

Methods for Performance Monitoring and Diagnosis of Multivariable Model-based Control Systems

Seungyong Lee, Seunghoon Yeom and Kwang Soon Lee[†]

Dept. of Chemical and Biomolecular Engng., Sogang University, 1-Shinsoodong, Mapogu, Seoul 121-742, Korea
(Received 14 October 2003 • accepted 10 January 2004)

Abstract—Methods for performance monitoring and diagnosis of multivariable closed loop systems have been proposed aiming at application to model predictive control systems for industrial processes. For performance monitoring, the well-established traditional statistical process control method is employed. To meet the underlying premise that the observed variable is univariate and statistically independent, a temporal and spatial decorrelation procedure for process variables has been suggested. For diagnosis of control performance deterioration, a method to estimate the model-error and disturbance signal has been devised. This method enables us to identify the cause of performance deterioration among the controller, process, and disturbance. The proposed methods were evaluated through numerical examples.

Key words: Process Monitoring, Process Diagnosis, Model Predictive Control, PCA, Identification

INTRODUCTION

Model predictive control (MPC) has now more than twenty years of history of industrial application. According to a recent survey [Qin and Badgwell, 2003], more than 4,500 cases of process implementation have been reported until 2001 worldwide. In industries, processes are necessarily subject to aging, modifications, and changes in operating conditions. All these result in performance deterioration of MPC which was optimized for the original process situation. Since MPC, unlike PID controller, is not easy to maintain by operation personnel, there has been a strong need for on-line performance monitoring and diagnosis systems for MPC and other advanced process control systems.

For the single-input single-output (SISO) case, the ratio of the control error variance under the present control loop to that under minimum variance control (MVC) has gained general acceptance as a standard control performance index after Harris [1989] proposed the concept first time. This idea has since attracted significant interests and has been developed further by many researchers [Stanfelj et al., 1993; Kozub and Garcia, 1993; Huang et al., 1995; Qin, 1997]. Later, Huang et al. [1997] and Harris et al. [1996] extended the MVC-based SISO assessment method to the multi-input multi-output (MIMO) case. Recently, Matrikon Application Co. has commercialized the technique in a software package, ProcessDocTM, and reported many successful industrial implementations [Matrikon, 2003]. In the next section, we will briefly review the MVC-based methods as implemented in ProcessDocTM.

While the above cited researches have been devoted only to monitoring, Kesavan and Lee [1997] proposed several MIMO control loop diagnosis tools based on the prediction error (PE) and other in-depth diagnosis techniques based on parallel filters. Since the PE can be calculated only when the disturbance model is available while most commercial MPC's still rely only on the input-output part

of the model, the Kesavan's method is thought to be still early for practical application.

Considering the above general background, the purpose of this research has been placed in developing a new line of closed loop performance monitoring and diagnosis methods for industrial MPC's or other model-based MIMO control systems. The performance monitoring method is devised based on the well-established traditional univariate SPC technique [Mamzic, 1995; Box and Luceno, 1997]. For this, a whitening filter and PCA are introduced to decorrelate the process variables temporally as well as spatially. For diagnosis, a method to identify the model-plant mismatch is proposed. Using this method, the cause of control performance deterioration can be identified up to the controller or process or disturbance, but not into more detail.

MVC-BASED CLOSED LOOP PERFORMANCE MONITORING METHODS

As was reviewed in Introduction, the MVC-based methods constitute the main stream of the current closed loop performance monitoring methods (CCPMM). To elucidate the status of the proposed

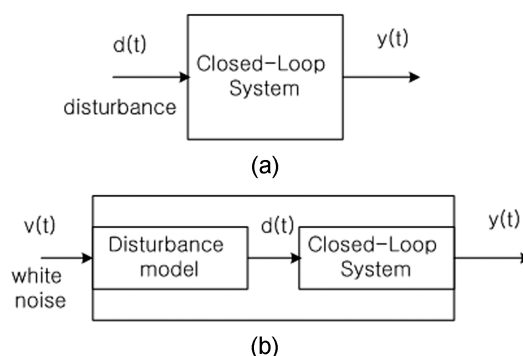


Fig. 1. Black box representations of a closed loop system under regulatory control.

[†]To whom correspondence should be addressed.
E-mail: kslee@sogang.ac.kr

methods in this paper, a brief review is made on the key concepts of the CCPMM.

A closed-loop system under regulatory control can be represented as in Fig. 1(a). Since virtually any stochastic disturbance can be represented as a filtered white noise, the closed-loop system can again be represented as in Fig. 1(b) with white noise as its input. A model for Fig. 1(b) can be identified using the output measurements $\{y(t)\}$, e.g.,

$$y(t) = h_0 z(t) + h_1 z(t-1) + \dots + h_n z(t-n) \quad (1)$$

where $\{z(t)\}$ is a zero-mean white noise sequence with variance σ_z^2 and $h_0=1$ without loss of generality. The CCPMM are centered around this model with various interpretations.

