# Multivariate Controller Performance Analysis: Methods, Applications and Challenges

Sirish L. Shah*, Rohit Patwardhan† and Biao Huang
Department of Chemical and Materials Engineering
University of Alberta
Edmonton, Canada T6G 2G6

**Abstract**

This paper provides a tutorial introduction to the role of the time-delay or the interactor matrix in multivariate minimum variance control. Minimum variance control gives the lowest achievable output variance and thus serves as a useful benchmark for performance assessment. One of the major drawbacks of the multivariate minimum variance benchmark is the need for *a priori* knowledge of the multivariate time-delay matrix. A graphical method of multivariate performance assessment known as the Normalized Multivariate Impulse Response (NMIR), *that does not require knowledge of the interactor*, is proposed in this paper. The use of NMIR as a performance assessment tool is illustrated by application to two multivariate controllers. Two additional performance benchmarks are introduced as alternatives to the minimum variance benchmark, and their application is illustrated by a simulated example. A detailed performance evaluation of an industrial MPC controller is presented. The diagnosis steps in identifying the cause of poor performance, e.g. as due to model-plant mismatch, are illustrated on the same industrial case study.

**Keywords**

Multivariate minimum variance control, Time delay, Normalized multivariate impulse response, Model predictive control, Model-plant mismatch

## Introduction

The area of performance assessment is concerned with the analysis of existing controllers. Performance assessment aims at evaluating controller performance from routine data. The field of controller performance assessment stems from the need for optimal operation of process units and from the need of getting value from immense volumes of archived process data. The field has matured to the point where several commercial algorithms and/or vendor services are available for process performance auditing or monitoring.

Conventionally the performance estimation procedure involves comparison of the existing controller with a theoretical benchmark such as the minimum variance controller (MVC). Harris (1989) and co-workers (1992; 1993) laid the theoretical foundations for performance assessment of single loop controllers from routine operating data. Time series analysis of the output error was used to determine the minimum variance control for the process. A comparison of the output variance term with the minimum achievable variance reveals how well the controller is doing currently. Subsequently Huang et al. (1996; 1997) and Harris et al. (1996) extended this idea to the multivariate case. In contrast to the minimum variance benchmark, Kozub and Garcia (1993), Kozub (1997) and Swanda and Seborg (1999) have proposed user defined benchmarks based on settling times, rise times, etc. Their work presents a more practical method of assessing controller performance. A suitable reference settling time or rise time for a process can often be chosen based on process knowledge.

The increasing acceptance of the idea of process and performance monitoring has also grown from the awareness that control software, and therefore the applications that arise from it, should be treated as capital assets and thus maintained, monitored and revisited routinely. Routine monitoring of controller performance ensures optimal operation of regulatory control layers and the higher level advanced process control (APC) applications. Model predictive control (MPC) is currently the main vehicle for implementing the higher level APC layer. The APC algorithms include a class of model based controllers which compute future control actions by minimizing a performance objective function over a finite prediction horizon. This family of controllers is truly multivariate in nature and has the ability to run the process close to its limits. It is for the above reasons that MPC has been widely accepted by the process industry. Various commercial versions of MPC have become the norm in industry for processes where interactions are of foremost importance and constraints have to be taken into account. Most commercial MPC controllers also include a linear programming stage that deals with steady-state optimization and constraint management. A schematic of a mature and advanced process control platform is shown in Figure 1. It is important to note that the bottom regulatory layer consisting mainly of PID loops forms the typical foundation of such a platform followed by the MPC layer. If the bottom layer does not perform and is not maintained regularly then it is futile to implement advanced control. In the same vein, if the MPC layer does not perform then the benefits of the higher level optimization layer, that may include real-time optimization, will not accrue.

The main contribution of this paper is in its general-

*sirish.shah@ualberta.ca
†Matrikon Consulting Inc., Suite 1800, 10405 Jasper Avenue, Edmonton, Canada , T5J 3N4

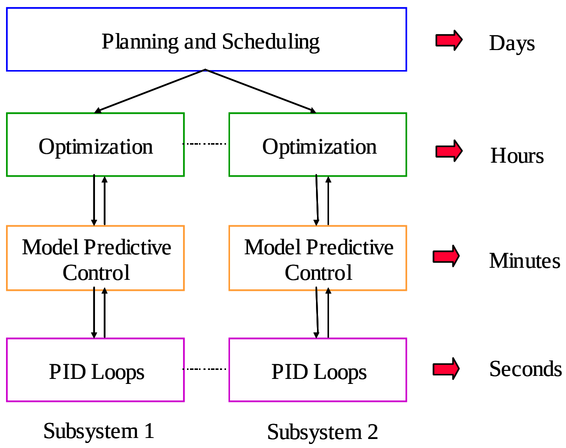


**Figure 1:** The control hierarchy.

ization of the univariate impulse response (between the process output and the whitened disturbance variable) plot to the multivariate case as the 'Normalized Multivariate Impulse Response' plot. A particular form of this plot, that does not require knowledge of the process time-delay matrix, is proposed here. Such a plot provides a graphical measure of the multivariate controller performance in terms of settling time, decay rates etc. Simple time and frequency domain measures such as multivariate autocorrelation and spectral plots are used to illustrate the interactions arising in multivariable control systems. Two relatively new multivariate performance evaluation ideas are also explored in detail: (1) the use of LQG as a benchmark based on the knowledge of the open loop process and noise models for the soft-constrained performance assessment problem (Huang and Shah, 1999) and (2) the use of the design performance as a benchmark (Patwardhan, 1999; Patwardhan et al., 2001). Both of these benchmarks can be applied to any type of controller. The LQG benchmark applies to all class of linear controllers, *irrespective of the controller objective function*, and is of use when input and/or output variance is of concern. The LQG benchmark represents the 'limits of performance' for a linear system, is more general and has the minimum variance as a special case. However, it needs a model of the linear process. The design objective function based approach can be applied to constrained MPC type controllers and is therefore a practical measure. However, it does not tell you how close the performance is relative to the lowest achievable limits. Issues related to the diagnosis of poor performance are discussed in the context of MPC controllers. Performance assessment of the general MPC is as yet an unresolved issue and presents a challenging research problem. A constrained MPC type controller is essentially a nonlinear controller, especially when operating at the constraints. Conventional MVC or linear controller benchmarking is infeasible and alternative techniques have to be developed. The development of new MPC performance monitoring tools thus represents an area of future challenges. The challenges associated with MPC performance evaluation are illustrated by considering an industrial case study of a $7 \times 6$ problem.

This paper is organized as follows. The next section provides a tutorial introduction to the concept of the time-delay matrix or the interactor. This is an important entity, particularly if one wants to evaluate MPC performance using multivariate minimum variance as a benchmark. The following two sections, respectively, discuss the tools required in the analysis of multivariate control loops such as the normalized multivariate impulse response, and alternative benchmarks for multivariate performance assessment. Applications are used to demonstrate the proposed techniques in each section. A discussion on the challenges in performance analysis and diagnosis and issues in MPC performance evaluation are outlined in the penultimate section, followed by a detailed industrial case study of an industrial MPC evaluation.

## The Role of Delays for Univariate and Multivariate Processes

Time delays play a crucial role in performance assessment particularly when the minimum variance benchmark is used. The concept of the multivariate delay is explained below in a tutorial manner by first defining the univariate delay term and then generalizing this notion to the multivariate case.

### Definition of a Delay Term for a Univariate Process:

The time-delay element in a univariate case is characterized by several different properties. For example, it represents the order of the first, non-zero (or *non-singular*) impulse response coefficient (also characterized by the number of infinite zeros of the numerator portion of the transfer function). It is important to fully understand the definition of a delay term for the univariate case in order to generalize the notion to a multivariate system. From a system theoretic point, the delay for a univariate system is characterized by the properties listed below. Consider a plant with the discrete transfer function or an impulse response model given by:

$$\begin{aligned} G(q^{-1}) &= \frac{q^{-d}B(q^{-1})}{A(q^{-1})} \\ &= 0q^{-1} + 0q^{-1} + \cdots + 0q^{-d+1} \\ &\quad + h_d q^{-d} + h_{d+1}q^{-d-1} + h_{d+2}q^{-d-2} \\ &\quad + \cdots \end{aligned}$$

The delay term for such a univariate system is defined by:

- the minimum integer $r$ such that

$$\lim_{q^{-1}\to 0} q^r \left(\frac{q^{-d}B(q^{-1})}{A(q^{-1})}\right) = k \quad \neq 0$$

(i.e. a non-singular coefficient) which in the case considered above, for $r = d$ gives:

$$\lim_{q^{-1}\to 0} q^r \left(\frac{q^{-d}B(q^{-1})}{A(q^{-1})}\right) = h_d \quad \neq 0$$

Note that for the univariate case, the number of infinite zeros of the process as obtained by setting the numerator of the process transfer function to zero, i.e. $q^{-d}B(q^{-1}) = 0$, also yields $d$ infinite zeros and $n$ finite zeros given by the roots of $B(q^{-1}) = 0$.

- the static or steady-state value of the delay term should be equal to 1, i.e. at steady-state (when $q^{-1} = 1$), $q^{-d} = 1$.

**Definition of a Delay Matrix for a Multivariate Process:**

Analogous to the univariate case, it is possible to factorize the open-loop transfer matrix into two elements: the delay matrix, $D(q^{-1})^{-1}$ containing all the infinite zeros of the system, and the 'delay-free' transfer-function matrix, $T^*(q^{-1})$, containing all the finite zeros and poles.

$$\begin{aligned} T(q^{-1}) &= D(q^{-1})^{-1} \cdot T^*(q^{-1}) \\ &= H_1 q^{-1} + H_2 q^{-2} + \cdots \\ &\quad + H_d q^{-d} + H_{d+1} q^{-d-1} + \cdots \end{aligned}$$

where $H_i$ are the impulse response or Markov matrices of the system parallel to the definition of the univariate delay, the multivariate delay matrix is defined by:

- Fewest number of linear combinations of the impulse response matrices that give a nonsingular matrix, i.e. (a finite and *nonsingular* matrix)

$$\lim_{q^{-1}\to 0} D(q^{-1})(D(q^{-1})^{-1} \cdot T^*(q^{-1})) = K$$

(a finite and nonsingular matrix)

Unlike the univariate case, a nonzero $H_i$ may not necessarily indicate the delay order. Instead, for the multivariate case, it is the fewest linear combination of such non-zero $H_i$ to give a non-singluar $K$ that defines the delay matrix, $D(q^{-1})^{-1}$. Applying this idea to the univariate case will reveal that $K = h_d$, which is the first or leading non-zero coefficient of the impulse response or the Markov parameter of the scalar system. Such an interpretation makes the choice of $D(q^{-1})^{-1}$, as a multivariate generalization of the univariate delay term, a very meaningful one. Note that $\det(D(q)) = cq^m$, where $m$ is the number of infinite zeros of the system and $c$ is a constant. For the multivariate case the number of infinite zeros may not be related to the order $d$ of the time-delay matrix.

