2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

Maximum Likelihood Estimation of Linear Disturbance Models for Offset-free Model Predictive Control*

Steven J. Kuntz[1] and James B. Rawlings[1]

Abstract— **The performance of industrially successful model predictive control (MPC) and offset-free MPC is reliant on identifying an adequate linear state-space model using plant data. While the models for MPC can be identified using one of many subspace identification methods, there are no methods for identifying the linear disturbance models used in offset-free MPC. Here we formulate a series of maximum likelihood estimation (MLE) problems for identifying linear disturbance models. To formulate the first problem, the state is estimated as a linear combination of past inputs and outputs, and the state-space model is then written as a linear estimation problem. The second problem is formulated as a linear estimation problem relating the long-range prediction error sequence to the disturbance and noise sequences. The last problem is simply a covariance estimation problem for the noises in the linear disturbance model. Each MLE problem has a closed-form solution. While size of the second MLE problem makes it computationally demanding, it can be simplified considerably in the case where the system has no integrators. Hardware experiments (TCLab, an Arduino-based heat transport laboratory) demonstrate that the proposed method generates offset-free performance under realistic conditions on systems without integrators. Numerical simulation experiments demonstrate that the results also generalize to systems with integrators.**

I. INTRODUCTION

Model predictive control (MPC) is the most successful advanced control method in the chemical process industries [1]. MPC is an advanced feedback control technique in which an optimal control problem is solved on-line [2]. Since MPC is formulated as an optimization problem, it can handle physical and safety constraints and optimize economic objectives, which are key requirements for operating a safe and profitable chemical plant. In offset-free MPC the plant model is modified to include additional states that account for disturbances and model error, effectively adding integral control to the MPC algorithm. Offset-free MPC can achieve offset-free tracking of setpoints even under significant plant-model mismatch, which is crucial for profitability in the modern chemical industry. The performance of both MPC and offset-free MPC relies on identifying an adequate model using data from the plant. For MPC, the model can be identified using subspace identification (SID) methods [3]. Prior to this work, there were no existing methods for identifying the models used in offset-free MPC.

Subspace methods were developed as an extension of Ho and Kalman's seminal paper on realization theory [4], [5]. All of these methods formulate an extended state-space model and approximate low-rank projections of vectors of future data on vectors of past data [3]. In canonical variate analysis (CVA) methods, the state is estimated from the low-rank projection and the model parameters follow by linear estimation [6]. In N4SID and MOESP, the model parameters are extracted from the projections using matrix factorizations [7], [8]. While the CVA methods estimate the complete process and measurement noise covariances, the other methods (N4SID and MOESP) produce only the Kalman filter gain. All three of these methods fall under a unifying theorem proposed in [9]. These methods cannot enforce the special structure found in linear disturbance models and are thus insufficient for offset-free MPC.

Some researchers have proposed using auto-regressive integrating moving average (ARIMA) models to allow MPC to reject disturbances [10], [11], [12]. The integrating terms allow for correction to plant-model mismatch. However these proposals focus on either dynamic matrix control applications or SISO systems, which excludes the vast majority of control systems that are relevant to the chemical process industries. Others have proposed a semi-infinite program approach to select the disturbance model for offset-free nonlinear MPC [13]. In this approach, the disturbance model which maximizes the size of the set of observable steady-states is selected. Such an approach only makes sense in the case of offset-free nonlinear MPC, however, as every steady-state is observable for linear systems. Moreover, the selected disturbance model is not necessarily an accurate model of the disturbances and may produce a suboptimal estimator.

We propose a first-of-its-kind method for identifying the disturbance models used in offset-free MPC. This method is formulated as a series of three maximum likelihood estimation (MLE) problems. In the first, the state is estimated as a linear combination of a finite number of past inputs and outputs, and the state-space model is written as a linear estimation problem. In the second, the long-range prediction error sequence is related to the disturbance and noise sequences through a linear model. The last problem is a covariance estimation problem for the noises in the disturbance model. Closed form solutions for these problems are provided. While the second MLE problem is shown to be computationally demanding, it is nonetheless considerably cheaper when the system itself has no integrators. Finally, the method's application to offset-free MPC for systems without integrators is validated with hardware experiments on the TCLab, an Arduino-based heat transport laboratory [14]. The generalization to systems with integrators is demonstrated through numerical simulations of a tank draining system.

*This work was sponsored by the industrial members of the Texas Wisconsin California Control Consortium (TWCCC).

