Recent Developments in Controller Performance Monitoring and Assessment Techniques

Thomas J. Harris *
Department of Chemical Engineering
Queen's University
Kingston, ON, K7L 3N6, Canada

Christopher T. Seppala
Automation Engineering
Equilon Enterprises LLC
Houston, TX 77251

Abstract

In the past several years there has been considerable commercial and academic interest in methods for analyzing the performance of univariate and multivariate control systems. This focus is motivated by the importance that control systems have in enabling companies to achieve goals related to quality, safety and asset utilization. Control system performance cannot be adequately described by simple statistics, such as the mean and variance of manipulated and controlled variables, the percentage of time that constraints are satisfied, and the on-stream time. Although these are important performance measures, a comprehensive approach for controller performance monitoring usually includes the following elements: i) determination of the capability of the control system, ii) development of suitable statistics for monitoring the performance of the existing system, and iii) development of methods for diagnosing the underlying causes for changes in the performance of the control system. In this paper, recent developments related to these items will be reviewed for both univariate and multivariate systems. Some multivariate time series methods helpful in supporting these controller performance assessment techniques in practice will be discussed, and an industrial example will be provided. Finally, the future direction of commercial applications of controller performance assessment will be briefly discussed, as will the issue of whether controller performance assessment is destined to be offered as a product or a service.

Keywords

Controller performance, Minimum variance control, Multivariable systems, Subspace methods, Vector autoregressions

Introduction

Early interest in theory and methods for the on-line analysis of control systems can be traced to papers by Åström (1967) and De Vries and Wu (1978). Since that time work in this area has continued, with considerable development taking place during the 1990s. Reviews and critical analyses of several approaches for assessing control loop performance can be found in Huang and Shah (1999), Harris et al. (1999), and Qin (1998). Control system capability statistics based on the performance benchmark of minimum variance control for single-input single-output (SISO) systems were the initial underlying concept for much of this work. Regulation of stochastic and deterministic disturbances, setpoint tracking, extensions to multiple-input-single-output (MISO) systems (i.e., single output systems with feedforward variables) are readily accommodated in this framework, and these aspects have been described in the aforementioned references. Recently, Ko and Edgar (2000a) have extended these ideas to evaluate cascade control systems. Horch and Isaksson (1999) have proposed a modification to the basic performance measures to more closely connect the monitoring and control objectives. Thornhill et al. (1999) provide comprehensive guidelines for the application of control loop assessment and Miller and Desborough (2000) describe a commercial product/service for control loop assessment.

Extensions of minimum variance performance assessment techniques to multiple-input-multiple-output (MIMO) systems for general time delay systems was initially considered by Harris et al. (1996), and Huang et al. (1997b). The challenges in evaluating a multivariate control system (as opposed to analyzing single loops in a complex control system) are considerable. These challenges arise primarily from: i) the requirement for *a priori* knowledge of the time delay structure of the process, ii) the time-varying nature of control loops which arises from the constraint handling requirements of multivariate controllers, and iii) the requirement to use sophisticated identification methods to obtain meaningful estimates of the closed loop impulse weights. There are a considerable number of challenges, both theoretical and practical, in the assessment of multivariate schemes.

The purpose of this paper is to provide: i) a concise summary of results in univariate performance monitoring, ii) an overview of challenges and recent developments in the assessment of multivariate control systems, and iii) a brief discussion on the future direction of industrial applications of controller performance assessment, including the role of supporting technologies (including an industrial example), and the consumer's perspective on commercial performance assessment solutions—are they products or services?

Univariate Performance Assessment

Process Description

To introduce the concepts of control performance monitoring and assessment, consider a process whose behavior about a nominal operating point can be modeled by

*harrist@post.queensu.ca

a linear transfer function with an additive disturbance:

$$Y_t = \frac{\omega(q^{-1})q^{-b}}{\delta(q^{-1})}U_t + D_t \tag{1}$$

where Y_t denotes the difference between the process variable and a nominal operating point. U_t denotes the difference between the manipulated variable and its nominal value, and $\omega(q^{-1})$ and $\delta(q^{-1})$ are polynomials in the backshift operator, q^{-1}. b whole periods of delay elapse between making a change in the input and first observing its effect on the process output. The process disturbance, D_t, is represented by an Autoregressive-Integrated-Moving-Average (ARIMA) time series model of the form:

$$D_t = \frac{\theta(q^{-1})}{\nabla^d \phi(q^{-1})}a_t \tag{2}$$

where $\theta(q^{-1})$ and $\phi(q^{-1})$ are stable polynomials in the backshift operator, and ∇ is a shortcut notation for $(1 - q^{-1})$. The integer d denotes the degree of differencing ($0 \leq d \leq 2$ in most applications). a_t denotes a sequence of independently and identically distributed random variables with mean zero and variance σ_a^2. This disturbance structure is capable of modeling commonly occurring stochastic and deterministic disturbances.

The process is controlled by a linear feedback controller of the form:

$$U_t = G_c(q^{-1})(Y_{sp} - Y_t) \tag{3}$$

where $G_c(q^{-1})$ is the controller transfer function and Y_{sp} denotes the deviation of the setpoint from its reference value. We will assume that these values are equal; the general case is considered in Desborough and Harris (1992). With these assumptions, the closed loop is given by:

$$Y_t = \left(\frac{1}{1 + \frac{\omega(q^{-1})q^{-b}}{\delta(q^{-1})}G_c(q^{-1})} \right) D_t \tag{4}$$

Substituting Equation 2 for D_t in Equation 4 and simplifying allows the closed-loop to be written in rational transfer function form as follows:

$$Y_t = \frac{\alpha(q^{-1})}{\beta(q^{-1})}a_t = \psi(q^{-1})a_t \tag{5}$$

The closed-loop impulse response coefficients are given by:

$$\psi(q^{-1}) = 1 + \psi_1 q^{-1} + \psi_2 q^{-2} + \cdots \tag{6}$$

Convergence of the series in Equation 6 is guaranteed if the closed-loop is stable; the expansion is valid for computation of the impulse weights, ψ_j. Tyler and Morari (1996) present a useful discussion on the duality between the impulse weights ψ_j and other classic measures of controller performance including settling time, decay rate, and desired reference trajectories.

Minimum Variance Performance Bounds and Performance Measures

If one were to design a controller to minimize the variance of the output, the impulse response coefficients beyond the process deadtime, ψ_j, $j = b, b+1, \ldots$, would equal zero. The output variance would equal (Åström, 1967, 1970; Box and Jenkins, 1976):

$$\sigma_y^2 = \sigma_{mv}^2 = (1 + \psi_1^2 + \cdots + \psi_{b-1}^2)\sigma_a^2. \tag{7}$$

If the minimum variance performance fails to meet the controller's design objectives, then reductions in the output variance can only be achieved by modifying the process to change the disturbance characteristics or by reducing the deadtime. Because σ_{mv}^2 provides a fundamental lower bound on performance, simply retuning the controller, or implementing a more sophisticated linear controller with the same manipulated variable and control interval, will not reduce process variability. This bound depends only on the process delay and is otherwise independent of the dynamic characteristics of the controller.

Implementation of a minimum variance controller that achieves the bound described in Equation 7 requires that the polynomials $\omega(q^{-1})$ and $\delta(q^{-1})$ and be stable. When these conditions are not satisfied, it is still possible to design a controller that minimizes the variance of the output subject to stability of both the closed-loop and manipulated variable. The output variance will, by necessity, exceed that described by Equation 7. This topic is discussed further in a subsequent section.

Desborough and Harris (1992), Stanfelj et al. (1993), Kozub and Garcia (1993) and Kozub (1996) have introduced a number of performance indices to provide an indication of the departure of the current performance from minimum variance control. Typical performance measures are:

$$\xi(b) = \frac{\sigma_y^2}{\sigma_{mv}^2} \tag{8}$$

and

$$\begin{aligned} \eta(b) &= 1 - \frac{1 + \psi_1^2 + \cdots + \psi_{b-1}^2}{1 + \psi_1^2 + \cdots + \psi_{b-1}^2 + \psi_b^2 + \cdots} \\ &= 1 - \frac{\sigma_{mv}^2}{\sigma_y^2} \end{aligned} \tag{9}$$

where $\xi(b) \geq 1$ and $0 \leq \eta(b) \leq 1$. The performance index $\xi(b)$ corresponds to the ratio of the actual variance to that which could theoretically be achieved under minimum variance control. The normalized performance index, $\eta(b)$, is a number between 0 (minimum variance performance) and 1 (far from minimum variance performance) that reflects the inflation of the output variance over the theoretical minimum variance bound. As indicated in Desborough and Harris (1992), it is more useful to replace σ_y^2 by the mean square error of y_t, thereby accounting for offset.