- $D^T(q^{-1})D(q) = I$ (As compared to the univariate case where $q^{-d}q = 1$). This is known as the unitary interactor matrix. This unitary property preserves the spectrum of the signal, which leads us to the result that the variance of the actual output and the interactor filtered outputs are the same, i.e. $E(Y_t^T Y_t) = E(\tilde{Y}_t^T \tilde{Y}_t)$, where $\tilde{Y}_t = q^{-d} D Y_t$ (see Huang and Shah, 1999).

**Example.** Consider the following transfer function matrix and its impulse response or Markov parameter model:

$$\begin{aligned} T(q^{-1}) &= \begin{bmatrix} \frac{q^{-2}}{1-q^{-1}} & \frac{q^{-3}}{1-2q^{-1}} \\ \frac{q^{-2}}{1-3q^{-1}} & \frac{q^{-3}}{1-4q^{-1}} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} q^{-1} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} q^{-2} \\ &\quad + \begin{bmatrix} 1 & 1 \\ .33 & 1 \end{bmatrix} q^{-3} + \begin{bmatrix} 1 & 0.5 \\ 0.109 & 0.25 \end{bmatrix} q^{-4} \\ &\quad + \cdots \end{aligned}$$

Note that even though $H_2 \neq 0$, a linear combination of $H_1$ and $H_2$ does not yield a nonsingular matrix. In this example, at least three impulse response matrices are required to define the delay matrix for this system. The delay matrix that satisfies the properties listed above, is given by:

$$D(q^{-1})^{-1} = \begin{bmatrix} -0.707q^{-2} & -0.707q^{-3} \\ -0.707q^{-2} & 0.707q^{-3} \end{bmatrix}$$

The order of the delay is 3, i.e. a linear combination of at least 3 impulse response matrices is required to have a non-singular $K$.

**Remark 1.** The interactor matrix $D(q)$ can be one of the three forms as described in the sequel. If $D(q)$ is of the form: $D(q) = q^d I$, then the process is regarded as having a simple interactor matrix. If $D(q)$ is a diagonal matrix, i.e., $D(q) = diag(q^{d_1}, q^{d_2}, \ldots, q^{d_n})$, then the process is regarded as having a diagonal interactor matrix. Otherwise the process is considered to have a general interactor matrix.

To factor the general interactor matrix, *one needs to have a complete knowledge of the process transfer function or at least the first few Markov matrices of the multivariate system.* This is currently the main drawback in using this procedure. Huang et al. (1997) have provided a closed loop identification algorithm to estimate the first few Markov parameters of the multivariate system and thus compute the unitary interactor matrix. However, this rank determination procedure is prone to errors as it requires one to check if a linear combination of matrices is of full rank or not. Ko and Edgar (2000) have also proposed the use of the first few Markov matrices for multivariate performance assessment based on the minimum variance benchmark. The factorization of the diagonal interactor matrix requires only time delays between each pair of the input and output variables. A diagonal interactor matrix by no means implies that the process has a diagonal transfer function matrix or that the process has no interaction. But the converse is true, i.e. a diagonal process transfer function matrix or a system with weakly interacting multivariate loops (a diagonally dominant system) does have a diagonal interactor matrix. In fact, experience has shown that many actual multivariable processes have the structure of the diagonal interactor, provided the input-output structuring has been done with proper engineering insight. This fact greatly simplifies performance assessment of the multivariate system.

## The Multivariate Minimum Variance Benchmark

In a univariate system, the first $d$ impulse response coefficients of the closed loop transfer function between the control error and the white noise disturbance term determine the minimum variance or the lowest achievable performance. In the same way, the first $d$ impulse response matrices of the closed loop multivariate system are useful in determining the multivariate minimum variance benchmark, where $d$ denotes the order of the interactor.

Performance assessment of univariate control loops is carried out, by comparing the actual output variance with the minimum achievable variance. The latter term is estimated by simple time series analysis of routine closed-loop operating data and knowledge of the process time delay. The estimation of the univariate minimum variance benchmark requires filtering and correlation analysis. This idea has been extended to multivariate control loop performance assessment and the multivariate filtering and correlation (FCOR) analysis algorithm has been developed as a natural extension of the univariate case (Huang et al., 1996, 1997; Huang and Shah, 1999). Harris et al. (1996) have also proposed a multivariate extension to their univariate performance assessment algorithm. Their extension requires a spectral factorization routine to compute the delay free part of the multivariate process and thus estimate the multivariate minimum variance or the lowest achievable variance for the process. The FCOR algortihm of Huang et al. (1996), on the other hand, is a term for term generalization of the univariate case to the multivariate case and also requires the knowledge of the multivariate time-delay or interactor matrix. Figure 2 summarizes the steps required in computing the mutivariate performance index. A quadratic measure of multivariate control loop performance is defined as:

$$J = E(Y_t - Y_t^{sp})^T(Y_t - Y_t^{sp})$$

(where $Y_t$ represnts an $n$ dimensional output vector).The lower bound or the quadratic measure of the multivariate control performance under minimum variance control is defined as

$$J_{\min} = E(Y_t - Y_t^{sp})^T(Y_t - Y_t^{sp})|_{mv}$$

It has been shown by Huang et al. (1997) that the lower bound of the performance measure $J_{\min}$can be estimated from routine operating data. In Huang et al. (1997), the multivariate performance index is defined as

$$\eta = \frac{J_{\min}}{J}$$

and is bounded by $0 \le \eta \le 1$. In practice, one may also be interested in knowing how each individual output (loop) of the multivariate system performs relative to multivariate minimum variance control. Performance indices of each individual output are defined as

$$\begin{bmatrix} \eta_{Y_1} \\ \vdots \\ \eta_{Y_n} \end{bmatrix} = \begin{bmatrix} \min(\sigma_{y_1}^2)/\sigma_{y_1}^2 \\ \vdots \\ \min(\sigma_{y_n}^2)/\sigma_{y_n}^2 \end{bmatrix} = diag(\tilde{\Sigma}_{mv}\tilde{\Sigma}_Y^{-1})$$

where $\tilde{\Sigma}_{mv} = diag(\Sigma_{mv})$ and $\tilde{\Sigma}_Y = diag(\Sigma_Y)$; $\Sigma_Y$ is the variance matrix of the output $Y_t$ and $\Sigma_{mv} = \min(\Sigma_Y)$ is the covariance matrix of the output $Y_t$ under multivariate minimum variance control. It has been shown by Huang et al. (1997) that $\Sigma_{mv}$ can also be estimated using routine operating data, and knowledge of the interactor matrix.

To summarize, a multivariate performance index is a single scalar measure of multivariate control loop performance relative to multivariate minimum variance control. Individual output performance indices indicate performance of each output relative to the loop's performance under multivariate minimum variance control. If a particular output index is smaller than other output indices, then some of the other loops may have to be de-tuned in order to improve this poorly tuned loop.

## Alternative Methods for Performance Analysis of Multivariate Control Systems

### Autocorrelation Function:

The autocorrelation function (ACF) plots may be used to analyze individual process variable performance. A typical example of the ACF plots for the two output variables of the simulated Wood-Berry (Wood and Berry, 1973) column control system is shown in Figure 3. A decentralized control system comprising two PI controllers was

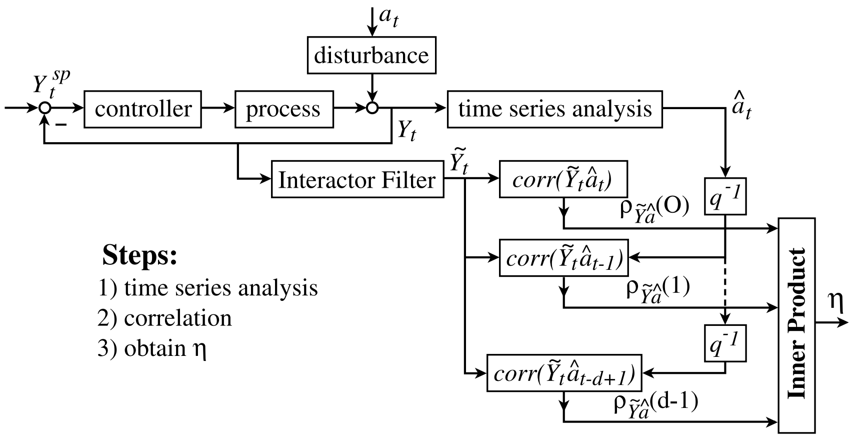


**Figure 2:** Schematic diagram of the multivariate FCOR algorithm.

used on the Wood-Berry column. The diagonal plots are autocorrelations of each output variable, while the off-diagonal plots are cross-correlation plots. The diagonal plots typically indicate how well each loop is tuned. For example, a slowly decaying autocorrelation function implies an under-tuned loop, and an oscillatory ACF typically implies an over-tuned loop. Off-diagonal plots can be used to trace the source of disturbance or the interaction between each process variables. Figure 3 clearly indicates that the first loop has relatively poor performance while the second loop has very fast decay dynamics and thus good performance. Interaction between the two loops can also be observed from the off-diagonal subplots. Note that the autocorrelation plot of the multivariate system is not necessarily symmetric.

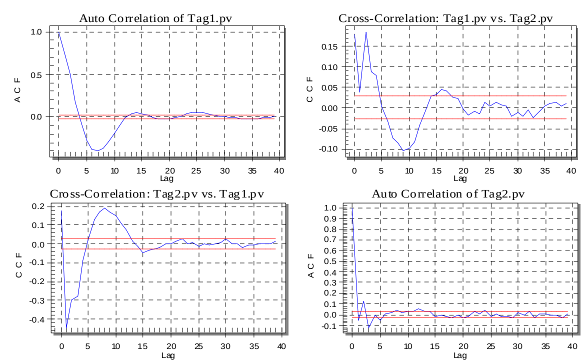


**Figure 3:** Autocorrelation function of multivariate process.

**Spectral Analysis:**

Frequency domain plots provide alternative indicators of control loop performance. They may be used to assess individual output dynamic behavior, interactions and effects of disturbances. For example, peaks in the diagonal plots typically imply oscillation of the variables due to an over-tuned controller or presence of oscillatory disturbances. Frequency domain plots also provide information on the frequency ranges over which the oscillations occur and the amplitude of the oscillations. Like time domain analysis, off-diagonal plots provide one with information on the correlation or interaction between the loops. Figure 4 is the power spectrum and the cross-power spectrum plot of the simulated Wood-Berry column. The first diagonal plot indicates that there is a clear mid-frequency harmonic in the 1$^{st}$ output. This could be due to an overtuned controller. Off-diagonal plots show a peak in the cross-spectrum at the same frequency. The poor performance of loop 1 can then be attributed to significant interaction effects from loop 2 to loop 1. In other words, the satisfactory or good performance of loop 2 could be at the expense of transmitting disturbances or upsets to loop 1 via the interaction term.. The power spectrum plots of a multivariate system are symmetric.

