

The Autocovariance Least-Squares Method for Estimating Covariances: Application to Model-Based Control of Chemical Reactors

Brian J. Odelson, Alexander Lutz, and James B. Rawlings

Abstract—This paper demonstrates the autocovariance least-squares (ALS) technique on two chemical reactor control problems. The method uses closed-loop process data to recover the covariances of the disturbances entering the process, which are required for state estimation. The data used for this purpose may be collected with or without the controllers running. We do not assume that the plant is at steady state nor that only nonzero disturbances are affecting the plant at the time of data collection. The ALS method also accounts for integrated white noise disturbances, which are required for offset-free control. Two examples are provided in this paper: a carefully controlled laboratory reactor and an industrial reactor controlled by a state-of-the-art advanced model predictive control (MPC) system. A variety of control scenarios are tested and the results demonstrate that the ALS method has the potential to improve the best industrial practice of process control by a factor of three to five.

Index Terms—Covariance matrices, disturbance model, industrial control, optimal control, state estimation.

I. INTRODUCTION

CONSIDER the following discrete-time state-space model:

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

$$y_k = Cx_k + v_k \quad (1b)$$

in which k is time, x is the state of the system, u is the manipulated input, y is the measured output, w is the disturbance to the process, and v is the disturbance to the measurement. The dimensions of the system are given by $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $G \in \mathbb{R}^{n \times g}$, and $C \in \mathbb{R}^{p \times n}$. The two disturbances, $\{w_k\}_{k=0}^{N_d}$ and $\{v_k\}_{k=0}^{N_d}$, are modeled as zero-mean Gaussian noise sequences with covariances Q_w and R_v , respectively. Estimates of the states of the system are reconstructed using the standard Kalman filter

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + AL_k(y_k - C\hat{x}_{k|k-1}). \quad (2)$$

Manuscript received December 3, 2003; revised April 19, 2005. Manuscript received in final form August 24, 2005. Recommended by Associate Editor D. Rivera. This brief was supported in part by the industrial members of the Texas-Wisconsin Modeling and Control Consortium and by the National Science Foundation under Grant CTS-0105360.

B. J. Odelson is with BP, Research and Technology, Naperville, IL 60540 USA (e-mail: odelbj@bp.com).

A. Lutz is with the Institute for System Dynamics and Control Engineering, University of Stuttgart, Stuttgart 70174, Germany (e-mail: alexander.lutz@gmx.de).

J. B. Rawlings is with the Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: jbraw@che.wisc.edu).

Digital Object Identifier 10.1109/TCST.2005.860519

The estimate error is defined as $\varepsilon_k \equiv x_k - \hat{x}_{k|k-1}$, with covariance $P_{k|k-1}$. This covariance $P_{k|k-1} = E[\varepsilon_k \varepsilon_k^T]$ is the solution to the Riccati equation

$$P_{k+1|k} = AP_{k|k-1}A^T + GQ_wG^T - AP_{k|k-1}C^T[CP_{k|k-1}C^T + R_v]^{-1}CP_{k|k-1}A^T \quad (3)$$

and the Kalman gain L_k is defined as

$$L_k = P_{k|k-1}C^T[CP_{k|k-1}C^T + R_v]^{-1}. \quad (4)$$

A. Current Industrial Approach: Finding the Filter Gain by Tuning

If the covariances of the disturbances Q_w and R_v are known, it is a simple matter to find the *optimal* filter gain L using the steady-state forms of (3) and (4). This situation is the best case scenario. In practical model predictive control (MPC) implementations, however, the covariances of the disturbances entering the process generally are not known. If one does not use process data to determine these covariances in some systematic fashion, then practitioners are forced to adopt an on-line “tuning” strategy. Tuning in this context means choosing an initial L , observing the closed-loop performance behavior of the manipulated input and the measured output, and then changing the value of L in an attempt to improve this performance. One adjusts L in this adhoc fashion until one is unable to further improve the closed-loop control performance. This tuning approach can be time consuming and prone to failure. The resulting performance is often far from the best achievable, as we demonstrate in Section III of the paper by examining industrial data provided by Eastman Chemical.

B. Overview of the Autocovariance Least Squares (ALS) Technique [1]

The more systematic and preferable approach to determine the filter gain is to estimate the covariances from data. That is the approach of adaptive filtering. The methods in this field can be divided into four general categories: Bayesian [2], [3], maximum likelihood [4], [5], covariance matching [6], and correlation techniques. Bayesian and maximum likelihood methods have fallen out of favor because of their sometimes excessive computation times. Covariance matching is the computation of the covariances from the residuals of the state estimation problem, but has been shown to give biased estimates of the true covariances. The fourth category is correlation techniques, largely pioneered by Mehra [7], [8] and Carew and Bélanger [9], [10]. The correlation method is the most popular and highly cited method for determining these covariances. These methods have been applied only to open-loop data,

and have not addressed disturbance models, which we discuss subsequently. In [11], a variety of techniques were developed to solve the covariance estimation problem using closed-loop plant data. The problem can be divided into two cases: 1) when the disturbance covariances can be estimated uniquely from the data and 2) when the covariances cannot be estimated uniquely. The former category is treated in detail in [1], which we briefly summarize here. Assume we process the $y(k)$ to obtain state estimates using a linear filter with gain L , which is not necessarily the optimal L for the system. The state estimation error $\varepsilon(k) = x_k - \hat{x}_{k|k-1}$ evolves according to

$$\varepsilon_{k+1} = \underbrace{(A - ALC)}_{\bar{A}} \varepsilon_k + \underbrace{[G \quad -AL]}_{\bar{G}} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k}. \quad (5)$$

We define the state-space model of the innovations $\mathcal{Y}_k \equiv y_k - C\hat{x}_{k|k-1}$ as

$$\begin{aligned} \varepsilon_{k+1} &= \bar{A}\varepsilon_k + \bar{G}\bar{w}_k \\ \mathcal{Y}_k &= C\varepsilon_k + v_k \end{aligned}$$

and we require subsequently that the system is detectable and the chosen filter is stable.

Assumption 1.1: (A, C) is detectable.

Assumption 1.2: $\bar{A} = A - ALC$ is stable.

A stable filter gain L exists because of Assumption 1.1. In this formulation, the state and sensor noises are correlated

$$E[\bar{w}_k(\bar{w}_k)^T] \equiv \bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix} \quad E[\bar{w}_k v_k^T] = \begin{bmatrix} 0 \\ R_v \end{bmatrix}.$$

Because the filter is stable (Assumption 1.2), as k increases, the mean converges to zero

$$E(\varepsilon(k)) \rightarrow 0 \quad \text{cov}(\varepsilon(k)) \rightarrow P$$

and the covariance approaches a steady state given by the solution to the following Lyapunov equation:

$$P = \bar{A}P\bar{A}^T + \bar{G}\bar{Q}\bar{G}^T. \quad (7)$$

We, therefore, assume that we have chosen k sufficiently large so that the effects of the initial conditions can be neglected or, equivalently, we choose the steady state as the initial condition

Assumption 1.3: $E(\varepsilon(0)) = 0$ $\text{cov}(\varepsilon(0)) = P$.

