next applied to five operating data sets provided by *ExxonMobil Chemical Company*. To apply the ALS technique, the innovations sequence defined as the difference between the measured outputs and the simulated model outputs should be zero mean. To remove the modelling errors and to ensure zero mean in the innovations, integrating disturbances are added to the inputs and the outputs in the model (Tenny, Rawlings, and Wright, 2004; Pannocchia and Rawlings, 2002). A state estimator is then used to estimate the integrating disturbance vector, which then compensates for the plant/model mismatch. To ensure detectability of the integrating disturbance model the number of disturbances is chosen as $n_d = 3$ (Muske and Badgwell, 2002; Pannocchia and Rawlings, 2002) (Also see Remark 4).

Remark 4. *Note that the level in the tank is an integrator and the model of the plant is not perfect. Hence, when the inputs provided in the data sets is used in the model simulation, the simulated $\hat{y}_{k,3}$ and the measured h from the data can never match. This issue is resolved by adding an integrating input disturbance to the state h. Also note that since the dynamics for h are integrating in the open-loop, an output disturbance model cannot be used to remove this mismatch (Qin and Badgwell, 2003; Dougherty and Cooper, 2003).*

The model in Equation 17 augmented with the disturbance model is given by:

$$x_{k+1} = f(x_k, u_k) + B_d d_k + G w_k \tag{18}$$

$$d_{k+1} = d_k + \xi_k \tag{19}$$

$$y_k = x_k + C_d d_k + v_k \tag{20}$$

where, d_k is the integrating disturbance and $\xi \sim N(0, Q_\xi)$ is the noise model for the disturbance. We make the following choice for B_d and C_d:

$$B_d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad C_d = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The model is simulated with the inputs and measurements given in the data sets and the extended Kalman filter (EKF) used as the initial state estimator. Since the covariances affecting the operating data are unknown, the following initial guesses for the covariances are used to calculate the initial estimator gain sequence in the EKF:

$$Q_w = 1 \times 10^{-14}, \quad Q_\xi = 10^{-9} \times I_3, \quad R_v = 2 \times \text{diag}\,(10^{-7}, 10^{-5}, 10^{-3}) \tag{21}$$

These choices are made by assigning the variance in the operating data measurements to the corresponding diagonal elements of R_v and choosing small values for Q_w, Q_ξ. This choice follows the rough industrial guidelines of assigning the noise covariances to the output when calculating state estimator gains.

The initial estimator gain sequence L_k obtained by implementing the EKF satisfies Assumption 1 and hence the ALS technique can be applied to the operating data. With the above model and the industrial data, the covariances Q_w, Q_ξ and R_v are estimated. The diagonal elements of the estimated covariances Q_w, Q_ξ, R_v using the ALS technique for the five data sets are shown in Figures 5 and 6. The figures show that the estimated covariances have consistent values across all the five industrial data sets provided. These estimated

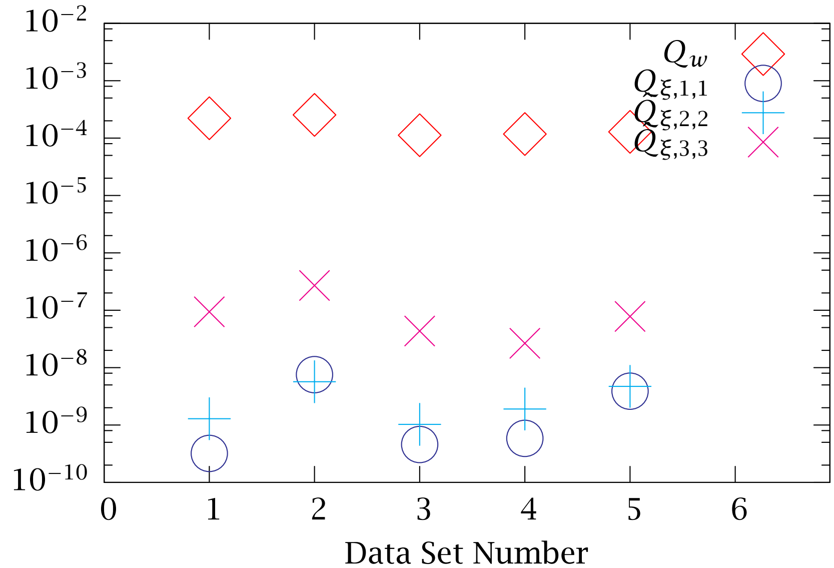

Figure 5: Comparison of the diagonal elements of Q_w and Q_ξ for the 5 industrial data sets

