15th Triennial World Congress, Barcelona, Spain

CONTROL SYSTEM PERFORMANCE MONITORING: NEW DEVELOPMENTS AND PRACTICAL ISSUES

G.A. Dumont [*] **L. Kammer** [**] **B.J. Allison** [**]
L. Ettaleb [***] **and A.A. Roche** [***]

[*] *Department of Electrical and Computer Engineering c/o Pulp and Paper Centre, 2385 East Mall University of British Columbia, Vancouver, BC V6T 1Z4 Canada*
email: guyd@ppc.ubc.ca
[**] *Paprican Vancouver Laboratory, 3800 Wesbrook Mall, Vancouver, BC V6S 2L9 Canada*
[***] *Paprican, 570 Boul. St Jean, Pointe-Claire, Que, H9R 3J9 Canada*

Abstract: This paper reviews some of the techniques and measures used in control loop performance monitoring systems. Basic tools are briefly reviewed. The Harris index, which for a single-input, single-output loop quantifies the distance from minimum variance, requires knowledge of the process dead time. Recently developed techniques for dead time estimation from operating data are discussed. A novel model invalidation technique is illustrated via an industrial example. Finally, issues of data quality are touched upon. *Copyright ©2002 IFAC*

Keywords: Delay estimation; Minimum variance control; Modelling errors; Performance monitoring; Predictive Control; Time series analysis.

1. INTRODUCTION

The last ten years have seen the emergence and rapid adoption of control loop performance monitoring techniques, (Harris (1988), Åström (1991), Perrier and Roche (1992), Kozub (1996), Huang and Shah (1999)). The purpose of on-line monitoring of control loops is to obtain information on the performance of a plant under control while it is operating with minimal disturbances to its normal operation. Various studies indicate that the half-life of control loop performance is about six months, Bialkowski (1992). Typically it takes two hours to manually audit the performance of a control loop. The typical process plant has between 2000 and 4000 loops and many plants do not have personnel with required skills. Because the average process engineer is responsible for 400 loops the availability of a set of tools to automatically estimate the performance of control loops is of great importance. For regulatory loops, minimum variance provides a benchmark. Harris (1988) showed that it is possible to estimate the minimum variance from closed-loop operating data without knowledge of the plant dynamics. After reviewing basic control loop performance monitoring methods, a recently developed model invalidation technique and its industrial application will be described. Finally, some important implementation issues will be discussed.

2. THE BASIC TOOLS

Figure 1 shows the components of a typical performance monitoring system.

2.1 *The Harris Index*

Consider the system

$$y(t) = \frac{B(q^{-1})}{A_1(q^{-1})} q^{-k} u(t) + \frac{C(q^{-1})}{A_2(q^{-1})\Delta^d} e(t)$$

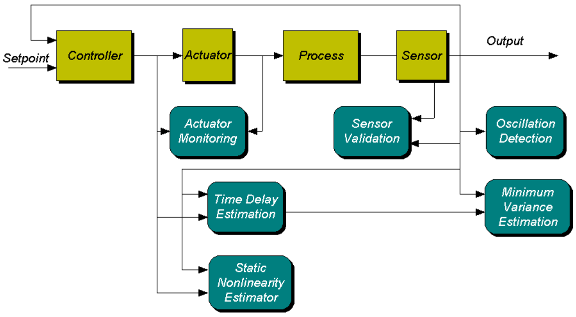

Fig. 1. A typical performance monitoring system

If the process is minimum phase, the minimum variance controller is

$$u(t) = -\frac{A_1(q^{-1})G(q^{-1})}{B(q^{-1})F(q^{-1})A_2(q^{-1})\Delta^d}y(t)$$

with

$$C(q^{-1}) = F(q^{-1})A_2(q^{-1})\Delta^d + q^{-k}G(q^{-1})$$

The minimum variance is then

$$\begin{aligned}\sigma^2_{\text{mv}} &= E[y^2(t+k)] = E\{[F(q^{-1})e(t+k)]^2\}\\ &= \sigma^2_e(1+f_1^2+f_2^2+\ldots+f_{k-1}^2)\end{aligned}$$

