Application of Dual-Rate Modeling to CCR Octane Quality Inferential Control

Dongguang Li, Sirish L. Shah, *Member, IEEE*, Tongwen Chen, *Senior Member, IEEE*, and Kent Z. Qi

***Abstract*—Octane quality control at Shell Canada's continuous catalytic reforming (CCR) units is typically done manually due to infrequent measurements of the research octane number (RON). The goal of this paper is to study automating the control loop by developing a dual-rate inferential control scheme. In particular, for a dual-rate process with fast input updating and slow output sampling, we propose a polynomial domain method to identify a fast single-rate linear model based on dual-rate input–output data; using the fast model to supply missing samples, we extend a popular model-based predictive control algorithm to the inferential control framework; the identification and control algorithms are applied to a Shell Canada's CCR reactor, and the inferential controller is implemented in real time, resulting in 40% reduction in octane quality variance—a significant improvement.**

***Index Terms*—Identification, industrial applications, inferential control, multirate systems, process control.**

I. Introduction

THIS PAPER was motivated by an industrial process control problem—improving octane quality control in continuous catalytic reforming (CCR) reactors at Shell Canada's Scotford plant in Alberta, Canada.

The octane content is an important quality variable in gasoline production; the CCR unit at Shell Canada is responsible for upgrading the research octane number (RON) of a naphtha feed stream, an indicator of the severity of the CCR reaction. Taking this RON as the process output, the reaction process consists of one manipulated variable, the reactor weighted average inlet temperature (WAIT), and several measured disturbances. The RON measurement is obtained through an octane GC analyzer, which is able to provide a reading every 2.5 h; based on this the operator adjusts the WAIT target manually. (More details of this process will be given in Section VI.)

In order to improve octane quality control, it is desirable to automate the control operation by closing the loop from RON to WAIT. Though WAIT can be updated relatively fast, say, every 30 min, the speed of the RON measurement is hardware-constrained. How to model and control such dual-rate systems with fast control input updating and slow output sampling presents a challenge, mainly due to *infrequent* output measurement. Our approach is to develop an inferential model predictive control (MPC) scheme; the inferential controller makes use of a fast-rate model for the process to compute a fast-rate output. The key element is the fast-rate model—we will propose a frequency domain method to estimate such models based on given dual-rate input–output data, i.e., the output sampled data used in identifying the fast-rate models is at the slow rate.

Multirate systems also arise in many other chemical processes such as polymer reactors [12] and fermentation processes [4], where the composition, density and molecular weight distribution measurements are typically obtained after several minutes of analysis, whereas the manipulated variables can be adjusted at relatively fast rates, the only limitation being the load on the distributed control system. Modeling and control of such systems are main objectives in this paper.

The study of multirate systems goes back to the early 1950s. The first important work was performed by Kranc on the switch decomposition technique [7]; later Friedland further developed the concept of lifting which converts a periodic discrete-time system into a time-invariant system [3]. Since then, the lifting technique has become an important and widely used tool for analysis of multirate systems. In the process control literature, multirate systems have also been studied. For example, Lu and Fisher [11] proposed an algorithm for estimating the inter-sample outputs by using slowly measured outputs and fast control inputs; Lee and Morari developed a generalized inferential control scheme and discussed various optimal control problems in the LQG, MPC, and IMC framework [8]; Gudi *et al.* developed an enhanced observability method to estimate the inter-sample outputs and used these for fast-rate control [5]. We note that most inferential control algorithms reported in the literature, e.g., [5] and [8], assume that fast single-rate models of the processes are available, which is usually not the case. One of the purposes of this paper is to fill the gap by providing a method of obtaining such fast single-rate models based on multirate data.

System identification techniques in the polynomial domain have been applied successfully in industries for the last two decades [10], [13]. For periodically time-varying systems, Verhaegen and Yu have presented a state-space based technique to estimate models of periodically time-varying systems [14]. Our approach is to estimate the fast single-rate model from multirate data in the polynomial domain. The results in this paper differ from those by Verhaegen and Yu [14] in that this method identifies not only a lifted model for the multirate process but also a fast single-rate model, based on *multirate* input–output data.

Manuscript received July 19, 2001; revised March 19, 2002. Manuscript received in final form April 19, 2002. Recommended by Associate Editor F. Doyle.

D. Li is with Honeywell Inc., Calgary, AB T2H 0C2, Canada (e-mail: dongguang.li@honeywell.com).

S. L. Shah is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada (e-mail: sirish.shah@ualberta.ca).

T. Chen is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: tchen@ee.ualberta.ca).

K. Z. Qi is with Shell Canada Products Ltd., Fort Saskatchewan, AB T8L 3T2, Canada (e-mail: kent.qi@shell.ca).

Digital Object Identifier 10.1109/TCST.2002.806433

As we discussed earlier in the CCR octane control system, this paper will focus on one special class of multirate systems where the output sampling interval is nT, but the control updating interval is T. This paper develops an algorithm to estimate a single-rate model of the system with sampling period T from dual-rate data. Several papers such as [5] and [8] developed many multirate inferential MPC algorithms. In this paper, we will extend an existing MPC algorithm which is widely accepted in industries into the dual-rate inferential control framework.

This paper is organized as follows. In Section II the lifting technique and lifted dual-rate systems are discussed in both state-space and transfer function forms. In Section III, a method for extracting fast-rate models from the lifted models is given and analyzed in the frequency domain. An inferential control algorithm for multirate systems is presented in Section IV. The inferential control algorithm in Section IV uses only step-response coefficients for the fast-rate model; in Section V, a method to estimate the fast step-response models directly from the dual-rate input–output data is briefly discussed. The results of the paper are applied and validated in Section VI on the industrial CCR unit with Shell Canada. Finally, conclusions are given in Section VII.

II. Preliminaries

In this section, the basic idea of the lifting operation will be presented first; then the lifting technique will be explored to show how to transform a dual-rate system into a single-rate system.