**Normalized Multivariate Impulse Response (NMIR) Curve as an Alternative Measure of Performance:**

As shown in Figure 2, the evaluation of the multivariate controller performance has to be undertaken on the interactor filtered output and not on the actual output. The reason for this is that the interactor filtered output vector, $\tilde{Y}_t = q^{-d}DY_t$, is a special linear combination of the actual output, lagged or otherwise, and this fictitious output preserves the spectral property of the system and facilitates simpler analysis of the multivariate minimum variance benchmark.

This new output ensures, as in the univariate case, that the closed loop output can be easily factored into two terms, a controller or feedback-invariant term and a second term that depends on the controller parameters. In the ensuing discussion, we consider an alter-

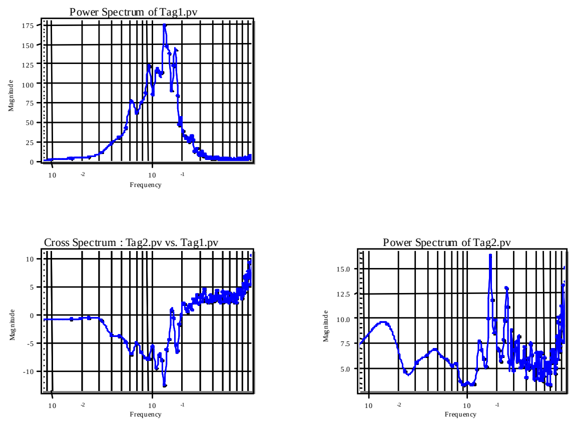


**Figure 4:** Frequencey response of multivariate process.

native graphical measure of multivariate performance as obtained from the interactor filtered output. So unless specified otherwise, the reader should assume that the operations elucidated below are on the interactor filtered output, $Y_t$.

An impulse response curve represents dynamic relationship between the whitened disturbance and the process output. This curve typically reflects how well the controller regulates stochastic disturbances. In the univariate case, the first $d$ impulse response coefficients are feedback controller invariant, where $d$ is the process time-delay. Therefore, if the loop is under minimum variance control, the impulse response coefficients should be zero after $d-1$ lags. The Normalized Multivariate Impulse Response (NMIR) curve reflects this idea. The first $d$ NMIR coefficients are feedback controller invariant, where $d$ is the order of the time-delay matrix or the interactor. If the loop is under multivariate minimum variance control, then the NMIR coefficients should decay to zero after $d-1$ lags. The sum of squares under the NMIR curve is equivalent to the trace of the covariance matrix of the data. In fact the NMIR is a graphical representation of the quadratic measure of the output variance as given by:

$$\begin{aligned} E(Y_t^T Y_t) &= E(\tilde{Y}_t^T \tilde{Y}_t) \\ &= tr(F_0 \Sigma_a F_0^T) + tr(F_1 \Sigma_a F_1^T) + \cdots \end{aligned}$$

where

$$\tilde{Y}_t = F_0 a_t + F_1 a_{t-1} + \cdots + F_{d-1} a_{t-d+1} + F_d a_{t-d} + \cdots$$

is an infinite series impulse response model of the interactor filtered output with respect to the whitened disturbance and matrices $F_i$ represent the estimated Markov matrices of the closed loop multivariate system. In the new measure, the first NMIR coefficient is given by $\sqrt{tr(F_0 \Sigma_a F_0^T)}$, the second NMIR coefficient is given by $\sqrt{tr(F_1 \Sigma_a F_1^T)}$, and so on. The multivariate performance index is then equal to the ratio of the sum of the squares of the first $d$ NMIR coefficients to the sum of squares of all NMIR coefficients (see top plot in Figure 5). Care has to be taken when interpreting the normalized impulse response curve. The NMIR represents a compressed scalar metric for a multi-dimensional system. It is a graphical representation of the weighted 2-norm multivariate impulse response matrix and provides a graphical interpretation of the multivariate performance index in much the same way as the univariate impulse response gives an indication of the level of damping afforded to a unit impulse disturbance.

The NMIR as outlined above and first described by Huang and Shah (1999) requires *a priori* knowledge of the interactor matrix. Since this NMIR curve is suitable for obtaining a graphical measure of the overall closed-loop response, we suggest an alternative measure that does not require knowledge of the interactor. We propose to use a similar normalized multivariate impulse curve without interactor filtering to serve a similar purpose. The NMIR *without interactor filtering* is calculated as before by computing the correlation coefficients between the pre-whitened disturbance and the actual output with lags $0, 1, 2, \ldots, d-1, d, d+1, \ldots$.

$$E(Y_t^T Y_t) = tr(E_0 \Sigma_a E_0^T) + tr(E_1 \Sigma_a E_1^T) + \cdots$$

where

$$Y_t = E_0 a_t + E_1 a_{t-1} + \cdots + E_{d-1} a_{t-d+1} + E_d a_{t-d} + \cdots$$

Note that unlike the original NMIR measure as proposed by Huang and Shah (1999), the new measure proposed here does *not* require interactor filtering of the output, i.e. an explicit knowledge of the interactor is not required in computing the new graphical and qualitative measure. From here onwards this new measure is denoted as NMIR$_{wof}$.

In the new measure, the first NMIR$_{wof}$ coefficient is given by $\sqrt{tr(E_0 \Sigma_a E_0^T)}$, the second NMIR$_{wof}$ coefficient is given by $\sqrt{tr(E_1 \Sigma_a E_1^T)}$, and so on. Note that the NMIR$_{wof}$ measure (without interactor filtering) is similar to NMIR with interactor filtering in the sense that both represent the closed-loop infinite series impulse response model of the output with respect to the whitened disturbance, one for the actual output and the other one for the interactor filtered output respectively. The newly proposed NMIR$_{wof}$ is physically interpretable, but does not have the property that the first $d$ coefficients are feedback control-invariant. The main rationale for using the newly proposed NMIR$_{wof}$ is that the following two terms are asymptotically equal:

$$\lim_{n \to \infty} \left\{ tr(E_0 \Sigma_a E_0^T) + tr(E_1 \Sigma_a E_1^T) + \cdots + tr(E_n \Sigma_a E_n^T) \right\} = \left\{ tr(F_0 \Sigma_a F_0^T) + tr(F_1 \Sigma_a F_1^T) + \cdots + tr(F_n \Sigma_a F_n^T) \right\}$$

This follows from the equality: $E(Y_t^T Y_t) = E(\tilde{Y}_t^T \tilde{Y}_t)$ (Huang and Shah, 1999). It is then clear that the NMIR curves with and without filtering will coincide with each other for $n$ sufficiently large. Alternately, the cumulative sum of the square of the impulse response coefficients can also be plotted and as per the above asymptotic equality, one would expect that the two terms will coincide for a sufficiently large $n$. These curves are reproduced here for the illustrative Wood-Berry column example. The ordinate in the bottom plot in Figure 5 gives the actual output variance when the curves converge for a sufficiently large $n$. Note that, unlike the NMIR curve with interactor filtering (solid line in Figure 5), the $\text{NMIR}_{wof}$ curve (dashed line in Figure 5) can *not* be used to calculate a numerical value of the performance index. However, it has the following important properties that are useful for assessment of multivariate processes:

1. The new $\text{NMIR}_{wof}$ represents the normalized impulse response from the white noise to the true output.
2. The new $\text{NMIR}_{wof}$ curve reflects the predictability of the disturbance in the original output. If the impulse response decays slowly, then this is clear indication of a highly predictable disturbance (e.g. an integrated white noise type disturbance) and relatively poor control. On the other hand a fast decaying impulse response is a clear graphical indication of a well-tuned multivariate system (Thornhill et al., 1999).
3. The new $\text{NMIR}_{wof}$ also provides a graphical measure of the overall multivariate system performance with information regarding settling time, oscillation, speed of response etc.

NMIRs with and without interactor filtering are calculated for the simulated Wood-Berry distillation column with two multiloop PI controllers. With sampling period 0.5 second, the interactor matrix of the process is found to have a diagonal structure and is given by

$$D(q) = \begin{bmatrix} q^3 & 0 \\ 0 & q^7 \end{bmatrix}$$

Since the order of the interactor is 7 sample units, the first 7 NMIR coefficients are feedback control invariant and depend solely on the disturbance dynamics and the interactor matrix. The sum of squares of these 7 coefficients is the variance achieved under multivariate minimum variance control. In fact as shown in Figure 6, the scalar multivariate measure of performance is equal to the sum of squares of the first 7 NMIR coefficients divided by the total sum of squares. Notice that for sufficiently large $n$ or prediction horizon, the two curves, as expected, coincide with each other. In this simulation example, observe that the NMIR and the $\text{NMIR}_{wof}$ curves decay to zero relatively quickly after 7 sample units, indicating relatively good control performance.

**Industrial MIMO Case study 1: Capacitance Drum Control Loops at Syncrude Canada Ltd.** Capacitance drum control loops of Plant AB in Syncrude Canada Ltd. were analyzed for this study. The primary objective of Plant AB is to further reduce the water in the (Plant A) diluted bitumen product prior to it reaching the next plant (Plant B) storage tanks. As the grade of the feed entering plant A reduces, the water required to process the oilsands increases proportionately. A large portion of this excess water ends up in the Plant A froth feed tank and ultimately increases the % volume of water in the Plant A product. Aside from degrading the quality of the product, the increased volume of water means reduction in the amount of bitumen that can be piped to the diluted bitumen tanks. In addition, the higher water content means more of the chloride compound present in the oilsands is dissolved and finds its way to the diluted bitumen tanks and eventually to plant B. The higher chloride concentration increases the corrosion rate of equipment in the Upgrading units. Plant AB was developed as a means of reducing the water content, and ultimately, the chlorides sent to Upgrading. This reduction is achieved by centrifuging the Plant A Product.