Now consider the autocovariance, defined as the expectation of the data with some lagged version of itself [12]

$$\mathcal{C}_j = E[\mathcal{Y}_k \mathcal{Y}_{k+j}^T]. \quad (8)$$

Using (5) and the steady-state initial condition (Assumption 1.3) gives for the autocovariance

$$E(\mathcal{Y}_k \mathcal{Y}_k^T) = CPC^T + R_v \quad (9)$$

$$E(\mathcal{Y}_{k+j} \mathcal{Y}_k^T) = C\bar{A}^j PC^T - C\bar{A}^{j-1} ALR_v \quad (10)$$

which are independent of k because of our assumption about the initial conditions. The symmetric autocovariance matrix (ACM) is then defined as

$$\mathcal{R}(N) = \begin{bmatrix} \mathcal{C}_0 & \cdots & \mathcal{C}_{N-1} \\ \vdots & \ddots & \vdots \\ \mathcal{C}_{N-1}^T & \cdots & \mathcal{C}_0 \end{bmatrix}. \quad (11)$$

The number of lags used in the ACM is a user-defined parameter N . The off-diagonal autocovariances are not assumed zero, because we do not process the data with the optimal filter, which is unknown. The ACM of the innovations can be written as follows:

$$\begin{aligned} \mathcal{R}(N) &= OPC^T + \Gamma \left[\bigoplus_{i=1}^N \bar{G} \bar{Q}_w \bar{G}^T \right] \Gamma^T + \Psi \left[\bigoplus_{i=1}^N R_v \right] \\ &+ \left[\bigoplus_{i=1}^N R_v \right] \Psi^T + \bigoplus_{i=1}^N R_v \end{aligned} \quad (12)$$

in which

$$\begin{aligned} \mathcal{O} &= \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix} \quad \Psi = \Gamma \left[\bigoplus_{j=1}^N (-AL) \right] \\ \Gamma &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \\ C\bar{A}^{N-2} & \cdots & C & 0 \end{bmatrix}. \end{aligned}$$

In this result and those to follow, we employ the standard definitions of the Kronecker product, Kronecker sum, and the direct sum [13]. In order to use the ACM relationship in a standard least-squares problem, we apply the “vec” operator, which is the columnwise stacking of a matrix into a vector [13]. If z_i is the i th column of the Z matrix

$$\text{vec}(Z) = Z_s = [z_1^T \quad \cdots \quad z_n^T]^T.$$

Throughout this paper, we use the s subscript to denote the outcome of applying the vec operator. Applying the vec operator to (12) and using the result of applying the vec operator on (7) yields

$$\begin{aligned} & \left[\mathcal{R}(N) \right]_s \\ &= [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma) \mathcal{I}_{n,N}] (G \otimes G) (Q_w)_s \\ &+ \{ [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma) \mathcal{I}_{n,N}] (AL \otimes AL) \\ &+ [\Psi \otimes \Psi + I_{p^2 N^2}] \mathcal{I}_{p,N} \} (R_v)_s \end{aligned} \quad (13)$$

in which $\mathcal{I}_{p,N}$ is a permutation matrix to convert the direct sum to a vector, i.e. $\mathcal{I}_{p,N}$ is the $(pN)^2 \times p^2$ matrix of zeros and ones satisfying

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

Ideally, we would like to compute the autocovariance as the expectation of the product $\mathcal{Y}_k \mathcal{Y}_{k+j}^T$. Practically, we approximate the expectation from the data using the time average, a valid procedure since the process is ergodic [12]. The estimate of the autocovariance is computed as

$$\hat{\mathcal{C}}_j = \frac{1}{N_d - j} \sum_{i=1}^{N_d - j} \mathcal{Y}_i \mathcal{Y}_{i+j}^T \quad (14)$$

which is the so-called unbiased autocovariance estimator. The estimated ACM $\hat{\mathcal{R}}(N)$ is analogously defined using the com-

puted $\hat{\mathcal{C}}_j$. At this point we can define a least-squares problem to estimate Q_w, R_v . We summarize (13) as $\mathcal{A}x = b$, in which

$$\begin{aligned} \mathcal{A} &= [D(G \otimes G)]D(AL \otimes AL) + [\Psi \oplus \Psi + I_{p^2 N^2}] \mathcal{I}_{p,N} \\ D &= [(\mathcal{O} \otimes \mathcal{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma) \mathcal{I}_{n,N}] \\ x &= \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} \quad b = \mathcal{R}(N)_s. \end{aligned} \quad (15)$$

We define the ALS estimate as

$$\hat{x} = \left[(\hat{Q}_w)_s^T (\hat{R}_v)_s^T \right]^T = \arg \min_x \left\| \mathcal{A}x - \hat{\mathcal{R}}(N)_s \right\|_2^2 \quad (16)$$

and the solution is the well-known

$$\hat{x} = \mathcal{A}^\dagger b \quad \mathcal{A}^\dagger = (\mathcal{A}^T \mathcal{A})^{-1} \mathcal{A}^T.$$

The uniqueness of the estimate is a standard result of least-squares estimation [14]. The covariances can be found uniquely when matrix \mathcal{A} has full column rank. For this case, it is shown in [1] that the ALS estimates of the covariances are unbiased for all sample sizes, and converge to the true values with increasing sample size.

C. Comments on Previous Work

The basic approach outlined here is similar to that of Mehra [8], Bélanger [9], and Isaksson [15], but there are significant differences. Mehra uses a three-step procedure for estimation that may result in estimates with much higher variance than the proposed method. Mehra provides conditions for unique estimates of the covariances that we show are neither necessary nor sufficient [1], [16]. Bélanger's method is similar to Mehra's in that he parameterizes the unknown covariances element by element. Symmetry of the estimated covariances are not preserved in this formulation, and must be enforced with equality constraints. None of the classic methods can maintain positive semidefiniteness of the estimated covariances. Indefinite covariance estimates is a significant problem that may arise in industrial applications because of model error or small sample size. We use semidefinite programming to add the positive semidefiniteness constraints and maintain a convex program. This convex program can be solved efficiently with Newton's method and the gradient can be determined analytically [1]. This approach is a significant simplification of the strategy to maintain positive semi-definiteness proposed by [17]. None of the earlier approaches have been tested on closed-loop data coming from controllers containing integrated white noise disturbance models. The integrating disturbance models are ubiquitous in industrial applications as a mechanism for ensuring offset-free control in the presence of nonzero disturbances, which we discuss next.