Note that the normalized performance index is independent of the magnitude of the disturbance driving force (a_t in Equation 2). It may happen that $\eta(b) = 0$, i.e., the system is operating at minimum variance control performance, yet σ_y^2 still exceeds process or product requirements. In this case the process—not the control system—is not capable.

Estimation of Minimum Variance Performance Bounds and Performance Measures

It is possible to calculate the minimum variance performance bound, σ_{mv}^2, by estimating a process plus disturbance model obtained from a designed experiment. This performance bound can then be used to determine the process capability. If the performance bound fails to meet process specifications, then the process modification remedies described above must be sought to rectify the situation. Such an approach would severely limit the usefulness of this bound because obtaining a process and disturbance model is labour intensive and intrusive, i.e., it requires a perturbation signal to be introduced. Furthermore, the disturbance structure may change during the period of data collection, potentially changing the results of the analysis.

Why then, has the minimum variance performance benchmark in Equation 7 proven to be so useful in practice? Its usefulness stems from two important properties:

1. **Autocorrelation Test**: Under minimum variance control, the autocorrelations of the y's are zero beyond lag $(b-1)$ since the closed-loop is a moving average process of order $(b-1)$, or an ARIMA$(0, 0, b-1)$ process (Åström, 1967, 1970; Box and Jenkins, 1976). Conversely, if any (stable) controller results in a closed-loop, which is an ARIMA$(0, 0, b-1)$ process, then the controller is a minimum variance controller. If the controller is unstable, except for the presence of p integrators, then the observed closed-loop may appear to be a moving average process of order less than $(b-1)$, (Foley and Harris, 1992). Except in these rare cases, the sample autocorrelation function of the y's, or a portmanteau test on the autocorrelations of y can be used to provide a simple, convenient, and useful method for testing whether any SISO controller is giving minimum variance performance (Harris, 1989; Stanfelj et al., 1993; Kozub and Garcia, 1993; Kozub, 1996).

2. **Invariance Property**: σ_{mv}^2 can, under mild conditions, be estimated from *routine operating data* when the time delay is known (Harris, 1989). It is straightforward to show that the first $(b-1)$ ψ_j coefficients of the closed-loop equal the first $(b-1)$ impulse coefficients of the disturbance transfer function. The remaining coefficients are functions of the controller, process, and disturbance transfer functions. Since the first $(b-1)$ ψ_j coefficients are not affected by any feedback controller they can collectively be interpreted as a system invariant (Harris, 1989; Tyler and Morari, 1996). These can be estimated by fitting a time series model to the closed loop error:

$$\alpha(q^{-1})(y_t - \bar{y}) = \beta(q^{-1})a_t \tag{10}$$

where $\alpha(q^{-1})$ and $\beta(q^{-1})$ are stable polynomials in the backshift operator, of order n_a and n_b, respectively. The term $\bar{y}$ accounts for non-zero mean data. The coefficients of the polynomials and their orders can be estimated using standard time series analysis techniques. Once these parameters have been estimated, the impulse weights are calculated by long division of $\alpha(q^{-1})$ into $\beta(q^{-1})$. Computational details, and variants, are discussed in Harris (1989), Desborough and Harris (1992) and Huang et al. (1997b).

It is important to note that the calculation of σ_a^2 and σ_{mv}^2 does not require separate identification of the process transfer function and disturbance transfer functions since $\eta(b)$ corresponds to the fraction of the output variance reduction that can be achieved by implementing a minimum variance controller. As a result of the above properties, σ_{mv}^2 and $\eta(b)$ can be estimated from routine operating data if the delay is known.

Exact distributional properties of the estimated performance indices are complicated, and not amenable to a closed-form solution. Desborough and Harris (1992) approximated first and second moments for the estimated performance indices and resorted to a normal theory to develop approximate confidence intervals. Asymptotically, the performance indices are ratios of correlated quadratic forms, and as such the distributions of the performance indices are non-symmetric. Refinements to the confidence intervals developed in Desborough and Harris (1992) can be obtained with little extra computational effort, by resorting to the extensive statistical literature on the distributional properties of quadratic forms (Harris, 2001).

Extensions and Modifications

The development thus far has been based on the simple process description given by Equation 4. Performance monitoring and assessment methods have been extended to include variable setpoints (Desborough and Harris, 1992), feedforward/feedback systems (Desborough and Harris, 1993; Stanfelj et al., 1993), processes with interventions (Harris et al., 1999), and cascade systems (Ko and Edgar, 2000a).

The performance bounds described above have been presented under idealized assumptions. The actual, as opposed to lower bound on performance, is also limited by the presence of non-invertible zeros, the requirement for smooth-manipulated variable movement, and the presence of hard constraints on the manipulated variable. A number of modifications have been proposed to accommodate these issues, and these will be reviewed in the following paragraphs.

Non-invertible systems. When the process transfer function is non-invertible, it is possible to design a modified minimum variance controller using spectral factorization methods (Bergh and MacGregor, 1987; Harris and MacGregor, 1987). This modified minimum variance controller has the lowest variance among all stable controllers. The following identities hold:

$$\sigma_a^2 \leq \sigma_{mv}^2 \leq \sigma_{mv^\star}^2 \leq \sigma_{Åström}^2 \tag{11}$$

where $\sigma_{mv^\star}^2$ denotes the variance of the modified minimum variance controller and $\sigma_{Åström}^2$ denotes the closed-loop variance of a simple pole placement algorithm proposed by (Åström, 1970). This latter controller is particularly easy to design; the limitation being that the non-invertible zeros of the process transfer function cannot be canceled. With this design the process output is a moving average process of order $(b-1+n^\star)$, where $n^\star$ is the number of zeros of $\omega(q^{-1})$, in q, outside the unit circle. When the location of the non-invertible zero is known, in addition to the time delay, $\sigma_{mv^\star}^2$ can be estimated from routine operating data (Harris et al., 1996; Tyler and Morari, 1995; Huang and Shah, 1999). These latter results use linear-quadratic-control theory to determine the achievable performance bound. The performance results can be sensitive to the location of the non-invertible zero (Tyler and Morari, 1995). Estimation of $\sigma_{mv^\star}^2$ requires considerably more process knowledge than is required to estimate σ_{mv}^2. Although not as rigorous, a number of alternate approaches, which retain the simplicity of the minimum variance bounds and calculations, can be used. These are discussed in subsequent sections. Recently, Ko and Edgar (2000c,b) have used fundamental results of Furuta and Wongsaisuwan (1993, 1995) to show how algorithms such as Dynamic Matrix Control (DMC) can be used to obtain several different performance bounds. This approach will be discussed further in the multivariate performance assessment section.

Excessive control action and robustness concerns. Minimum variance controllers may call for unacceptably large changes in manipulated variable action. This happens when the process is sampled "quickly" relative to its dominant time constant. In these circumstances minimum variance controllers (or deadbeat controllers) may be sensitive to process model mismatch (Åström, 1970; Bergh and MacGregor, 1987). In these instances, it has been found useful to modify the performance indices so that the latter more closely reflects the controller design requirements. Two modified controller performance indices have been proposed to deal with these issues: The extended horizon performance index and the user-defined benchmark performance index.

Extended-horizon performance index. Desborough and Harris (1992, 1993), Kozub (1996), Harris et al. (1996), and Thornhill et al. (1999) utilize an extended horizon performance index defined as:

$$\eta(b+h) = 1 - \frac{1+\psi_1^2+\cdots+\psi_{b-1}^2+\cdots+\psi_{b+h-1}^2}{1+\psi_1^2+\cdots+\psi_{b-1}^2+\psi_b^2+\cdots} \tag{12}$$

This normalized performance index gives the proportion of the variance arising from non-zero impulse coefficients ψ_j, $j > b+h$. $\eta(b+h)$ can also be interpreted as the square of the correlation between the current error and the least squares estimate of the prediction made $(b+h)$ control periods in the past (Harris et al., 1999). The extended horizon predictor closely matches control objectives of model based control strategies, such as Dynamic Matrix Control (DMC). It is important to note that when $h > 0$, the prediction error variance is affected by the structure and tuning of the feedback controller (in contrast to the case when $h = 0$). The use of the extended horizon performance index indirectly acknowledges the fact that minimum variance control may not be desirable or feasible. One obvious advantage of using $\eta(b+h)$ instead of $\eta(b)$ is that the former does not require a precise estimate of the process delay. Kozub (1996) and Thornhill et al. (1999) indicate that many problems in diagnosing the performance of controllers can be solved by estimating both $\eta(b)$ and $\eta(b+h)$.