[1]University of California–Santa Barbara, Department of Chemical Engineering (skuntz@ucsb.edu, jbraw@ucsb.edu)

978-1-6654-5197-0/$31.00 ©2022 AACC

II. SYSTEMS OF INTEREST

We refer to the general stochastic linear system as the *standard model*,

$$x^+ = Ax + Bu + w \tag{1a}$$

$$y = Cx + Du + v \tag{1b}$$

$$\begin{bmatrix} w \\ v \end{bmatrix} \sim N(0, S) \tag{1c}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^{n_u}$ is the input, $y \in \mathbb{R}^{n_y}$ is the measured output, and $w \in \mathbb{R}^n$, $v \in \mathbb{R}^{n_y}$ are the process and measurement noises, respectively. We assume (w, v) is uncorrelated in time. Assuming the standard model is observable, there exists a Kalman filter gain K such that the plant equations (1) can be rewritten in so-called *predictor form*,

$$\hat{x}^+ = A_K \hat{x} + B_K \begin{bmatrix} u \\ y \end{bmatrix} \tag{2a}$$

$$y = C\hat{x} + Du + e_y \tag{2b}$$

where $A_K = A - KC$, $B_K = [B - KD, \ K]$, and the innovations $e_y := y - C\hat{x} - Du \sim N(0, R_e)$ are uncorrelated in time. The predictor form is particularly important for both subspace identification methods and our proposed method. We make the following assumptions about the system (1).

Assumption 1: The system (1) is minimal, i.e., (A, C) is observable and $(A, [B, \ K])$ is controllable. The estimator is stable, i.e. the eigenvalues of A_K lie within the unit circle.

We refer to the following special case of the stochastic linear system as the *disturbance model*,

$$x^+ = Ax + Bu + B_d d + w \tag{3a}$$

$$d^+ = d + w_d \tag{3b}$$

$$y = Cx + Du + C_d d + v \tag{3c}$$

$$\begin{bmatrix} w \\ w_d \\ v \end{bmatrix} \sim N(0, S_d) \tag{3d}$$

where $d \in \mathbb{R}^{n_y}$ is the integrating disturbance state and $w_d \in \mathbb{R}^{n_d}$ is the driving noise for the disturbances. Again, we assume (w, w_d, v) is uncorrelated in time.

Instead of directly identifying the disturbance model (3), we will first identify the standard model (1), augment that system with parameters $B_d \in \mathbb{R}^{n \times n_d}$ and $C_d \in \mathbb{R}^{n_y \times n_d}$, and finally estimate the augmented noise covariance S_d of (3). Two of the sufficient conditions for offset-free MPC to work are that $n_d = n_y$ and the following rank condition is satisfied [15], [16]:

$$\operatorname{rank} \begin{bmatrix} A - I & B_d \\ C & C_d \end{bmatrix} = n + n_d \tag{4}$$

It is worth noting that for each (A, B, C, D) *all* disturbance models (B_d, C_d) that satisfy (4) are equivalent up to a similarity transformation [17], making it unnecessary to estimate the parameters (B_d, C_d) from data. In the special case where A contains no integrators and $n_d = n_y$, the so-called *output disturbance model* $(B_d, C_d) = (0, I_{n_y})$ satisfies the rank condition (4). More generally, if A contains no integrators and $n_d = n_y$, then (4) is satisifed by $B_d = 0$ and any C_d invertible. This special case is shown to be a numerically advantageous choice in our algorithm.

III. DISTURBANCE MODEL IDENTIFICATION

The goal of disturbance model identification is to find parameters $(\hat{A}, \hat{B}, \hat{C}, \hat{D}, B_d, C_d, \hat{S}_d)$ that capture the behavior of available input/output data $\{u(0), y(0), \ldots, u(N), y(N)\}$. In this section, we identify the parameters with three consecutive MLE problems.

A. Maximum likelihood estimation of the standard model

It is well known from subspace identification methods that the state can be well-approximated by a linear combination of sufficient past information (Thm. 1 in Appendix),

$$x(k) \approx \mathcal{C}_p Z(k)$$

where $Z(k) := [u(k-1)', \ y(k-1)', \ \cdots, \ u(k-p)', \ y(k-p)']'$ and $\mathcal{C}_p := [B_K, \ A_K B_K, \ \cdots, \ A_K^{p-1} B_K]$ is full rank. Given the SVD of the data matrix,

$$H = \begin{bmatrix} Z(p) & \cdots & Z(N) \end{bmatrix} \approx U_1 S_1 V_1' \tag{5}$$

we can define the estimated state as

$$\hat{x}(k) = U_1' Z(k)$$

which we show in the Appendix (Thm. 2) is an approximation of the state up to a similarity transformation,

$$x(k) \approx T\hat{x}(k)$$

for some nonsingular $T \in \mathbb{R}^{n \times n}$. Using the estimated state, one can stack the equations (1) to write a simple linear estimation problem with i.i.d. Gaussian noise

$$s(k) = \Theta t(k) + e(k), \quad e(k) \overset{iid}{\sim} N(0, S) \tag{6}$$

for $k = p, \ldots, N-1$, where

$$s = \begin{bmatrix} \hat{x}^+ \\ y \end{bmatrix}, \quad t = \begin{bmatrix} \hat{x} \\ u \end{bmatrix}, \quad e = \begin{bmatrix} w \\ v \end{bmatrix}, \quad \Theta = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