Consider the dual-rate sampled-data system in Fig. 1. Here, P is a continuous-time process with additive noise. The noise is generated by a continuous-time model N with a white noise input e, while the output of P is corrupted by that of N, and is sampled by a sampler S_{nT} with period nT, yielding the sampled output $y(k)$. The input to P is generated by a zero-order hold with period T. This is a fairly common and a practical situation in the process industry—see our discussion on the CCR octane quality control problem in Section I. From here on, we will focus on this case.

Note that both $u(k)$ and $y(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \ldots\}$; but their underlying periods are T and nT, respectively. Throughout the paper P is assumed to be linear and time-invariant (LTI). However, the discrete-time system from $u(k)$ to $y(k)$, with $e = 0$, is linear and periodically time-varying. In order to get a tractable model for this dual-rate system, the lifting technique is used.

Let $u(k)$ be a discrete-time signal defined on $\mathcal{Z}_+$. The n-fold lifting operator L_n maps u to u_L (subscript L denotes lifting) as follows: If

$$u = \{u(0), u(1), u(2), \ldots\} \tag{1}$$

then

$$u_L = Lu = \left\{ \begin{bmatrix} u(0) \\ u(1) \\ \cdots \\ u(n-1) \end{bmatrix}, \begin{bmatrix} u(n) \\ u(n+1) \\ \cdots \\ u(2n-1) \end{bmatrix}, \ldots \right\}. \tag{2}$$

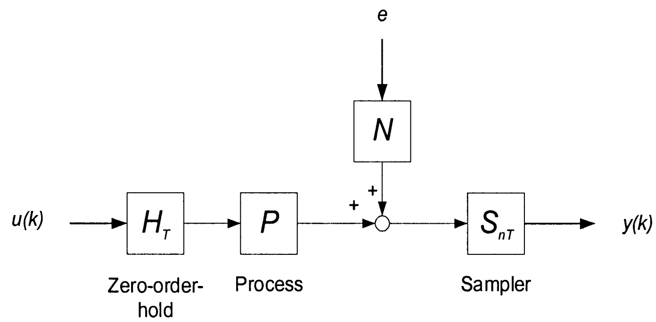


Fig. 1. Dual-rate sampled-data system.

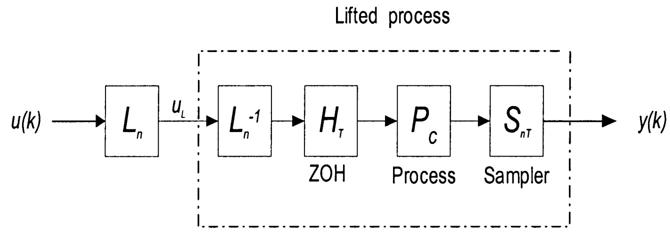


Fig. 2. Lifted dual-rate system.

Note that the dimension of the lifted signal u_L is n times that of u, and the underlying period of u_L again is n times that of u. It is easy to see that the inverse lifting process, L_n^{-1}, mapping u_L back to u, is well defined. It is straightforward to obtain the following identities:

$$L_n^{-1} L_n = I, \qquad L_n L_n^{-1} = I. \tag{3}$$

The dual-rate system from $u(k)$ to $y(k)$ in Fig. 1 can be transformed into a single-rate system with underlying period nT by lifting the input; and hence u needs to be lifted by L_n to arrive at Fig. 2.

In order to derive a state-space model for the lifted system, one can discretize P via the zero-order hold method with the sampling period T to get $P_T := S_T P H_T$, S_T and H_T being the sampler and zero-order hold with period T, and assume that P_T has a state-space model of the form:

$$D + C(zI - A)^{-1}B =: \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{4}$$

Let P_L represent the lifted system; it is shown that P_L is LTI [1]. Denote a state-space model for P_L by

$$P_L(z) = D_L + C_L(zI - A_L)^{-1}B_L. \tag{5}$$

These matrices can be related to those of P_T in (4) as follows [1]:

$$\left[\begin{array}{c|c} A_L & B_L \\ \hline C_L & D_L \end{array}\right] = \left[\begin{array}{c|cccc} A^n & A^{n-1}B & A^{n-1}B & \cdots & B \\ \hline C & D & 0 & \cdots & 0 \end{array}\right]. \tag{6}$$

Discretizing the continuous-time noise model N in Fig. 1 with sampling period nT gives N_{nT}. Thus, one can get the overall lifted model as follows:

$$y(k) = P_L u_L(k) + N_{nT} e(k). \tag{7}$$

Most chemical processes are strictly proper or have a time delay, hence D can be set to a zero matrix without loss of generality. Since the lifted system is LTI, most existing identification algorithms in either the polynomial domain or the state-space domain can be used to estimate the lifted model.

III. Dual-Rate System Identification

If the model of a lifted dual-rate system can be estimated, then the natural question to ask is: Can one determine the model of the fast-rate system from the estimated lifted model? The answer to this question is affirmative. This problem has been solved in the state-space domain [9]; and in this section a method for extracting a fast model in the transfer function domain will be presented.

One can start with a model for P_L in (5). Knowing the transfer function $P_L(z)$, we would like to compute a model for $P_T = S_T P H_T$; specifically, from (6), we would like to compute the transfer function $D + C(zI - A)^{-1}B$. Since P_L has n inputs, P_L has n subsystems:

$$P_L = [\, P_0 \quad P_1 \quad \cdots \quad P_{n-1} \,]. \tag{8}$$

Equation (6) indicates that systems $P_0, P_1, \ldots, P_{n-1}$ can be represented by the following models:

$$\begin{cases} P_0(z) = D + C(zI - A^n)^{-1}A^{n-1}B \\ P_1(z) = C(zI - A^n)^{-1}A^{n-2}B \\ \vdots \\ P_{n-1}(z) = C(zI - A^n)^{-1}B. \end{cases} \tag{9}$$

After estimating the transfer function of the lifted system, one can form a new transfer function $\varphi(z)$ as follows:

$$\varphi(z) = P_0(z^n) + zP_1(z^n) + \cdots + z^{n-1}P_{n-1}(z^n). \tag{10}$$

We will show that $\varphi(z) = P_T(z)$. First note from (10) and (9)

$$\varphi(z) = D + C(z^nI - A^n)^{-1} \cdot [A^{n-1} + zA^{n-2} + \cdots + z^{n-1}I]B. \tag{11}$$

Now observe that

$$z^nI - A^n = (A^{n-1} + zA^{n-2} + \cdots + z^{n-1}I)(zI - A)$$

which implies immediately from (11) that

$$\varphi(z) = D + C(zI - A)^{-1}B = P_T(z).$$

Let us summarize the steps involved to extract a fast single-rate model $P_T(z)$ based on the lifted transfer function $P_L(z)$: First, partition $P_L(z)$ as in (8) to get the subsystem transfer functions $P_0(z), \ldots, P_{n-1}(z)$. Second, compute the function $\varphi(z)$ as in (10). Finally, set $P_T(z) = \varphi(z)$.