When the system is under feedback control with the controller

$$u(t) = \frac{N(q^{-1})}{D(q^{-1})}y(t)$$

he closed-loop system is described by

$$y(t) = \frac{CA_1D}{\Delta^d[A_1A_2D - BNA_2q^{-k}]}e(t) \triangleq H(q^{-1})e(t)$$

Using the Diophantine equation, one can write

$$y(t) = Fe(t) + q^{-k}\frac{BNA_2\Delta^dF + GA_1D}{(A_1A_2D - BNA_2q^{-k})\Delta^d}e(t)$$

Note that the first k terms of H, represented by F are unaffected by the controller, i.e. are feedback invariant and allow the computation of the system's minimum variance. Estimating H in $y(t) = He(t)$ from closed-loop data, it is then easy to compute the minimum variance σ^2_{mv} from the first k terms of the impulse response of H. The performance of the regulatory loop can then be measured by the ratio of the actual variance to the minimum one:

$$P = \frac{\sigma^2_y}{\sigma^2_{\text{mv}}}$$

Thus, one only needs identifying the output of the process as a time-series and computes the first k terms of its impulse response to find F. This does not require perturbating the plant and does not present any closed-loop identifiability problem. Many different structures can be chosen for $H(q)$, e.g. an AR model, an ARMA model or a Laguerre model. For example with the latter, the time series used to represent the closed-loop system can be modelled as white noise $e(t)$ filtered through a Laguerre network, Lynch and Dumont (1996):

$$\begin{aligned}l(t+1) &= Al(t) + be(t)\\ y(t) &= c^Tl(t) + e(t)\end{aligned}$$

The vector c is easily estimated from closed-loop data using least-squares, and the F-polynomial coefficients are computed as:

$$f_0 = 1, \quad f_1 = c^Tb, \quad f_2 = c^TAb, \quad f_3 = c^TA^2b, \quad \cdots$$

Generally, a comparison to minimum variance control is overly pessimistic. It does not take into account the existence of input constraints or the presence of non-minimum phase zeros. Most importantly, it does not take into account the structure of the present controller. Various researchers have proposed a performance index that allows the closed-loop response to a load disturbance to decay exponentially after the dead time. For example, Horch (2000) proposes to place one pole $q = \mu$ in which case the benchmark variance σ^2_{bench} becomes

$$\sigma^2_{bench} = \sigma^2_{mv} + f^2_{k-1}\frac{\mu^2\sigma^2_e}{1-\mu^2} = \sigma^2_{mv} + \sigma^2_\mu$$

Various ways to choose μ are possible, but according to Horch (2000) $\mu = 0.5$ is reasonable in most situations. Other simply propose to extend the moving average used to compute the benchmark variance beyond the dead time of the process. Note that the terms beyond the dead-time are then not feedback-invariant.

2.2 *Estimation of Time Delay*

The original Harris index relies on an estimate of the process time delay. Although in practice a very accurate knowledge of the delay is not crucial, an automated way of estimating it is necessary to facilitate the commissioning of a performance monitoring system. Keeping in mind the non-intrusive nature of performance monitoring tools, it is of interest to have a reliable method for closed-loop estimation of time delay using normal operating data. A method that has proven reliable in practice is described in Isaksson *et al.* (2001).To ensure proper excitation, it uses transient data following a setpoint change. A discrete Laguerre model of the plant is then estimated. The time delay can then be calculated directly from the zeros of the discrete Laguerre model.

The Laguerre model of the plant can be factored as

$$G(z) = G_{\text{mp}}(z)G_{\text{ap}}(z)$$

where $G_{\text{mp}}(z)$ is minimum phase and $G_{\text{ap}}(z)$ is all-pass. Let $\varphi = \arg(G_{\text{ap}})$. The dead-time estimate is then given as

$$\hat{T}_d = \lim_{\omega\to 0}\left(-\frac{\varphi(\omega)}{\omega}\right)$$

For the sampled system, if T is the sampling interval, the dead time in number of samples is

$$\hat{k} = 1 + \frac{\hat{T}_d}{T}$$

2.3 *Oscillation Detection*

Oscillations occur in many control loops. The presence of oscillation renders the Harris index invalid, and so there is a need to detect oscillations before proceeding to the computation of the Harris index. Furthermore, oscillations are very detrimental and need to be diagnosed. Major causes for loop oscillations are i) Load disturbances near the ultimate frequency that are too fast to be treated by the controller and too slow to be filtered out, ii) Poor controller tuning, especially on nonlinear systems, iii) Valve friction resulting in stick-slip motion and a limit cycle which is often far from sinusoidal. Hägglund (1994) has proposed a simple yet elegant way of detecting oscillations in control loops. Once a loop has been flagged as oscillatory, finding the cause of oscillation can be a challenge. To detect the presence of stiction in control valves, Horch (1999) has proposed a simple method based on cross-correlation between control input and process output.