The MLE solution for this model is then

$$\hat{\Theta} = \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} = \left(\sum_{k=p}^{N-1} s(k) t(k)' \right) \left(\sum_{k=p}^{N-1} t(k) t(k)' \right)^{-1} \tag{7a}$$

$$\hat{S} = \frac{1}{N_s} \sum_{k=p}^{N-1} (s(k) - \hat{\Theta} t(k))(s(k) - \hat{\Theta} t(k))' \tag{7b}$$

where $N_s = N - p$ [18, Thm. 8.2.1], [19, pp. 404-411].

It is worth noting that the solution derived here is similar to the Larimore-type subspace method [3], [6], [20]. In particular, Larimore's CVA algorithm treats states as canonical variables of the past data vectors $Z(k)$ projected onto the space of future data $Y(k) = [y(k)', \ \cdots, \ y(k+f-1)']'$ and derives the parameters based on these approximated states. In fact, any subspace method which estimates parameters (A, B, C, D, S) can be substituted for this step. The method above was chosen for its simplicity.

B. Maximum likelihood estimation of the disturbances

To estimate the disturbance sequence, we treat the estimated state $\hat{x}$ and parameters $(\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{S})$ as the true parameters in the model (3) and a the disturbance sequence that corrects for the model error. Let (B_d, C_d) be any matrices that satisfy the offset-free rank condition (4). The *true* long-range output is

$$y(k) = \hat{C}\hat{A}^{k-p}\hat{x}(p) + \sum_{j=p}^{k-1} \hat{C}\hat{A}^{k-j-1}\hat{B}u(j) + \hat{D}u(k) + \sum_{j=p}^{k-1} \hat{C}\hat{A}^{k-j-1}(B_d d(j) + w(j)) + C_d d(k) + v(k)$$

and the *predicted* long-range output is

$$\hat{y}(k) = \hat{C}\hat{A}^{k-p}\hat{x}(p) + \sum_{j=p}^{k-1} \hat{C}\hat{A}^{k-j-1}\hat{B}u(j) + \hat{D}u(k)$$

for each $k = p, \ldots, N-1$. Next, we define the long-range prediction error as $z(k) := y(k) - \hat{y}(k)$ which gives

$$z(k) = \sum_{j=p}^{k-1} \hat{C}\hat{A}^{k-j-1}(B_d d(j) + w(j)) + C_d d(k) + v(k)$$

Rewriting this as a linear model,

$$\mathbf{z} - \mathbb{A}\mathbf{d} = \mathbb{B}\mathbf{e} \sim N(0, \mathbb{V}) \tag{8}$$

where $\mathbf{z} := [z(p)', \ \ldots, \ z(N-1)']'$ is the sequence of long-range prediction errors, $\mathbf{d} := [d(p)', \ \ldots, \ d(N-1)']'$ is the sequence of disturbances, $\mathbf{e} := [e(p)', \ \ldots, \ e(N-1)']'$ is the noise sequence, and

$$\mathbb{A} := \begin{bmatrix} C_d & & & \\ \hat{C}B_d & C_d & & \\ \vdots & \ddots & \ddots & \\ \hat{C}\hat{A}^{N-p-2}B_d & \ldots & \hat{C}B_d & C_d \end{bmatrix},$$

$$\mathbb{B} := \begin{bmatrix} B_0 & & & \\ B_1 & B_0 & & \\ \vdots & \ddots & \ddots & \\ B_{N-p-1} & \ldots & B_1 & B_0 \end{bmatrix},$$

$$B_0 := \begin{bmatrix} 0 & I_{n_y} \end{bmatrix},$$
$$B_j := \begin{bmatrix} \hat{C}\hat{A}^{j-1} & 0 \end{bmatrix} \quad \forall j \geq 1,$$
$$\mathbb{V} := \mathbb{B}(I_{N_s} \otimes \hat{S})\mathbb{B}'$$

The model (8) has a MLE solution due to [21] and [22, p. 313]:

$$\hat{\mathbf{d}} = (\mathbb{A}'\mathbb{V}_0^\dagger\mathbb{A})^\dagger\mathbb{A}'\mathbb{V}_0^\dagger\mathbf{z} \tag{9}$$

where $\mathbb{V}_0 := \mathbb{V} + \mathbb{A}\mathbb{A}'$. This is an $O(N^3)$ computation with $O(N^2)$ memory requirements. Notice that when $B_d = 0$ and C_d is invertible, $\mathbb{A} = I_N \otimes C_d$ and $\mathbb{V}_0$ are also invertible and

$$(\mathbb{A}'\mathbb{V}_0^\dagger\mathbb{A})^\dagger\mathbb{A}'\mathbb{V}_0^\dagger = \mathbb{A}^{-1} = I_N \otimes C_d^{-1}$$

Therefore (9) is equivalently written

$$\hat{d}(k) = C_d^{-1}z(k) \tag{10}$$

which is an $O(N)$ computation without additional memory requirements. It is clear that whenever the system is free of integrators, the simplified solution (10) should be used. A similarity transformation can be used to find the desired disturbance model after the output disturbance model is found [17].