Adding another dimension to the process insights, analysis in the frequency domain is of great importance. Next, we will briefly analyze the proposed dual-rate identification method in the frequency domain.

Assume that a process model

$$y(k) = \hat{P}_L u_L(k) + \hat{N}_{nT} e(k) \tag{12}$$

is estimated from the dual-rate input–output data, where $\hat{P}_L$ is the estimated deterministic model P_L, and $\hat{N}_{nT}$ the estimated disturbance model N_{nT}. The model in (12) can be used to predict the output. Letting $\hat{y}(k|\theta)$ represent the predicted output, we have

$$\begin{aligned} \hat{y}(k|\theta) &= \hat{P}_L u_L(k) + \left(\hat{N}_{nT} - 1\right) e(k) \\ &= \hat{P}_L u_L(k) + \left(\hat{N}_{nT} - 1\right) \frac{1}{\hat{N}_{nT}} \left(y - \hat{P}_L u_L(k)\right) \\ &= y - \frac{1}{\hat{N}_{nT}} \left(y - \hat{P}_L u_L(k)\right) \end{aligned} \tag{13}$$

so the prediction error is

$$\hat{e}(k) = y(k) - \hat{y}(k|\theta) = \frac{1}{\hat{N}_{nT}} \left[y(k) - \hat{P}_L u_L(k)\right]. \tag{14}$$

Substituting (7) and (13) into (14) gives

$$\begin{aligned} \hat{e}(k) &= \frac{1}{\hat{N}_{nT}} \left[P_L u_L(k) + N_{nT} e(k) - \hat{P}_L u_L(k)\right] \\ &= \frac{1}{\hat{N}_{nT}} \left[\left(P_L - \hat{P}_L\right) u_L(k) + N_{nT} e(k)\right] \\ &= \frac{1}{\hat{N}_{nT}} \left(P_L - \hat{P}_L\right) u_L(k) + \frac{N_{nT}}{\hat{N}_{nT}} e(k). \end{aligned}$$

In prediction error methods, one tries to find the parameter vector θ by minimizing the sum of the square of the prediction error

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{M} \sum_{k=1}^{M} [\hat{e}(k)]^2 \tag{15}$$

where M is the number of data points. Define

$$R_{u_L}(\tau) = \lim_{M\to\infty} \frac{1}{M} \sum_{t=1}^{M} u_L(t) u_L^T(t - \tau)$$

$$\phi_{u_L}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{u_L}(\tau) e^{-j\omega\tau}$$

similarly for $R_e(\tau)$ and $\phi_e(\omega)$. For simplicity, we assume that $u_L(k)$ is independent of $e(k)$. As $M \to \infty$, we can apply Parseval's theorem to (15) and get that $\hat{\theta}$ is obtained by minimizing over θ the following quantity:

$$\int_{-\pi}^{\pi} \frac{1}{\left|\hat{N}_{nT}(e^{j\omega})\right|^2} \left\{\left[P_L(e^{j\omega}) - \hat{P}_L(e^{j\omega})\right] \phi_{u_L}(\omega) \cdot \left[P_L^T(e^{j\omega}) - \hat{P}_L^T(e^{j\omega})\right] + |N_{nT}(e^{j\omega})|^2 \phi_e(\omega)\right\} d\omega. \tag{16}$$

Since u_L is an $n \times 1$ vector, $\phi_{u_L}(\omega)$ is an $n \times n$ matrix. Singular value decomposition (SVD) of $\phi_{u_L}(\omega)$ yields

$$\phi_{u_L}(\omega) = U(\omega) S(\omega) V^T(\omega)$$

where $S(\omega)$ is a diagonal matrix of the same dimensions as $\phi_{u_L}(\omega)$ and with nonnegative diagonal elements in the decreasing order, $U(\omega)$ and $V(\omega)$ are unitary matrices. Since

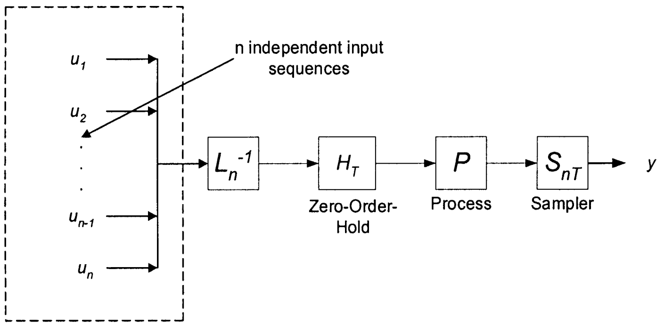


Fig. 3. Dual-rate system with *vectorized* excitation.