All product from plant A is directed to the Plant AB feed storage tank. The IPS portion of the product is routed through 5 Cuno Filters prior to entering the Plant AB feed storage tank. Feed from the feed storage tank is then pumped through the feed pumps to the Alfa Laval centrifuges. The Alfa Laval centrifuges remove water and a small amount of solids from the feed. Each centrifuge has its own capacitance drum and product back pressure valve. This arrangement allows for individual centrifuge "E-Line" control and a greatly improved product quality. Heavy phase water from plant A is used as Process Water in plant AB.

The cap drum pressure controller controls the capacitance drum pressure by adjusting the nitrogen flow into the drum. The Cap Drum Primary level controller maintains the cap drum water level by adjusting water addition into the drum. Control of these two variables is essential to maintain the E-Line in the centrifuges. Currently these two loops are controlled by multiloop PID controllers. The two process variables, pressure and level, are highly interacting. The objective of the performance assessment is to evaluate the existing multiloop PID controllers' performance, and to identify opportunities, if any, to improve performance by implementing a multivariate controller

*Discussion of Performance Analysis.* Process data with a 5-second sampling interval are shown in Figure 7 . These are typical (representative) process data encoun-

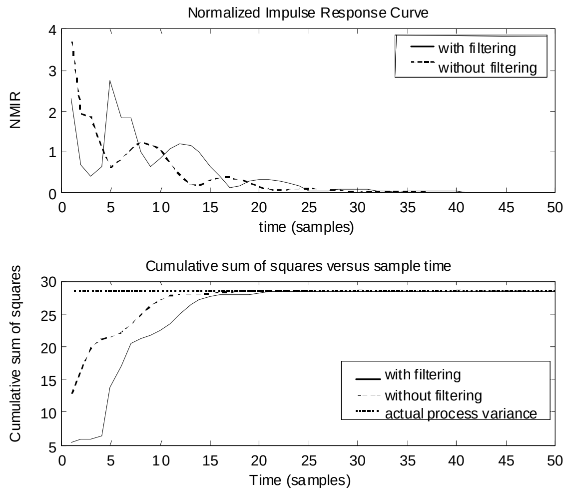


**Figure 5:** NMIR and NMIR$_{wof}$ curves (top plot) and the cumulative sum of squares plots of the impulse response coefficients (bottom plot).

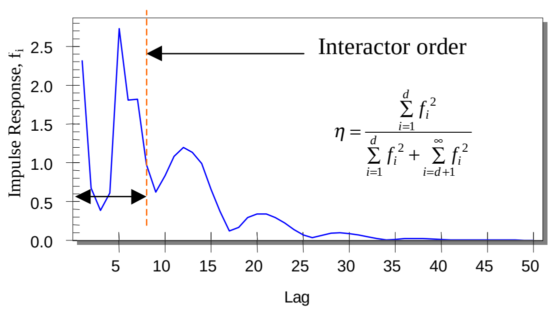


**Figure 6:** Normalized multivariate impulse response.

tered in this process. By assuming that both pressure and level loops have no time delay except for the delay induced by the zero-order-hold device, a scalar multivariate performance index was calculated as 0.022 and individual output indices are shown in Figure 8 . Based on these indices, one may conclude that controller performance is poor and may be improved significantly by re-tuning the existing controllers or re-designing a multivariate controller. However, since the exact time delays for these loops are unknown, further analysis of performance in both time domain and frequency domain is necessary. For example, the NMIR$_{wof}$ response shown in Figure 9 does indicate that the disturbance persists for about 50 samples before it is compensated by the feedback controllers. Overall the decay in the NMIR$_{wof}$ curve is rather slow indicating a predictable disturbance and generally ineffective regulatory control in dealing with such disturbances. This is equivalent to a settling time of 4 minutes for the overall system. To check which loop causes such long settling time, one can look at the auto- and cross-correlation plots.

The individual loop behavior can be observed from the auto and cross correlation plots shown in Figure 10 . It is observed that the pressure response does not settle down even after 40 samples. This is clearly unacceptable for a pressure loop. In addition some oscillatory response is observed in the level response as evident from the spectrum plot shown in Figure 11. Notice that the peak (oscillation) appears in both the pressure and level responses as well as in the cross spectrum plot. This indicates that both loops interact and oscillate at the same frequency. Thus, this analysis indicates that 1) the existing multiloop controller has relatively poor performance primarily due to the long settling time and oscillatory behavior or presence of oscillatory disturbances; 2) the two loops are strongly interacting and a multivariate controller may be able to improve performance significantly.

The final recommendation for this system was that performance of the two loops individually as well as a multivariate system is relatively poor. For predictable disturbances, there is insufficient integral action in the pressure loop resulting in a slowly decaying ACF plot as noticeable in the top left hand plot in Figure 10 . The performance is poor mainly due to interaction between the two loops. Because of the interaction, the multiloop retuning exercise may be futile. If however only a simple control solution is desired then the level loop can be detuned and the pressure loop can have larger gains with smaller integral action to reduce oscillations. If the system warrants, then a multivariate control loop could be designed.

As would be evident from the above discussion and case study, there remains much to be desired in obtaining practically meaningful measures of multivariate control performance. The minimum variance control is a useful benchmark as it requires little *a priori* information about the process. If however, more detailed performance measures are desired then, as would be expected, more process information is needed. For example, it would be desirable to include the control 'cost' or effort in the performance evaluation of a controller or answers to questions such as the following may be required: What is the best available control subject to soft constraints on the controller output variance. Two relatively new benchmarks are presented next as alternative measures of practical multivariate controller performance.

## LQG Benchmarking

Preliminary results on the LQG benchmark as an alternative to the minimum variance benchmark were proposed in Huang and Shah (1999). These results are reviewed here and a new benchmark that takes the control cost into account is proposed. The main advantage of the minimum variance benchmark is that other than the time-delay, it requires little process information. On the other hand if one requires more information on controller performance such as how much can the output variance

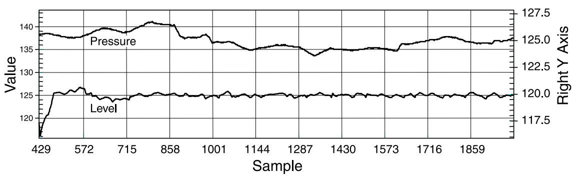


**Figure 7:** Pressure and level data with sampling interval 5 seconds.

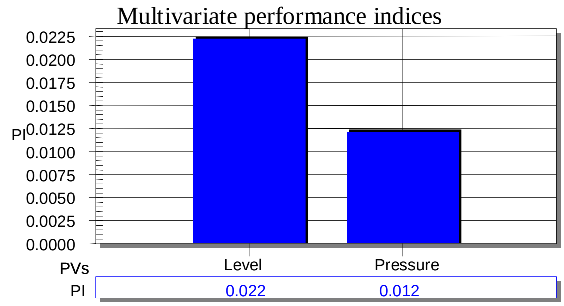


**Figure 8:** Individual output performance indices.

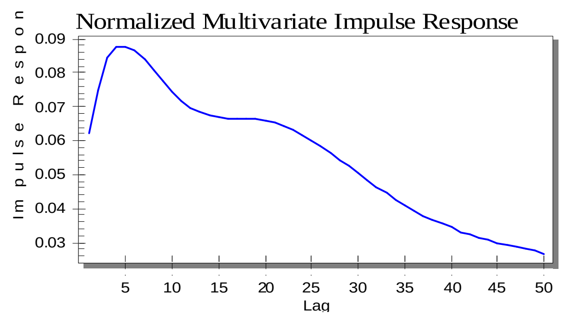


**Figure 9:** Normalized multivariate impulse response without interactor filtering.

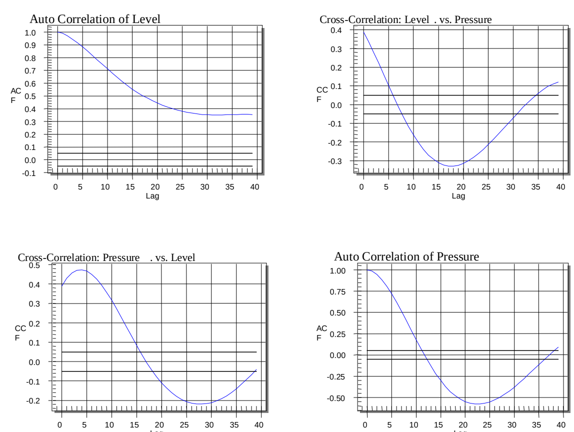


**Figure 10:** Auto and cross-correlation of process output.

be reduced without significantly affecting the controller output variance then one needs more information on the process. In short it is useful to have a benchmark that explicitly takes the control cost into account. The LQG cost function is one such benchmark. This benchmark does not require that an LQG controller be implemented for the given process. Rather the benchmark provides the 'limit of performance' for any linear controller in terms of the input and output variance. As remarked earlier, it is a general benchmark with the minimum variance as a special case. The only disadvantage is that the computation of the performance limit curve as shown in Figure 12 requires knowledge of the process model. For MPC type controllers these models may be readily available. Furthermore the benchmark cannot handle hard constraints, but it can be used to compare the performance of unconstrained and constrained controllers (see Figure 13).

In general, tighter quality specifications lead to smaller variations in the process output, but typically require more control effort. Consequently one may be interested in knowing how far away the control performance is from the "best" achievable performance with the same effort, i.e., in mathematical form the resolution of the following problem may be of interest:

> Given that $E(u^2) \leq \alpha$, what is the lowest achievable $E(y^2)$?

The solution is given by a tradeoff curve as shown in Figure 12. This curve can be obtained by solving the LQG problem (Kwakernaak and Sivan, 1972), where the LQG objective function is defined by:

$$J(\lambda) = E(y^2) + \lambda E(u^2)$$

By varying $\lambda$, various optimal solutions of $E(y^2)$ and $E(u^2)$ can be calculated. Thus a curve with the optimal output variance as ordinate, and the incremental manipulative variable variance as the abscissa can be plotted from these calculations. Boyd and Barratt (1991) have shown that any linear controller can only operate in the region above this curve. In this respect this curve defines the limit of performance of all linear controllers, as applied to a linear time-invariant process, including the minimum variance control law. If the process is modelled as an ARIMAX process then the resulting LQG benchmark curve due to the optimal controller will have an integrator built into it to asymptotically track and reject step type setpoints and disturbances respectively. Five optimal controllers may be identified from the tradeoff curve shown in Figure 12. They are explained as follows:

- Minimum cost control: This is an optimal controller identified at the left end of the tradeoff curve. The minimum cost controller is optimal in the sense that

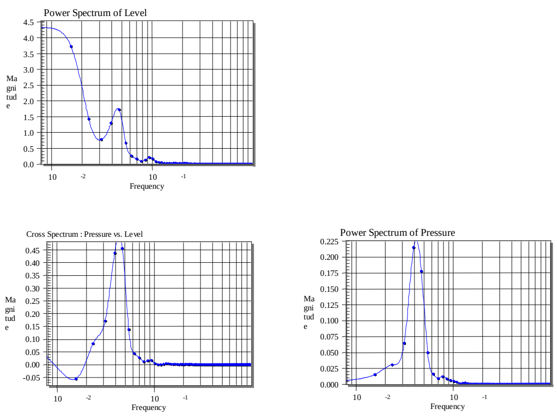


**Figure 11:** Frequency domain analysis of process output.

it offers an offset-free control performance with the minimum possible control effort. It is worthwhile pointing out that this controller is different from the open-loop mode since an integral action is guaranteed to exist in this controller.