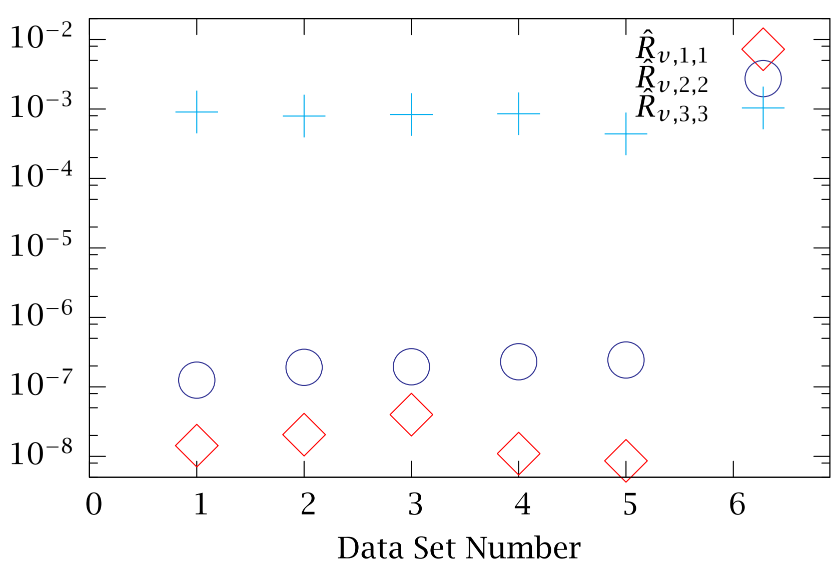

Figure 6: Comparison of the diagonal elements of R_v for the 5 industrial data sets

covariances are then used to calculate new filter gains for the EKF. A snapshot of the measurements compared against the model estimates with the initial and the ALS covariances for Data Set I is shown in Figures 7, 8 and 9. The y-axis on the figures have been arbitrarily scaled to disguise the original industrial operating data. The figures also show the estimated disturbance vector $\hat{d}_k$ with the initial and the ALS covariances.