2.4 *Multivariate Extensions*

Various research groups have developed multivariable extensions to the original Harris index, see Huang and Shah (1999) and the references therein. The performance assessment of constrained model predictive controllers has been looked at in Ko and Edgar (2001). In the process industries, such controllers typically use proprietary model-based predictive control algorithms. A fundamental question with such systems, beyond the suitability of a minimum-variance based benchmark, is whether in case of poor performance the controller tuning or the model used to design the controller is at fault. A recently developed technique that attempts to answer this question will now be reviewed and illustrated by means of an application to an industrial lime kiln.

3. MODEL INVALIDATION FOR MODEL-BASED PREDICTIVE CONTROLLERS

Suppose a multivariable closed-loop system suffers an upset that causes some signal variances to increase and remain at a higher level from that point onwards. It is important to know whether the problem is due to an increase in the noise/disturbance level, process-model mismatch or controller tuning. It is usual to put some control loops in manual (off the cascade control), and see if the process variables settle down. For model-based controllers, the problems are possibly related to bad models. These problems could be fixed by re-identifying the whole multivariable model and re-tuning the controller, but this is an expensive procedure.

The model-invalidation mechanism, recently developed by Kammer *et al.* (2001), deals with the scenario described above and reveals if the model embedded in the controller is no longer valid. That mechanism indicates which part of the multivariable model is wrong, so that only part of the model needs to be re-identified. We assume that the only signals available from the system are collected after the problem is detected and no measurable external excitation occurs during data collection, that is, the only excitation driving the loop comes from stationary noise. Moreover, the actual control law is not used by the mechanism since for most of the commercially available model-based predictive controllers the details of the controller design are unavailable to the plant personnel.

The model-invalidation mechanism compares two time series associated with each output: the model output error $(y(t)-\hat{y}(t))$, collected during normal operation, and the open-loop output, collected when the controller is put in manual. If the process model is correct, then those two time series present the same behaviour. In order to compare these signals, a certain "distance" is measured between the two independent time series. All developments are made on a probabilistic framework, therefore there is a significance level associated with acceptance/rejection of the hypothesis that a particular part of the model is correct. Given that uncertainties are always present in models identified from measured data, the mechanism for comparison of time series also includes a provision for accommodating model uncertainties, or more accurately, uncertainties in the closed-loop response.

3.1 *Model Output Error*

Although the full process dynamics could encompass sources of measurable disturbance, the scenario analyzed here is restricted to unmeasurable sources of stochastic noise. Therefore the process dynamics is given by

$$y(t) = G(q)\,u(t) + H(q)\,e(t), \tag{1}$$

where $y(t)$ is the vector of n_y output signals (controlled variables), $u(t)$ is the vector of n_u input signals (manipulated variables) and $\{e(t)\}$ is a vector of n_e independent zero-mean Gaussian white-noise processes. The transfer-function matrices $G(q)$ and $H(q)$ are stable and have dimensions $n_y \times n_u$ and $n_y \times n_e$, respectively.

The reference signal is kept at a constant value (taken as zero without loss of generality) so the control action is described as:

$$u(t) = -C(q)\,y(t), \tag{2}$$

where $C(q)$ is an unknown operator designed from a known model of the process:

$$\hat{y}(t) = \hat{G}(q)\,u(t). \tag{3}$$

From (1) and (3) we define the model output error as

$$\begin{aligned}\varepsilon(t) &\triangleq y(t) - \hat{y}(t) \\ &= [G(q) - \hat{G}(q)]\, u(t) + H(q)\, e(t) \qquad (4) \\ &= [I + \hat{G}(q)C(q)][I + G(q)C(q)]^{-1} H(q)\, e(t).\end{aligned}$$