- Least cost control: This optimal controller offers the same output error as the current or existing controller but with the least control effort. So if the output variance is acceptable but actuator variance has to be reduced then this represents the lowest achievable manipulative action variance for the given output variance.
- Least error control: This optimal controller offers least output error for the same control effort as the existing controller. If the input variance is acceptable but the output variance has to be reduced then this represents the lowest achievable output variance for the given input variance.
- Tradeoff controller: This optimal controller can be identified by drawing the shortest line to the tradeoff curve from the existing controller; the intersection is the tradeoff control. Clearly, this tradeoff controller has performance between the least cost control and the least error control. It offers a tradeoff between reductions of the output error and the control effort.
- Minimum error (variance) control: This is an optimal controller identified at the right end of the tradeoff curve. The minimum error controller is optimal in the sense that it offers the minimum possible error. Note that this controller may be different from the traditional minimum variance controller due to the existence of integral action.

The challenges with respect to the LQG benchmark lie in the estimation of a reasonably accurate process model. The uncertainty in the estimated model then has to be 'mapped' onto the LQG curve, in which case

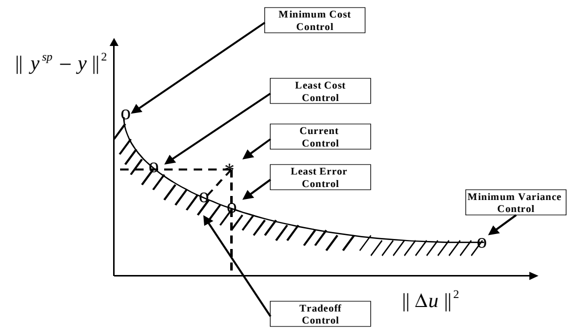


**Figure 12:** The LQG tradeoff curve with several optimal controllers.

it would become a fuzzy trade-off curve. Alternately the uncertainty region can be mapped into a region around the current performance of the controller relative to the LQG curve (see Patwardhan et al., 2000).

**An Alternative Method for Multivariate Performance Assessment Using the Design Case as a Benchmark**

An alternative approach is to evaluate the controller performance using a criterion commensurate with the actual design objective(s) of the controller and then compare the achieved performance. This idea is analogous to the method of Kammer et al. (1996), which was based on frequency domain comparison of the achieved and design objective functions for LQG. For a MPC controller with a quadratic objective function, the design requirements are quantified by:

$$\hat{J}(k) = \sum_{i=1}^{p} (y_{sp}(k+i|k) - \hat{y}(k+i|k))^T \Gamma_{k,i}(y_{sp}(k+i|k) - \hat{y}(k+i|k)) + \sum_{i=1}^{M-1} \Delta u(k+i-1)^T \Lambda \Delta u(k+i-1)$$

where

$\hat{y}(k+i|k)$ is the $i$-step ahead predictor of the outputs based on the process model

$y_{sp}(k+i|k)$ is the setpoint trajectory

$\Delta u(k+i-1)$ are the future moves of the inputs

$\Gamma_{k,i}$ are the output weightings that, in general, can depend upon the current time and the prediction horizon

Details of the model predictive control calculations can be found in any standard references (Garcia et al., 1989; Mayne et al., 2000; Qin and Badgwell, 1996). Here we restrict ourselves to the performance assessment aspects. The model predictive controller calculates the optimal control moves by minimizing this objective function over the feasible control moves. If we denote the optimal control moves by $\Delta u^*(k)$, the optimal value of the design objective function is given by

$$\hat{J}^*(k) = \hat{J}(\Delta u^*(k))$$

The actual output may differ significantly from the predicted output due to inadequacy of the model structure, nonlinearities, modeling uncertainty etc. Thus the achieved objective function is given by:

$$\hat{J}(k) = \sum_{i=1}^{p}(y_{sp}(k+i|k) - y(k+i|k))^T \Gamma_{k,i}(y_{sp}(k+i|k) - y(k+i|k)) + \sum_{i=1}^{M-1} \Delta u(k+i-1)^T \Lambda \Delta u(k+i-1)$$

where $y(k)$ and $\Delta u(k)$ denote the measured values of the outputs and inputs at corresponding sampling instants appropriately vectorized. The inputs will differ from the design value in part due to the receding horizon nature of the MPC control law. The value of the achieved objective function cannot be known *a priori*, but only $p$ sampling instants later. A simple measure of performance can then be obtained by taking a ratio of the design and the achieved objective functions as:

$$\eta(k) = \frac{\hat{J}^*(k)}{J(k)}$$

This performance index will be equal to one when the achieved performance meets the design requirements. The advantage of using the design criterion for the purpose of performance assessment is that it is a measure of the deviation of the controller performance from the expected or design performance. Thus a low performance index truly indicates changes in the process or the presence of disturbances, resulting in sub-optimal control. The estimation of such an index does not involve any time series analysis or identification. The design objective is calculated by the controller at every instant and only the measured input and output data is needed to find the achieved performance. The above performance measure represents an instantaneous measure of performance and can be driven by the unmeasured disturbances. In order to get a better overall picture the following measure is recommended:

$$\alpha k = \frac{\sum_{i=1}^{k} \hat{J}^*(i)}{\sum_{i=1}^{k} J(k)}$$

$\alpha(k)$ is the ratio of the average design performance to the average achieved performance up to the current sampling instant. Thus $\alpha(k) = 1$ implies that the design performance is being achieved on an average. $\alpha(k) < 1$ means that the achieved performance is worse than the design. This alternative metric of multivariate controller performance has been applied towards the evaluation of a QDMC and another MPC controller. Further details on the evaluation of this algorithm can be found in (Patwardhan, 1999).

The motivation for a lumped performance index is that the MPC controllers in the dynamic sense, attempt to minimize a lumped performance objective. The lumped objective function and subsequently the performance index, therefore does reflect the true intentions of the controller. The motivation for this idea was to have a performance statistic for MPC that is commensurate with its constrained and time-varying nature. The idea of comparing design with achieved performance has been common place in the area of control relevant identification (also known as iterative identification and control, identification for control)—see the survey by Van den Hof and Schrama (1995). Performance degradation is measured as a deviation from design performance and becomes the motivation for re-design/re-identification.

**Simulation Example: A Mixing Process.** The above approach was applied to a simulation example. The system under consideration is a $2 \times 2$ mixing process. The controlled variables are temperature ($y_1$) and water level ($y_2$) and the manipulated inputs are inlet hot water ($u_1$) and inlet cold water ($u_2$) flow rates. The following model is available in discrete form,

$$P(z^{-1}) = \begin{bmatrix} \frac{0.025z^{-1}}{1-0.8607z^{-1}} & \frac{-0.1602z^{-1}}{1-0.8607z^{-1}} \\ \frac{0.2043z^{-1}}{1-0.9827z^{-1}} & \frac{0.2839z^{-1}}{1-0.9827z^{-1}} \end{bmatrix}$$

A MPC controller was used to control this process in the presence of unmeasured disturbances. The controller design parameters were:

$$p = 10, \quad , m = 2, \quad , \lambda = diag([1, 4]), \quad \Gamma = diag([1, 2])$$

White noise sequences at the input and output with covariance equal to 0.1I served as the unmeasured disturbances. First the LQG benchmark was found, and the performance of a constrained and unconstrained MPC was evaluated against this benchmark (see Figure 13). Constraints on input moves were artificially imposed in order to activate the constraints frequently. The unconstrained controller showed better performance, compared to the constrained controller, with respect to the LQG benchmark.

A plot of the LQG objective function compared to the achieved objective function is shown in Figure 14. A performance measure of $J_{lq}/J_{ach} = 0.467/0.71 = 0.6579$ was

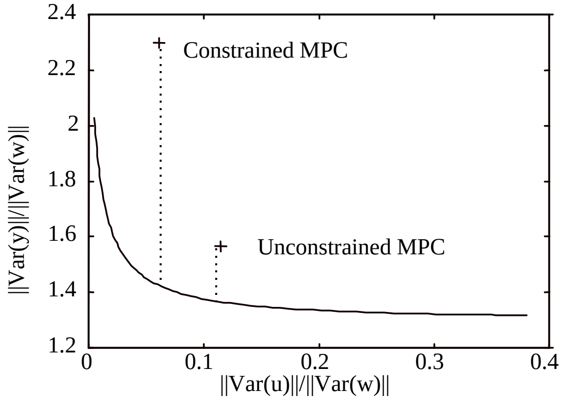


**Figure 13:** MPC performance assessment using LQG benchmark.

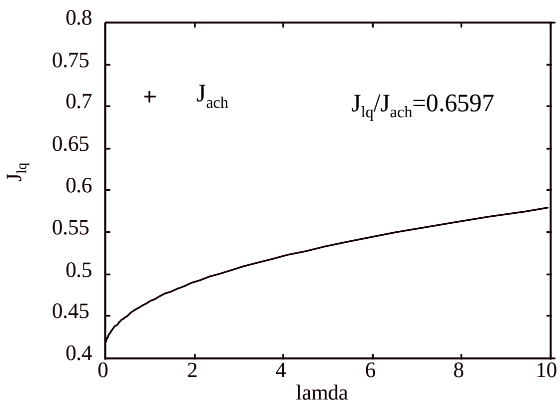


**Figure 14:** Comparison of the achieved performance with the LQG objective function.

| | LQG | $\alpha(k)$ |
|---|---|---|
| Unconstrained | 0.6579 | 0.8426 |
| Constrained | 0.4708 | 1.00 |

**Table 1:** Effect of constraints on MPC performance.

obtained for the unconstrained controller. Performance assessment of the same controller using the design case benchmarking approach yields contrasting results (Table 1). For the unconstrained controller a performance index 0.8426 revealed satisfactory performance while the imposition of constraints led to a performance index of 1. The constrained controller showed improvement according to one benchmark and deterioration with respect to the LQG benchmark. The design case approach indicates that the controller is doing its best under the given constraints while the LQG approach which is based on comparison with an unconstrained controller shows a degradation in performance.