Given the estimated states and disturbances, one can stack the equations (3) to write a simple covariance estimation problem of i.i.d. Gaussian noise

$$\tilde{e}(k) = \tilde{s}(k) - \tilde{\Theta}\tilde{t}(k) \overset{iid}{\sim} N(0, S_d) \tag{11}$$

for $k = p, \ldots, N-1$, where

$$\tilde{s} = \begin{bmatrix} \hat{x}^+ \\ \hat{d}^+ \\ y \end{bmatrix}, \quad \tilde{t} = \begin{bmatrix} \hat{x} \\ \hat{d} \\ u \end{bmatrix}, \quad \tilde{\Theta} = \begin{bmatrix} \hat{A} & B_d & \hat{B} \\ 0 & I & 0 \\ \hat{C} & C_d & \hat{D} \end{bmatrix} \tag{12}$$

The MLE solution for this model is then

$$\hat{S}_d = \frac{1}{N_s}\sum_{k=p}^{N-1} \tilde{e}(k)\tilde{e}(k)' \tag{13}$$

where $N_s = N - p$ [18, Thm. 8.2.1], [19, pp. 404-411]. Thus, we have found the complete set of parameters for the model (3) which concludes our description of the algorithm.

IV. TCLAB EXPERIMENTS

The TCLab hardware (Fig. 1) is a simple two-input, two-output heat transport laboratory system on which MIMO controllers can be quickly implemented [14]. Moreover, it is a prototypical system of the form (3) with the state $x = [T_{S,1}, \ T_{S,2}, \ T_{H,1}, \ T_{H,2}]'$ as the temperatures near the sensors and heaters, the disturbance $d = [T_{a,1}, \ T_{a,2}]'$ as environmental temperatures, the input $u = [\dot{Q}_1, \ \dot{Q}_2]'$ as the heater duties, and the measured output $y = [T_1, \ T_2]'$ as the temperatures measured by the sensors. The method described herein is evaluated via experiments on this hardware.

A. Standard model

To gather data on the TCLab hardware, we first warmed up the hardware by running both heaters at 30% voltage for 7 minutes, then manipulated the input with a PRBS signal (119 minutes, $N = 7139$). An additional pulse was added at the end to ensure the input data had a mean of 30%. From the input-output data generated by this experiment, we fit the model parameters (7a) using $p = 50$ and $n = 6$.[1] Noting that the system started at equilibrium $x(0) = 0$, we then computed the long-range predictions as $\hat{y}(k) = \sum_{j=0}^{k-1} \hat{C}\hat{A}^{k-j-1}\hat{B}u(j) + \hat{D}u(k)$, which are presented in Fig. 2. Notice that there is a drift in the underlying data (a result of the TCLab not being fully warmed up and a slow change in the ambient temperature). Step responses of the identified model show moderate interaction between the two heater-sensor units (Fig. 3). We posit the inverse responses in

[1] The parameters n and p were tuned by hand, although they can also be selected by minimizing either the Aikaike or Bayesian information criterion [23], [24]. In general, we found larger values of n and p produce better model fit and state approximations, whereas computation time and model complexity increase correspondingly.

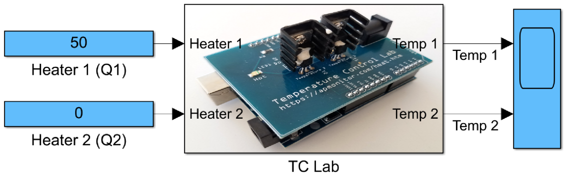

Fig. 1. Temperature Control Lab developed by Hedengren [14].

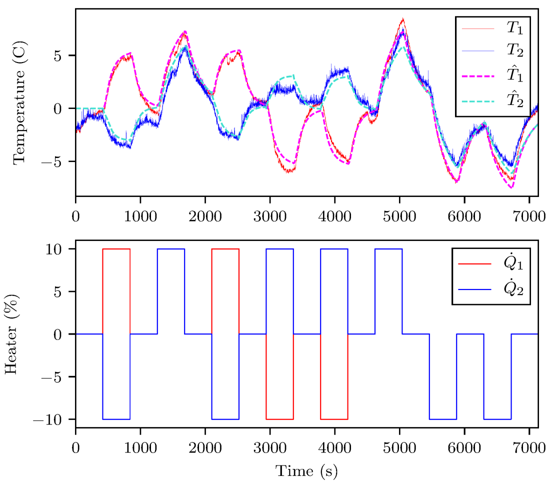

Fig. 2. Input-output data (red and blue points and lines) and estimated model fits (magenta and cyan lines) for the TCLab hardware. The model was fit according to Theorems 1 and 2 and Equation (7a) using $p = 50$ and $n = 6$.

the interactions are numerical artifacts of the state estimation procedure. The singular values of H are shown in Fig. 4. It is worth noting that the singular values are large even at the chosen cutoff, which motivates future work on developing the method for computing the state from the past data.