Assume that the process has d -step delay. Under this condition, $\{h_0, \dots, h_{d-1}\}$ are independent of feedback control and determined only by the process and disturbance models while $\{h_d, \dots, h_n\}$ vary depending on feedback control. Feedback control that gives $h_d = \dots = 0$ is called minimum variance control (MVC) since the output variance under this situation takes the minimum achievable value.

$$\sigma_{y,mv}^2 \triangleq \sigma_z^2 (h_0^2 + \dots + h_{d-1}^2) \quad (2)$$

From Eq. (1), the real output variance is given as

$$\sigma_y^2 \triangleq \sigma_z^2 \sum_{k=0}^n h_k^2 \quad (3)$$

The closed loop performance assessment index is defined as

$$\eta(d) \triangleq \frac{\sigma_{y,mv}^2}{\sigma_y^2} = \frac{\sum_{k=0}^{d-1} h_k^2}{\sum_{k=0}^n h_k^2} \quad (4)$$

which is between 0 and 1. In general, d is not accurately known. Hence, performance assessment is carried out for different different values of d .

Sometimes, $G(q^{-1}) = \sum_{k=0}^n h_k q^{-k}$ is represented in the frequency domain together with $G_{MVC}(q^{-1}) = \sum_{k=0}^{d-1} h_k q^{-k}$. Comparing these two, one may get more insight on the cause of poor closed loop performance.

In addition to the above, statistical tests such as autocorrelation function of control error, which should be zero for more than $d-1$ lag under MVC, and residual test for validation of Eq. (1) are conducted as supplementary assessment tools.

In the MIMO control loop assessment, the above methods are applied to individual outputs.

Though many successful applications have been reported, the CCPMM have some shortcomings. First, it lacks the diagnosis capability. When poor closed loop performance is detected, it cannot determine if the culprit is control loop itself or a large or unmanageable disturbance. The reason for this is that the impulse response model $\{h_i\}$ represents the combined closed loop and disturbance model (Fig. 1(b) instead of Fig. 1(a)). Hence, even when the closed loop is all right, $\{h_i\}$ may have poor values. Second, in MIMO systems, the statistical tests are conducted on individual outputs without considering the possible correlation between the outputs. Thirdly, the time delay, a prerequisite information for analysis, is not easy to know, especially in MIMO cases.

SPC-BASED PERFORMANCE MONITORING

May, 2004

The statistical monitoring typified by the Schwart chart and CUSUM chart monitoring is a mature field having a long history of industrial practice. This technique was developed only for univariate random variables and gives valid statistical interpretation when the monitored variables are results of independent experiments. Such a premise can be very often satisfied for observed variables in manufacturing industries. In process industries, however, observed process variables are subject to dynamics, *i.e.*, have temporal correlations. This violates the assumption of the outcome of independent experiments. Moreover, in case of multivariate observation, monitored variables may have spatial correlations. This implies that blind application of SPC technique to process variables may easily fail.

In this research, we devised a decorrelation procedure and propose to employ the existing rich SPC technique as a MIMO closed loop monitoring tool. The decorrelation is conducted into two steps: temporal decorrelation and then spatial decorrelation.

Temporal Decorrelator - Whitening Filter: Using control error $\{\varepsilon(t) \in \mathbb{R}^n; t=1, \dots, N\}$ measured under in-control state, a multivariable ARMA model is identified in the state space form using the N4SID [Overschee and DeMoore, 1994] or other standard identification techniques.

$$\begin{aligned} \varepsilon(t) = F(q)z(t) &\Leftrightarrow x(t+1) = Ax(t) + Kz(t) \\ \varepsilon(t) &= Cx(t) + z(t) \end{aligned} \quad (5)$$

The temporal decorrelator (whitening filter) can be constructed from this model through the following rearrangement:

$$\begin{aligned} z(t) = W(q)\varepsilon(t) = F(q)^{-1}z(t) &\Leftrightarrow x(t+1) = (A - KC)x(t) + K\varepsilon(t) \\ z(t) &= \varepsilon(t) - Cx(t) \end{aligned} \quad (6)$$

Hence, by processing $\varepsilon(t)$ with $W(q)$, a temporally decorrelated signal $z(t)$ is obtained.

Spatial Decorrelator: To the collection of whitened signal $Z = [z(1) \dots z(N)]$ for the in-control state measurements, the principal component analysis (PCA) is applied such that

$$Z \approx PS \quad (7)$$

where $P = [p_1 \dots p_a]$ and $S = [s(1) \dots s(N)]$ with $a \ll N$ represent loading and score matrices for the major principal components of Z , respectively. The elements $s_i(t)$'s of $s(t)$ are temporally as well as spatially uncorrelated. The loading matrix is stored for future on-line monitoring and the score values are used to determine the control limits.

When a new $z(t)$ is obtained during on-line monitoring, it is projected on P to get the score such that

$$s(t) = P^T z(t) \quad (8)$$

and each $s_i(t)$ is monitored according to the SPC method.