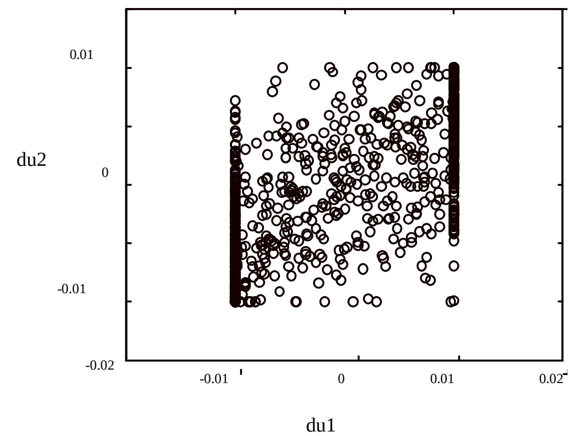


**Figure 15:** The input moves for the constrained controller during the regulatory run.

Figure 15 shows the input moves during the regulatory run for the constrained controller. The constraints are active for a large portion of the run and are limiting the performance of the controller in an absolute sense (LQG). On the other hand the controller cannot do any better due to design constraints as indicated by the design case benchmark.

## Challenges in Performance Analysis and Diagnosis: General Comments and Issues in MPC Performance Evaluation

A single index or metric by itself may not provide all the information required to diagnose the cause of poor performance. Considerable insight can be obtained by carefully interpreting all the performance indices. For example, in addition to the minimum variance benchmark performance index, one should also look at the cross-correlation plots, normalized multivariate impulse response plots, spectrum analysis, etc., to determine causes of poor performance. As an example, if the process data is 'white' then the performance index will always be close to 1 irrespective of how large the variance is. On the other hand, if the data is highly correlated (highly predictable), then the performance index will be low irrespective of how small the output variance is. In this respect the performance index plus the impulse response or the auto-correlation plot would provide a complete picture of the root cause of the problem. (The auto-correlation plot would have yielded information on the predictability of the disturbance). In summary then, each index has its merits and its limitations. Therefore, one should not just rely on any one specific index. It would be more appropriate to check all relevant indices that reflect performance measures from different aspects.

As mentioned earlier, the multivariate extension of the minimum variance benchmark requires a knowledge of the time-delay or the interactor *matrix*. This requirement of *a priori* information on the interactor has been regarded by many as impractical. However, from our experience this benchmark, when applied with care, can yield meaningful measures of controller performance. Yet, many outstanding issues remain open before one can confidently apply MIMO assessment techniques for a wide-class of MIMO systems. Some of the issues related to the evaluation of multivariate controllers are listed below:

1. To calculate a *general* interactor matrix, one needs to have more *a priori* information than just the time delays. However, experience has shown us that a significant number of MIMO processes do have the diagonal interactor structure. In fact, a properly designed MIMO control structure will most likely have a diagonal interactor structure (Huang and Shah, 1999). A diagonal interactor depends only on the time delays between the paired input and output variables.
2. Since models are available for all MPC based controllers, *a priori* knowledge of the time delay matrix is surely not an issue at all.
3. A more important issue in the analysis and diagnosis of control loops is the accuracy of the models and their variability over time. How does the model uncertainty affect the calculation of performance index? This question has not been answered so far. It is a common problem in both MIMO and SISO performance assessment. Therefore, one of the many outstanding issues remaining is the robustness of performance assessment, i.e. how to transfer the model uncertainty onto the uncertainty in the calculation of the performance index? This issue has been addressed to some extent in Patwardhan (1999; 2001), where SISO and MIMO examples, relating modeling uncertainty to uncertainty in performance measures, are given.

Industrial MPC is a combination of a dynamic part and a steady state part, which often comprises a linear programming (LP) step. The dynamic component consists of unconstrained minimization of a dynamic cost function, comprising of the predicted tracking errors and future input moves, familiar to academia. The steady state part focuses on obtaining economically optimal targets, which are then sent to the dynamic part for tracking as illustrated in Figure 16. This combination of the dynamic and steady state parts and constraint handling via the linear programming or the LP step renders the MPC system as a nonlinear multivariate system. Patwardhan et al. (1998) have illustrated the difficulties caused by the LP step on an industrial case study involving a demethanzier MPC (see Figure 17). In that particular

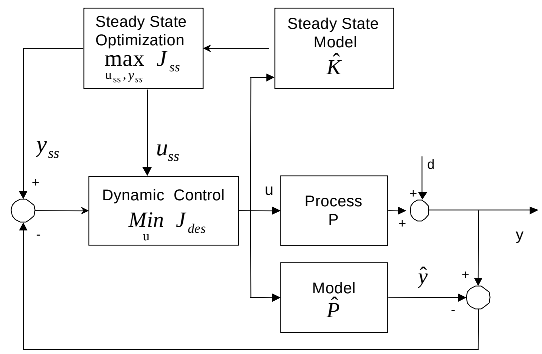


**Figure 16:** Schematic of a typical commercial MPC with a blended linear programming module that sets targets for the controlled and manipulative variables.

application as in a number of other MPC applications, when the LP stage is activated fairly routinely, significant correlation exists between the LP targets and the measurements that the LP relies upon. In such instances when the LP stage is activated at the same frequencies as the control frequencies, the controller structure is no longer linear. Situations such as these preclude the use of conventional performance assessment methods such the LQ or the minimum variance benchmark. We believe that the variable structure nature of industrial MPC can be captured by the objective function method since it takes into account the time varying nature of the MPC objective. Patwardhan (1999) has applied this method successfully on an industrial QDMC application. Even though one may argue that QDMC is devoid of the LP step, it is a variable structure MIMO controller that allows different inputs and outputs to swap into active and inactive states relative to the active constraint set. In this respect, the lumped objective function and subsequently the performance index proposed here does 'measure' the true intentions of the controller relative to the design case. It thus provides a useful performance metric. The only limitation being that the access to the actual design control objective has to be available in the MPC vendor software.

Establishing the root causes of performance degradation in industrial MPCs is indeed a challenging task. Potential factors include models, inadequately designed LP in that the LP operates at the control frequencies, inappropriate choice of weightings, ill-posed constraints, steady-state bias updates etc. In practice, these factors combine in varying degrees to give poor performance. Thus the issues and challenges related to the diagnosis aspects of MPC performance assessment are many. Some of these issues are listed below and one 'quantifiable' diagnosis issue related to model-plant mismatch is discussed. The diagnosis stage for poor performance involves a trial and error approach (Kesavan and Lee,

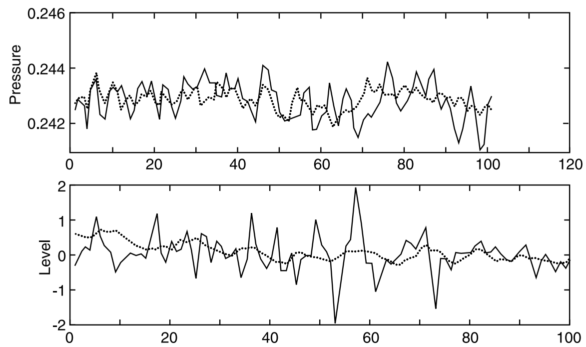


**Figure 17:** An example of the interaction between the steady state optimization and the dynamic layer in industrial MPC. Note that setpoints have higher variation compared to controlled variables!

1997). For example, the diagnosis or the decision support system has to investigate the cause of poor performance as being due to:

- Poor or incorrect tuning.
- Incorrect controller configuration, e.g. choice of MVs may not be correct.
- Large disturbances, in which case the sources of measured disturbances have to be identified and potential feedforward control benefits should be investigated.
- Engineering redesign , e.g. is it possible to reduce process time delays?
- Model-plant mismatch in the case of MPC controllers and how does model uncertainity affect the calculation of the performance index (e.g. if the index has been obtained from an uncertain interactor).
- Poor choice of constraint variables and constrained values.

Some of the above referenced issues have been already dealt with in the literature, e.g. (ANOVA analysis to investigate the need for feedforward control by Desborough and Harris (1993), Vishnubhotla et al. (1997), and Stanfelj et al. (1993); others are open problems. The diagnosis issues related to the model-plant mismatch is briefly discussed below in a theoretical framework and illustrated on an industrial MPC evaluation case study that follows. A discussion of the poor performance diagnosis steps leading to guidelines for tuning and controller design issues is beyond the scope of this paper.

### Model-Plant Mismatch

MPC controllers rely heavily on process models. In particular an accurate model is required if the process is to be regulated very tightly. On the other hand performance can be detuned in favour of robustness if an accurate model is not available or should the process change over a period of time. The extent of MPM can not be easily discerned by simply examining the closed loop prediction error. As shown in Figure 18, the prediction error under closed loop conditions is a function of the MPM, setpoint changes and measured and unmeasured disturbances. Thus the cause of large prediction errors may not necessarily be attributed to a large MPM. Consider Figure 18, where the prediction error is denoted as $e$.

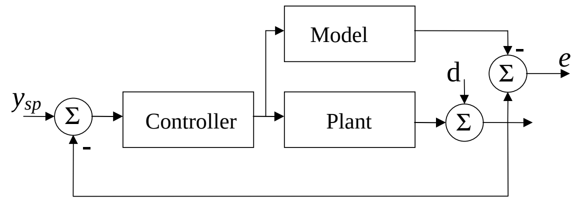


**Figure 18:** Schematic of a closed loop system in which the prediction error is monitored.

Under open loop conditions, the prediction errors is: $e = (P - \hat{P})u + d$. Under closed loop control the prediction error expression is:

$$e = \left( \frac{(P - \hat{P})C}{1 + CP} \right) y_{sp} + \left( \frac{1 + C\hat{P}}{1 + CP} \right) d$$

It is clear from the above expression that a large prediction error signal could be due to a large MPM term, or a large disturbance term or setpoint changes. Thus the question of attributing a large prediction error as being due to model-plant-mismatch needs careful scrutiny. Huang (2000) has studied the problem of detecting significant model plant mismatch or process parameter changes in the presence of disturbances.