B. Disturbance model

Since the TCLab system has no intrinsic integrators, we used an output disturbance model $(B_d, C_d) = (0, I_2)$ and utilized the simplification (10). The disturbance sequence for the TCLab data (Fig. 2) was estimated as the long-range prediction errors (Fig. 5, top). The disturbance sequence is clearly correlated in time. However, when the disturbance estimates are differenced, the differences are uncorrelated in time (Fig. 5, bottom). These results are expected, and suggest the integrating disturbance model (3) is a good description of the TCLab system.

C. Closed-loop offset-free MPC

For brevity, we refer the reader to [15], [16], [17] for a description of the components of the offset-free MPC algorithm. The result of the closed-loop experiment of offset-free MPC on the TCLab is presented in Fig. 6. The TCLab was warmed up for 5 minutes prior to the experiment, and two difficult setpoint changes were applied. For this experiment, the controller is able to achieve nearly offset-free performance.

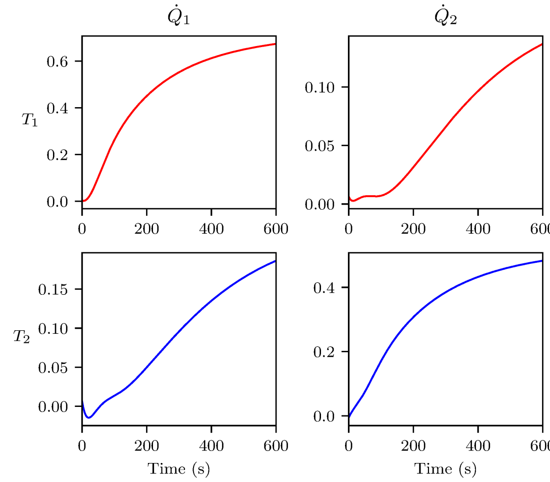

Fig. 3. The estimated model's unit step responses (in deviation variables) of the two measured temperature outputs to the two heater inputs. Top row: output T_1, bottom row: output T_2, left column: input $\dot{Q}_1$, right column: input $\dot{Q}_2$.

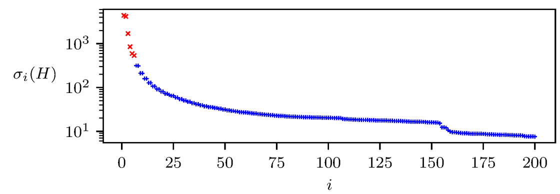

Fig. 4. Singular values from the SVD of the H matrix (5) generated by the TCLab data (Fig. 2). The first 6 singular values are kept (red $\times$) and the remaining are discarded (blue $+$).

V. DRAINING TANK SIMULATION

To evaluate the method on a system with integrators, we considered the following draining tank problem

$$\dot{h}_i = F_i - E_i, \quad i = 1, 2$$

where the h_i are tank levels, F_i are the feed rates, and E_i are the effluent rates. Assume the first tank effluent and the second tank feed are linked, $F_2 = E_1$. In deviation variables, the tank levels are the states $h_i - h_{i,s} = x_i$, the tank feeds are the inputs with input disturbances $F_i - F_{i,s} = u_i + d_i$, and the second tank drains on its own $E_2 = x_2$. This gives the following discrete-time system of the form (3),

$$\begin{aligned} x^+ &= \begin{bmatrix} 1 & 0 \\ 0 & e^{-1} \end{bmatrix} x + \begin{bmatrix} 1 & -1 \\ 0 & 1 - e^{-1} \end{bmatrix} (u + d) + w \\ d^+ &= d + w_d \\ y &= x + v \\ \begin{bmatrix} w \\ w_d \\ v \end{bmatrix} &\sim N(0, S_d) \end{aligned}$$

where the noise covariance was chosen as

$$S_d = \frac{1}{4} \operatorname{diag}\left(\begin{bmatrix} 1 & \frac{1}{4} \\ \frac{1}{4} & 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{1000} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}, \begin{bmatrix} 1 & \frac{1}{4} \\ \frac{1}{4} & 1 \end{bmatrix} \right)$$

Data was gathered in a similar manner to the TCLab hardware. The simulation was started at steady-state, then the