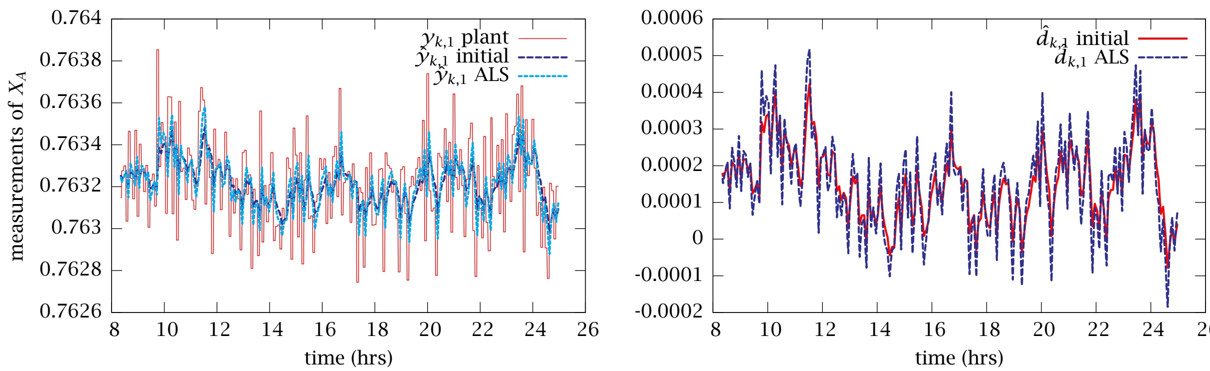

Figure 7: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

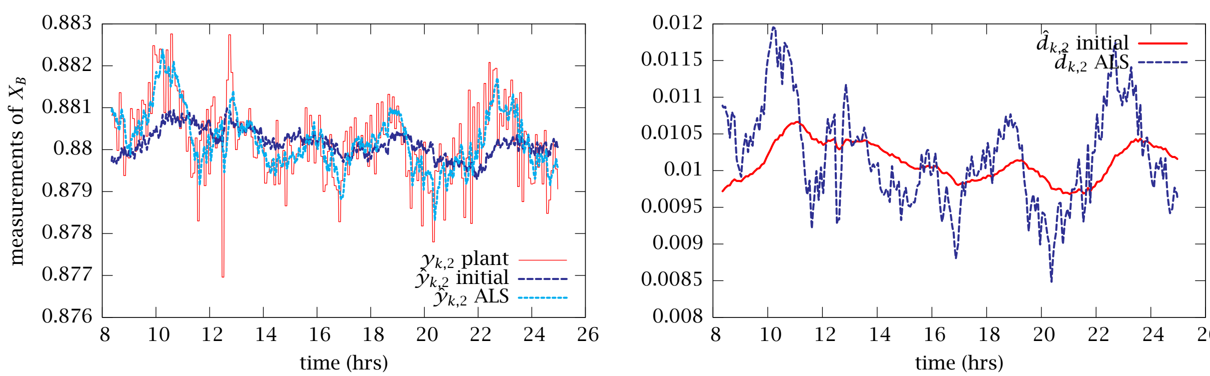

Figure 8: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

Remark 5. *If the operating data is collected close to the steady state, the time-varying linear approximation of the nonlinear model given in Equation 3 can be further simplified to a linear time-invariant model since the linearized terms in Equation 4 are evaluated close the steady state and have approximately the*

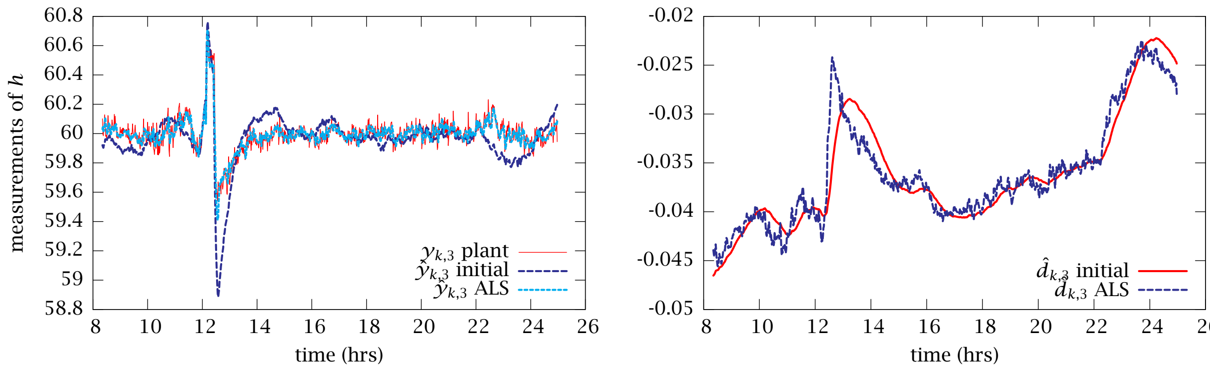

Figure 9: A snapshot of data comparing the model estimates using the initial and the ALS covariances with the operating data in Data Set I (the y-axis is arbitrarily scaled to disguise the original data)

same values. When there are multiple set point changes in the operating data, the transience can be ignored and the linearization at the new set point can be approximated as a new time-invariant linear model. This time-invariant approximation is useful to compare the state estimator specified with the ALS technique against other choices when the actual states are unknown.

A check for the optimality of the state estimator is to test the correlations of the innovations $\mathscr{Y}_k = (y_k - h(\hat{x}_{k|k-1}))$ for whiteness and zero mean (Gelb, 1974, p. 319). For the linear time-invariant model, when the optimal state estimator gains are implemented, the autocovariance matrix in Equation 7 is given as:

$$\mathscr{R}_k(N) = \begin{bmatrix} CPC^T + R_v \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where, P is the time-invariant solution to the Riccati equation with the actual covariances Q_w, R_v. Note that the lagged autocovariances are zero.

We use a similar approach to test the state estimator specified with the covariances Q_w, Q_ξ, R_v estimated using the ALS technique. Following Remark 5 and the above analysis of the autocovariances for the time-invariant model, we only consider innovations of the steady state operating data and neglect the transience between the set point changes. If the values of the estimated covariances are close to the actual covariances in the operating data, we expect the innovations approximated from a linear time-invariant model to be uncorrelated in time and the lagged autocovariances of the data to be close to zero.