## Industrial MIMO Case study 2: Analysis of Cracking Furnace Under MPC Control

This section documents the results of the controller performance analysis carried out on an ethane cracking furnace. The control systems comprises of (1) a regulatory layer and (2) an advanced MPC control layer. The first pass of performance assessment revealed some poorly performing loops. Further analysis revealed that these loops were in fact well tuned but were being affected by high frequency disturbances and setpoint changes. The furnace MPC application considered here, however, is unlike conventional MPC applications. The steady state limits were set in such a way that the setpoints for the controlled variables were held constant, i.e. the focus of the evaluation was on the models and the tuning of the dynamic part.

The MPC layer displays satisfactory performance levels when there are no rate changes. During rate changes, the MPC model over predicts thus causing poor performance. Re-identification of the model gains was found to be necessary to improve MPC performance.

### Control Strategy Overview

The purpose of the Furnace MPC controllers is to maintain smooth operation, maximize throughput and minimize energy consumption to the furnaces while simultaneously honoring all constraints. There is one MPC controller per furnace. The conversions, the total dry feed, the wet feed bias, the steam/feed ratio and the oxygen to fuel ratio is controlled by MPC. These variables are manipulated by moving the north and south fuel gas duty setpoints, the north and south wet feed flow setpoints, the steam pressure setpoint, the fan speed controller setpoint and the induced fan draft. Thus there are 6 primary controlled variables and 7 manipulated variables. There are a number of secondary controlled variables, which MPC is required to maintain within a constraint region. These secondary CVs include valve constraints, constraints on critical variables such as the coil average temperatures (COTS). Considering the degrees of freedom (MVs) available, it may not always be possible to satisfy all the constraints. In such cases, a ranking mechanism decides what constraints are least important and could be let go.

A critical component of the MPC controllers is the model describing the relationships between the MVs and the CVs—primary as well as secondary. These models are developed on the basis of open loop tests. Step response curves are used to parameterize the models. These models are used by MPC to predict the future process response. A portion of the step response model matrix is shown in Figure 19.

Performance analysis of the furnace control loops was conducted in two stages. Phase one loop analysis was performed on the lower regulatory layer. With the exception of a few loops, the first pass of performance assessment revealed satisfactory performance of most loops. The higher level MPC performance assessment commenced next.

Multivariable Performance Assessment for MPC. Table 2 summarizes the performance statistics. A diagonal interactor was used, based on the knowledge of the process models.

The performance metrics indicated satisfactory performance on all the variables except for tags 4 and 6. The closed loop settling time is approximately the same as the open loop settling time, which indicates a conservatively tuned application. Part of the reason for the slow response in control of tags 4 and 6 could be the presence of a measurement delay. During January 2000, the service factor for MPC was low due to some communication issues, which have been since resolved. This meant that there was an opportunity to compare the furnace performance with and without MPC. Based on data from Jan 14-16 when MPC was shut off for part of the time, performance metrics were obtained to compare the two control systems—MPC and conventional PID controls. The statistics indicate that the overall control is only slightly better with MPC turned on.

| Description | PI | Closed-Loop Settling Time (min) | Status |
|---|---|---|---|
| Tag 1 | 0.95 | 8 | |
| Tag 2 | 0.94 | 1 | |
| Tag 3 | 0.76 | 15 | |
| Tag 4 | 0.44 | 15 | LOW |
| Tag 5 | 0.71 | 15 | |
| Tag 6 | 0.49 | 15 | LOW |
| Tag 7 | 0.88 | 10 | |
| **Multivariable PI** | **0.71** | | |

**Table 2:** Summary of MPC performance.

### MPC Diagnostics

Is MPC doing its best? Can we improve the current performance levels of the furnace MPC controller? These questions lead us to two issues that are closely related to each other:

1. How good are the models used for predicting the process response?
2. How well tuned is the multivariable controller? "Tuning" includes a whole range of different of parameters—weightings, horizons, constraints, rankings . . .

We will try to illustrate a case where the model predictions can mislead the controller and hence cause poor performance. This Furnace was showing poor MPC performance, especially during rate changes. This motivated us to look more closely at the model prediction accuracy.

Before evaluating the current predictions, we establish a baseline when the open loop tests were conducted. Figures 21–23 compare the conversion predictions for the open loop case as conducted before commissioning the MPC. The model accuracy is reasonable. The average prediction error, for this data set was 3.18.

$$\text{Average Prediction Error} = \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N_y}\{y_i(k+1) - \hat{y}_i(k+1|k)\}^2$$

The predictions for other variables—COTs, Feed Flow and Bias, S/F ratio also fared well. The remaining variables are not shown for the sake of brevity.

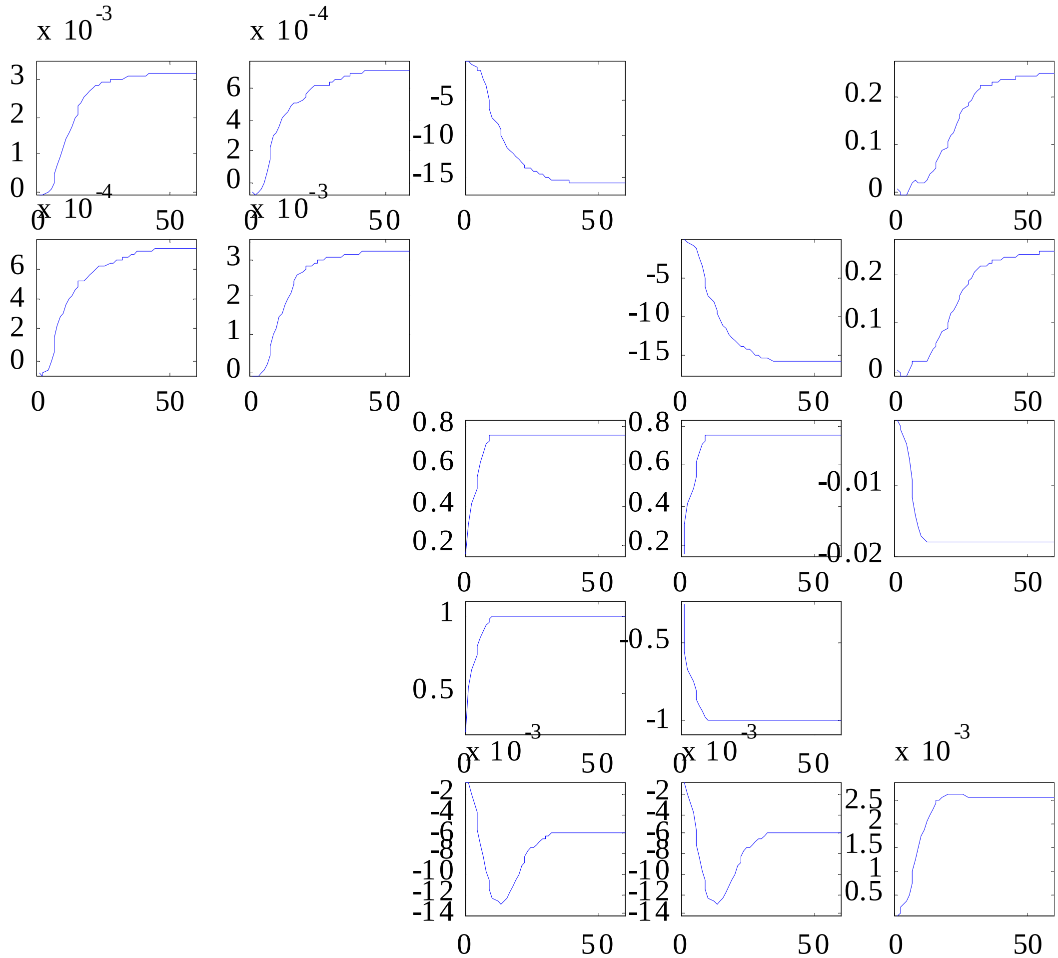

**Figure 19:** A portion of the step response models used by MPC.

The model accuracy can then be compared with the current predictions. There were two rate changes during this period, due to furnace decokes.

An average prediction error of 175.7 was observed during this period. The tags 3 to 6 predictions were much worse than the rest. Taking a closer look at the tags 3-6 predictions revealed that the models were over predicting by a factor of 2. Based on a combination of statistical analysis and process knowledge it was decided that the model gains were incorrect.

## Improving MPC Performance

The main stumbling block to improving MPC performance was its model accuracy. The models causing the large mismatch were identified. One of the reasons for the model plant mismatch is the fact that open loop tests were carried out in a operating region which is quite different from the current operating conditions (higher rates, conversions, duties). Fairly routine plant test, were used to identify the suspected changes in steady state gains. These tests were conducted under closed loop conditions and the gain mismatch has now been fixed with satisfactory MPC performance. The newly identified gains were indeed found to be significantly different from the earlier gains.

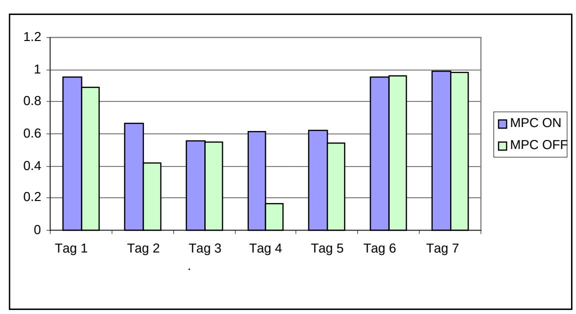


**Figure 20:** Comparison of performance, with and without MPC.

To illustrate the effect of the MPM on cracking efficiency, conversion control on the 3 furnaces was com-

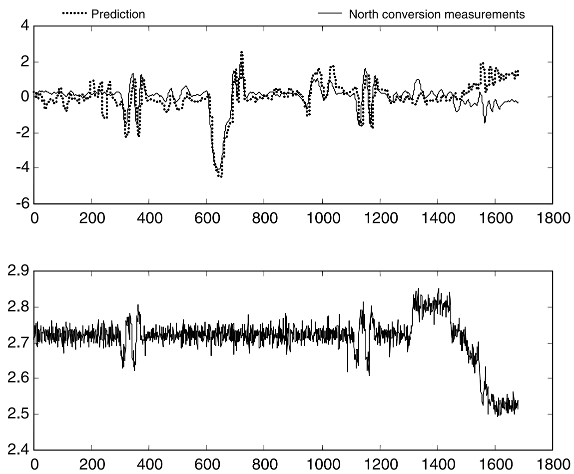


**Figure 21:** The predictions (green) and the North conversion measurements (blue) are shown in the top graph. The bottom graph shows the changes in fuel duty during this period.