$\phi_{u_L}(\omega)$ is a symmetric matrix, we have that $U(\omega)$ and $V(\omega)$ are identical. Hence

$$\phi_{u_L}(\omega) = U(\omega)S(\omega)U^T(\omega).$$

If

$$P_L(e^{j\omega}) - \hat{P}_L(e^{j\omega}) = [0 \quad \cdots \quad 0 \quad 1]U^{-1}$$

then

$$\begin{aligned}\left[P_L(e^{j\omega}) - \hat{P}_L(e^{j\omega})\right] \phi_{u_L}(\omega) \left[P_L^T(e^{j\omega}) - \hat{P}_L^T(e^{j\omega})\right] \\ = [0 \quad \cdots \quad 0 \quad 1]S(\omega)\begin{bmatrix}0\\ \vdots \\ 0 \\ 1\end{bmatrix} = \sigma_n\end{aligned}$$

which is the smallest singular value of $\phi_{u_L}(\omega)$. Thus, the quantity in (16) equals to

$$\int_{-\pi}^{\pi} \frac{\{\sigma_n + |N_{nT}(e^{j\omega})|^2 \phi_e(\omega)\}}{\left|\hat{N}_{nT}(e^{j\omega})\right|^2}\, d\omega.$$

Recall

$$u_L(k) = \begin{bmatrix} u(nk) \\ \vdots \\ u(nk+n-1)\end{bmatrix}.$$

If

$$u(nk) = u(nk+1) = \cdots = u(nk+n-1)$$

for most of the time $(k = 0, 1, \ldots, M)$, then there is not enough excitation, and $\sigma_n \to 0$. When there is not enough excitation, those identification methods in the polynomial domain may give poor results. In order to avoid this, we introduce the so called *vectorized* excitation method: Design a vector input sequence $u_L(k)$ whose components are independent to each other; inverse lift $u_L(k)$ by L_n^{-1} before it is applied to the dual-rate process. This is illustrated in Fig. 3.

The quantity ϕ_{u_L} for the vectorized excitation is a proper diagonal matrix, and hence giving good signal-to-noise ratio. We can then expect good identification results.

IV. Inferential MPC

In this section, we will focus on one specific class of processes, namely, the processes where the process output is corrupted by an additive integrated white noise. Implementing MPC to dual-rate systems, slow output sampling (with period nT) normally limits the performance. However, using the fast model with sampling period T, we can estimate the inter-sample outputs $y(nk+1), y(nk+2), \ldots, y(nk+n-1)$, and design an MPC based on this fast single-rate model using both the measured outputs at sampling instants nT and the estimated outputs at the inter-sample instants. This is the basic idea of the model based-inferential control algorithm.

The model of the process sampled with period T is

$$y(k) = P_T(q^{-1})u(k) + \frac{e(k)}{1-q^{-1}} \tag{17}$$

where q^{-1} is the backward shift operator, $P_T(q^{-1})$ is a strictly proper transfer function, $u(k)$ is the control signal, $y(k)$ is the sampled output, $e(k)$ is a white noise input, and $u(k)$, $y(k)$, and $e(k)$ are discrete-time signals defined on the time set $\mathcal{Z}_+ := \{0, 1, 2, \ldots\}$ with underlying period T. In the rest of this section, we will develop a dual-rate inferential control algorithm for the system in (17). Consider the system in (17) at time instant nkT, the future output can then be expressed as

$$\begin{aligned} y(nk+j) &= P_T(q^{-1})u(nk+j) \\ &\quad + (1 + q^{-1} + q^{-2} + \cdots)e(nk+j) \\ &= P_T(q^{-1})u(nk+j) + e(nk+j) + \cdots \\ &\quad + e(nk+1) + \frac{1}{1-q^{-1}}\, e(nk). \end{aligned} \tag{18}$$

Since

$$e(nk) = (1-q^{-1})[y(nk) - P_T(q^{-1})u(nk)] \tag{19}$$

substituting (19) into (18) gives

$$\begin{aligned} y(nk+j) = P_T(q^{-1})u(nk+j) + e(nk+j) + \cdots \\ + e(nk+1) + y(nk) - P_T(q^{-1})u(nk). \end{aligned}$$

Because $e(nk)$ is a white-noise signal, the minimum variance prediction of $y(nk+j)$ at time nk is

$$\begin{aligned} &\hat{y}(nk+j|nk) \\ &\quad = P_T(q^{-1})u(nk+j) + y(nk) - P_T(q^{-1})u(nk). \end{aligned} \tag{20}$$

Rewrite (20) as

$$\begin{aligned} \hat{y}(nk+j|nk) &= \frac{P_T(q^{-1})}{\Delta}\, \Delta u(nk+j) \\ &\quad + y(nk) - \frac{P_T(q^{-1})}{\Delta}\, \Delta u(nk) \end{aligned} \tag{21}$$

where $\Delta = 1 - q^{-1}$. Write

$$\frac{P_T(q^{-1})}{\Delta} = a_1 q^{-1} + a_2 q^{-2} + \cdots. \tag{22}$$

Substituting (22) into (21) yields

$$\begin{aligned}\hat{y}(nk+j|nk) &= a_1\Delta u(nk+j-1)+\cdots+a_j\Delta u(nk)\\ &\quad + a_{j+1}\Delta u(nK-1)+\cdots+y(nk)\\ &\quad - a_1\Delta u(nk-1)-a_2\Delta u(nk-2)-\cdots\\ &= \sum_{i=1}^{j} a_i\Delta u(nk+j-i)+y(nk)\\ &\quad + \sum_{i=1}^{\infty}(a_{i+j}-a_i)\Delta u(nk-i). \qquad (23)\end{aligned}$$

The first term is the effect of future inputs, and the rest represents the expression that defines the free response.