Schwart Chart Monitoring: Monitoring $s_i(t)$'s can be conducted according to the standard Schwart and CUSUM chart methods. For example, in the Schwart chart, \bar{x} -bar which is defined for a disjoint subgroup for m -consecutive $s_i(t)$'s as

$$\bar{x}_i(k) = \frac{s_i(m(k-1)+1) + \dots + s_i(mk)}{m} \quad (9)$$

is monitored for each i . The two control limits, UCL and LCL, are determined using the in-control state data such that

$$\begin{aligned} UCL_i &= \bar{x}_i + \alpha_i \bar{R}_i \\ LCL_i &= \bar{x}_i - \alpha_i \bar{R}_i \end{aligned} \quad (10)$$

where

$$\bar{x}_i = \frac{\bar{x}_i(1) + \dots + \bar{x}_i(N)}{N} \text{ for a sufficiently large } N \quad (11)$$

$$\bar{R}_i = \frac{R_i(1) + \dots + R_i(N)}{N} \text{ for a sufficiently large } N \quad (12)$$

$$R_i(k) = \max_t s_i(t) - \min_t s_i(t), t \in [m(k-1)+1, mk]$$

α_i is given in relation to a specified risk level and can be found in a standard textbook like Box and Luceno [1997].

Remark: In CCPMM, $F(q)$ (in terms of the impulse response) in Eq. (5) is used for the closed loop monitoring while the present study monitors $z(t)$ (in the form of $s(t)$). In this respect, the two methods are complementary instead of competing. Note that CCPMM estimates $F(q)$ for each monitoring occasion while the proposed method in the present form estimates $F(q)$ only once in a certain in-control state.

DIAGNOSIS

The monitoring technique only gives a clue that something goes wrong in the closed loop but doesn't show which part in the closed loop is responsible for the performance deterioration.

The cause of poor control performance can be classified into three: inadequate controller design or tuning, large plant-model mismatch, and large and/or unmanageable disturbance. There may be different ways to identify the cause up to the above level. However, a method based on closed-loop identification is thought to give the most lucid conclusion. In this research, we propose a method to estimate the model error and the disturbance signal at the same time. From this result, one can determine which one is the most probable cause of the performance degradation.

Identification of plant-model mismatch: Fig. 2 shows a block-diagram that represents the situation of the proposed identification experiment. For unbiased model estimate, a zero-mean dither signal $\hat{u}(t)$ is superimposed at the input port. An important assumption of this method is that only the input-output part of the model, $G(q)$, except the disturbance model is used for MPC design, which is a general situation of the present commercial MPC's.

The block diagram analysis shows that the output error is given by

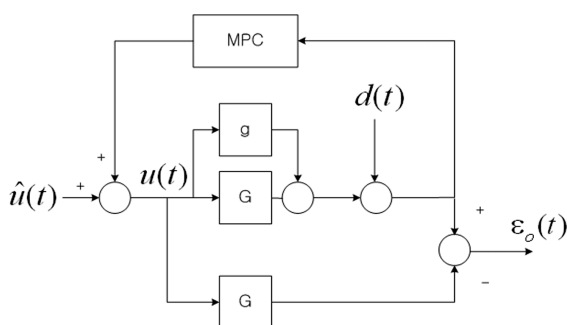


Fig. 2. A diagram to show the situation of the model error identification experiment.

$$\varepsilon_o(t) = g(q)(u(t) + \hat{u}(t)) + d(t) \quad (13)$$

Attempt to identify $g(q)$ via the least square method using $u(t) + \hat{u}(t)$ as the regressor results in a bias estimate since $u(t)$ and $d(t)$ are correlated through the feedback loop. To avoid this problem, we use the correlation method [Ljung, 1999] utilizing the fact that \hat{u} is uncorrelated with $d(t)$, i.e., $R_{\hat{u}d}(\tau) = 0$ for all τ . Then, we can derive the relation

$$R_{\varepsilon_o \hat{u}}(\tau) = g(q)(R_{u\hat{u}}(\tau) + R_{d\hat{u}}(\tau)) \quad (14)$$

and estimate $g(q)$ through the standard least squares method. Once the estimate $\hat{g}(q)$ is obtained, the disturbance signal can be reproduced according to

$$\hat{d}(t) = \varepsilon_o(t) - \hat{g}(q)(u(t) + \hat{u}(t)) \quad (15)$$

Now, through further investigation we can conclude who among the model error, disturbance, and controller is the culprit of the performance degradation. For this, we may carry out closed loop simulation for two cases where the process is assumed to be $G(q)$ and $G(q) + g(q)$, respectively, while injecting the estimated disturbance for both cases. Then the cause can be easily revealed by inspecting the resulting closed loop responses.

$g(q)$ can be represented as either a rational function matrix or finite impulse response matrix.

The model error estimate can be visualized in the frequency domain as an array of relative model error $[\hat{g}_{ij}(e^{j\omega})/G_{ij}(e^{j\omega})]$, $\omega \in [0, \pi]$. $\hat{g}(q)$ can also be used to correct the process model $G(q)$ on which the present MPC is based.