This shows that if the model is perfect ($\hat{G}(q) \equiv G(q)$) then the model output error presents the same dynamics as the process noise, $H(q)e(t)$. This fact is essential to the model-invalidation mechanism, as it compares the model output error, $\varepsilon(t)$, and the open-loop output signal,

$$y^o(t) = H(q)\, e(t). \qquad (5)$$

The need for collecting open-loop data implies that the plants are open-loop stable. This is a minor issue since for unstable processes the common industrial practice is to form an inner loop with a stabilizing feedback controller and have the model predictive controller at the outer loop. It is imperative that the noise dynamics, $H(q)$, be time invariant, at least for the duration of the experiment. The invalidation test thus becomes a trial of the assumption that, for each process output $j = 1, \ldots, n_y$, the independent time series $\{\varepsilon_j(t)\}$ and $\{y_j^o(t)\}$ are realizations of the same stationary process. The outcome of this test is probabilistic in nature since there are no hard bounds assumed on $\{e(t)\}$. The highest sensitivity to mismatches between $G(q)$ and $\hat{G}(q)$ occurs in the direction and frequency range where $G(q)C(q)$ is closest to $-I$. Therefore the invalidation test is increasingly more robust to uncertainties in $\hat{G}(q)$ as these uncertainties are less relevant for control purposes.

3.2 *Comparison of Two Time Series*

The invalidation test addresses the following problem statement: *Given two independent time series, $\{z_1(t)\}$ and $\{z_2(t)\}$, test the assumption that they are realizations of the same stationary process.*

The solution developed by Kammer *et al.* (2001) computes the “distance” between smoothed periodograms of the time series. A very important characteristic of that “distance” is that if $\{z_1(t)\}$ and $\{z_2(t)\}$ are realizations of the same stationary process, then its (cumulative) distribution function depends only on the number of samples in the time series. From this particular situation it is possible to derive a threshold level for testing the null hypothesis with a given significance level.

From a mathematical point of view, the hypothesis that $\{z_1(t)\}$ and $\{z_2(t)\}$ are realizations of the same stationary process is not rejected as long as the “distance” between smoothed periodograms is less than the threshold level. This test has a given probability of committing a Type I error (false alarm). Since this type of error should be avoided at all costs, the significance level is typically a small value, e.g. 1%. Hence any violations of the threshold are attributed to model uncertainty and/or model mismatch. A complete description of the invalidation test and the analysis of the model uncertainty is contained in the original paper where this mechanism was developed (Kammer *et al.*, 2001).

3.3 *Industrial Example—Lime Kiln*

The model-invalidation mechanism was employed in the analysis of an industrial lime kiln under model-based predictive control. The process under evaluation comprises two inputs, amount of fuel (F) and fan speed (S), and two outputs, cold-end temperature (T_c) and percentage of oxygen (O_2). A third output, the hot-end temperature, is not included in this analysis due to the special structure of the controller implementation.

The data used in the analysis are shown in Figure 2. Observe that the open-loop control actions are not held constant at all times, as an operator was keeping the controlled variables within a specified range. This implies that the test is going to contrast pairs of model output errors: one time series obtained with the model-based predictive controller, $\{\varepsilon_j(t)\}$, and the second one obtained with manual control, $\{\varepsilon_j^o(t)\}$. The model used in

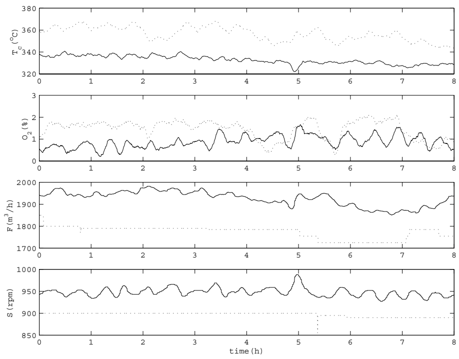

Fig. 2. Lime kiln data: closed loop (solid lines) and open loop (dotted lines)

the controller design is provided in its continuous-time form:

$$\begin{aligned}T_c(s) &= \frac{0.16}{140s + 1} F(s) + \frac{0.21}{10s + 1} S(s) \\ O_2(s) &= -\frac{0.013}{6.1s + 1} F(s) + \frac{0.026}{3.5s + 1} S(s)\end{aligned}$$

where all time constants are in minutes and the sampling time for all variables is 10 seconds. The ratio between the smallest time constant and the sampling time suggests that only a limited range of frequencies needs attention. This conclusion is also enforced by the spectral analysis of the open-loop signals, Figure 3, which shows a very strong concentration of energy at low frequencies.