User-defined benchmark performance index. Recently, Horch and Isaksson (1999) have introduced a normalized performance index:

$$\xi_{mod}(b) = \frac{\sigma_y^2}{\sigma_{mod}^2} \tag{13}$$

where:

$$\sigma_{mod}^2 = \left(1+\psi_1^2+\cdots+\psi_{b-1}^2+\psi_{b-1}^2\frac{\upsilon^2}{1-\upsilon^2}\right)\sigma_a^2 \tag{14}$$

and $0 \leq \upsilon < 1$. The motivation for this modified performance index is very simple; a minimum variance controller can be interpreted as a requirement that all of the closed-loop system poles be placed at the origin. If instead, one of the closed-loop poles is moved to a location specified by the designer, then the variance of the closed-loop is given by σ_{mod}^2 in Equation 13. Horch and Isaksson (1999) show that this design is equivalent to a requirement that the closed-loop have an exponential decay to target rather than the dead-beat response required of minimum variance control. With this interpretation, specification of υ is not difficult. Horch and Isaksson call the modified performance index a *user-defined benchmark*. They point out that the basic simplicity of the original performance index is retained, while offering greater flexibility. The authors do not require that the controller be designed using this technique; rather they point out that the analysis of closed-loop data is facilitated by the choice of υ. Statistical properties of the performance index are proposed, and the relationship between the modified index and specifications on the autocorrelation function (suggested in Kozub and Garcia (1993) Huang and Shah (1998)) are also discussed.

Note the following properties of this modified controller performance index: i) if the process is operating at the desired user-defined benchmark, $\xi_{mod}(b) = 1$ ii) if the performance is "better" than the user-defined benchmark, $\xi_{mod}(b) < 1$, iii) if the process variance exceeds the user-defined benchmark, then $\xi_{mod}(b) > 1$, and iv) $\xi_{mod}(b) \geq 1 - \upsilon^2$. These properties provide a convenient "normalization" for the performance index.

Hard constraints. When the manipulated variable is at a hard constraint, the closed-loop is no longer described by Equation 4. However, it is possible to estimate σ^2_{mv} from routine data by including inputs and outputs in the time series model (Desborough and Harris, 1992). It is necessary to keep record of when the constraints are active, so that the model structure properly reflects the status of the control system. Manipulated variable constraints usually result in offset between controlled variables and their setpoint. Under such conditions, controller performance assessment can still be possible if the output(s) of interest are part of a multivariate predictive control scheme. A working solution in this case is to substitute the reachable target associated with the constrained output, say $Y^{\star}_{sp,t}$, for the setpoint in the calculation of the closed-loop error, i.e., $y_t = Y^{\star}_{sp,t} - Y_t$. The reachable target is internally calculated by the control algorithm and is simply a feasible value for the output of interest conditional on the active constraints. The estimates of σ^2_{mv} derived under such conditions may be suspect due to the influences of other input variables. In this situation, inspection of the closed-loop impulse response coefficients, which provide dynamic information on the output's tracking of the reachable target is recommended. In any case, when a controller is regularly switching between different sets of active constraints, benchmarking the dynamic performance may not be as important as monitoring how well the controller is meeting its overall design objectives, e.g., output prioritization and the distribution of offset.

Performance assessment with fixed controller structure. Most controllers employ a fixed structure, i.e., a Proportional-Integral-Derivative (PID) controller. It is of interest to develop performance monitoring and performance assessment methods for these widely used systems. Isaksson (1996), Ko and Edgar (1998) and Harris et al. (1999) have investigated these topics. Performance limitations arising from a fixed controller structure can only be determined if a process model is available. If opportunities for significant performance improvements are indicated using the minimum-variance methods, then one can determine the achievable limitations that arise from using a particular controller structure only by identifying a process and disturbance model. The use of previously identified models for assessment in a predictive control environment is discussed in a later section.

Detection of oscillations, valve stiction and other maladies. A number of researchers and practitioners have indicated that more realistic estimates of the achievable performance are obtained when one detects, diagnoses and "removes" the effect of oscillations (Owen et al., 1996; Owen, 1997; Horch, 2000). Methods for detecting oscillation and stiction are described in Hãgglund (1995), Bittanti et al. (1997), Horch and Isaksson (1998, 1999), Seborg and Miao (1999), and Forsman (2000). Oscillations and valve stiction can be viewed as faults. There are other faults that beset control loops; the purpose of this paper is not to review this extensive literature (Isermann and Ballé, 1997). Rather, we indicate that automated procedures for control loop assessment using the methods proposed here, or descriptive statistics, must have proper data segmentation so that the presence of faults do not lead to improper interpretations or conclusions.

Nonlinear and time varying processes. In deriving the minimum variance controller, we assumed the process admits the description given in Equations 1 and 2. When the process is described by a nonlinear difference equation, either for the dynamics or disturbances, development of the nonlinear minimum variance controller may be very difficult or essentially impossible. This of course depends upon the structural form of the nonlinearity. For those descriptions which admit a nonlinear description and closed-form expressions for the minimum variance control law, it is possible to construct examples that show that the feedback invariance property does not exist. To ascertain performance bounds from routine operating data, one must assume that the process admits a local linear representation. The performance assessment results are "locally" valid. If changes in operating point cause changes in the process model, then the data must be properly segmented prior to analysis. Methods for detecting changes in model structure are discussed in Basseville (1998). If the disturbances are time-varying or consist of a mixture of stochastic and deterministic type disturbances, which is often the case, then the process description in Equations 1 and 2 must be expanded to account for this behavior. Again, methods for detecting these interventions must be part of the data analysis. The performance assessment techniques reviewed in this paper can then be applied to these types of processes (Harris et al., 1999).

Discussion

We point out that σ^2_{mv} may often not be a realizable performance bound due to the practical limitations described above. It has been pointed out by a number of authors (Desborough and Harris, 1992; Huang et al., 1997b,c) that if $\sigma^2_{mv} < \sigma^2_y$ there *may* be opportunities to reduce the output variance. However, a diagnosis of the control system is required to investigate the cause(s) of variance inflation. If it suspected that non-invertible zeros or restrictions on the manipulated variables are limiting performance, then a process plus disturbance model must be identified to calculate $\sigma^2_{mv^\star}$, $\sigma^2_{\text{Åström}}$, or any other performance measure which requires knowledge of the process dynamics and disturbances. Alternatively, a number of the modified performance indices described above can be used to aid in the diagnosis of performance and detect changes in performance from a specified target value (Kozub, 1996).

Although the performance bounds and performance measures described in this section were originally introduced to ascertain how far the current performance was from minimum variance, they have found widespread use as a component of a more comprehensive performance monitoring and assessment methodology. Typically, industrial controller performance monitoring packages include some minimum variance-based performance statistics but also elementary descriptive statistics (such as mean, standard deviation, % uptime), histograms, power spectra, autocorrelation functions, impulse response functions and even non-linear valve diagnostics. Continuous performance monitoring applications also have significant information technology requirements such as access to historized data, dedicated servers, scheduling algorithms, and rule-based event notification and exception reporting (Jofriet et al., 1996). Guidelines for implementing univariate performance monitoring methods in practice are discussed in the references contained in Harris et al. (1999), Vishnubhotla et al. (1997), Thornhill et al. (1999), and Miller and Desborough (2000).

Multivariate Performance Assessment

The extension of performance assessment to multivariable systems has been studied by Harris et al. (1996), Huang et al. (1997a,b,c), and Huang and Shah (1998, 1999). Assessment of minimum variance performance bounds arising from deadtimes in MIMO systems requires knowledge of the interactor matrix. The interactor matrix allows a multivariate transfer function to be factored into two terms; one having its zeros located at infinity and another containing the finite zeros. To introduce this concept, consider a linear time-invariant process with n outputs and m inputs having transfer function $T(q^{-1})$. The interactor is a square matrix polynomial having the following properties (Dugard et al., 1984):

$$\lim_{q \to \infty} \xi(q)T(q^{-1}) = K \tag{15}$$

and

$$|\xi(q)| = q^B \tag{16}$$

where K is a non-singular matrix and B is the number of zeros of the transfer function located at infinity. In the univariate case, $\xi(q) = q^b$ and $B = b$. Other properties of the interactor matrix are discussed in Dugard et al. (1984), Goodwin and Sin (1984), Tsiligiannis and Svoronos (1988), Mutho and Ortega (1993), and Mutho (1995). It is important to note that the interactor matrix is not unique, and that it cannot always be constructed solely from knowledge of the delay structure. The interactor matrix can be constructed using linear algebra techniques from the process transfer function in the aforementioned references and Rogozinski et al. (1987). Huang et al. (1997a) have shown that the interactor matrix can be estimated from the Markov parameters of the process transfer function.

It is convenient to define the inverse-interactor matrix as follows:

$$\xi^{-1}(q^{-1}) = [\xi(q)]^{-1} = \xi_k q^{-k} + \cdots + \xi_d q^{-d} \tag{17}$$

where k is the minimum delay in the first row of the process transfer function, and d is not less than the maximum delay in the transfer function. Note that the bound on k shows that the interactor matrix is not unique; it can be altered by re-ordering the inputs and outputs.