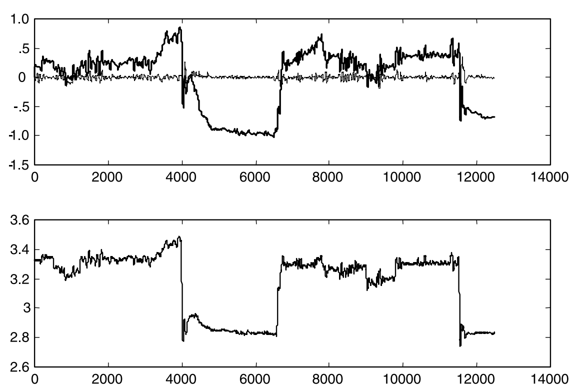

**Figure 22:** The scaled conversion predictions for Feb 11-15.

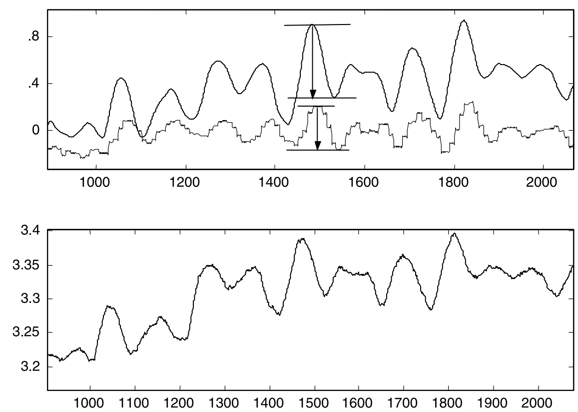

**Figure 23:** The scaled conversion predictions—apparent gain mismatch.

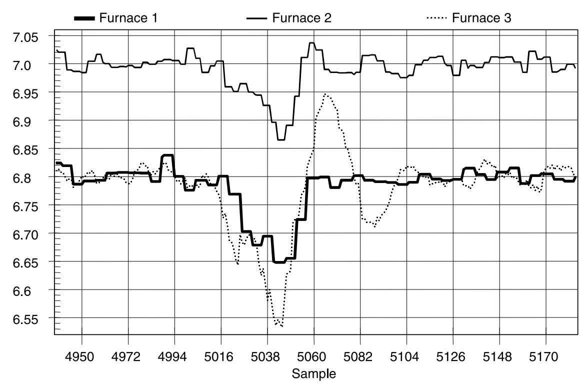


**Figure 24:** Comparison of Furnace control with updated models.

pared and is shown in Figure 24. Furnace 3 with the older models takes almost 45 minutes more to settle out, compared to furnaces 1 and 2, which have the updated models. The overshoot for Furnace 3 is 2.5%, as opposed to 1.5% on Furnace 1 and 2 (a significant 40% decrease). Thus the performance and subsequent diagnosis analysis of this MPC application illustrates the value in routine monitoring and maintenance of MPC applications.

## Concluding Remarks

In summary, industrial control systems are designed and implemented or upgraded with a particular objective in mind. We hope that the new controller loop performance assessment methodology proposed in the literature and illustrated here, will eventually lead to automated and repeated monitoring of the design, tuning and upgrading of the control loops. Poor design, tuning or upgrading of the control loops will be detected, and repeated performance monitoring will indicate which loops should be re-tuned or which loops have not been effectively upgraded when changes in the disturbances, in the process or in the controller itself occur. Obviously better design, tuning and upgrading will mean that the process will operate at a point closer to the economic optimum, leading to energy savings, improved safety, efficient utilization of raw materials, higher product yields, and more consistent product qualities. Results from industrial applications have demonstrated the applicability of the multivariate performance assessment techniques in improving industrial process performance.

Several different measures of multivariate controller performance have been introduced in this paper and their applications and utility have been illustrated by simulation examples and industrial case studies. The multivariate minimum variance benchmark allows one to compare the actual output performance with the minimum achievable variance. However it requires knowledge of the process time-delay matrix or the interactor. On the other hand the newly proposed $\text{NMIR}_{wof}$ measure of performance provides a graphical 'metric' that requires little or no *a priori* information about the process and gives a graphical measure of mutivariate performance in terms of settling time, rate of decay etc.. The challenges related to MPC performance evaluation are illustrated by an industrial case study of an ethylene cracker. It is shown how routine monitoring of MPC applications can ensure good or 'optimal' control. The lumped objective function based method of monitoring MPC performance is shown to work well on the industrial case study. The study illustrates how controllers, whether in hardware or software form, should be treated like 'capital assets'; how there should be routine monitoring to ensure that they perform close to the economic optimum and that the benefits of good regulatory control will be achieved

## Acknowledgments

The authors of this paper wish to express sincere gratitude and appreciation to the following persons who have provided assistance by participating in industrial evaluation studies at their respective work sites: Jim Kresta and Henry Tse (Syncrude Research), Janice Vegh-Morin and Harry Frangogiannopoulus (both of TDCC), Genichi Emoto (Mitsubishi Chemical Corporation, Japan) and Anand Vishnubhotla (Matrikon Consulting) Their feedback and industrial insight were helpful in enhancing the practicality and utility of the proposed methods to make it better suited to the process industry.

Financial support in the form of research grants from the Natural Science and Engineering Research Council of Canada and Equilon Enterprises (Shell-USA) is gratefully acknowledged.

## References

Boyd, S. and C. Barratt, *Linear Control Design*. Prentice Hall, Englewood Cliffs, New Jersey (1991).

Desborough, L. and T. Harris, "Performance Assessment Measures for Univariate Feedback Control," *Can. J. Chem. Eng.*, **70**, 1186–1197 (1992).

Desborough, L. D. and T. J. Harris, "Performance Assessment Measures for Univariate Feedforward/Feedback Control," *Can. J. Chem. Eng.*, **71**, 605–616 (1993).

Garcia, C. E., D. M. Prett, and M. Morari, "Model Predictive Control: Theory and Practice—A Survey," *Automatica*, **25**(3), 335–348 (1989).

Harris, T. J., F. Boudreau, and J. F. MacGregor, "Performance Assessment of Multivariate Feedback Controllers," *Automatica*, **32**(11), 1505–1518 (1996).

Harris, T. J., "Assessment of Control Loop Performance," *Can. J. Chem. Eng.*, **67**, 856–861 (1989).

Huang, B. and S. L. Shah, *Performance Assessment of Control Loops*. Springer-Verlag, London (1999).

Huang, B., S. L. Shah, and K. K., How Good Is Your Controller? Application of Control Loop Performance Assessment Techniques to MIMO Processes, In *Proc. 13th IFAC World Congress*, volume M, pages 229–234, San Francisco (1996).

Huang, B., S. L. Shah, and E. K. Kwok, "Good, Bad or Optimal? Performance Assessment of Multivariable Processes," *Automatica*, **33**(6), 1175–1183 (1997).

Huang, B., S. L. Shah, and H. Fujii, "The Unitary Interactor Matrix and its Estimation Using Closed-Loop Data," *J. Proc. Cont.*, **7**(3), 195–207 (1997).

Huang, B., "Model Validation in the Presence of Time-variant Disturbance Dynamics," *Chem. Eng. Sci.*, **55**, 4583–4595 (2000).

Kammer, L. C., R. R. Bitmead, and P. L. Barlett, Signal-based testing of LQ-optimality of controllers, In *Proc. of the 35th IEEE CDC*, pages 3620–3624, Kobe, Japan (1996).

Kesavan, P. and J. H. Lee, "Diagnostic Tools for Multivariable Model-Based Control Systems," *Ind. Eng. Chem. Res.*, **36**, 2725–2738 (1997).

Ko, B. and T. F. Edgar, Performance Assessment of Multivariable Feedback Control Systems (2000). Submitted to *Automatica*.

Kozub, D. J. and C. E. Garcia, Monitoring and Diagnosis of Automated Controllers in the Chemical Process Industries, AIChE Annual Meeting, St. Louis (1993).

Kozub, D., Controller Performance Monitoring and Diagnosis Experiences and Challenges, In Kantor, J. C., C. E. García, and B. Carnahan, editors, *Chemical Process Control—V*, pages 156–164. CACHE (1997).

Kwakernaak, H. and R. Sivan, *Linear Optimal Control Systems*. John Wiley and Sons, New York (1972).

Mayne, D. Q., J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained Model Predictive Control: Stability and Optimality," *Automatica*, **36**(6), 789–814 (2000).

Patwardhan, R. S. and S. L. Shah, Issues in Performance Diagnostics of Model-based Controllers, Accepted for publication, J. Proc. Control (2001).

Patwardhan, R. S., G. Emoto, and S. L. Shah, Model-based Predictive Controllers: An Industrial Case Study, AIChE Annual Meeting, Miami (1998).

Patwardhan, R., S. L. Shah, and K. Qi, Techniques for Assessment of Model Predictive Controllers, Submitted for publication (2000).

Patwardhan, R., S. L. Shah, and G. Emoto, Experiences in Performance Annalysis of Industrial Model Predictive Controllers, Submitted for publication (2001).

Patwardhan, R. S., *Studies in the Synthesis and Analysis of Model Predictive Controllers*, PhD thesis, University of Alberta (1999).

Qin, S. J. and T. A. Badgwell, An Overview of Industrial Model Predictive Control Technology, In Kantor, J. C., C. E. Garcia, and B. Carnahan, editors, *Fifth International Conference on Chemical Process Control—CPC V*, pages 232–256. American Institute of Chemical Engineers (1996).

Stanfelj, N., T. E. Marlin, and J. F. MacGregor, "Monitoring and Diagnosing Process Control Performance: The Single-Loop Case," *Ind. Eng. Chem. Res.*, **32**, 301–314 (1993).

Swanda, A. P. and D. E. Seborg, Controller Performance Assessment Based on Setpoint Response Data, In *Proceedings of the American Control Conference*, pages 3863–3867, San Diego, California (1999).

Thornhill, N. F., M. Oettinger, and P. Fedenczuk, "Refinery-wide control loop performance assessment," *J. Proc. Cont.*, **9**, 109–124 (1999).

Van den Hof, P. M. J. and R. J. P. Schrama, "Identification and control—closed-loop issues," *Automatica*, **31**(12), 1751–1770 (1995).

Vishnubhotla, A., S. L. Shah, and B. Huang, Feedback and Feedforward Performance Analysis of the Shell Industrial Closed Loop Data Set, In *Proc. IFAC Adchem 97*, pages 295–300, Banff, AB (1997).

Wood, R. K. and M. W. Berry, "Terminal Composition Control of a Binary Distillation Column," *Chem. Eng. Sci.*, **28**, 1707–1717 (1973).