Figure 10 shows a plot comparing the autocovariances estimated from steady state data at different time lags using filter gains from the ALS technique and the initial covariances in Equation 21. As seen in the figure, the values of the lagged autocovariances are close to zero after a time lag of 4 and remain zero with increasing time lags when the ALS covariance estimates are used. On the other hand, the lagged autocovariances remain nonzero when the state estimator is specified with the rough initial covariances. Clearly the identified covariances with the integrating disturbance model have improved the accuracy of the state estimates. We can also see there remains some undermodelling of the plant disturbance given the nonzero autocovariance at small lags.

An estimate of $E[\mathscr{Y}_k \mathscr{Y}_k^T] = CPC^T + R_v$ for Data Set I, calculated using a time-invariant linear approximation by averaging the time-varying matrices and covariances $\hat{Q}_w, \hat{Q}_\xi, \hat{R}_v$ specified using the ALS technique is given by:

$$\begin{aligned} C_m \hat{P} C_m^T &\approx \text{diag}\,(2.3 \times 10^{-9}, 1.3 \times 10^{-8}, 6.0 \times 10^{-4}) \\ \hat{R}_v &\approx \text{diag}\,(1.43 \times 10^{-8}, 1.25 \times 10^{-7}, 9.05 \times 10^{-4}) \\ \text{giving,} \qquad C_m \hat{P} C_m^T + \hat{R}_v &\approx \text{diag}\,(1.6 \times 10^{-8}, 1.4 \times 10^{-7}, 1.5 \times 10^{-3}) \end{aligned}$$

where, C_m is the average of the time varying linearization C_k for Data Set I and $\hat{P}$ is calculated as the solution of the Riccati equation from the average time-varying system. These values are in good agreement with the estimated zero-lag autocovariances shown in Figure 10.

The plots for the other data sets and similar and presented in Appendix A.

4 Conclusions

The Autocovariance Least-Squares (ALS) procedure was extended to nonlinear models to estimate the noise covariances from operating data. A semidefinite constrained optimization with a least-squares objective was solved to give the covariance estimates. The ALS technique was then shown to give good covariance estimates from data simulated with known noise statistics. The practical benefits of the technique was illustrated by an application to blending drum operating data sets provided by *ExxonMobil Chemical Company*. Consistent values for the covariances were estimated in all the industrial data sets provided by *ExxonMobil Chemical Company*. The state estimator specified with covariances calculated using ALS technique produced white innovations implying optimality in comparison to an initial choice for the covariances. The ALS technique was also used to estimate the covariances for integrating output disturbance models to remove the plant/model mismatch. The improvement in the estimator performance implies cost benefits in the implementation of advanced control schemes (see for eg. Odelson, Lutz, and Rawlings (2006a) for numerical values) when the ALS technique is used to estimate the noise covariances.