D. Disturbance Models

Model predictive controllers typically use an integrated white noise disturbance model to ensure offset-free control. The augmented system of state plus disturbance is

$$\begin{bmatrix} x \\ d \end{bmatrix}_{k+1} = \begin{bmatrix} A & B_d \\ 0 & I \end{bmatrix} \begin{bmatrix} x \\ d \end{bmatrix}_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \begin{bmatrix} G \\ G_d \end{bmatrix} w_k \quad (17a)$$

$$y_k = [C \quad C_d] \begin{bmatrix} x \\ d \end{bmatrix}_k + v_k \quad (17b)$$

and the dimension of the driving noise w may or may not need to be increased to describe the integrated white noise disturbance. In many MPC applications, the disturbance matrices (B_d, C_d) are assumed to be of a fixed, simple structure. For example, $C_d = I, B_d = 0$ is a common choice (pure output disturbance), as is $C_d = 0, B_d = B$ (pure input disturbance). The noise shaping matrix G_d can either be assumed to be an identity or must be identified from data. If we assume $G_d = I$, it is unlikely that unique estimates of the covariances Q_w, R_v can be found. In the examples in this paper, we use the following disturbance model with the dimension of the driving noise w increased to the most general form

$$\begin{bmatrix} x \\ d \end{bmatrix}_{k+1} = \begin{bmatrix} A & B_d \\ 0 & I \end{bmatrix} \begin{bmatrix} x \\ d \end{bmatrix}_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + w_k \quad (18a)$$

$$y_k = [C \quad C_d] \begin{bmatrix} x \\ d \end{bmatrix}_k + v_k. \quad (18b)$$

This is the second case of the ALS problem, in which there is not enough information to find the covariances uniquely. In this case, there may be enough information to find the optimal filter gain uniquely, so we estimate the optimal L (composed of the state estimate part, L_x , and the disturbance estimate part, L_d) and R_v from data as follows. Consider the first column block of the autocovariance matrix

$$\mathcal{R}_1(N) = \begin{bmatrix} \mathcal{C}_0^T & \mathcal{C}_1^T & \cdots & \mathcal{C}_{N-1}^T \end{bmatrix}^T. \quad (19)$$

From (12), we have

$$\underbrace{\begin{bmatrix} C & I \\ C\bar{A} & -CAL \\ \vdots & \vdots \\ C\bar{A}^N & -C\bar{A}^{N-1}AL \end{bmatrix}}_{\mathcal{A}_1} \begin{bmatrix} PC^T \\ R_v \end{bmatrix} = \mathcal{R}_1(N). \quad (20)$$

Notice that both \mathcal{R}_1 and \mathcal{A}_1 depend on the L used to form the innovations. As in the previous case, we can estimate \mathcal{R}_1 from the data. From (4) the optimal steady-state filter gain also satisfies

$$L = PC^T [CPC^T + R_v]^{-1}. \quad (21)$$

So we can solve the following optimization problem to estimate R_v, PC^T, L :

$$\begin{aligned} \min_{R_v, PC^T, L} & \left\| (I_p \otimes \mathcal{A}_1) \begin{bmatrix} PC^T \\ R_v \end{bmatrix}_s - \hat{\mathcal{R}}_{1s} \right\|_2^2 \\ \text{s.t.} & \quad L = PC^T [CPC^T + R_v]^{-1}. \end{aligned} \quad (22)$$

Because the constraint is nonlinear, an iterative approach is suggested. One can choose an L , process the data and compute \mathcal{A}_1 and $\hat{\mathcal{R}}_{1s}$, and then solve the linear least-squares minimization over only R_v and PC^T . Substituting these values into the constraint then gives a new value for L . This procedure can be repeated until convergence. A variation on this strategy was first proposed by Mehra [7], [8], and a similar approach was presented by Carew and Bélanger [10]. A discussion of the convergence properties of this procedure is outside the scope of this paper. The procedure converged reliably and quickly for both applications considered in this paper. Using this procedure on the augmented system gives the optimal filter gains for both the state, L_x , and disturbance estimates, L_d . Notice that while the

disturbance is modeled as an integrated white noise, it is unlikely that such a disturbance would be present in the plant data since it is an unstable process. Unless the covariance of the white noise is small for the length of time being considered, the inputs or outputs of the system would become excessively large. Therefore, modeling the disturbance as an integrated white noise automatically introduces plant/model mismatch. We show in the two applications that the ALS method works well in spite of this mismatch.

Remark 1.1: We note that the method of Carew and Bélanger would give similar results to those presented in this paper. The main focus of this paper, however, is to demonstrate the regulator benefits that may be realized by including a disturbance model in an adaptive estimation technique, which can be used to account for significant plant/model mismatch

E. Covariance Matching

An initial tuning might be computed using covariance matching, which proceeds as follows. Given a stable state estimator, the covariances of the disturbances are computed from the residuals of the estimator

$$\hat{Q}_w = \text{cov}(\hat{x}_{k+1|k} - A\hat{x}_{k|k-1} - Bu_k - Bd\hat{d}_{k|k-1}) \quad (23a)$$

$$\hat{R}_v = \text{cov}(y_k - C\hat{x}_{k|k-1} - Cd\hat{d}_{k|k-1}). \quad (23b)$$

The covariances of the residuals of the filtering equations give biased estimates of the disturbance covariances. Nevertheless, this method of tuning the estimator represents a careful, data-based approach to finding the unknown covariances, and is compared to the ALS method in the laboratory reactor application.

F. Quantifying the Control System Performance Benefits

In order to compare the control performance using different state estimators, we define the performance measure for the closed-loop system. The controller penalizes the deviation of the outputs from setpoint, r_k , and the differential control action used. The average stage cost is

$$\begin{aligned} \bar{\Phi} = & \underbrace{\frac{1}{N} \sum_{k=0}^{\infty} (y_k - r_k)^T Q (y_k - r_k)}_{\Phi_y} \\ & + \underbrace{\frac{1}{N} \sum_{k=0}^{\infty} (u_k - u_{k-1})^T S (u_k - u_{k-1})}_{\Phi_u}. \end{aligned} \quad (24)$$

We also define an effectiveness factor η , similar to the one used in [18], to measure the closed-loop control benefit

$$\eta = \frac{E[\bar{\Phi}|\text{original estimator tuning}]}{E[\bar{\Phi}|\text{ALS estimator tuning}]}. \quad (25)$$

II. UW LABORATORY REACTOR APPLICATION

To explore the effectiveness of the ALS method in a highly reproducible environment, we constructed a laboratory reactor in which acetic anhydride and water react irreversibly to acetic acid

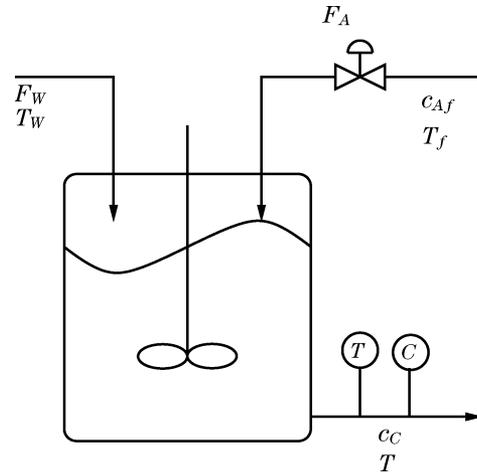


Fig. 1. Schematic of UW laboratory reactor.