We know that when i is large enough, $a_{i+j}\approx a_i\approx K_g$ (K_g is the steady-state gain of the system), so assume that $a_i\approx K_g$ when $i\geq M$ (M is a sufficiently large integer). Equation (23) can then be approximated as

$$\begin{aligned}&\hat{y}(nk+j|nk)\\ &\quad\approx \sum_{i=1}^{j} a_i\Delta u(nk+j-i)+y(nk)\\ &\qquad + \sum_{i=1}^{M}(a_{i+j}-a_i)\Delta u(nk-i)\\ &\quad\approx \sum_{i=1}^{j} a_i\Delta u(nk+j-i)+\hat{f}(nk+j|nk) \qquad (24)\end{aligned}$$

where

$$\hat{f}(nk+j|nk) := y(nk)+\sum_{i=1}^{M}(a_{i+j}-a_i)\Delta u(nk-i)$$

denotes the output at time $(nk+j)$ due to the free response starting from time nk. Assume that the control action u is constant after time instant $(nk+m-1)T$, i.e.,

$$u(nk+m-1)=u(nk+m)=\cdots=u(nk+h-1) \qquad (25)$$

where h and m are the so called prediction horizon and control horizon, respectively. Now consider the prediction of the output trajectories over the intervals $(nk+1)T$ to $(nk+h)T$

$$\begin{aligned}\hat{y}(nk+1|nk) &= \hat{f}(nk+1|nk)+a_1\Delta u(nk)\\ &\;\;\vdots\\ \hat{y}(nk+h|nk) &= \hat{f}(nk+h|nk)+a_h\Delta u(nk)+\cdots\\ &\quad + a_{h-m+1}\Delta u(nk+m-1).\end{aligned}$$

The above equations can be written into a compact vector form as

$$\hat{Y}=\hat{F}+A\Delta U \qquad (26)$$

where

$$\hat{Y}=\begin{bmatrix}\hat{y}(nk+1|nk)\\ \vdots\\ \hat{y}(nk+h|nk)\end{bmatrix},\quad \hat{F}=\begin{bmatrix}\hat{f}(nk+1|nk)\\ \vdots\\ \hat{f}(nk+h|nk)\end{bmatrix}$$

$$\Delta U=\begin{bmatrix}\Delta u(nk)\\ \vdots\\ \Delta u(nk+m-1)\end{bmatrix}$$

and

$$A=\begin{bmatrix}a_1 & 0 & 0 & \cdots & 0\\ a_2 & a_1 & 0 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_m & a_{m-1} & a_{m-2} & \cdots & a_1\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_h & a_{h-1} & a_{h-2} & \cdots & a_{h-m+1}\end{bmatrix}.$$

A set of control actions can be calculated by minimizing the following quadratic objective function:

$$J(\Delta U)=\left(\overline{r}-\hat{Y}\right)^T Q\left(\overline{r}-\hat{Y}\right)+\Delta U^T R\Delta U \qquad (27)$$

subject to (26), where Q and R are symmetric output weighting and control weighting matrices, respectively, and $\overline{r}$ is a vector containing the h desired future outputs.

The solution of this optimization problem is

$$\begin{aligned}\Delta U &= \left(A^TQA+R\right)^{-1}A^TQ^T\left(\overline{r}-\hat{F}\right)\\ &=: K_c\left(\overline{r}-\hat{F}\right)\\ K_c &= \left(A^TQA+R\right)^{-1}A^TQ^T.\end{aligned}$$

Since the MPC is a receding horizon based algorithm, only the first control action is implemented. (This MPC algorithm, which was first developed by Cutler and Ramaker [2], is now widely practiced in the process industry.)

From the MPC algorithm we can see that the output is needed every time the control action is calculated. But in the dual-rate system the output is measured only every period nT; if we want to implement MPC every period T, one option is to estimate the $(n-1)$ inter-sample outputs between two successive output measurements. The minimum variance estimation of the $(n-1)$ inter-sample outputs from time nkT to $(n+1)kT$ can be derived from (24) as follows:

$$\begin{aligned}\hat{y}(nk+1) &= y(nk)+\sum_{i=1}^{M}(a_{i+1}-a_i)\Delta u(nk-i)\\ &\quad + a_1\Delta u(nk)\\ &\;\;\vdots\\ \hat{y}(nk+n-1) &= y(nk)+\sum_{i=1}^{M}(a_{i+n-1}-a_i)\Delta u(nk-i)\\ &\quad + \sum_{i=0}^{n-2}a_{n-1-i}\Delta u(nk+i).\end{aligned}$$

The unconstrained inferential MPC algorithm then consists of the following steps.

Step 1) Calculate the controller matrix $K_c = (A^TQA+R)^{-1}A^TQ^T$.

Step 2) Calculate the prediction of the output trajectories over the prediction horizon of interest. At time instant knT the output is measured, and the free response can be calculated as follows:

$$\hat{f}(nk+j|nk)=y(nk)+\sum_{i=1}^{M}(a_{i+j}-a_i)\Delta u(nk-i)$$
$$j=1,2,\ldots,h.$$

At time instant $t = (kn + l)T$, $1 \le l < n$, the output is not measured, but we can estimate the inter-sample output

$$\hat{y}(nk + l) = y(nk) + \sum_{i=1}^{M} (a_{i+l} - a_i)\Delta u(nk - i) + a_l \Delta u(nk + 1) + \cdots + a_1 \Delta u(nk + l - 1)$$

and then compute the free response based on $\hat{y}(nk + l)$

$$\hat{f}(nk + l + j|nk + l) = \hat{y}(nk + l) + \sum_{i=1}^{M} (a_{i+j} - a_i)\Delta u(nk + l - i)$$

for $j = 1, 2, \ldots, h$.

Step 3) Calculate only the next control action

$$\Delta u(t) = k_1 \left(\overline{r} - \hat{F}\right)$$

where k_1 is the first row of the controller matrix K_c.

V. Identification of Fast Step-Response Models

In the inferential MPC strategy developed in the last section, one can see that only fast-rate step-response models are needed. Clearly, fast step-response models can be calculated after the transfer functions for the fast-rate models are computed using the method in Section III. But there is also a more direct way to estimate the coefficients in the step response of a fast-rate model based on dual-rate input–output data; we will briefly discuss this approach here.

Setting $j = n$ in (24), we have

$$\hat{y}(nk + n|nk) \approx \sum_{i=1}^{n} a_i \Delta u(nk + n - i) + y(nk) + \sum_{i=1}^{M} (a_{i+n} - a_i)\Delta u(nk - i).$$

The prediction error is

$$\begin{aligned} v(k) &= y(nk + n) - \hat{y}(nk + n|nk) \\ &\approx y(nk + n|nk) - y(nk) - \sum_{i=1}^{n} a_i \Delta u(nk + n - i) \\ &\quad - \sum_{i=1}^{M} (a_{i+n} - a_i)\Delta u(nk - i). \end{aligned}$$

A least-squares optimization problem

$$\min_{a_i} \sum_{k=1}^{J} v^2(k)$$

can be formulated after the dual-rate data being collected, where J is the length of the collected output data. The fast step-response model can be computed directly by solving this optimization problem. Though this method is simpler, prior knowledge about the settling time of the process is required in order to determine the truncation number M.