One thing to note is that the disturbance estimation is possible only when the whole $g(q)$, not a part of $g(q)$, is estimated. This can be illustrated using Fig. 3. If only $g_{11}(q)$ is estimated using $\{u_1(t), \hat{u}_1(t), y_1(t)\}$, the disturbance estimate according to Eq. (15) represents

$$y_1(t) - \hat{g}_{11}(q)u_1(t) = d_1(t) + q_{12}(q)u_2(t) \neq d_1(t) \quad (16)$$

The above method concerns only the input-output part of the process model. If the disturbance model H is available, too, methods based on prediction error can be used together to draw a more concrete conclusion. We introduce two such methods that are considered to be useful.

Cross-correlation Test: The cross-correlation test can be conducted without closed loop perturbation. Instead, it requires both $G(q)$ and the disturbance mode $H(q)$.

The test is conducted between the input $u(t)$ and the prediction error $\varepsilon(t)$. Fig. 4 shows the experimental situation.

The prediction error is represented as

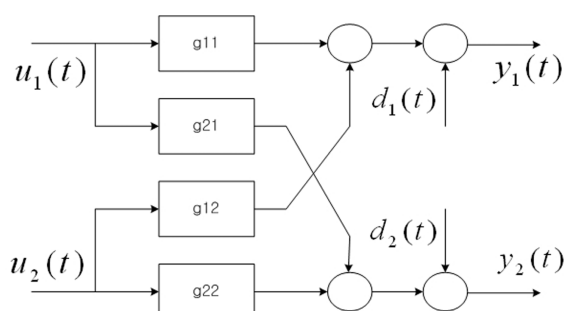


Fig. 3. A 2x2 system with output disturbance.

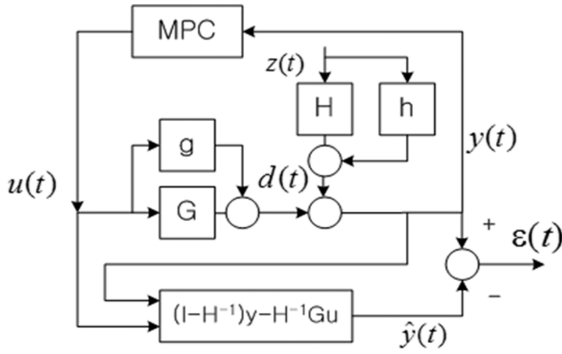


Fig. 4. A diagram to show the situation of the cross-correlation test.

$$\varepsilon(t) = y(t) - \hat{y}(t) = H(q)^{-1}(y(t) - G(q)u(t)) \quad (17)$$

From the fact that $y(t) = (G(q) + g(q))u(t) + (H(q) + h(q))z(t)$, the effect of model error on the prediction error becomes

$$\varepsilon(t) = (I + H(q)^{-1}h(q))z(t) + H(q)^{-1}g(q)u(t) \quad (18)$$

Under closed-loop control, $u(t)$ is necessarily affected by $z(t - \tau)$, $\tau \geq 0$. Thus $\varepsilon(t)$ and $u(t - \tau)$, $\tau \geq 0$ are independent only when $g(q) = h(q) = 0$. The independence can be checked by the following hypothesis testing [Lung, 1999]:

Under the hypothesis that $\varepsilon(t)$ and $u(t - \tau)$, $\tau \geq 0$ are independent, it holds that

$$\hat{R}_{\varepsilon u}^N(\tau) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t) u(t - \tau) \quad (19)$$

satisfies

$$\sqrt{N} \hat{R}_{\varepsilon u}^N(\tau) \rightarrow \mathcal{N}(0, P_1) \text{ where } P_1 = \sum_{k=-\infty}^{\infty} R_{\varepsilon}(k) R_u(k) \quad (20)$$

From $\mathcal{N}(0, P_1)$ (zero-mean normal density with variance P_1), we can determine a critical value for $\sqrt{N} \hat{R}_{\varepsilon u}^N(\tau)$ with a specified risk level and test if the hypothesis is acceptable or not, or equivalently, the model errors for both $G(q)$ and $H(q)$ can be negligible or not.

Detuning test: It is obvious that the (one step ahead) prediction error $\varepsilon(t)$ is unaffected by any change in the controller tuning as far as $g(q) = h(q) = 0$ because the process has at least one-step delay. The detuning test proposed by Kesavan and Lee [1997] is considered as another option that supplements the model error identification method.

NUMERICAL EXAMPLES

1. Linear System

In this example, we demonstrate the performance of the model-error identification method for a 2×2 linear system. The process $(G(q) + g(q))$ and the model $(G(q))$ are zero-order hold sampled equivalents of

$$\begin{bmatrix} \frac{2}{10s^2 + 7s + 1} & \frac{1}{24s^2 + 10s + 1} \\ \frac{1}{6s^2 + 5s + 1} & \frac{3}{40s^2 + 13s + 1} \end{bmatrix} \text{ and } \begin{bmatrix} \frac{1.8}{10s^2 + 7s + 1} & \frac{0.9}{24s^2 + 10s + 1} \\ \frac{0.9}{6s^2 + 5s + 1} & \frac{2.9}{40s^2 + 13s + 1} \end{bmatrix}, \quad (21)$$

respectively, with sampling period of $h=0.5$. Two independent integrated white noise sequences were added to each output as a disturbance. A standard MIMO MPC is installed to regulate the process output against the disturbance.