Using the inverse interactor matrix, the process may be represented in right matrix fraction form as follows:

$$\begin{aligned} Y_t &= L(q^{-1})R^{-1}(q^{-1})U_t + D_t \\ &= \xi^{-1}(q^{-1})\tilde{L}(q^{-1})R^{-1}(q^{-1})U_t + D_t \end{aligned} \tag{18}$$

where $\xi^{-1}(q^{-1})$ represents the inverse interactor matrix and D_t represents the process disturbance, which can often be modeled by a multivariate ARIMA process.

Once the interactor matrix is known, the multivariate extension of the univariate performance bounds can be established. Several methods can be used, all leading to equivalent results. Harris et al. (1996) define the performance bound:

$$\eta = 1 - \frac{E[Y^T_{mv} W Y_{mv}]}{E[Y_t^T W Y_t]} \tag{19}$$

where $E[\cdot]$ denotes mathematical expectation and $E[Y^T_{mv} W Y_{mv}]$ denotes the weighted multivariate minimum variance performance. W is a positive definite weighting matrix, which allows for differential weights on specific outputs. Determination of the multivariate minimum variance control performance requires that an all-pass representation of the interactor matrix be constructed. There are two general approaches for performing this: spectral factorization and construction of a spectral interactor.

Spectral Factorization

The spectral factor of the interactor matrix, $\gamma(q^{-1})$, is defined as the solution to the spectral factor equation:

$$\gamma^T(q)W\gamma(q^-1) = \xi^{-T}(q)W\xi^{-1}(q^{-1}). \tag{20}$$

Since $\xi^{-1}(q^{-1})$ is unimodular, the spectral factor $\gamma(q^{-1})$ is also unimodular. A property of a unimodular spectral factor is that $\gamma^{-1}(q^{-1})$ exists, and is a finite polynomial in q^{-1}. Efficient methods for the construction of the spectral factor involve solution of a bilinear set of equations, for which iterative application of Cholesky decompositions are very efficient (Kucera, 1979; Harris and Davis, 1992). These algorithms have quadratic convergence in a finite number of iterations when the polynomial matrix for which the spectral factor is to be obtained is unimodular.

Construction of the Spectral Interactor Matrix

Huang et al. (1997b) exploit the fact that the interactor matrix is not unique. They use a spectral interactor, $\tilde{\xi}^{-1}(q^{-1})$, introduced by Peng and Kinnaert (1992) and Bittanti et al. (1994) having the property:

$$\tilde{\xi}^{-T}(q)\tilde{\xi}^{-1}(q^{-1}) = I. \tag{21}$$

Linear algebra techniques can be used to construct the spectral interactor from the process transfer function or Markov parameters.

Once the all-pass filter representation has been obtained, it is possible to express the closed-loop system in the following form (Harris et al., 1996; Huang et al., 1997b):

$$S_t = \Psi_1(q^{-1})a_t + q^{-(d-1)}\Psi_2(q^{-1})a_t \tag{22}$$

where S_t is a filtered output, having the property that $E[S_t^T W S_t] = E[Y_t^T W Y_t]$. The terms $\Psi_1(q^{-1})a_t$ and $q^{-(d-1)}\Psi_2(q^{-1})a_t$ are uncorrelated. The first term on the right hand side of Equation 22 is a function only of the disturbance and the all-pass interactor matrix, and is otherwise independent of the dynamics of the process. This term is the multivariate equivalent of the system invariant, $(1 + \psi_1 q^{-1} + \cdots + \psi_{b-1}q^{-(b-1)})a_t$, encountered in univariate performance assessment. The term $\Psi_1(q^{-1})a_t$ represents the dynamics of the multivariate minimum variance controller. The second term in Equation 22 is a function of the controller, the process transfer function and the disturbances. In the derivation of Equation 22, it was assumed that a linear time-invariant controller was used.

Once the decomposition in Equation 22 has been affected, it is possible to calculate the performance index from Equation 19 as follows:

$$\eta = 1 - \frac{Tr\left(W\sum_{j=0}^{d-1}\Psi_j\Sigma_A\Psi_j^T\right)}{Tr\left(W\sum_{j=0}^{\infty}\Psi_j\Sigma_A\Psi_j^T\right)} \tag{23}$$

The two important properties encountered in the univariate case, namely the autocorrelation test and the invariance property, are also found in the multivariate extension. Once the spectral factor is obtained, one can also construct a portmanteau test for multivariate minimum variance control that is similar to the autocorrelation function (Harris et al., 1996). Performance bounds can be determined regardless of the number of inputs and outputs; there is no need that the process be "square". In the multivariate case it can also be shown that the minimum variance performance can be estimated from routine operating data if the interactor matrix is known, and there are several different methods for calculating the minimum variance performance bounds (Harris et al., 1996; Huang et al., 1997b). In the process of calculating the performance bound, it is necessary to fit a multivariate time series to the observations (when a linear controller is used). When constraints are active it is necessary to fit a predictive model to both the inputs and outputs. Haarsma and Nikolaou (2000) tested several identification methods in an application of multivariate performance assessment in the food processing industry. Other examples of the application of multivariate performance assessment and monitoring are given Harris et al. (1996), Huang et al. (1997a,b,c), Huang and Shah (1998, 1999), Miller and Huang (1997), and Huang et al. (2000).

The minimum variance controller described in Goodwin and Sin (1984) Dugard et al. (1984) is a sequential minimum variance controller that is dependent on the order of the inputs and outputs and choice of interactor representation. The construction of the all-pass filter representation of the interactor matrix leads to a "true" minimum variance controller, which is independent of these factors.

Remarks

1. Univariate and multivariate performance assessment are conceptually similar, however in the latter case knowledge of the time delay structure alone does not guarantee that the performance bounds can be calculated. Knowledge of the interactor matrix is an impediment to using multivariate techniques. Huang et al. (1997a) have shown that the interactor matrix may be calculated from the impulse coefficients of the process transfer function and have proposed a technique to estimate this from process data. This is akin to estimating the delay in an online fashion for SISO systems. However, this method requires that a dither signal be added to the process during the period of data collection. Furthermore, the method assumes that a linear, time-invariant controller be used during the period of data collection.

2. In the SISO case, an extended prediction horizon can be used for performance monitoring (Equa-

tion 12). This extended horizon serves two purposes: i) it provides an indication of the sensitivity of the performance index to the selection of delay, and ii) it indirectly addresses the issue that minimum variance control may not be the desired control objective. In multivariate analysis, one can also use a similar concept (Harris et al., 1996). Essentially, one replaces the interactor matrix by the term $\xi^{p+h}I$, where p is an estimate of the maximum order of the inverse-interactor matrix and $h > 0$. This approach does not enable calculation of the lower bound on performance; rather it is more useful in monitoring changes in the predictive structure of the process.

3. The performance bounds calculated using the interactor matrix are not restricted to those processes for which there are an equal number of inputs and outputs. In most cases, multivariable controllers are used where constraints are a factor. In these cases, the structure of the time-series model changes as constraints are engaged. One can adapt the structure of the time series model to reflect the evolving constraint set structure. There can be rather dramatic changes in the minimum variance performance bound when the set of constraint variables changes. The utility of the minimum variance performance bounds in these instances has yet to be determined. One can imagine that other performance measures may be more appropriate.

   Ko and Edgar (2000c,b) have proposed several methods to address performance assessment in the presence of constraints. Their work is based on the fundamental results of Furuta and Wongsaisuwan (1993, 1995) who show that a receding horizon controller, with input and output weightings, i.e., soft constraints, can be used to obtain the solution to the infinite horizon linear quadratic controller. In these papers, Furuta and Wongsaisuwan use the Markov parameters of the controller and disturbance (i.e., the impulse coefficients) to design the controller. Ko and Edgar (2000c,b) have used these results to provide a number of performance bounds. The method requires that an input/output model relating Y and U be available; a step response model used in the design of a predictive controller would suffice for this purpose. As usual, it is assumed that this model adequately describes the process. Given a record of $\{Y_t, U_t\}, t = 1..N$, the process disturbance can be reconstructed from the measurements as follows:

$$D_t = Y_t - T(q^{-1})U_t. \quad (24)$$

   A time series model is then fit to the D's. Once this time-series model has been determined, a number of performance bounds can be determined using the results of Furuta and Wongsaisuwan (1993, 1995).

   By applying zero weight to the inputs, Ko and Edgar (2000c,b) demonstrate that the minimum variance performance bound that corresponds to solution of the unconstrained linear quadratic minimum variance bound discussed in Harris et al. (1996) can be estimated. This bound equals the bound obtained from using the interactor matrix when the process transfer function has no non-invertible zeros, other than those associated with the time delay.

