Adaptive NO_X predictive emission monitoring for industrial process heaters

Young-Hak Lee[1], Minjin Kim[2], Chonghun Han[1]*
[1]Seoul National University, Seoul 151-744, Korea
[2]Pohang University of Science and Technology, Pohang, Gyeongbuk 790-784, Korea

1. Introduction

Process heaters are major sources of nitrogen oxides (NO_X), which are ozone precursors and known sources of health and environmental problems. Most governments regulate NO_X emissions from industrial plants. To comply with these regulations, NO_X emissions must be monitored and controlled. Typically, NO_X emissions are monitored using on-line hardware analyzers called continuous emissions monitoring systems (CEMS). Generally, CEMS are costly to install and difficult to maintain. Predictive emissions monitoring systems (PEMS) have been used as alternatives to CEMS. One type of PEMS predicts NO_X emissions using detailed kinetic models that include approximately 3,000 reactions and 200 species along with calculated temperature and concentration fields for the major gas components (e.g., CH_4, O_2, N_2, CO_2, H_2O). The accuracy of developed models has been found to strongly depend on the quality of the assumptions and the relevance of the physical/chemical models.

Another type of PEMS predicts NO_X emissions by using empirical predictive models based on engineering correlations of measured NO_X effluent data for a variety of operating conditions. Fuzzy logic, neural networks, and regression models have been used to estimate emissions (Collins & Terhune, 1994; Kocijan, 1997). Empirical predictive models for NO_X emissions are gaining acceptance and approval in industrial applications such as refinery and petrochemical plants, power plants, incineration facilities, and pulp and paper mills. In these PEMS, neural network models have been preferred because they can handle the several non-linear terms that must be included to maintain estimation accuracy over the entire operating range (Baines *et al*., 1997). However, although neural network models can cover the full operating range of the equipment, they tend to produce valid emission predictions for only a short period because (i) combustion profiles change due to variations in the fuel quality, heating load, slag/soot deposits, ambient conditions, and the conditions of the plant equipment; and (ii) the overall plant operation may change due to capital projects and process modifications undertaken to improve the operation of heaters and boilers. If the changes push the process outside the model, the predictive model must be rebuilt and revalidated. Multivariate statistical process control (MSPC) has been widely employed as an alternative of neural network models because it can offer the prompt model adaptation and robust estimation of emissions (Quinn, 2002, Wang *et al*., 2003). Partial least squares (PLS) regression is one of most the powerful and frequently applied

---

[*] Corresponding author. Tel.: +82-2-880-1887; Fax: +82-2-888-7295; E-mail: chhan@snu.ac.kr

techniques in multivariate statistical process control when process variables are highly correlated.

This paper aims to propose an adaptive predictive emission modeling method based on the performance assessment criteria of PLS models. To evaluate the performance of the regression models, multiple cumulative sum (CUSUM) charts for T^2 statistics and prediction errors are proposed based on a median filter in a moving window approach. The proposed CUSUM charts not only evaluate model performance in sample-wise or batch-wise, but also help to select the updating target or to take an updating action. The updating action is performed with either partial adaptation such as mean shift or complete one such as batch-type PLS remodeling. Once the update action is determined, the adaptive modeling can be executed in a robust and a computationally efficient manner so that the time points to update a model are detected through the charts and the model is updated only at the points, not recursively.

2. Process description

An industrial refinery fired heater is introduced to illustrate the proposed method. The fired heater plays a role preheating crude oil and lessening the heat duty of a main distillation column in the refinery plant. The process variables for the fired heater include (i) inlet feed conditions; (ii) outlet conditions; (iii) inlet conditions of fuel and air; (iv) flue gas conditions; and (v) combustion conditions within the heater box (temperatures, pressures, excess O_2 composition). Fuel gas and/or oil are supplied to each heater. Liquefied natural gas (LNG) is used as the fuel gas, and Bunker-C as the fuel oil. The concentration of NO_X emitted from the common stack should be below 200 ppm according to South Korean law. The concentration of NO_X is measured on-line with an infrared (IR) gas analyzer in CEMS. The pollutant emissions can also be monitored with a software-based analyzer called the predictive emissions monitoring system (PEMS).