A schematic of the reactor is shown in Fig. 1. The reaction consists of several steps with many intermediates [19], [20]. However, our data show that the complicated reaction expressions can be well approximated by a single reaction equation as long as the concentration of water is in excess. The controlled variable is the concentration of acetic acid (c_C). The water flow rate (F_W) is constant and far greater than the flow rate of acetic anhydride (F_A) which is used as input to the system. We measure the temperature in the reactor as well as the conductivity [21] of the solution, which are used to calculate the concentration of acetic acid. The output flow is driven by the pressure difference between top and bottom of the reactor, and we assume that the volume in the reactor is constant. This assumption is valid since the constant flow rate of water into the reactor is much greater than the varying flow rate of acetic anhydride, so the volume effect can be neglected. The concentrations of the acetic anhydride (c_A), and acetic acid (c_C), can be expressed with the following set of nonlinear differential equations:

$$\frac{dc_A}{dt} = \frac{F_A}{V} c_{A,f} - \left(\frac{F_W}{V} + \frac{F_A}{V} \right) c_A - k_0 e^{-\frac{E}{RT}} c_A \quad (27a)$$

$$\frac{dc_C}{dt} = 2k_0 e^{-\frac{E}{RT}} c_A - \left(\frac{F_W}{V} + \frac{F_A}{V} \right) c_C. \quad (27b)$$

We use a linear model predictive controller to control the acid concentration, so the nonlinear model is linearized around the steady states c_{As} , c_{Cs} and F_{As} to obtain a continuous-time state-space model in which

$$\begin{aligned} A = & \begin{bmatrix} -\left(\frac{F_W}{V} + \frac{F_{As}}{V} + k_0 e^{-\frac{E}{RT}} \right) & 0 \\ 2k_0 e^{-\frac{E}{RT}} & -\left(\frac{F_W}{V} + \frac{F_{As}}{V} \right) \end{bmatrix} \\ B = & \begin{bmatrix} \frac{c_{A,f}}{V} - \frac{c_{As}}{V} \\ -\frac{c_{Cs}}{V} \end{bmatrix} \quad C = [0 \quad 1] \end{aligned}$$

and then discretized with a sample time of 1 s. A number of nonisothermal experiments were performed in order to estimate the model parameters. A full discussion of these studies is available elsewhere [22]; the best estimated parameter values are summarized in Table I.

TABLE I
NUMERICAL VALUES FOR PARAMETERS

Parameter	Value	Unit
$c_{A,f}$	10.58	mol
$c_{A,s}$	0.034236	mol
$c_{C,s}$	0.04	mol
E	4.544×10^4	J/mol
R	8.314	J/mol/K
k_0	2.6×10^5	1/s
V	0.45	l
T	296	K
F_W	1.7611×10^{-3}	l/s
$F_{A,s}$	9.0745×10^{-6}	l/s

A. Robustness to Incorrect Disturbance Model

Since we have complete operational and environmental control over the laboratory-scale reactor, we are able to identify a highly-accurate physical process model. Considerable effort went into estimating the parameters of this physical model, and the temperature operating region was bounded to limit the effects of linearizing the model. Obtaining this accurate model allows us to distinguish the effects of two different kinds of modeling errors: the disturbance model (disturbance location, covariances) and the process model (u to y model). We conjecture that using the estimator to diagnose and compensate for modeling errors can have large benefits in terms of closed-loop control performance. The following two experiments support this conjecture. First we examine the effect of having an incorrect disturbance model. In the model we model the disturbance enters as an integrated white noise at the *measured output*, the acetic acid concentration. That is the usual disturbance model choice in industrial MPC applications. But in the experiments we inject into the process a series of step disturbances at the *process input*, the flow rate of acetic anhydride. We use these data in the ALS method to estimate a filter to compensate for this *incorrectly modeled disturbance*. We also use Bélanger's method to estimate a filter gain. The covariance matching, ALS, and Carew-Bélanger (CB) filter gains are calculated to be

$$L_{\text{nominal}} = \begin{bmatrix} 0.0021987 \\ 0.0049199 \\ 0.9853614 \end{bmatrix} \quad L_{\text{ALS}} = \begin{bmatrix} -0.57857 \\ -1.79762 \\ 1.86749 \end{bmatrix}$$

$$L_{\text{CB}} = \begin{bmatrix} 0.33247 \\ -2.17841 \\ 2.22646 \end{bmatrix}$$

and the eigenvalues of the estimators are

$$\lambda(A - AL_{\text{ALS}}C) = \begin{bmatrix} 0.98971 + 0.00000i \\ 0.96293 + 0.05524i \\ 0.96293 - 0.05524i \end{bmatrix}$$

$$\lambda(A - AL_{\text{CB}}C) = \begin{bmatrix} 0.99473 + 0.00000i \\ 0.96851 + 0.09651i \\ 0.96851 - 0.09651i \end{bmatrix}.$$

The eigenvalues of the ALS and CB estimators are close. In this example, the CB method required 3000 data points to achieve convergence, whereas the ALS method required only 2000. The covariances of the prediction errors for the two methods are

$$\text{cov}[\mathcal{Z}_k]_{\text{ALS}} = 8.359 \times 10^{-8} \quad \text{cov}[\mathcal{Z}_k]_{\text{CB}} = 8.639 \times 10^{-8}.$$

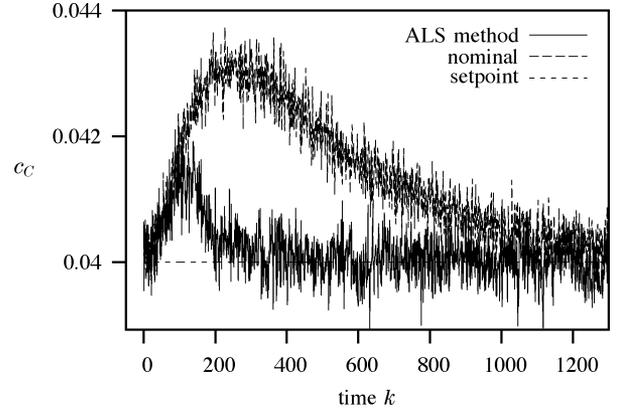


Fig. 2. Rejection of a nonzero deterministic disturbance. Setpoint tracking of the ALS method and a nominal controller. Acetic acid concentration versus time.