VI. Application to CCR Octane Quality Control

The octane content is an important quality criterion in the gasoline production units. The CCR unit at Shell Canada's Scotford plant is responsible for improving the RON (research octane number) of a naphtha feed stream. The increase in RON is accomplished by conversion of a naphthene and paraffin material to aromatic. Therefore, it is an indication of the severity of the CCR reaction. The CCR reactor consists of four beds with continuous catalyst circulation and regeneration. Because the reforming reactions are largely endothermic, heat must be added to the reactor feed prior to entering each bed. The heat is provided by a four-cell balanced draft fired heater with a common convection section. The reactor bed volumes are different and increase with each successive bed. Currently, a weighted average inlet temperature (WAIT) quadratic dynamic matrix control (QDMC) application is used to control the reactor severity.

Affecting the RON, there are also many other variables such as loop pressure, reactor temperatures inside the unit and the feed composition changes outside the unit. For instance, as the amount of feed precursors is decreased, the RON will also decrease for a given severity. If this change is measured, increasing the reactor WAIT can offset the decrease in RON. By implementing RON control, the effect of process disturbances can be minimized. The WAIT can also be manipulated to compensate for other critical operating variables such as decreased catalyst activity.

Usually the octane is sampled and tested in a plant laboratory on a daily basis. Good octane control needs frequent measurements of the octane content which requires expensive analytical equipment. There are limited technologies in the market but all require large capital investment and extensive maintenance efforts. As a tradeoff of performance and investment, an octane GC analyzer was installed in the process environment to measure the composition of the CCR product stream and the octane is validated and calculated for online measurement. It provides the octane reading every 2.5 h. Though it seems slow for a typical control application, it is certainly a big step forward toward plant optimization. Currently, the WAIT target is set by operators based on the slow sampled octane measurement and the desired octane requirement, i.e., the plant runs under manual closed-loop conditions but the control frequency is *slow*.

The main objective of this control application is to close the loop for octane control automatically at the *fast* rate. Due to the slow sampling rate of the analyzer, an inferential octane model has to be developed to allow the control application to make moves at a faster rate. Reactor WAIT is the manipulated variable and is desired to be adjusted every 30 min. All other disturbance measurements are also available to estimate the octane at inter-sampling intervals. Therefore, the first step is to identify the dynamic models from all input variables to the output octane variable, clearly this is a multirate model identification problem.

For the purpose of convenience, y is used to represent the output (RON), u to represent the manipulated variable (WAIT), and d_i, $i = 1, \ldots 7$, to represent the disturbances. A total of 205 output data are collected with a period of 150 min. The corresponding input and disturbances are collected at a shorter period of 30 min. The detrended output and manipulated variables are shown in Fig. 4. Three disturbances are shown in Fig. 5. (The reason why only three disturbances are shown will be stated later.)

In total there are one output and eight inputs (manipulated variable and disturbances), and the ratio between the output

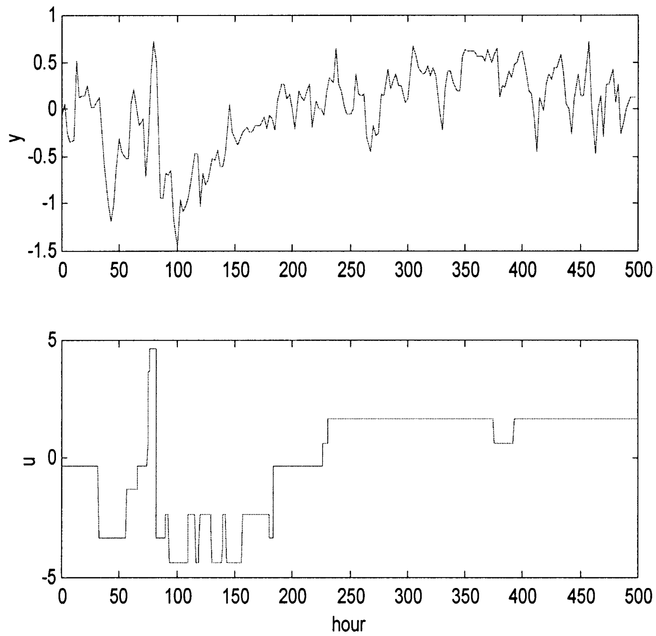


Fig. 4. Output and manipulated variable measurement.

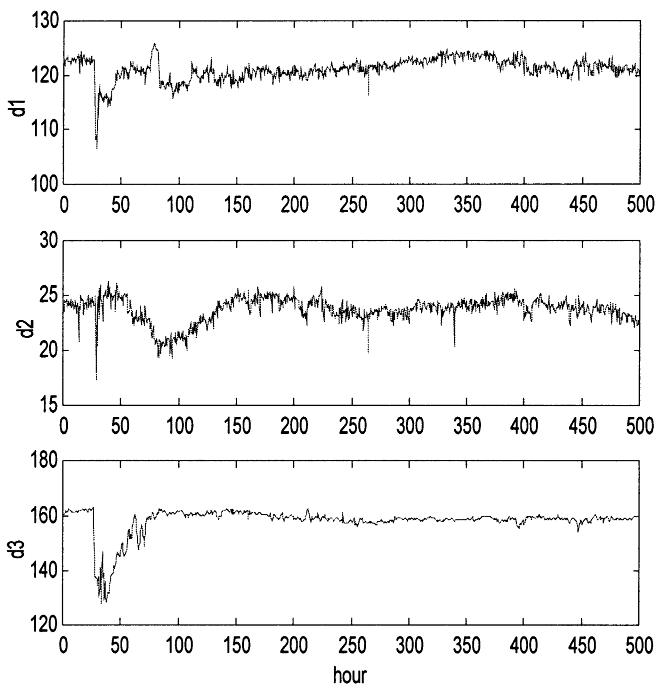


Fig. 5. Three disturbances.

sampling interval and control interval is $n = 5$; so if one lift the inputs in order to get a single-rate lifted system, there will be 40 (8 $\times$ 5) inputs in the lifted system. Our objective is to estimate the model of the lifted system and then extract a fast sampled model from the lifted model; clearly it is difficult to estimate the model of the system with 40 inputs from only 200 data points. So before we estimate the model, we ask a question: Do all the inputs affect the output significantly? If some of the inputs do not affect the output much, then we can ignore these inputs and estimate a model between the output and the important inputs.