For model error identification, independent PRBS's (with an increased clock period for signal spectrum adjustment) were applied to $\hat{u}_1(t)$ and $\hat{u}_2(t)$, respectively. For identification, $g(q)$ was parameterized as $D(q)^{-1}N(q)$ where D and N are 2×2 diagonal and full polynomial matrices, respectively.

In Fig. 5, the estimated impulse response coefficients of the model-plant mismatch are shown in comparison with the true values. We can see that the proposed method yields highly reliable results.

Fig. 6 shows a part of the disturbance signals reproduced according to Eq. (15). This time, too, a satisfactory result was obtained.

2. BTX Distillation Column

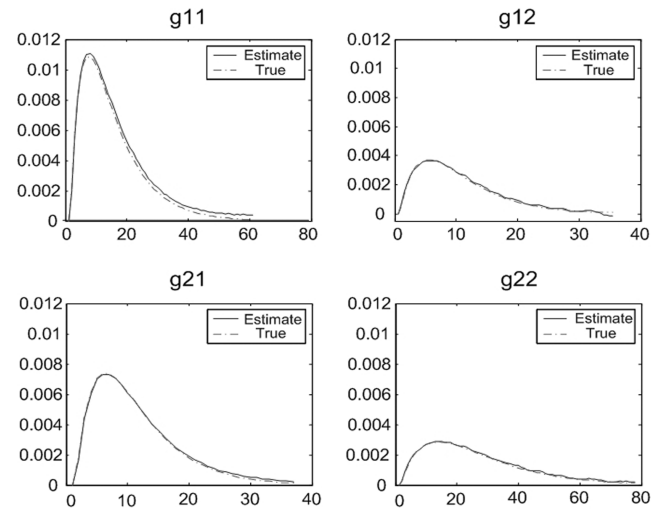


Fig. 5. Impulse response coefficients of the model-plant mismatch for the linear system example.

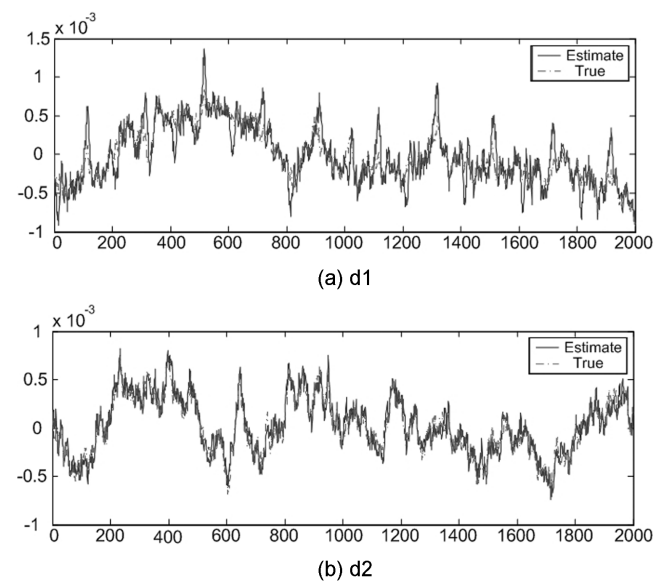
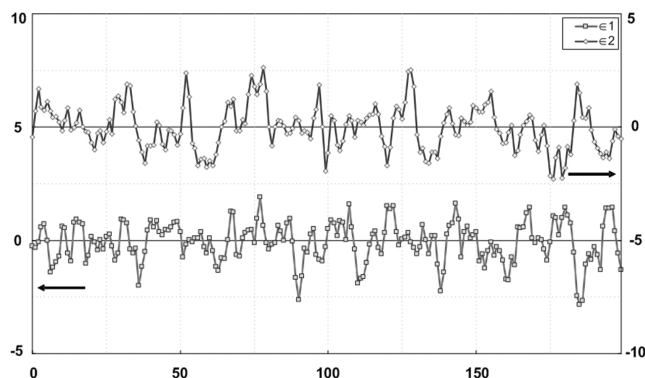
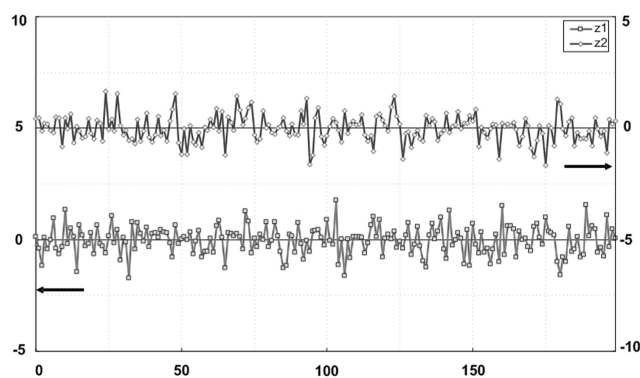