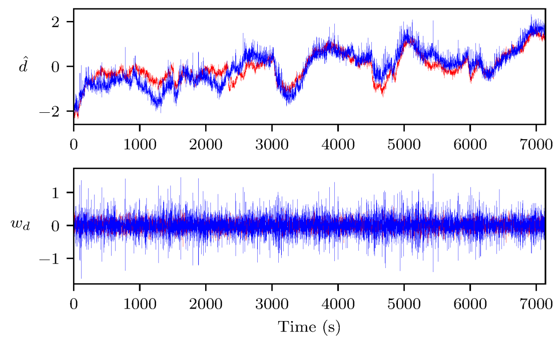

Fig. 5. (Top) Disturbance estimates as long-range prediction errors $\hat{d}(k) = y(k) - \hat{y}(k)$ for the two temperatures. (Bottom) Driving noise of the disturbance estimates $w_d(k) = \hat{d}(k+1) - \hat{d}(k)$.

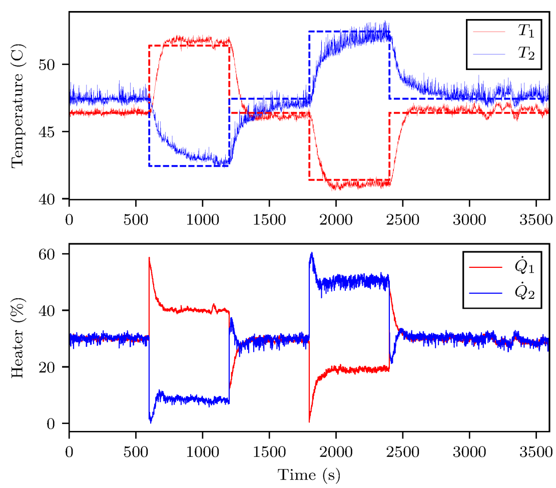

Fig. 6. Closed-loop MPC experiment on the TCLab using the model (3). All outputs are treated as controlled variables ($r_{sp}(k) = y_{sp}(k)$), the estimator initial guess is zero ($(\hat{x}(0), \hat{d}(0)) = (0, 0)$), and the following target calculator and regulator tuning was used: $N = 50$, $Q_s = I$, $R_s = R = 10^{-2}I$, and $Q = \hat{C}'Q_s\hat{C}$.

input was manipulated with a PRBS signal ($N = 170$). We fit the model parameters (7a) using $p = 4$ and $n = 4$. The long-range predictions are presented in Fig. 7. Notice that there is a drift in the underlying data due to the addition of an integrating disturbance. Since the tank system has integrating modes, we used an input disturbance model $(B_d, C_d) = (\hat{B}, 0)$, and the disturbance sequence for the simulated data (Fig. 7) was estimated using (9). The disturbance sequence is clearly correlated in time (Fig. 8, top). Moreover, when the disturbance estimates are differenced, the differences are uncorrelated in time (Fig. 8, bottom). These results validate the use of (9) to estimate the disturbance sequence.

The result of the closed-loop simulation on the tank system is presented in Fig. 9. The tanks started at the origin, and two difficult setpoint changes were applied. For this experiment, the controller appears to achieve zero expected offset.

VI. CONCLUSION

We formulate a series of maximum likelihood problems to estimate the parameters of the disturbance models used in

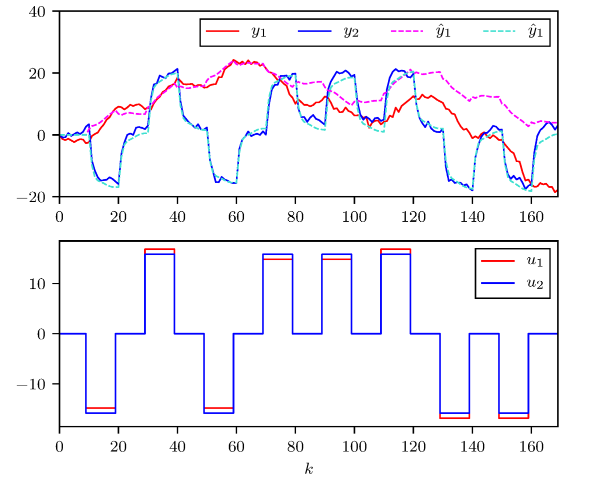

Fig. 7. Input-output data (red and blue points and lines) and estimated model fits (magenta and cyan lines) for the draining tank simulation. The model was fit according to Theorems 1 and 2 and Equation (7a) using $p = 4$ and $n = 4$.