   By using a time series model for the disturbance, it is possible to simulate a generalized predictive controller, over the data set used to estimate the disturbance. By applying the same inputs and output constraints used in the actual controller, as well as the same prediction horizons for the inputs and outputs, an estimate of the performance using the identified disturbance structure is obtained. Recall that most receding horizon controllers assume that the disturbance is adequately modeled by a multivariate random walk. This bound correctly accounts for the presence of constraints.

   Both of these approaches, and other variations which can be derived from this approach, enable one to use a previously identified process model as part of the multivariate performance assessment process. A fundamental assumption is that this model is accurate, and that the disturbance model identified from Equation 24 has no model mismatch component.

4. When a more comprehensive model identification is undertaken it is possible to use more sophisticated performance measures. Kendra and Çinar (1997) have developed a frequency domain identification and performance assessment procedure for closed-loop multivariable systems. *A priori* information, such as design stage transfer function specifications, can be incorporated into the analysis. This allows model mismatch to be assessed, and makes possible comparisons of current operating performance to design specifications for the sensitivity and complementary sensitivity functions. External excitation must be provided in the form of a dither signal to enable identification of the sensitivity functions. Gustafsson and Graebe (1998) have developed a procedure to ascertain whether changes in closed-loop performance arise from changes in disturbance structure or changes in the process transfer function. A test signal must be applied for this analysis.

5. Intervention analysis provides a framework to incorporate variable setpoint changes, feedforward variables and deterministic disturbances in the univariate case. Analogous methods for the multivariable case have not been developed extensively.

One possible criticism of recent research developments in controller performance assessment is that too much

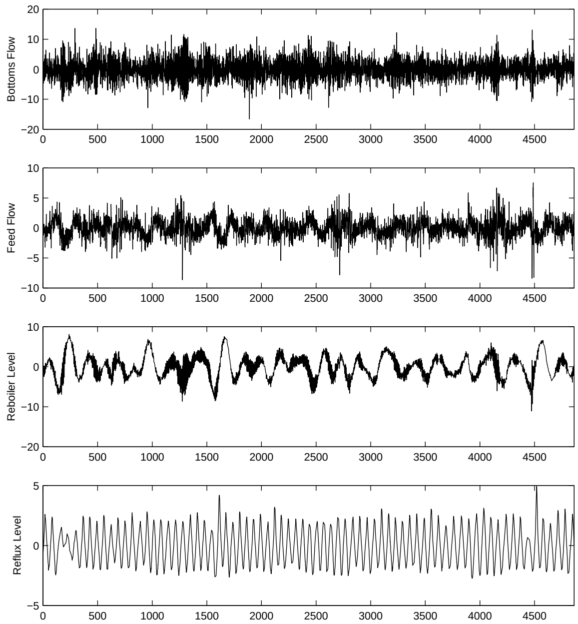

Figure 1: A time series plot showing the CV errors for the feed flow loop, the bottoms flow loop, the reboiler level loop, and the reflux level loop for the industrial distillation column. The trend plots show roughly five-thousand samples of one-minute data for each variable.

effort has been focused on estimating system invariants. While this is certainly one of the most interesting and challenging problems from an academic perspective, it is really only one of the many tools that an engineer might effectively use to monitor and/or analyze control system performance. The use of system invariants for multivariate performance assessment is a significant barrier to use due to the information requirements and the level of expertise needed to apply the methodology and interpret the results. Comprehensive methods for analyzing the interaction structure of the closed-loop are essential for diagnosing multivariate systems. In the next section we will demonstrate some analysis methods that can be derived using multivariable time series methods.

Example of Multivariate Process Analysis

In this section we will apply multivariate time series techniques to analyze an industrial data set. The objective of the analysis is to provide a qualitative and quantitative analysis of the closed-loop behavior. Our interest extends beyond the question as to whether or not the control system is operating at a desired performance benchmark.

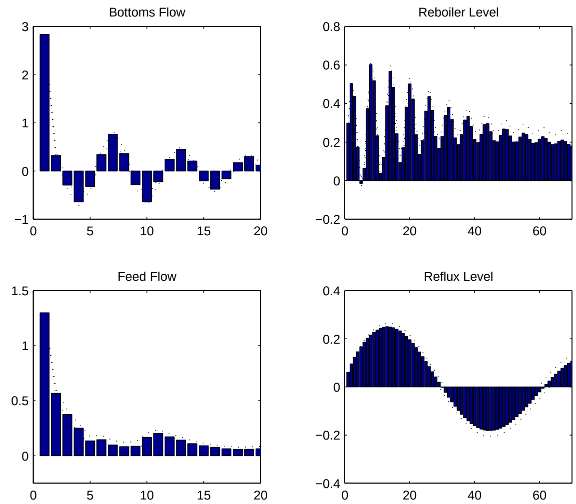

Figure 2: Univariate CV error impulse response functions. Clockwise, from the top left: the bottoms flow, the reboiler level, the feed flow, and the reflux drum level.

Process Description

In this example, multi-output impulse response analysis will be used to study the dynamic relationships between four controlled variable (CV) error variables sampled from an industrial distillation column. The CV error vector was calculated from setpoint and output observations sampled at one-minute intervals from the following control loops on the column: the feed flow controller, the bottoms flow controller, the reboiler level controller, and the reflux drum level controller. Time series plots of these variables are shown in Figure 1. It is assumed that no prior information is available concerning the multivariable delay structure. Note, all of the modeling methods used in this section are standard results that have been adapted from the multivariate time series analysis literature (Hamilton, 1994; Lütkepohl, 1991).

Univariate Impulse Response Analysis

The first step was to estimate the closed-loop impulse weights, Equation 6, for each of the process variables. This was accomplished by fitting a univariate autoregressive model (using a least squares approach) to each variable and calculating the impulse weights by long division. The estimated impulse response plots are shown in Figure 2. The time horizon for the plots has been set to twenty minutes for the flow controllers (column one), and seventy minutes for the level controllers (column two).

Some deductions regarding the dynamic performance of each of these control loops can be made from the univariate impulse response functions. For example, strong cyclical behavior is observed in all the tracking error variables except for the feed flow. There is a one hour cycle in the reflux drum level, a seven minute cycle in the bottoms flow, and a seven minute cycle combined with another, slower, response in the reboiler level. The feed flow controller seems reasonably well tuned; it is free from overshoot or cycles, and damps out quickly. These are all valid observations, but no information regarding possible interactions between these loops can be made unless a multivariate analysis is performed. With the exception of the feed flow, none of the variables is close to their individual minimum variance performance bounds.

Multi-Output Impulse Response Analysis

An impulse response plot is simply a graphical representation of a time series model in moving average form. The displayed impulse response coefficients are the weights that describe the dynamic relationship between the input and the output. When the input is assumed to be a unit impulse, the impulse response plot shows the predicted output response. In the multi-output case, the $(i, j)^{th}$ entry in the $(n \times n)$ impulse response matrix gives the model weights between input driving force j and output i.

Note that in the current context of analyzing CV error dynamics, the multi-output impulse response estimates are based on routine operating data. In contrast, step response data to be used for identification is collected under experimental conditions where input variables are manipulated. So while models based on the latter approach can be considered causal, the same is not true for the former. If the underlying data has not been collected during an experiment, the tracking error impulse response plots simply help the analyst interpret the correlation structure between the tracking error trends, not the true causal relationships.

Modeling Control Error Trends—Vector Time Series Approach

Multi-output control error trends can be considered a group of univariate control error trends of equal length that all share the same time stamp. Rather than being a scalar at time t, a multi-output control error trend is an n dimensional vector at time t, with one element for each of the n controlled variable (CV) error trends. Define the following vector time series:

$$\mathbf{y}_t = \mathbf{Y}_{sp,t} - \mathbf{Y}_t \tag{25}$$

where $\mathbf{y}_t$, $\mathbf{Y}_{sp,t}$, and $\mathbf{Y}_t$ are vectors representing the control error, the output, and the setpoint, respectively. In practice, one would typically be working with $\mathbf{y}$, an $(N \times n)$ array of CV error data, based on N samples of n CV error trends.