In principle this concentration of energy should not be a problem, but the whitening of such signals causes a huge relative amplification of low-amplitude high-frequency noise. Therefore we perform our test exclusively in the frequency range between 0 rad/s and the frequency corresponding to half of the smallest time constant, that is, 0.06 rad/s.

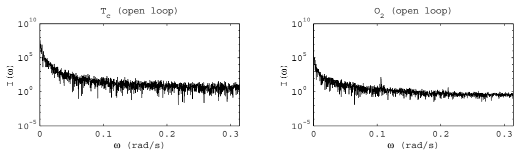

Fig. 3. Frequency spectra of the open-loop signals

Figure 4 shows the results of the test for each output. Notice that there is not enough evidence in these results to invalidate the model in use. The main reason for that is the high degree of conservativeness adopted in the control design. If one were to demand more performance from the control action, simple models would be invalidated more frequently.

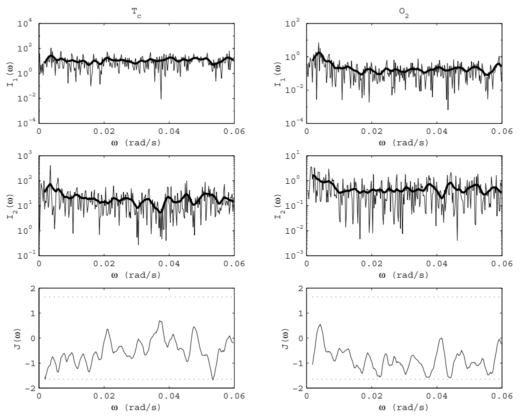

Fig. 4. Smoothed periodograms and their "distance", for each output

4. IMPLEMENTATION ISSUES

All previously described methods require raw data collected at a proper sampling frequency. Although availability of data suitable for control loop performance monitoring is usually taken for granted, accessing such data is not always a trivial task in an industrial setting. Our experience in the pulp and paper industry shows that there are indeed many limitations to obtaining useful operating data. Thornhill *et al.* (1999) discuss the implementation of a control loop performance assessment in a refinery, touching on the same issues at greater length.

Physical limitations: Old equipment abounds in the process industries. Some distributed control systems (DCS) were never updated and use proprietary communication protocols, making them hard to interface with or network and, in many cases, unable to provide high frequency time series.

Lack of standardization: Lack of standardization between instrument manufacturers and control software manufacturers is a major obstacle when collecting and transferring data to different users. In an attempt to avoid writing custom drivers for every supplier, a new standard for hardware and software was created, called OPC (OLE for Process Control). OPC may facilitate data collection and transfer. Today, in many process plants, acquiring an archival system seems to be the preferred solution to centralizing data.

Data storage: Some plants are equipped with the latest process equipments and information technology. Not only can they retrieve operating data but also store it in databases called archiving systems or sometime process information systems. These systems communicate with DCS from different vendors through OPC standards or sometimes through specially designed drivers and act as a central bank for all operating data. However, due to the large number of loops and tags often involved, to the limited archiving capacity and to the limited bandwidth of the plant network, data is often manipulated before archival. Because of that, archived data suffers from two major problems:

Scan frequency: Archival systems communicate and receive data from different DCS system according to a certain specific scan frequency. This frequency is different from the frequency at which the DCS computes the control action. The scan frequency is set by the user. It takes into account the network structure and the time constant of the process variables involved. A scan rate that is not properly chosen may induce aliasing, making it unsuitable for most of the techniques previously described.

Data manipulations: Typically, data is archived if it satisfies exception reporting and compression conditions. In the report by exception method, only values that have changed by a user-defined amount from their last archived values will be passed to the next test. Data that successfully passes the exception-reporting specifications is subject to another test to determine whether there exist a certain linear relationship between data previously stored and the newly received. In case of such a linear relationship, only two coordinates are stored and all other coordinates are to be determined from the linear relationship. Otherwise the received data is archived. This process is to ensure that only significant data are written to the archive. These manipulations introduce some filtering therefore the archived data loses its raw data form. If the parameters for exception reporting and compression are set too high then the data stored is not suitable for control loop performance monitoring.