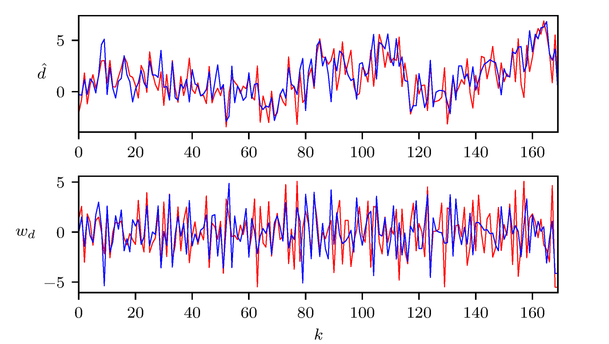

Fig. 8. (Top) Disturbance estimates for the two flowrates, computed according to (9). (Bottom) Driving noise of the disturbance estimates $w_d(k) = \hat{d}(k+1) - \hat{d}(k)$.

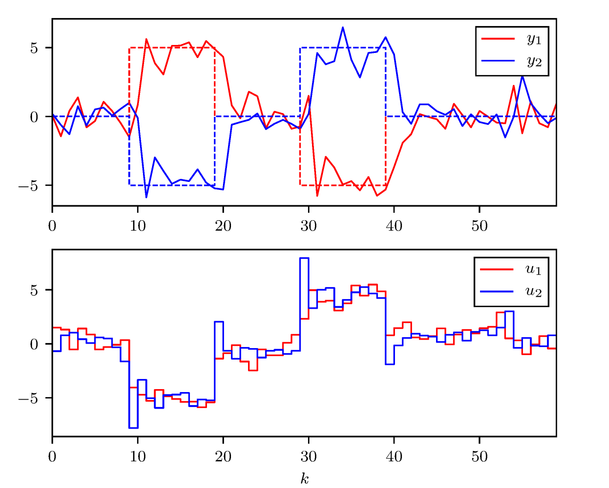

Fig. 9. Closed-loop offset-free MPC experiment on the simulated tank. All outputs are treated as controlled variables ($r_{sp}(k) = y_{sp}(k)$), the estimator initial guess is zero ($(\hat{x}(0), \hat{d}(0)) = (0, 0)$), and the following target calculator and regulator tuning was used: $N = 10$, $Q_s = I$, $R_s = R = I$, and $Q = \hat{C}'Q_s\hat{C}$.

offset-free MPC. We provide closed-form solutions to these problems and discuss numerical considerations in the design of the disturbance model. The method is tested on the TCLab hardware, which shows the estimated disturbance model is sufficient to achieve offset-free control in realistic conditions.

No method prior to this work has estimated the full disturbance model for offset-free MPC. As such, the properties of the closed-loop system, especially with regards to systems with explicit integrators, remain an open area of research. There are many possibilities for future research into variants on this method and for comparisons of various methods with hardware and numerical simulation experiments. Finally, while we tuned the parameters n and p by hand, future work could explore their selection via information criteria methods [23], [24].

VII. APPENDIX

Theorem 1: For the system (1), there exists a constant $\varepsilon \in [0, 1)$ such that for each $p \geq n$,

$$x(k) = \mathcal{C}_p Z(k) + O(\varepsilon^{p+1})$$

where $Z(k) := [u(k-1)',\ y(k-1)',\ \cdots,\ u(k-p)',\ y(k-p)']'$ and $\mathcal{C}_p := [B_K,\ A_K B_K,\ \cdots,\ A_K^{p-1} B_K]$ is full rank.

Proof: Starting with the predictor form (2) and applying recursion,

$$x(k) = A_K^p x(k-p) + \sum_{j=k-p}^{k-1} A_K^{k-j-1} B_K \begin{bmatrix} u(j) \\ y(j) \end{bmatrix}$$

If A_K is strictly stable and the state is bounded,

$$A_K^p x(k-p) = O(\varepsilon^{p+1})$$

where $\varepsilon = \sigma_1(A_K) \in [0, 1)$ and $\sigma_j(\cdot)$ denotes the j-th singular value of the argument $(\cdot)$. More importantly, the state can now be written approximately as a function of a *finite* number of past input-output pairs,

$$\begin{aligned} x(k) &= \sum_{j=k-p}^{k-1} A_K^{k-j-1} B_K \begin{bmatrix} u(j) \\ y(j) \end{bmatrix} + O(\varepsilon^{p+1}) \\ &= \mathcal{C}_p Z(k) + O(\varepsilon^{p+1}) \end{aligned}$$

Notice that the $\mathcal{C}_p$ contains the controllability matrix (which is full rank because (A_K, B_K) is controllable) and thus $\mathcal{C}_p$ is full rank. ■

Theorem 2: Consider the system (1). For each $p \geq n$, there exists a matrix $U_1 \in \mathbb{R}^{(n_y+n_u)p \times n}$ with orthonormal columns, nonsingular matrix $T \in \mathbb{R}^{n \times n}$, and constants $\varepsilon \in [0, 1)$ and $\delta \in \mathbb{R}_{>0}$ such that

$$x(k) = T\hat{x}(k) + O(\max\{\varepsilon^{p+1}, \delta\})$$

where $\hat{x}(k) = U_1' Z(k)$.