Treating the dynamic analysis of multi-output control error trends as an endogenous estimation problem with no *a priori* information has been explored by Seppala (1999). Linear dynamic approximations of endogenous system behavior with no *a priori* information and no assigned input/output structure had been previously used in the field of applied econometrics. The simplest multivariate dynamic model that can represent $\mathbf{y}_t$ is a vector autoregressive (VAR) model which is written as follows:

$$\Phi(q^{-1})\mathbf{y}_t = \mathbf{a}_t \tag{26}$$

where $\mathbf{a}_t$ is a vector of driving forces, and $\Phi(q^{-1})$ is an autoregressive matrix polynomial defined as:

$$\Phi(q^{-1}) = I_n + \Phi_1 q^{-1} + \cdots + \Phi_p q^{-p} \tag{27}$$

where each Φ_i is an $(n \times n)$ coefficient matrix. The expanded form of Equation 26 is clearly analogous to a scalar autoregressive model; each variable in $\mathbf{y}_t$ is expressed as a function of lagged values of itself and the other $(n-1)$ variables in $\mathbf{y}_t$:

$$\mathbf{y}_t = -\Phi_1\mathbf{y}_{t-1} - \cdots - \Phi_p\mathbf{y}_{t-p} + \mathbf{a}_t \tag{28}$$

where p is the autoregressive model order. The driving force covariance matrix, Σ_a, is given by:

$$\Sigma_a = E[\mathbf{a}_t\mathbf{a}_t^T]. \tag{29}$$

The diagonal elements of Σ_a are the driving force variances, and the off-diagonal elements are the driving force covariances.

To find the multi-output impulse responses, one proceeds in much the same fashion as in the univariate case. If the autoregressive matrix polynomial in Equation 27 is stable, then the VAR model for $\mathbf{y}_t$ may be expressed in vector moving average (VMA) form:

$$\begin{aligned}\mathbf{y}_t &= \Theta(q^{-1})\mathbf{a}_t \\ &= (1 + \Theta_1 q^{-1} + \cdots + \Theta_r q^{-r})\mathbf{a}_t \\ &= \sum_{i=0}^{r} \Theta_i \mathbf{a}_t - i\end{aligned} \tag{30}$$

where $\Theta(q^{-1})$ is the vector moving average matrix polynomial, defined such that $\Phi(q^{-1})\Theta(q^{-1}) = I_n$. The Θ_i's can be found using the recursion:

$$\Theta_i = \sum_{j=1}^{i} \Theta_{i-j}\Phi_j \tag{31}$$

where $i = 1, 2, \ldots,$ and $\Theta_0 = I_n$. The Θ_i coefficient matrices contain the multi-output impulse response coefficients.

The VMA model for $\mathbf{y}_t$ in Equation 30 is not unique; a property of many types of multivariate models. To illustrate the non-uniqueness property, consider any non-singular matrix P: the Θ_i matrices can be replaced by $\Psi = \Theta_i P$, and the driving forces can be replaced by $\mathbf{v}_t = P\mathbf{a}_t$, resulting in the following equivalent model for $\mathbf{y}_t$:

$$\mathbf{y}_t = \sum_{i=0}^{r} \Theta_i P P^{-1}\mathbf{a}_{t-i} = \sum_{i=0}^{r} \Psi_i \mathbf{v}_{t-i} \tag{32}$$

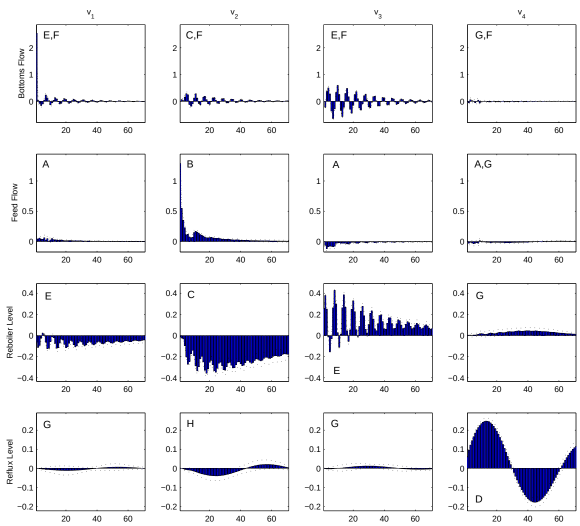

Figure 3: Multi-output CV error impulse response plot matrix for the four variable system consisting of the bottoms flow, the feed flow, the reboiler level and the reflux level. These responses were computed from an eighth order VAR model converted to orthogonal innovations form VMA via a Cholesky factorization of the covariance matrix. This plot shows the first seventy steps of the response.

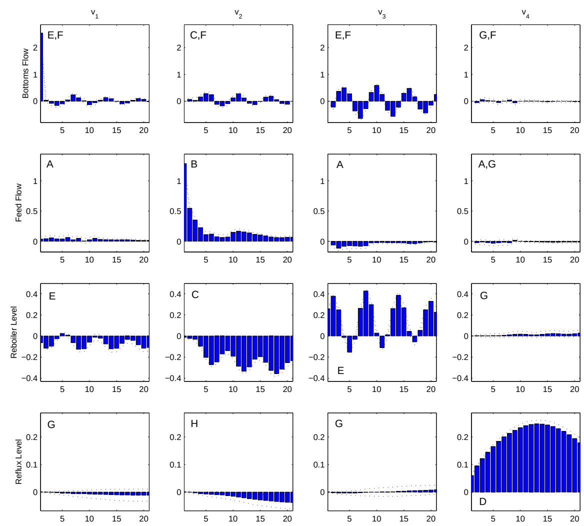

Figure 4: Multi-output CV error impulse response plot matrix for the four variable system consisting of the bottoms flow, the feed flow, the reboiler level and the reflux level. These responses were computed from an eighth order VAR model converted to orthogonal innovations form VMA via a Cholesky factorization of the covariance matrix. This plot shows the first twenty steps of the response.

The models in Equations 30 and 32 are equivalent in the sense that they produce identical estimates of the k-step ahead forecast error covariance. The non-uniqueness property of the VMA model can be used to choose a particular P that orthogonalizes the driving forces, thereby simplifying multi-output impulse response analysis and variance calculations. A common choice is to select P so that it is the Cholesky factor of the driving force covariance matrix, resulting in orthogonalized driving forces $\mathbf{v}_t$. When one interprets the impulse response coefficient matrices, the Ψ_i's, one can consider the effects of shocks to the driving force processes one-at-a-time because they are orthogonal. This topic and other methods for analyzing multi-output control systems are discussed in Seppala et al. (2001) and Seppala (1999).

Results of Multivariate Analysis

An eighth order VAR model was found to adequately model the control error data in Figure 1, i.e., Equation 28 was used with $n = 4$ and $p = 8$. Standard residual analysis showed the model to be adequate, and the residuals themselves to be nearly orthogonal. The impulse response form of the multivariate model for the distillation column data is shown in Figure 3 with a time horizon of seventy minutes, and in Figure 4 with a time horizon of twenty minutes. In these figures, the rows (from top to bottom) represent the four CV error variables (bottoms flow, feed flow, reboiler level, and reflux level) and the columns represent the system's driving forces. Because the driving forces are essentially orthogonal, the driving force in column i can be interpreted as a deviation from setpoint or disturbance in variable i, and the coefficients plotted in position (i, i) of the impulse response matrix can be considered the endogenous component of the response for variable i.

Non-significant interactions. The responses in the subplots labeled A and/or G in Figures 3 and 4 contain non-significant relationships because the 95% confidence intervals for the responses contain zero across the entire time horizon. This can be interpreted as a lack of significant correlation between a deviation from setpoint or a disturbance in variable i and the presence of any corresponding response in variable j. From Figures 3 and 4 it can be seen that the model shows non-significant interactions along the dynamic pathways from: i) the bottoms flow to the feed flow, ii) the bottoms flow to the reflux level, iii) the reboiler level to the feed flow, iv) the reboiler level to the reflux level, and v) the reflux level to any other system variable.

The feed flow error. Inspection of the second row of plots in Figures 3 and 4 reveals that only driving force v_2 (recall, v_2 is interpreted as an impulse-like upset in the feed flow) has a significant effect on the feed flow. This aspect of the model makes sense physically because in this example, only upstream properties affect the feed flow controller. If v_2 is the only driving force significantly correlated with the feed flow control error, then the univariate and multivariate impulse response plots should be similar for the feed flow tracking error.

How do the other system variables respond to a shock in the feed flow loop? The plots labeled C show that after a couple of units of delay, the bottoms flow and the reboiler level both show significant responses with a strong cyclical component at seven minutes. The plot labeled H shows that a shock in the feed flow has a small, perhaps negligible, effect on the reflux level.

The reflux level error. The reflux level tracking error exhibits a dominant cycle with a period of about 60 minutes. Since all the responses labeled G in Figures 3 and 4 are non-significant, it can be deduced that the reflux level control error is unrelated to the other three tracking error variables over the period of data collection. Note also that the impulse response plot in pane D is essentially identical to the univariate impulse response plot for the reflux level in Figure 2. The multi-output impulse response plots in column four provide statistical evidence that errors in the reflux level controller are unrelated to dynamics in the other three loops during the period of data collection.