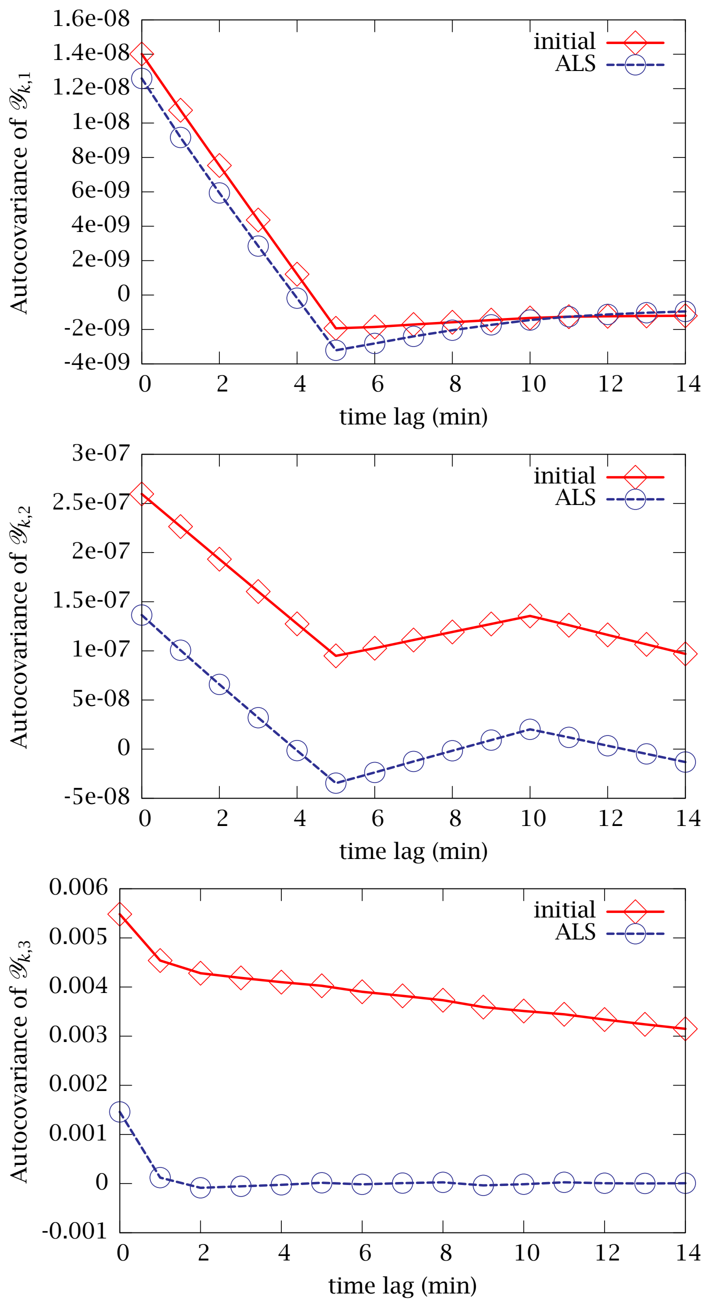

Figure 10: Data Set I: Diagonal elements of the innovations autocovariance matrix at different time lags

5 Acknowledgments

The authors acknowledge financial support from NSF through grant #CNS–0540147 and PRF through grant #43321–AC9 and the support of ExxonMobil Chemicals through the Texas-Wisconsin Modeling and Control Consortium (TWMCC).

References

A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1935.

B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J., 1979.

B. Carew and P. Bélanger. Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 18(6):582–587, 1973.

D. Dougherty and D. J. Cooper. Tuning guidelines of a dynamic matrix controller for integrating (non-self-regulating) processes. *Industrial and Engineering Chemistry Research*, 42(8):1739–1752, 2003.

A. Gelb, editor. *Applied Optimal Estimation*. The M.I.T. Press, Cambridge, Massachusetts, 1974.

A. Graham. *Kronecker products and matrix calculus with applications*. Ellis Horwood Limited, West Sussex, England, 1981.

G. Jenkins and D. Watts. *Spectral Analysis and its Applications*. Holden-Day, 500 Sansome Street, San Fransisco, California, 1968.

R. Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(12):175–184, 1970.

K. R. Muske and T. A. Badgwell. Disturbance modeling for offset-free linear model predictive control. *Journal of Process Control*, 12(5):617–632, 2002.

C. Neethling and P. Young. Comments on “Identification of optimum filter steady-state gain for systems with unknown noise covariances”. *IEEE Transactions on Automatic Control*, 19(5):623–625, 1974.

B. J. Odelson, A. Lutz, and J. B. Rawlings. The autocovariance least-squares methods for estimating covariances: Application to model-based control of chemical reactors. *IEEE Control Systems Technology*, 14(3):532–541, May 2006a.

B. J. Odelson, M. R. Rajamani, and J. B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, February 2006b. URL `http://www.elsevier.com/locate/automatica`.

G. Pannocchia and J. B. Rawlings. Disturbance models for offset-free MPC control. *AIChE Journal*, 49(2):426–437, 2002.

S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, 11(7):733–764, 2003.

M. R. Rajamani and J. B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. Technical Report 2007–02, TWMCC, Department of Chemical and Biological Engineering, University of Wisconsin-Madison (Available at http://jbrwww.che.wisc.edu/tech-reports.html), June 2007.

M. R. Rajamani and J. B. Rawlings. Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. Submitted for publication in *Automatica*, June, 2007.

W. Steeb. *Kronecker product of matrices and applications*. Mannheim, 1991.

M. J. Tenny, J. B. Rawlings, and S. J. Wright. Closed-loop behavior of nonlinear model predictive control. *AIChE Journal*, 50(9):2142–2154, September 2004. URL `http://www.aiche.org/aichejournal/`.