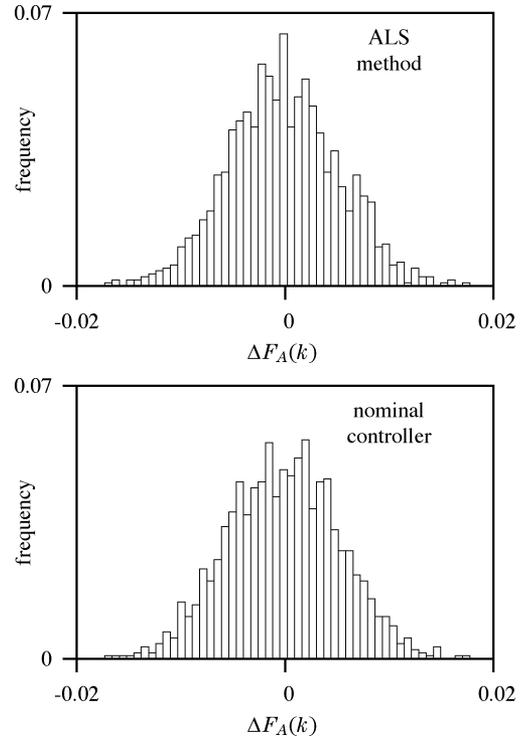


Fig. 3. Rejection of a nonzero deterministic disturbance. Frequency distribution of the change in the manipulated input, the acetic anhydride flow rate, F_A , for the ALS and a nominal controller.

We do not claim that the ALS method is better than the Carew-Bélanger method for calculating the filter gains for all cases, but ALS is better in this example.

Next, we close the loop with the estimated filter gain corresponding to this combination of (incorrect) disturbance model and ALS filter. We compare this controller's performance to one that uses the same disturbance model and covariances estimated by covariance matching. A comparison of the output performance is shown in Fig. 2. Using the ALS estimator, the regulator is able to reject the disturbance four time faster than using the nominal estimator. The frequency distributions for the flow rate changes ($F_A(k) - F_A(k-1)$) are shown in Fig. 3. We see they are about the same, and the improved output performance does *not* require a more active manipulated input. In fact, the ALS case gives a slightly smaller variance in the manipulated

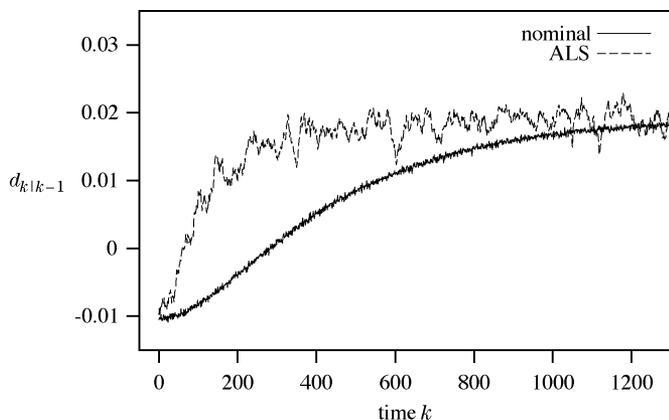


Fig. 4. Rejection of a nonzero deterministic disturbance. Estimation of the disturbance, ALS estimator versus nominal controller.

TABLE II
OBJECTIVE FUNCTION COSTS FOR INCORRECT DISTURBANCE
MODEL EXPERIMENTS

	Covariance Matching	Autocovariance Least-Squares	η
Φ_y	2.966938	0.40781	7.3
Φ_u	0.291027	0.15653	1.9
Φ	3.257964	0.56434	5.8

TABLE III
REPLICATES OF OBJECTIVE FUNCTION COSTS FOR INCORRECT DISTURBANCE
MODEL EXPERIMENTS

	ALS Replicate 1	ALS Replicate 2	ALS Replicate 3
Φ_y	0.23935	0.23493	0.23669
Φ_u	0.28427	0.26355	0.2568
Φ	0.52362	0.49848	0.49352

input. The estimated disturbances for the ALS and covariance matching methods are shown in Fig. 4. The ALS method is able to adapt quickly to the disturbance. The average objection function costs and efficiency factors are summarized in Table II. The fast rejection of the disturbance using the ALS estimator is reflected in the total objective function costs, which is six times better than the nominal estimator. The ALS results are replicated in Table III. Notice the experiment is highly reproducible.

B. Robustness to Process Model Error

In the second experiment, we introduce significant error in the model between the manipulated input and the measured output. The volume of the reactor is assumed to be in error and we choose a value for the model that is smaller than the reactor's actual value. Therefore, the linearization of the physical model is performed at an incorrect steady state. The concentration is depicted for a setpoint change in Fig. 5. The nominal controller (based on the covariance matching filter) displays oscillatory behavior as the regulator tries to drive the output to the setpoint using this incorrect dynamic model. Since the volume of the reactor is underestimated in the model, the controller contains a smaller reactor time constant, and the controller is too aggressive. The output using the ALS estimator is shown as well. The

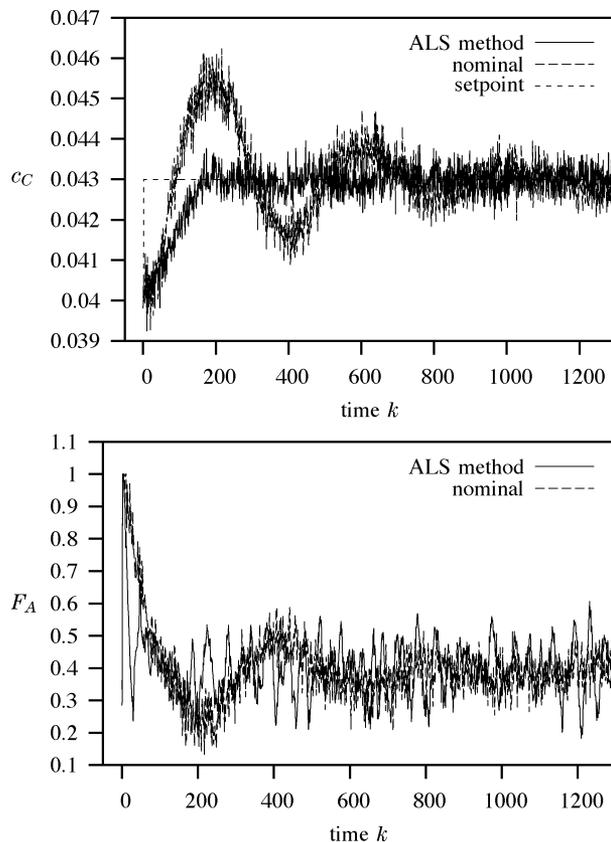


Fig. 5. Robustness of ALS method to model error. Top: tracking error of acetic acid concentration versus time. Bottom: manipulated input acetic anhydride flow rate versus time.