The reboiler level/bottoms flow pair. In Figures 3 and 4, the rows showing the responses for the bottoms flow (row 1) and the reboiler level (row 3) indicate that there is a strong seven minute cycle shared by both CV error variables in this cascaded pair. The multivariate model that has been estimated shows that an upset in any of the CV error variables except the reflux level error is related to this statistically significant response in the bottoms flow/reboiler level pair. As mentioned above, deductions regarding causality are out of the question, but the analysis shows that a common cycle exists between the reboiler level error and the bottoms flow error, and that the error in at least one external loop (the feed flow) is correlated with this pair.

Industrial Perspectives

At CPC V, controller performance monitoring was categorized as a new direction for academic research. Since then, there has been considerable research in this area, with a significant focus of this work directed towards developing controller performance monitoring (CPM) techniques for multivariable systems. Most of the work in this area has focused on developing multivariable system invariants, with more recent work addressing incorporation of constraints. There are other topics requiring attention, and we shall indicate a few that are of industrial interest.

As CPM matures as a technology, and as its acceptance becomes more widespread, the question of how to affect CPM solutions arises. Since 1996, a number of commercial products and services for CPM have appeared. There are an enormous number of challenges in developing, supporting and ensuring that these packages are used effectively. We will provide a short discussion of these challenges. Finally, a brief discussion on the relative merits of CPM solutions as vendor products or as vendor services is included.

New Areas for Research

Multivariate predictive controllers have an optimization layer in the form of a linear or quadratic program, and some plants have real time optimization (RTO) systems downloading targets to Multivariable Predictive Controllers (MPCs) or to the base level control system. What is there to be learned by monitoring the behavior of these targets? On the time scale of control systems, can this data be considered dynamic? RTO targets arrive on the order of hours and can be considered static; however, over-active optimization targets that cannot be considered static from the point of view of the control system have been observed. What effect does would the latter have on one's perception of control system performance?

The use of dynamic analysis of variance (ANOVA), i.e., studying the correlation and quantifying the variance propagation between key control system variables was used by Desborough and Harris (1993) to analyze multiple-input-single-output controlled systems. Ideally, one would use ANOVA methods to identify process variables that are chiefly responsible for variance inflation of key controlled variables. The technical challenges are well-known: causal ordering of upstream variables, the effects of feedback and recycle, collinearity of disturbances, and a dealing with the component of variance propagation due to invariants. Further study of this topic is warranted.

Several methods for modeling multivariate dynamic and/or static data have matured to the point where powerful software packages are now available for their application. A couple of examples well known in the control engineering community are: ADAPTX (Subspace ID), Simca-P (PCA/PLS), and the host of data analysis tools available for the Matlab$^{(R)}$ environment. In combination with process knowledge, these data analysis packages can be very useful for analysis and diagnosis. Note, however, that batches of control system data that have been gathered because they came from a previously identified problem area are good for analysis, but this does not count as monitoring. True performance monitoring requires constant, scheduled contact with the plant information system, and this has been known for some time. There are many practical challenges with real-time applications: data integrity, fault-detection, robust algorithms, data visualization and presentation, to name a few.

CPM in the Field: A Product or A Service?

Many practicing control specialists are now at least familiar with controller performance monitoring. A number of prototype industrial controller performance monitoring systems have been described in the open literature, Harris et al. (1999). Commercial products are available from Honeywell, Matrikon and Control Arts to name a few. When one considers purchasing CPM capabilities for a plant, besides the obvious issues of the level of technology required, the issue of whether to buy a CPM product, or a CPM service emerges. Using a CPM product would be like using any other piece of installed software; essentially one has access to on-line help and product support. Pursuing a CPM service could involve engineers visiting the site to perform control performance audits, service providers consulting on difficult CPM problems, or an electronic exchange of raw data for loop performance reports (Miller and Desborough, 2000).

CPM products and services will both be costly, and both will require support from the provider. Without trying to answer the question of which model is better, CPM as a product or CPM as a service, some of the important issues will be outlined below.

Whether CPM is used as a product or a service, proper training is required if plant operation going to benefit from CPM. One of the frustrations with applications of CPM, and other quality monitoring methods, is the level of training for individuals who are asked to use these methods. Although CPM technology is not as wide in scope or as complex as multivariate predictive control, the availability of training in the latter area far exceeds that which is available for CPM. As with most statistical methods, attention must be paid to the length of data and sampling interval used for analysis, the type of filtering used prior to analysis (such as compressed data) and other aspects of data integrity.

Process knowledge has long been known to be an essential ingredient to successful application of CPM in the field (Jofriet and Bialkowski, 1996; Haarsma and Nikolaou, 2000; Horch, 2000). In order to integrate CPM into engineers' work practices, regular hands-on experience with CPM is required to develop skills. CPM products are best suited for this, because the product becomes just another tool, one that does not rely on a third party to use successfully. A particular challenge is that advanced multivariate techniques, which require *a priori* structural information and advanced system identification techniques, may only be successful when used by experts in CPM. Widespread use by control engineers requires automation of most of the methodology, with an emphasis placed on interpretation and analysis. These requirements are not different than those encountered in applications of multivariate statistical process control. Finally, we note that control engineers working in environments where constraints on available funds, time, and support personnel are limiting, are the least likely to get involved in CPM. In this situation, if CPM is to be implemented at all, then the service model is probably more appropriate.

Conclusions

The use of controller performance monitoring and assessment tools in industrial settings has grown considerably in the past several years. Extensions and variations of minimum variance based methods have been used extensively, primarily due to ease of understanding, robust computational methods, and minimal requirements for *a priori* knowledge. Industrial versions of these packages are available as both products and services.

Since CPC V, there has been considerable development of the underlying theory for multivariable controller performance assessment methods. Two main approaches to multivariate controller performance assessment have emerged thus far. The first method requires the use of extensive *a priori* process knowledge. In particular, the use of previously identified process models, which enables constraint handling to be addressed in a logical and straightforward fashion. The outcome of enables one to ascertain performance bounds and thus to subsequently monitor changes in these bounds over time. Of course, the results of the analysis are interpreted presuming that the process model is correct. The second approach largely dispenses with the requirement for *a priori* knowledge. Empirical models are built and used to analyze the predictive structure of the data. In particular, process interactions, and variance-decompositions over time can be used to help diagnosis process interactions. With such an approach, one is not restricted to using time series models; many multivariate statistical methods can be used if they are modified to include lagged data to account for serial correlation. The two approaches share common features, and it is clear that they are non-trivial generalizations of the univariate measures. It remains to be shown whether the more demanding and complex multivariate methods can be successfully integrated into a plant-wide monitoring and assessment strategy.

Most approaches for performance assessment use data collected in a passive mode or use data generated when significant events occur (Isaksson et al., 2000; Stanfelj et al., 1993). This is one of the key attributes of the performance measures—one performs the analysis with representative data. When poor performance is detected, a combination of statistical tools and process knowledge is required to analyze and diagnose the underlying problems. The role of designed experiments, in either closed-loop or open-loop, to aid in the diagnosis and analysis is an area requiring attention. Preliminary results have been reported by Kendra and Çinar (1997) and Gustafsson and Graebe (1998). Since many of the proposed methods for accommodating constraints require that the process transfer function model be known, on-line and off-line methods for model validation are essential for these techniques to be used with confidence. The development of diagnostics for model-based control is an open area for research (Kesavan and Lee, 1997).

The focus of much of this paper has been on the performance measures themselves. Large-scale industrial applications require incorporation of such performance measures into a plant-wide monitoring and performance assessment package. Industrial experience indicates that many of the challenges to broader application of performance measures lie in the successful development and maintenance of such systems.

Acknowledgments

TJH would like to thank the Natural Sciences and Engineering Research Council of Canada and Equilon Enterprises LLC for their financial support.

References

Åström, K. J., "Computer Control of a Paper Machine—An Application of Linear Stochastic Control Theory," *IBM Journal, July 1967*, pages 389–404 (1967).

Åström, K. J., *Introduction to Stochastic Control Theory*. Academic Press, London (1970).

Basseville, M., "On-Board Component Fault Detection and Isolation Using the Statistical Local Approach," *Automatica*, **34**, 1391–1415 (1998).

Bergh, D. and J. F. MacGregor, "Constrained Minimum Variance Controllers: Internal Model Structure and Robustness Propperties," *Ind. Eng. Chem. Res.*, **26**, 1558–1564 (1987).

Bittanti, S., P. Colaneri, and M. F. Mongiovi, The Spectral Interactor Matrix for the Singular Riccatti Equation, In *Proc. 33rd CDC*, pages 2165–2169 (1994).

Bittanti, S., M. Campi, and S. M. Savaresi, "Unbiased Estimation of a Sinusoid in Colored Noise via Adapted Notch Filters," *Automatica*, **33**(2), 209–215 (1997).