Proof: Take an economic SVD of the data matrix,

$$H = \begin{bmatrix} Z(p) & \dots & Z(N) \end{bmatrix} = U_1 S_1 V_1' + O(\delta)$$

where $\delta = \sigma_{n+1}(H)$. We can estimate the states as $\hat{x}(k) = U_1' Z(k)$ and therefore $Z(k) = U_1 \hat{x}(k) + O(\delta)$. Finally, we have the result

$$x(k) = T\hat{x}(k) + O(\max\{\varepsilon^{p+1}, \delta\})$$

by Theorem 1 where $T = \mathcal{C}_p U_1 \in \mathbb{R}^{n \times n}$ is invertible because it is the product of full row and column rank matrices. ■

REFERENCES

[1] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Eng. Pract.*, vol. 11, no. 7, pp. 733–764, 2003.

[2] J. B. Rawlings, D. Q. Mayne, and M. M. Diehl, *Model Predictive Control: Theory, Design, and Computation*, 2nd ed. Madison, WI: Nob Hill Publishing, 2020, 770 pages, ISBN 978-0-9759377-5-4.

[3] S. J. Qin, "An overview of subspace identification," *Comput. Chem. Eng.*, vol. 30, pp. 1502–1513, 2006.

[4] B. L. Ho and R. E. Kalman, "Efficient construction of linear state variable models from input/output functions," *Regelungstechnik*, vol. 14, pp. 545–548, 1966.

[5] M. Gevers, "A personal view of the development of system identification," *IEEE Ctl. Sys. Mag.*, vol. 26, no. 6, pp. 93–105, Dec 2006.

[6] W. E. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control," in *Proceedings of the 29th Conference on Decision and Control*, 1990, pp. 596–604.

[7] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[8] P. Van Overschee and B. De Moor, "N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[9] ——, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 31, no. 12, pp. 1853–1864, Dec 1995.

[10] F. Zhao and Y. P. Gupta, "Disturbance Rejection Using a Simplified Predictive Control Algorithm," *Ind. Eng. Chem. Res.*, vol. 44, no. 2, pp. 381–390, Jan 2005.

[11] Z. Sun, Y. Zhao, and S. J. Qin, "Improving industrial mpc performance with data-driven disturbance modeling," in *2011 50th IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 1922–1927.

[12] J. K. Huusom, N. K. Poulsen, S. B. Jørgensen, and J. B. Jørgensen, "Tuning SISO offset-free Model Predictive Control based on ARX models," *J. Proc. Cont.*, vol. 22, no. 10, pp. 1997–2007, Dec 2012.

[13] A. Caspari, H. Djelassi, A. Mhamdi, L. T. Biegler, and A. Mitsos, "Semi-infinite programming yields optimal disturbance model for offset-free nonlinear model predictive control," *J. Proc. Cont.*, vol. 101, pp. 35–51, May 2021.

[14] J. Park, R. A. Martin, J. D. Kelly, and J. D. Hedengren, "Benchmark temperature microcontroller for process dynamics and control," *Comput. Chem. Eng.*, vol. 135, Apr 2020.

[15] K. R. Muske and T. A. Badgwell, "Disturbance modeling for offset-free linear model predictive control," *J. Proc. Cont.*, vol. 12, no. 5, pp. 617–632, 2002.

[16] G. Pannocchia and J. B. Rawlings, "Disturbance models for offset-free MPC control," *AIChE J.*, vol. 49, no. 2, pp. 426–437, 2003.

[17] M. R. Rajamani, J. B. Rawlings, and S. J. Qin, "Achieving state estimation equivalence for misassigned disturbances in offset-free model predictive control," *AIChE J.*, vol. 55, no. 2, pp. 396–407, Feb 2009.

[18] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. New York: John Wiley & Sons, 2003.

[19] M. D. Graham and J. B. Rawlings, *Modeling and Analysis Principles for Chemical and Biological Engineers*. Madison, WI: Nob Hill Publishing, 2013, 552 pages, 978-0-9759377-1-6.

[20] W. E. Larimore, "Optimal reduced rank modeling, prediction, monitoring, and control using canonical variate analysis," in *Proceedings of the IFAC 1997 International Symposium on Advanced Control*, Banff, Canada, June 1997, pp. 61–66.

[21] C. R. Rao, "Unified theory of linear estimation," *Sankhyā A*, vol. 33, no. 4, pp. 371–394, 1971.

[22] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Hoboken, NJ: John Wiley & Sons Ltd., 2019.

[23] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in *Selected Papers of Hirotugu Akaike*, ser. Springer Series in Statistics, E. Parzen, K. Tanabe, and G. Kitagawa, Eds. New York, NY: Springer, 1998, pp. 199–213.

[24] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar 1978.