A Plots showing the Improvement in the State Estimates for the Industrial Data Sets II, III, IV and V

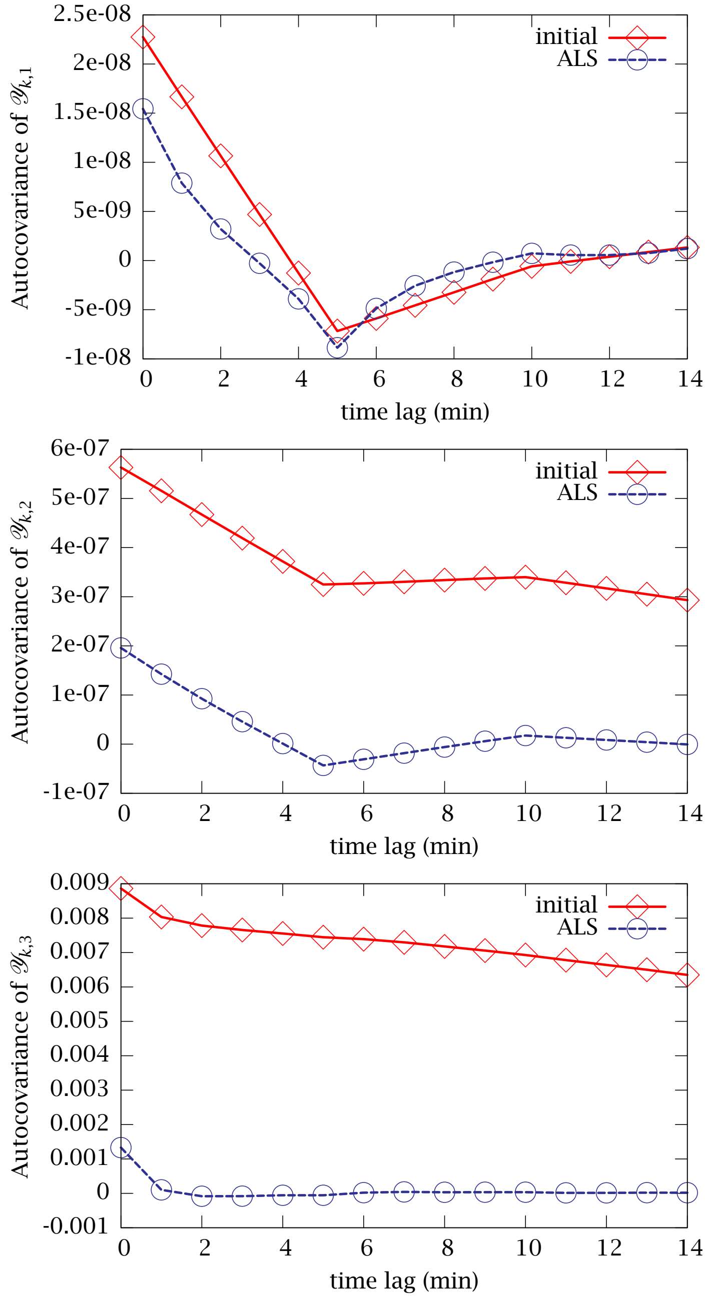

Figure 11: Data Set II: Diagonal elements of the innovations autocovariance matrix at different time lags