Box, G. E. P. and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco (1976).

De Vries, W. R. and M. Wu, "Evaluation of Process Control Effectiveness and Diagnosis of Variation in Paper Basis Weight via Multivariate Time-Series Analysis," *IEEE Trans. Auto. Cont.*, **23**, 702–708 (1978).

Desborough, L. D. and T. J. 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).

Dugard, L., G. C. Goodwin, and X. Xianya, "The Role of the Interactor Matrix in Multivariable Stochastic Adaptive Control," *Automatica*, **20**, 701–709 (1984).

Foley, M. and T. J. Harris, "Performance and Structure of H_∞ Controllers," *J. Optim. Cont. Appl. Meth.*, **13**, 1–28 (1992).

Forsman, K., On Detection and Classification of Valve Stiction, In *TAPPI Conf. on Process Control*, Williamsburg, VA (2000).

Furuta, K. and M. Wongsaisuwan, "Closed-Form Solutions to Discrete-Time LQ Optimal Control and Disturbance Attenuation," *Sys. Cont. Let.*, **20**, 427–437 (1993).

Furuta, K. and M. Wongsaisuwan, "Discrete-Time LQG Dynamic Controller Design Using Plant Markov Parameters," *Automatica*, **31**, 1317–1324 (1995).

Goodwin, G. C. and K. S. Sin, *Adaptive Filtering Prediction and Control*. Prentice-Hall, Englewood Cliffs, N. J. (1984).

Gustafsson, F. and S. F. Graebe, "Closed-Loop Performance Monitoring in the Presence of System Changes and Disturbances," *Automatica*, **34**, 1311–1326 (1998).

Hãgglund, T., "A Control-Loop Performance Monitor," *Control Eng. Practice*, **5**(11), 1543–1551 (1995).

Haarsma, G. and M. Nikolaou, Multivariate Controller Performance Monitoring: Lessons from an Application to a Snack Food Process (2000). Submitted to *Journal of Process Control*.

Hamilton, J. D., *Time Series Analysis*. Princeton University Press, Princeton (1994).

Harris, T. J. and J. H. Davis, "An Iterative Method for Matrix Spectral Factorization," *SIAM J. Sci. Statist. Comput.*, **13**, 531–540 (1992).

Harris, T. J. and J. F. MacGregor, "Design of Multivariable Linear-Quadratic Controllers Using Transfer Functions," *AIChE J.*, **33**, 1481–1495 (1987).

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

Harris, T. J., C. T. Seppala, and L. D. Desborough, "A Review of Performance Monitoring and Assessment Techniques for Univariate and Multivariate Control Systems," *J. Proc. Cont.*, **9**(1), 1–17 (1999).

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

Harris, T. J., Confidence estimation for performance assessment measures (2001). In Preparation, Queen's University, Kingston.

Horch, A. and A. J. Isaksson, A method for detection of stiction in control valves, In *IFAC Workshop on On-Line-Fault Detection and Supervision in the Chemical Process Industry*, Lyon, France (1998).

Horch, A. and A. J. Isaksson, "A Modified Index for Control Performance Assessment," *J. Proc. Cont.*, **9**, 475–483 (1999).

Horch, A., *Condition Monitoring of Control Loops*, Ph. D. Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden (2000).

Huang, B. and S. L. Shah, "Practical Issues in Multivariable Feedback Control Performance Assessment," *J. Proc. Cont.*, **8**, 421–430 (1998).

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

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

Huang, B., S. L. Shah, K. E. Kwok, and J. Zurcher, "Performance Assessment of Multivariate Control Loops on a Paper-Machine Headbox," *Can. J. Chem. Eng.*, **75**, 134–142 (1997c).

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 (1997a).

Huang, B., S. L. Shah, and R. Miller, "Feedforward Plus Feedback Controller Performance Assessment of MIMO Systems," *IEEE Trans. Cont. Sys. Tech.*, **3**, 580–587 (2000).

Isaksson, A. J., A. Horch, and G. A. Dumont, Event-triggered deadtime estimation: A comparison of methods, In *Proc. Control Systems 2000*, pages 209–215, Victoria, BC (2000).

Isaksson, A. J., PID Controller Performance Assessment, In *Proc. Control Systems '96*, pages 187–194, Halifax, NS (1996).

Isermann, R. and P. Ballé, "Trends in the Application of Model-Based Fault Detection and Diagnosis of Technical Processes," *Control Eng. Practice*, **5**(5), 709–719 (1997).

Jofriet, P. and B. Bialkowski, Process Knowledge: The Key to On-Line Monitoring of Process Variability and Control Loop Performance, In *Proc. Control Systems 96*, pages 163–169, Halifax, NS (1996).

Jofriet, P., C. Seppala, M. Harvey, B. Surgenor, and T. Harris, "An Expert System for Control Loop Performance," *Pulp and Paper Canada*, **6**, 207–211 (1996).

Kendra, S. J. and A. Çinar, "Controller Performance Assessment by Frequency Domain Techniques," *J. Proc. Cont.*, **7**(3), 181–194 (1997).

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, Assessment of Achievable PI Control Performance for Linear Processes with Deadtime, In *Proc. American Control Conference*, pages 1548–1552, Philadelphia, PA (1998).

Ko, B. and T. F. Edgar, "Performance Assessment of Cascade Loops," *AIChE J.*, **46**, 281–291 (2000a).

Ko, B. and T. F. Edgar, Performance Assessment of Constrained Model Predictive Control Systems (2000b). Submitted to *AIChE Journal*.

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

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

Kozub, D. J., Monitoring and Diagnosis of Chemical Processes with Automated Process Control, In *CPC V Proceedings*, Lake Tahoe (1996).

Kucera, V., *Discrete Linear Control: The Polynomial Equation Approach*. John Wiley & Sons, New York (1979).

Lütkepohl, H., *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin (1991).

Miller, R. M. and L. D. Desborough, "Web-enabled control loop assessment," *Chemical Engineering* (2000).

Miller, R. M. and B. Huang, Perspectives on Multivariate Feedforward/Feedback Controller Performance Measures for Process Diagnosis, In *Proc. IFAC Adchem 97*, pages 435–440, Banff, AB (1997).

Mutho, Y. and R. Ortega, "Interactor Structure Estimation for Adaptive Control of Discrete-Time Multivariable Nondecouplable Systems," *Automatica*, **29**, 635–647 (1993).

Mutho, Y., "A note on the invariant properties of ineractors," *Int. J. Control*, **62**, 1247–1252 (1995).

Owen, J. G., D. Read, H. Blekkenhorst, and A. A. Roche, A Mill Prototype for Automatic Monitoring of Control Loop Performance, In *Proc. Control Systems 96*, pages 171–178, Halifax, NS (1996).

Owen, J. G., Automatic control loop monitoring and diagnostics (1997). Patent Application. Int. Publ No WO 97/41494, Int. Appl. No PCT/CA97/00266.

Peng, Y. and M. Kinnaert, "Explicit Solution to the Singular LQ Regulation Problem," *IEEE Trans. Auto. Cont.*, **37**, 633–636 (1992).

Qin, S. J., "Control Performance Monitoring—A Review and Assessment," *Comput. Chem. Eng.*, **23**, 173–186 (1998).

Rogozinski, M. W., A. P. Paplinski, and M. J. Gibbard, "An Algorithm for Calculation of a Nilpotent Interactor Matrix for Linear Multivariable Systems," *IEEE Trans. Auto. Cont.*, **32**, 2234–2237 (1987).

Seborg, D. and T. Miao, Automatic detection of excessively oscillatory feedback control loops, In *IEEE Int. Conf. on Control Applications*, pages 359–364, Hawaii, USA (1999).

Seppala, C. T., T. J. Harris, and D. W. Bacon, Time Series Methods for Dynamic Analysis of Multiple Controlled Variables (2001). Submitted to *Journal of Process Control*.

Seppala, C. T., *Dynamic Analysis of Variance Methods for Monitoring Control System Performance*, Ph. D. Thesis, Queen's University, Canada (1999).

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).

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

Tsiligiannis, C. A. and S. A. Svoronos, "Dynamic Interactors in Multivariable Process Control—I: The General Time Delay Case," *Chem. Eng. Sci.*, **43**, 339–347 (1988).

Tyler, M. L. and M. Morari, Performance assessment for unstable and nonminimum-phased systems, In *Proc. IFAC Workshop on On-Line Fault Detection and Supervision in the Chemical Process Industries*, pages 200–205, Newcastle-upon-Tyne, UK (1995).

Tyler, M. L. and M. Morari, "Performance Monitoring of Control Systems Using Likelihood Methods," *Automatica*, **32**, 1145–1162 (1996).

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).