TABLE IV
OBJECTIVE FUNCTION COSTS FOR PROCESS MODEL ERROR EXPERIMENTS

	Covariance Matching	Autocovariance Least-Squares	η
Φ_y	1.13182	0.47902	2.3
Φ_u	36.83336	12.49991	3.0
Φ	37.96518	12.97893	2.9

control cost is three times better using the ALS estimator compared to the nominal case, and is summarized in Table IV. Replicates of the ALS results are summarized in Table V, and the second experiment also is shown to be highly reproducible. We have therefore demonstrated that estimating the filter gain from the data allows the controller to compensate better for incorrect input/output models.

III. INDUSTRIAL APPLICATION

As we move away from the highly controlled laboratory environment to an industrial facility, we would like to test two critical assumptions we have made in deriving the ALS method. The first is the assumed time invariance of the statistics of the disturbances. The second is that improved estimation leads to improved closed-loop control. We have worked with Eastman Chemical Company to apply this method to data from a gas-phase reactor, represented in Fig. 6. As shown, the composition in the reactor is the controlled variable, and the feed fraction to the reactor is the manipulated variable. The temperature drop across the reactor is also measured. While the

TABLE V
REPLICATES OF OBJECTIVE FUNCTION COSTS FOR PROCESS MODEL
ERROR EXPERIMENTS

	ALS Replicate 1	ALS Replicate 2	ALS Replicate 3
Φ_y	0.49004	0.46385	0.49291
Φ_u	11.941	10.435	11.514
Φ	12.431	10.899	12.007

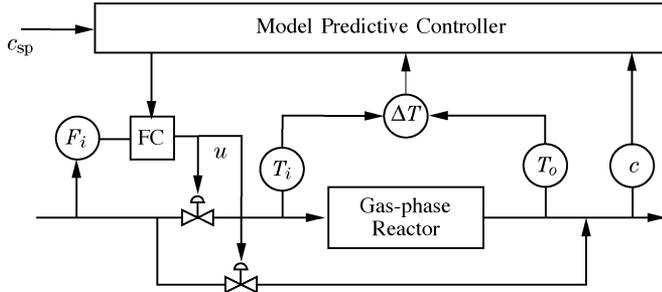


Fig. 6. Eastman gas-phase reactor control system.

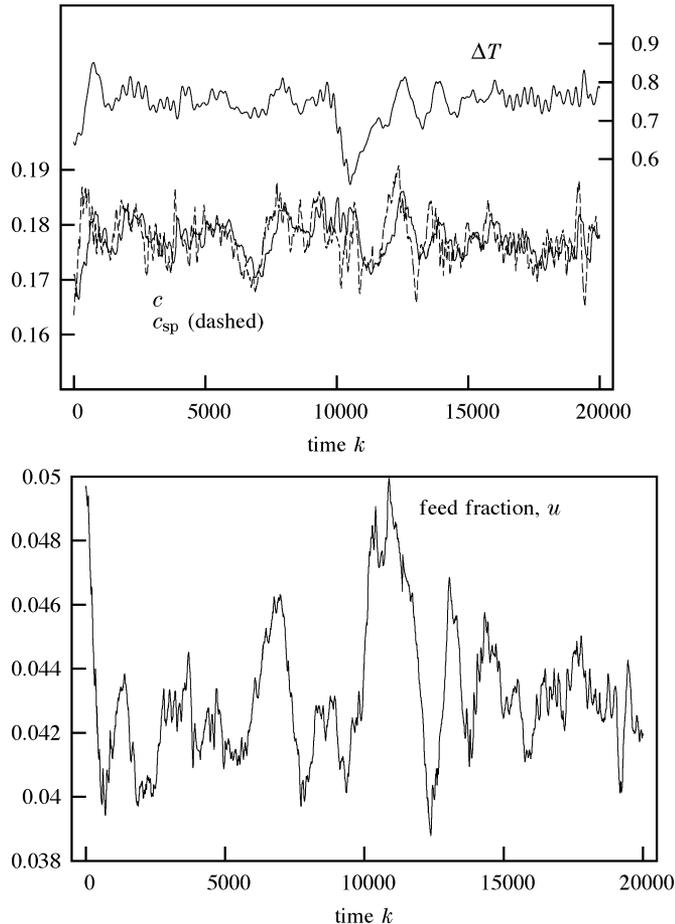


Fig. 7. Normalized process data. Top: effluent composition, composition setpoint, and temperature drop versus time. Bottom: manipulated variable, feed fraction, versus time.

temperature drop does not have a setpoint, it does have upper and lower constraints. The setpoint for the composition comes from a downstream process and is time-varying. All of the data presented have been scaled to remove information about the operating conditions of the reactor. There are n states in

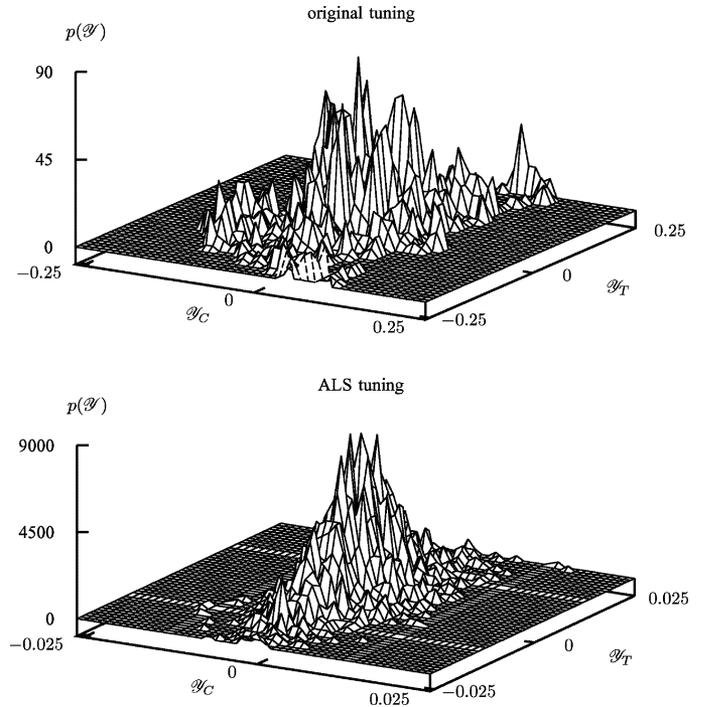


Fig. 8. Frequency distribution of the innovations using the original filter gain (top) and the ALS method's filter gain (bottom). Notice the ALS estimator has decreased the variance of the innovations by a factor of 100.