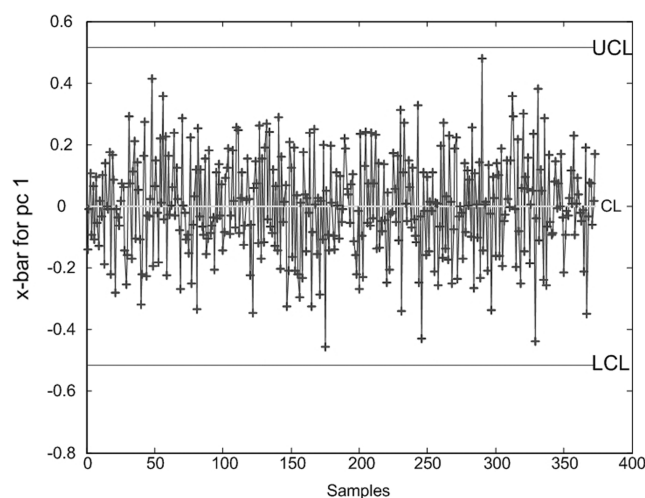
Fig. 6. Comparison of the estimated disturbance signals with the true ones for the linear system example.



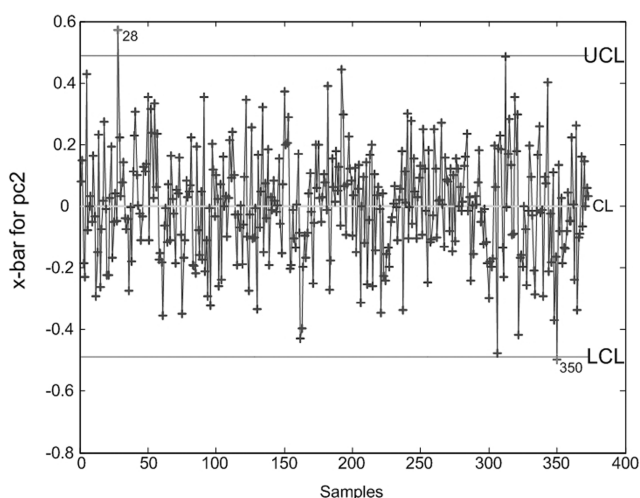
(a) Output error under the in-control state



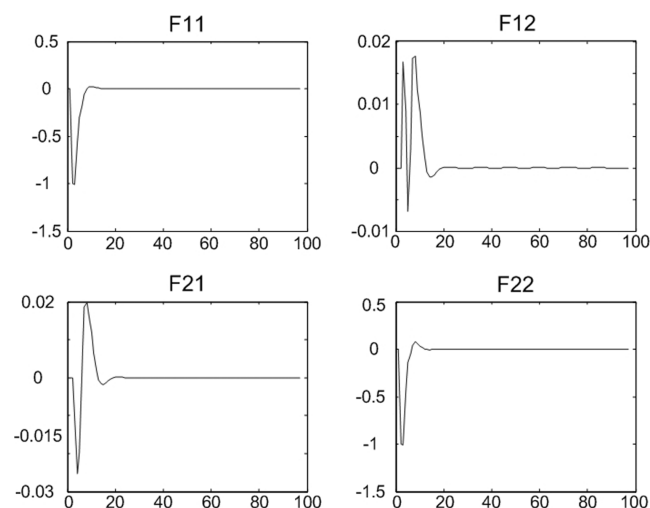
(b) Whitenized signal of the output error under the in-control state

Fig. 7. Control error and whitenized signals for the distillation column under in-control state.

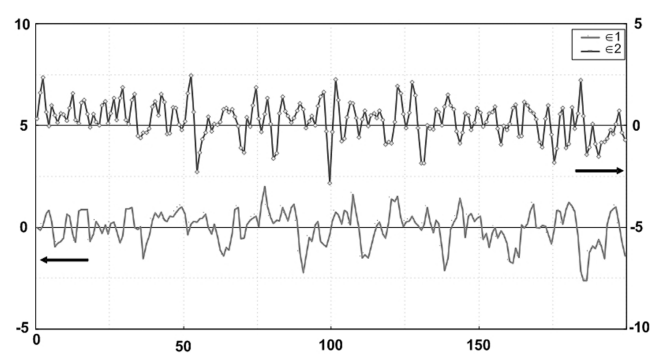
(a) For principal component 1



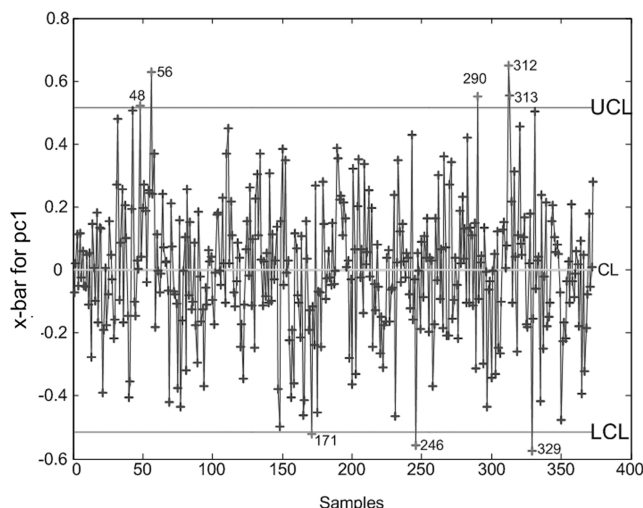
(b) For principal component 2

Fig. 8. x-bar chart of the score values for the two principal components (subgroup size $m=20$, risk level=1%).**Fig. 9. Impulse response coefficients of $F(q)$.**