Example: Data shown in Figure (5) is from an inner cascaded loop. That figure shows raw data for process variable (pv) and set-point (sp) while Figure (6) shows the compressed data where the report by exception parameter is set to 2% and the compression parameter set to 4% Although no significant difference is apparent, the actual variance of pv-sp for the raw data is $\sigma^2_{pv-sp} = 32.15$ while for the compressed data it is $\sigma^2_{(pv-sp)_{comp}} = 53.85$. The increase in variance has nothing to do with the performance of this loop, it simply results from archiving manipulations. Such archival systems may soon become the only source of operating data. For successful control loop performance monitoring it is important to ensure quality data by properly setting the parameters involved in data manipulations. This may also grant the development of more efficient data compression algorithms that result in less data distortion.

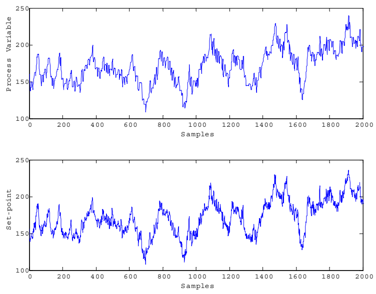

Fig. 5. Raw data (non-compressed).

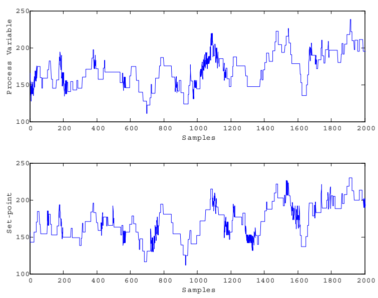

Fig. 6. Compressed data.

5. CONCLUSIONS

We have seen the basic tools of control performance monitoring. A simple technique for invalidating models in model based predictive control was proposed and tested on industrial data. Finally, some implementation issues, which tend to be overlooked and sometimes dismissed as trivial were touched upon.

6. REFERENCES

K.J. Åström (1991). Assessment of Achievable Performance of Simple Feedback Loops. *International Journal of Adaptive Control and Signal Processing*. 5(1):3-19.

W.L. Bialkowski. Dreams vs. Reality: A View from Both Sides of the Gap. In Control Systems '92: Dream vs. Reality, Whistler, B.C., September, 1992.

T. Hägglund (1994). Automatic monitoring of control loop performance monitoring. *Control Systems'94*. Stockholm, 190-196.

T.J. Harris (1988). Assessment of Control Loop Performance. *The Canadian Journal of Chemical Engineering*. 67:857–862.

A. Horch (1999). A simple method for detection of stiction in control valves. *Control Engineering Practice*. 7(10):1221-1231.

A. Horch (2000). Condition Monitoring of Control Loops. Doctoral Thesis, Royal Institute of Technology, Stockholm.

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

A. Isaksson, A. Horch and G. Dumont 2001). Event-triggered deadtime estimation from closed-loop data. *American Control Conference*. Arlington, VA.

Kammer, L. C., D. Gorinevsky and G. A. Dumont (2001). Semi-intrusive multivariable model invalidation. *Proc. of the European Control Conference*. Porto, Portugal.

Ko and T. Edgar (2001). Performance Assessment of Constrained Model Predictive Control Systems. *AIChE Journal*. 47(6):1363-1371.

D.J. Kozub (1996). Controller performance monitoring and diagnosis: experiences and challenges. *Chemical Process Control-V*. Tahoe City, CA.

C. Lynch and G.A. Dumont (1996). Control Loop Performance Monitoring. *IEEE Transactions on Control Systems Technology*. v. 4, no. 2, pp. 185-192.

M. Perrier and A.A. Roche (1992). Towards Mill-Wide Evaluation of Control Loop Performance. *In Control Systems '92: Dream vs. Reality*. Whistler, B.C.

N.F. Thornhill, M. Oettinger and P. Fedenczuk (1999). Refinery-wide control loop performance assessment. *J. of Process Control*. 9:109–124.