the system, including a large number, n_d , of time-delay states. The data set provided spans several days of operating data, and is shown in Fig. 7. The first 5000 points are used as the training set for the ALS approach, and the next 5000 are used as a validation set. The ALS method is used to estimate a filter based on a pure output disturbance model ($B_d = 0, C_d = I$). Recall that the covariance estimation techniques are based on the properties of the process innovations. The covariance of the innovations (prediction errors) in the validation set using the original filter and ALS filter are

$$\text{cov}[\mathcal{Y}_k] = \begin{bmatrix} 3.59 \times 10^{-3} & 1.67 \times 10^{-3} \\ 1.67 \times 10^{-3} & 8.18 \times 10^{-3} \end{bmatrix} \quad \text{original filter}$$

$$\text{cov}[\mathcal{Y}_k] = \begin{bmatrix} 1.00 \times 10^{-5} & 3.14 \times 10^{-6} \\ 3.14 \times 10^{-6} & 7.25 \times 10^{-5} \end{bmatrix} \quad \text{ALS filter.}$$

The frequency distributions of the innovations for both the original and ALS filters are shown in Fig. 8. The variances of the measurement innovation errors have been reduced by two orders of magnitude. Fig. 8 indicates that implementing ALS in industry has *large potential for improving the quality of the state estimates*.

A. Time Invariance of the Disturbance Statistics

To test the time-invariance of the disturbance statistics, the overall data set is divided into eleven subsets as shown in Fig. 9, and we use each of these subsets to estimate eleven filter gains L . Two eigenvalues of $(A - ALC)$, i.e., the estimator's poles, are shown in Fig. 10 for these eleven data subsets. With the exception of subsets four and five, the results are consistent. Referring back to Fig. 9, subsets four and five contain a large temperature excursion. The temperature innovations are shown in Fig. 11. We see two significant outliers in the innovations around sample

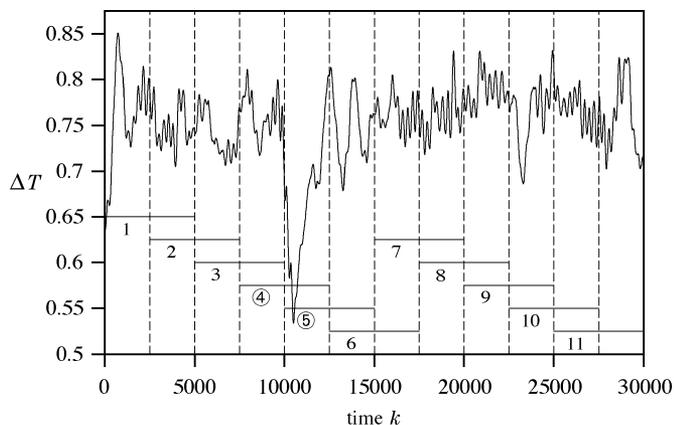


Fig. 9. Reactor temperature drop versus time. Eleven data sets used to test the time-invariance of the disturbance statistics. Large deterministic disturbance present in overlapping portion of data sets ④ and ⑤.

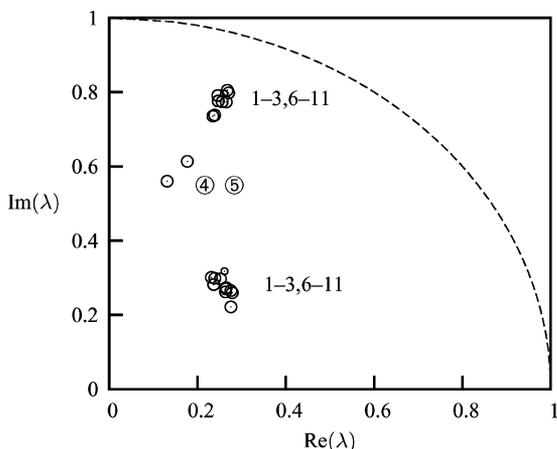


Fig. 10. Eigenvalues of $A - ALC$ for the eleven data sets. The disturbance in data sets ④ and ⑤ cause inconsistent estimates of filter gain L .

time 11 391. Next we display the *nonzero mean* estimated disturbance \hat{d}_T around the time of the large excursion. Notice that the disturbance estimate adjusts at sample time 11 391 when the disturbance enters the process. Once the estimator accounts for the unmodeled deterministic disturbance in the system, the innovations return to normal. Also if we delete the two outliers in Fig. 11, then the eigenvalues shown in Fig. 10 at ④ and ⑤ join the other groups. We conclude that the industrial data displays disturbance statistics that are remarkably time *invariant* with occasional outliers. Several routine statistical methods are available to remove such outliers so they present no special problems to applying the ALS method. If the character of the disturbances does change *slowly* with time, the ALS disturbance estimation procedure can be run continuously to monitor the situation and periodically adjust to the new conditions.

B. Control System Performance Payoff

We next demonstrate the benefits of adaptive estimation in terms of closed-loop control performance. The challenge of using industrial data is that the true plant is not available for simulation. In order to accurately quantify the regulator payoff,

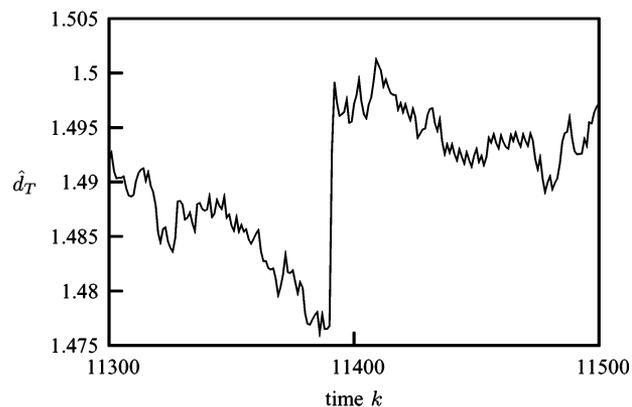
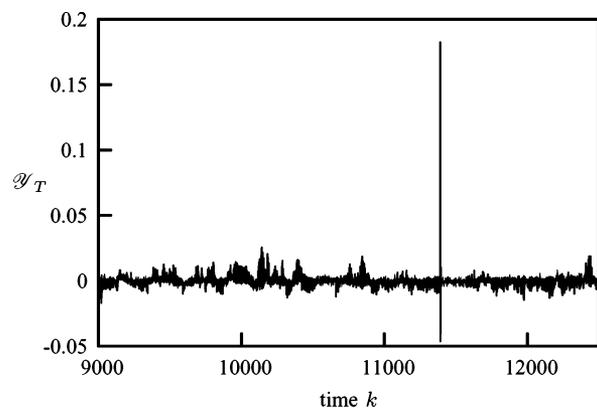


Fig. 11. Top: temperature innovations versus time. The large innovation is caused by an unmodeled disturbance occurring at sample time 11 391. Bottom: estimated temperature disturbance versus time.