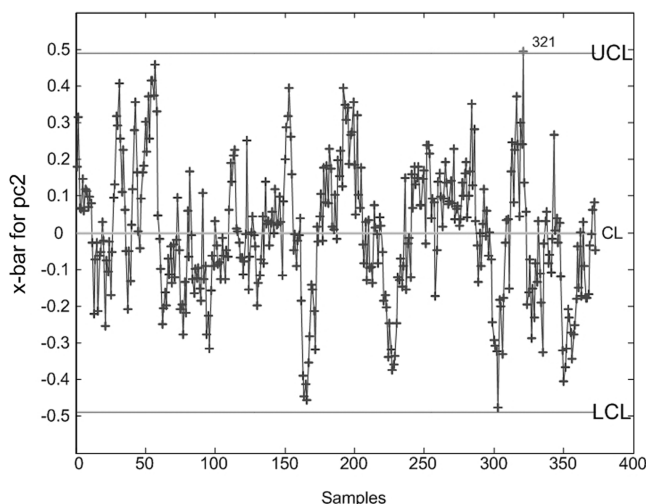
In this example, we consider a numerical distillation column separating xylene at the bottom from a benzene-toluene-xylene mix-

**Fig. 10. Control error of the distillation column when there is a large model error.**

ture [Ramirez, 1997]. The column has 5 stages and treats 250 lbmol/hr of feed mixture (mole ratio of B : T : X=0.1 : 0.6 : 0.3) at steady state. It is assumed that the benzene mole fraction at the top and xylene mole fraction at the bottom are measured and regulated at 0.23 and 0.9, respectively, using the reflux ratio and reboiler duty as manipulated variables. The process model for MPC was deter-



(a) Principal component 1



(b) Principal component 2

Fig. 11. \bar{x} -bar charts under an out-of-control state for the distillation column (subgroup size $m=20$, risk level=1%).

mined through identification through PRBS test. To each output port we added independent integrated white noise signals as disturbance. Process variables are measured and collected at every 3.6 min.

Figs. 7 to 9 summarized the results for the in-control state. First, Fig. 7 shows the control errors (a) and the corresponding whitened signals (b) under a state that we chose to consider *in-control*. Then we took PCA on Z . As a result, it was found that contributions of the first and second principal directions to ZZ^T are 52.5% and 47.5%, respectively, thus both are important. Fig. 8 shows the \bar{x} -bar charts (Schwartz charts) for each score value when the subgroup size is 20 and the risk level was chosen 1%. It can be observed that the chance of outlier is around 1% level, which verifies that the temporal and spatial decorrelations are appropriate. For reference, we plot impulse response coefficients of $F(q)$ in Fig. 9, which are used for closed

loop assessment in CCPMM.

This time we change the tray efficiency of the distillation model from 0.9 to 0.8. It is assumed that there is no change in the disturbance signal. Fig. 10 shows typical output error response under original MPC after transient. It can be seen that the variance of output error grows compared to the in-control state shown in Fig. 7(a). When we monitor the \bar{x} -bar values in the Schwartz charts (Fig. 11), a number of outliers was found to exceed the normal value, which indicates that something is wrong in the control system. To diagnose the system, model-error and the disturbance signal were estimated as given in Figs. 12 and 13. Further closed loop simulation study injecting the estimated disturbance shows that the model error (consequently, poorly designed MPC) is the reason of the trouble.

CONCLUSIONS

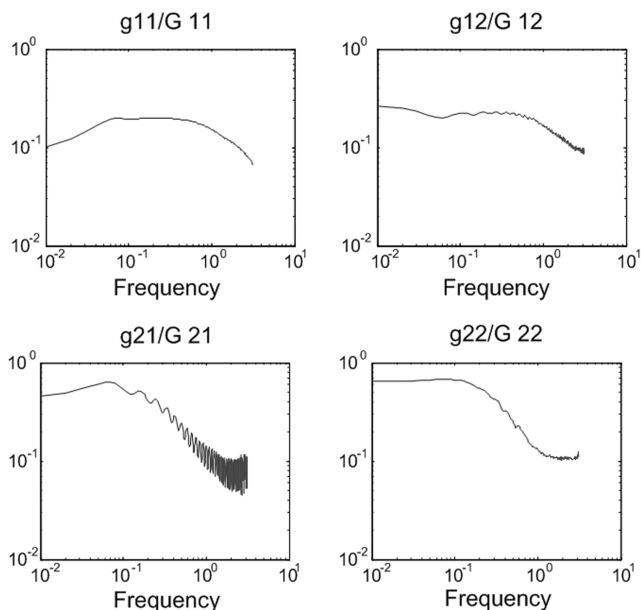


Fig. 12. Frequency domain plot of the relative model errors for the distillation column.