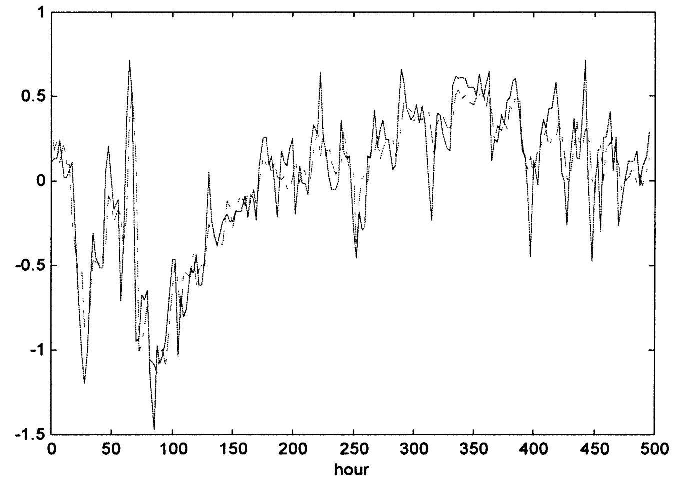


Fig. 6. Output measurement (solid line) and the prediction of the fast model (dash-dot line).

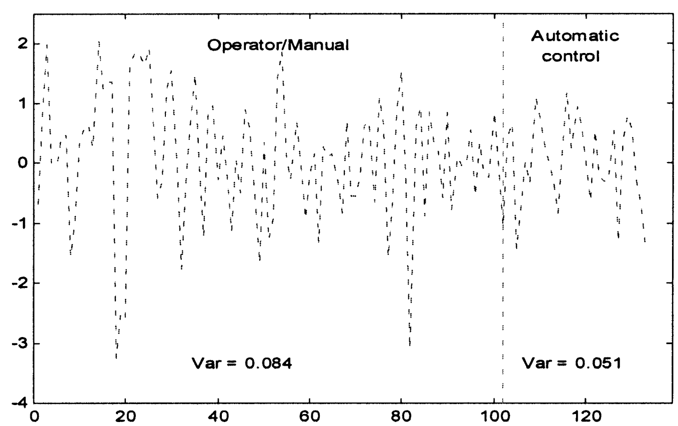


Fig. 7. Comparison of performance before and after implementing automatic control.

In order to answer this question, we apply the canonical correlation analysis (CCA) which was initially developed by Hotelling [6]. The purpose of CCA is to identify and quantify the association between two sets of variables: Find a linear combination of variables in the first set and a linear combination in the second set such that the correlation between the two linear combinations is maximized. Thus, we formulate two data sets, the output set Y and the lifted input set U, in the following way:

$$\begin{aligned} Y &= y(5k) \\ U &= [u(5k)\ \cdots\ u(5k+4)\ \cdots\ d_7(5k)\ \cdots\ d_7(5k+4)]. \end{aligned}$$

Then, we compute the CCA relations between Y and U, which show that only the manipulated variable and three out of the seven disturbances affect the output significantly; this is why only three disturbances were shown in Fig. 5. Hence, we estimate the model between the output and four inputs (the manipulated variable and three disturbances), giving us 20 inputs to the lifted system.

The lifted model is estimated; it is a lower-order model—a few time-delays at inputs combined with a first-order system. Based on this, we have extracted a fast sampled model, using the method discussed in Section III. Because the data set is too small, again we use the same data set to fit the estimated fast model. (Ideally another data set should be used to validate the model if there are enough data.) The comparison between the output measurement and the prediction of the model is shown in Fig. 6.

The modeling result is reasonable with variance of the prediction error being 0.04. The models can be further improved by collecting more data with more input excitations. As mentioned above, a data set of 205 output sample points is relatively small for estimating a model with 20 inputs. Figs. 4 and 5 also show that the changes in both the manipulated variable and disturbance variables are not rich enough—using all the 205 sampled points, we can compute the lifted input covariance matrix for the four inputs (u, d_1, d_2, and d_3), and then find that the condition number for this covariance matrix is 31 400; poor input excitation is justified. There are other unmeasured disturbance/noise in the data set as well. So, all these factors would affect the quality of the model obtained.

The inferential model was implemented into a QDMC control application in real time with a control interval of 30 min, using the algorithms developed in Section IV. The controller is dual-rate which makes optimal moves at a shorter period of 30 min while the output variable is measured every 2.5 h. When the slow sampled output measurement is available, it is used for bias or feedback correction. If not, the model gives an output estimate for control. After several days of online operation, the output variance is shown in Fig. 7 and compared with manual/operator control in the past, where the x-axis is in terms of number of samples and the sampling interval is 2.5 h. (We stress that Fig. 7 is based on real-time operation data.) It can be seen that the controller has reduced the output variance by 40%. Therefore, significant economic benefits have been achieved.

VII. Conclusion

In many industrial applications, some key quality control variables are measured infrequently due to various limitations on the sensor technology, whereas the control (manipulated) variables can be adjusted at relatively fast speed. The question arises: Based on such infrequent measurement, is closed-loop control possible with acceptable performance? We have demonstrated through an industrial case study involving a Shell Canada's CCR unit that closed-loop control at the fast control rate is possible. The solution we proposed is an inferential control scheme, making use of fast-rate models, which can be estimated based on dual-rate input–output data. In the CCR application, the inferential controller was implemented in real time with significant benefits in terms of reduced quality variance.