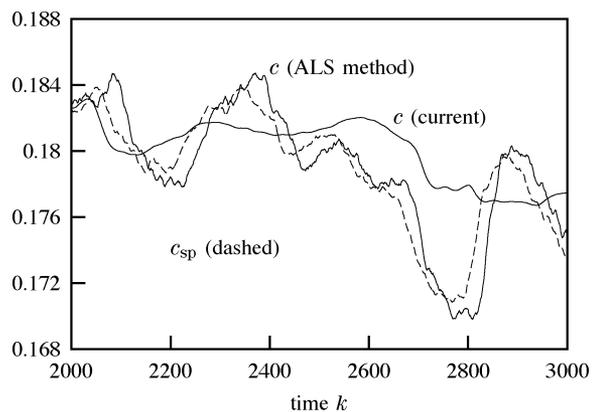


Fig. 12. ALS method tracks composition setpoint (dashed line) more closely than the current industrial practice.

we first attempt to simulate the system with an additional disturbance

$$x_{k+1}^m = \hat{A}x_k^m + \hat{B}u_k^m \quad (28a)$$

$$y_k^m = \hat{C}x_k^m + b_k. \quad (28b)$$

We have used an output disturbance model to account for the differences between (A, B, C) and $(\hat{A}, \hat{B}, \hat{C})$, as well as any unmodeled disturbances in the system that cause differences between the plant outputs and the predicted outputs y_k^m . Put another way, we use b_k to make the prediction errors of the simulated data look identical to the process data using the original

estimator. In Fig. 12, a snapshot of the composition tracking is shown, which has significantly improved. The expected average objective function cost using the original filter is 8.2 and the average cost using the ALS filter is 2.9. The regulator performance using the updated estimator is nearly three times better than the performance using the original estimator.

IV. CONCLUSION

We have demonstrated that the autocovariance least-squares method is effective in diagnosing and correcting the estimates of the disturbance statistics in a plant using data from an industrial chemical reactor as well as a laboratory reactor. When plant/model mismatch is present in the laboratory process, the ALS method can account for some of this mismatch, which results in control performance three to five times better than a typically tuned state estimator. The mismatch can occur when the input–output connection has been mismodeled, or when an inappropriate disturbance model is chosen. In the industrial reactor, we have shown that the innovations variance can be reduced by two orders of magnitude using the ALS method. The estimated controller performance is three times better using the ALS method compared to the estimator currently in use.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. J. Downs and Dr. S. M. Miller of Eastman Chemical Company for their assistance with the data, and M. Rajamani for helpful discussion. All simulations were performed using Octave (<http://www.octave.org>). Octave is freely distributed under the terms of the GNU General Public License.

REFERENCES

- [1] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," *Automatica*, vol. 42, no. 2, pp. 303–308, Feb. 2005.
- [2] D. Alspach, "A parallel filtering algorithm for linear systems with unknown time-varying noise statistics," *IEEE Trans. Autom. Control*, vol. AC-19, no. 5, pp. 552–556, Oct. 1974.
- [3] C. Hilborn and D. Lainiotis, "Optimal estimation in the presence of unknown parameters," *IEEE Trans. Syst., Sci., Cybern.*, vol. SSC-5, no. 1, pp. 38–43, Jan. 1969.
- [4] T. Bohlin, "Four cases of identification of changing systems," in *System Identification: Advances and Case Studies*, 1st ed, R. Mehra and D. Lainiotis, Eds. New York: Academic, 1976.
- [5] R. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Trans. Autom. Control*, vol. AC-15, no. 1, pp. 25–34, Feb. 1970.
- [6] K. Myers and B. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Trans. Autom. Control*, vol. AC-21, no. 4, pp. 520–523, Aug. 1976.
- [7] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. AC-15, no. 12, pp. 175–184, Dec. 1970.
- [8] —, "Approaches to adaptive filtering," *IEEE Trans. Autom. Control*, vol. AC-17, no. 5, pp. 903–908, Oct. 1972.
- [9] P. Bélanger, "Estimation of noise covariance matrices for a linear time-varying stochastic process," *Automatica*, vol. 10, pp. 267–275, 1974.
- [10] B. Carew and P. Bélanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariances," *IEEE Trans. Autom. Control*, vol. AC-18, no. 6, pp. 582–587, Dec. 1973.
- [11] B. J. Odelson, "Estimating disturbance covariances from data for improved control performance," Ph.D. dissertation, Univ. Wisconsin-Madison, Madison, WI, 2003.
- [12] G. Jenkins and D. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.
- [13] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.
- [14] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Philadelphia, PA: SIAM, 1995.
- [15] A. Isaksson, "Identification of time varying systems through adaptive Kalman filtering," in *Proc. 10th IFAC World Congr.*, 1987, pp. 305–310.
- [16] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," TWMCC, Dept. Chem. Eng., Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 2003-04, Sep. 2003.
- [17] G. Noriega and S. Pasupathy, "Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1146–1159, Sept. 1997.
- [18] M. L. Tyler and M. Morari, "Estimation of cross directional properties: scanning versus stationary sensors," *AIChE J.*, vol. 41, no. 4, pp. 846–854, Apr. 1995.
- [19] F. A. Carey, *Organic Chemistry*. New York: McGraw-Hill, 1996.
- [20] J. J. Shatynski and D. Hanesian, "Adiabatic kinetic studies of the cytidine/acetic anhydride reaction by utilizing temperature versus time data," *Ind. Eng. Chem. Res.*, vol. 32, pp. 594–599, 1993.
- [21] S. P. Asprey, B. W. Wojciechowski, N. M. Rice, and A. Dorcas, "Applications of temperature scanning in kinetic investigations: The hydrolysis of acetic anhydride," *Chem. Eng. Sci.*, vol. 51, no. 20, pp. 4681–4692, 1996.
- [22] B. J. Odelson, A. Lutz, and J. B. Rawlings, "Application of autocovariance least-squares methods to laboratory data," TWMCC, Dept. Chem. Eng., Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 2003-03, Sep. 2003.