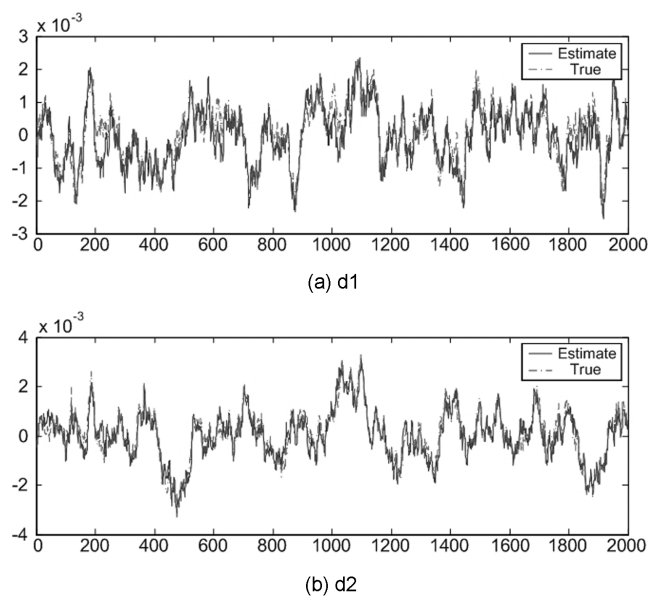


Fig. 13. Estimated disturbance signals for the distillation column.

In this paper, we presented new tools for monitoring and diagnosing the performance of multivariable control systems. The monitoring part is based on the statistical monitoring technique (SPC) whereas the diagnosis part utilizes the closed-loop identification as the key technique. To enable the SPC technique for the process variable under closed control, a special procedure that removes temporal and spatial correlations of the process variable has been proposed.

Investigation shows that the proposed monitoring technique is indeed complementary to the widely accepted MVC-based technique in that both techniques rely on the ARMA model determined from the control error sequence. However, the latter inquires the estimated impulse response coefficients whereas the former investigates the input white noise of the ARMA model for closed loop performance monitoring. It is considered that the two methods can be synergistically combined to produce a more powerful monitoring tool. In addition, if the proposed diagnosis tool and others are supplemented, the whole system will be able to provide quite sensitive and accurate performance assessment and diagnosis for MIMO MPC and other model-based advanced control systems.

ACKNOWLEDGEMENT

This was supported by the IMT2000 (project number: 00015993) in 2003 by MOICE and Research Institute for Applied Science and Technology of Sogang University. This paper is dedicated to Professor Hyun-Ku Rhee on the occasion of his retirement from Seoul National University.

REFERENCES

- Box, G. and Luceno, A., "Statistical Control by Monitoring and Feedback," John Wiley, NY (1997).
- Harris, T. J., "Assessment of Control Loop Performance," *Can. J. Chem. Engng.*, **67**(10), 856 (1989).
- Harris, T. J., Boudreau, F. and MacGregor, J. F., "Performance Assessment of Multivariable Feedback Controllers," *Automatica*, **32**, 1505 (1996).
- Huang, B., Shah, S. L. and Kwok, E. K., "On-Line Control Performance Monitoring of MIMO Processes," *Proc. ACC*, 1250-1254 (1995).
- Huang, B., "Multivariable Statistical Methods for Control Loop Performance Assessment," Ph.D thesis, Univ. of Alberta, Edmonton, Canada (1997).
- Kesavan, P. and Lee, J. H., "Diagnostic Tools for Multivariable Model-Based Control Systems," *Ind. Eng. Chem.*, **36**, 2725 (1997).
- Kozub, D. J. and Garcia, C. E., "Monitoring and Diagnosis of Automated Controllers in the Chemical Process Industries," 1993 AIChE Annual Meeting, Nov. (1993).
- Ljung, L., "System Identification: Theory for the User," Prentice-Hall, NJ (1999).
- Mamzic, C. L. (ed.), "Statistical Process Control," ISA, Research Triangle Park, NC, USA (1995).
- Matrikon homepage, www.matrikon.com (2003).
- Overschee, P. Van and DeMoore, B., "N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, **30**, 75 (1994).
- Ramirez, W. F., "Computational Methods for Process Simulation," 2nd ed., Butterworth-Heinemann, Boulder, CO (1997).
- Qin, S. J., "Control Performance Monitoring - a Review and Assessment," *Comp. Chem. Engng.*, **23**, 173 (1997).
- Qin, S. J. and Badgwell, T. A., "A Survey of Industrial Model Predictive Control Technology," *Control Eng. Practice*, **11**, 733 (2003).
- Stanfelj, N., Marlin, T. E. and MacGregor, J. F., "Monitoring and Diagnosing Process Control Performance: the Single-loop Case," *Ind. Eng. Chem.*, **32**, 301 (1993).