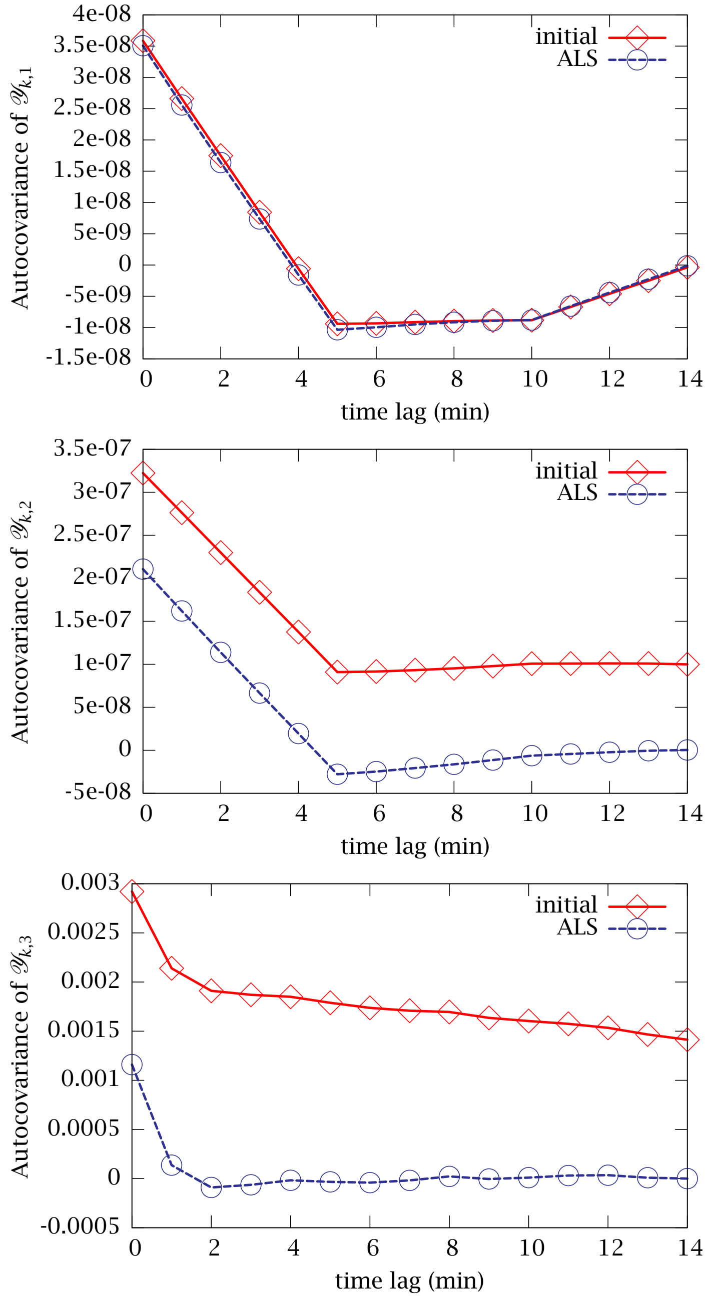

Figure 12: Data Set III: Diagonal elements of the innovations autocovariance matrix at different time lags