Future improvement for this industrial application includes obtaining a better model for the process, which requires longer data set, sufficiently excited input and disturbances, and so on.

Acknowledgment

The authors would like to thank the anonymous reviewers and the Associate Editor, Prof. F. J. Doyle, III, for their constructive comments and suggestions, which helped to improve the paper.

References

[1] T. Chen and B. Francis, *Optimal Sampled-Data Control Systems*. London, U.K.: Springer-Verlag, 1995.

[2] C. R. Cutler and B. L. Ramaker, "Dynamic matrix control—A computer control algorithm," *JACC*, 1980.

[3] B. Friedland, "Sampled-data control systems containing periodically varying members," in *Proc. 1st IFAC Congr.*, 1960.

[4] R. D. Gudi, S. L. Shah, and M. R. Gray, "Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements," *Biotechnol. Bioeng.*, vol. 44, pp. 1271–1278, 1994.

[5] ——, "Adaptive multirate state and parameter estimation strategies with application to a bioreactor," *AICHE J.*, vol. 41, pp. 2451–2464, 1995.

[6] H. Hotelling, "Relations between two sets of variables," *Biometrika*, vol. 28, pp. 321–377, 1936.

[7] G. M. Kranc, "Input–output analysis of multirate feedback systems," *IRE Trans. Automat. Contr.*, vol. 3, pp. 21–28, Jan. 1957.

[8] J. H. Lee and M. Morari, "Robust inferential control of multi-rate sampled-data systems," *Chem. Eng. Sci.*, vol. 47, pp. 865–885, 1992.

[9] D. Li, S. L. Shah, and T. Chen, "Identification of fast-rate models from multirate data," *Int. J. Contr.*, vol. 74, no. 7, pp. 680–689, 2001.

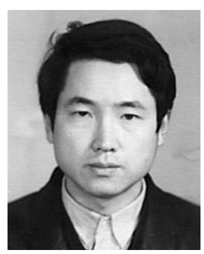
[10] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[11] W. P. Lu and D. G. Fisher, "Output estimation with multi-rate sampling," *Int. J. Contr.*, vol. 48, pp. 149–160, 1988.

[12] M. Ohshima, I. Hashimoto, M. Takeda, T. Yoneyama, and F. Goto, "Multi-rate multivariable model predictive control and its application to a semi-commercial polymerization reactor," in *Proc. ACC*, vol. 2, 1992, pp. 1576–1581.

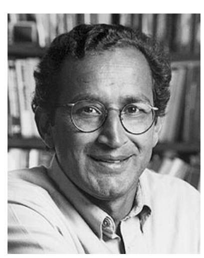
[13] T. Soderstrom and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[14] M. Verhaegen and X. Yu, "A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems," *Automatica*, vol. 32, pp. 201–216, 1995.



Dongguang Li received the B.Sc. and M.Sc. degrees in 1992 and 1995, respectively, from the Department of Automation, University of Science and Technology of China, and the Ph.D. degree in process control in chemical engineering from the University of Alberta, Edmonton, AB, Canada, in 2001.

He is now working as an advanced process control engineer with Honeywell, Calgary, AB, Canada. The main areas of his current research are model-based predictive control and system identification of multi-rate systems.



Sirish L. Shah (S'73–M'76) received the B.Sc. degree in control engineering from Leeds University Leeds, U.K., in 1971, a M.Sc. degree in automatic control from the University of Manchester Institute of Science and Technology, Manchester, U.K., in 1972 and the Ph.D. degree in process control in chemical engineering from the University of Alberta, Edmonton, AB, Canada, in 1976.

During 1977, he worked as a Computer Applications Engineer at Esso Chemicals in Sarnia, ON, Canada. Since 1978, he has been with the University of Alberta, where currently he is Professor of Chemical Engineering and the NSERC-Matrikon-ASRA Senior Industrial Research Chair in Computer Process Control. He has published extensively in academic journals and conference proceedings and recently coauthored a textbook, *Performance Assessment of Control Loops: Theory and Applications*.

Dr. Shah was awarded a SERC fellowship at Oxford University, Oxford, U.K., and appointed Visiting Fellow of Balliol College, Oxford, U.K., in 1985–1986. In 1989, he was the recipient of the Albright and Wilson Americas Award of the Canadian Society for Chemical Engineering in recognition of distinguished contributions to chemical engineering. He held the 1994 senior research fellowship of the Japan Society for the Promotion of Science (JSPS). He has also been a consultant with a number of different industrial organizations.

Tongwen Chen (S'86–M'91–SM'97) received the B.Sc. degree from Tsinghua University, Beijing, China, in 1984 and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 1988 and 1991, respectively, all in electrical engineering.

From 1991 to 1997, he was on faculty in the Department of Electrical and Computer Engineering at the University of Calgary, Canada. Since May 1997, he has been with the Department of Electrical and Computer Engineering at the University of Alberta, Edmonton, Canada, and is presently a Professor of Electrical Engineering. He held visiting positions at the Hong Kong University of Science and Technology and Tsinghua University. His current research interests include process control, multirate systems, robust control, digital signal processing, and their applications to industrial problems. He coauthored with B. A. Francis the book *Optimal Sampled-Data Control Systems* (New York: Springer-Verlag, 1995).

Dr. Chen received a University of Alberta McCalla Professorship for 2000/2001. He was an Associate Editor for IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 1998 to 2000. Currently, he is an Associate Editor for *Systems and Control Letters* and a member of the Editorial Board of *Dynamics of Continuous, Discrete and Impulsive Systems—Series B: Applications and Algorithms*. He is a registered Professional Engineer in Alberta, Canada.

Kent Z. Qi received the B.Sc. and M.Sc. degrees from Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in process control from the University of Alberta, Edmonton, Canada.

He is a Senior Advanced Process Control Engineer at Shell Canada Scotford refinery. He has over ten years experiences on DCS configuration, control software design, and advanced control applications. He has been very active on developing new control applications and applying new research results to the refinery.