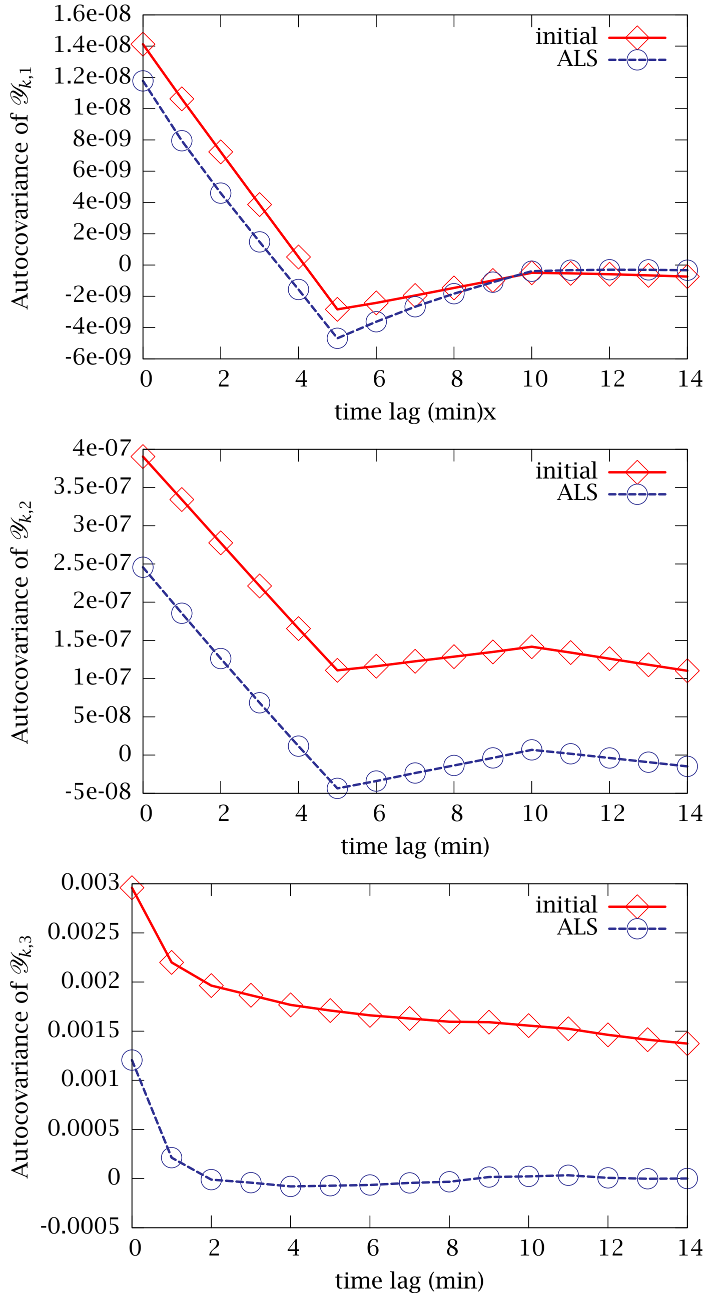

Figure 13: Data Set IV: Diagonal elements of the innovations autocovariance matrix at different time lags

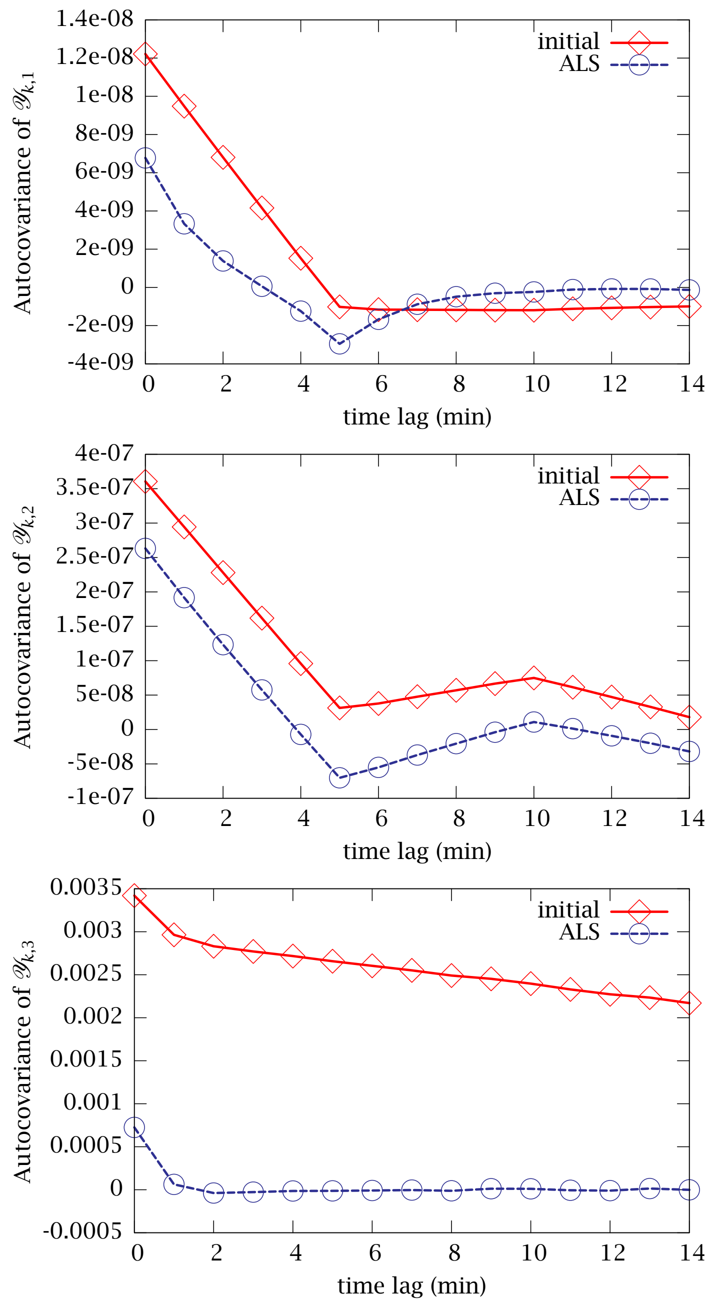

Figure 14: Data Set V: Diagonal elements of the innovations autocovariance matrix at different time lags