3. Proposed predictive modeling method for adaptive monitoring

3.1. Performance assessment of PLS models

Performance assessment method consists of two phases, namely the design of a robust multiple CUSUM chart and the assessment of the model performance using this chart. To evaluate PLS model performance, two performance measures are proposed: window-based robust T^2 statistics and absolute prediction error (APE). Each measure uses *median* known as one of robust statistics. For test data of a window size L, $\mathbf{X}_\mathrm{L}$, the robust T^2 statistics, T_R^2 is defined as:

$$T_R^2 = median\{\mathbf{X}_\mathrm{L}^\mathrm{T}\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^\mathrm{T}\mathbf{X}_\mathrm{L}\} \tag{1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the variances of the scores corresponding to the loadings, $\mathbf{P}$. APE_R is as follows:

$\text{APE}_R = \ median\{|\mathbf{Y}_\mathrm{L} - \mathbf{X}_\mathrm{L}\mathbf{B}|\}$ (2)

where $\mathbf{Y}_\mathrm{L}$ is response data matrix corresponding to the $\mathbf{X}_\mathrm{L}$.

Cumulative sum (CUSUM) charts have been used in industry to detect a change in the quality of a manufactured product. CUSUM charts, while not as intuitive and simple as Shewhart charts, have shown more efficient performance in detecting small shifts in the mean of a process. The tabular CUSUM, used in this study, works by accumulating derivations from zero for T_R^2 and APE_R, which represent the center of process measurements and perfect match between the predicted values and the observed ones. Let $T_{R,n}^2$ and $\text{APE}_{R,n}$ be the robust values of T^2 and APE for nth dataset. $C_n^{T_R^2}$ and $C_n^{\text{APE}_R}$ mean the CUSUM values for $T_{R,n}^2$ and $\text{APE}_{R,n}$. They are computed as following equations:

$C_n^{T_R^2} = \max\,[0,\ T_{R,n}^2 - K + C_{n-1}^{T_R^2}]$ (3)

$C_n^{\text{APE}_R} = \max\,[0,\ \text{APE}_{R,n} - K + C_{n-1}^{\text{APE}_R}]$ (4)

where K is usually called the reference value. Note that C_n accumulates deviations from the target value that are greater than K, which means that C_n signals only if the deviations is significant. Figure. 1 illustrates the multiple CUSUM plot of T_R^2 and APE_R for assessing model performance. The four cases classified with the decision limits of the two charts are identified according to the patterns of the degradation of model performance as shown in Figure 1.

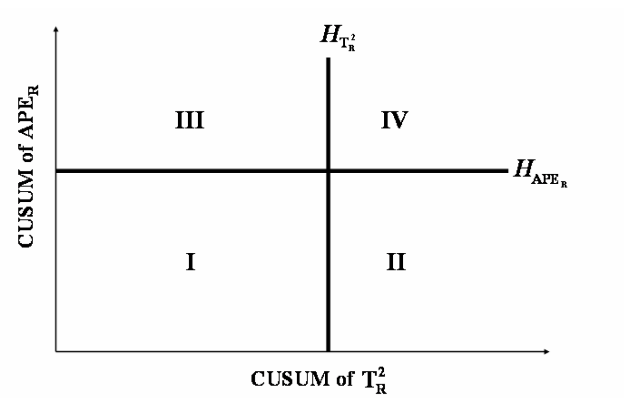

Figure 1. Multiple CUSUM plot of T_R^2 and APE_R used for the model performance assessment

The performance monitoring and assessment procedure for a PLS model are carried out as follows. (i) Collect the test data of window size, and then estimate the medians of the T2 and APE from Eq. (1) and (2) for the data. (ii) Compare the calculated measures with the decision limits. If the values exceed the limits, go to the next step.

Otherwise, estimate the medians of new test data including the new data of block size, and then iterate this step for the continuous monitoring of the model performance until detecting the model violation. (iii) Assess the model performance depending on the monitoring result and classify it into one of following four cases as seen in Figure 1. Region I represents the normal region of model performance. In the region II, the model has a good prediction power, but the predictor variables go out of the normal region due to the extrapolating feature of the regression model. Region III means the significant change in the relationship between the predictor variables and response variables or among the only predictor variables. Operating conditions are similar to those of the modeling data. For example, analyzer malfunction of response variables can be identified as one of causes to the signals in the region. Finally, region IV signals the significant changes in the process behaviors and/or in the analyzers' characteristics accompanied with the poor prediction power.

3.2 Adaptive PLS regression modeling

The proposed adaptive modeling method proceeds according to the following steps: (1) checking the model performance; (2) detecting model violation; and (3) updating the model. It is different from the recursive PLS modeling method, in that the adaptation process is executed only when model violation is detected. This method offers better model adaptation and lower update frequency.

First of all, the model performs partial adaptation depending on the results of the model performance assessment. In the partial adaptation of the model parameters, the update targets are the mean and variance of the predictor or response variables, and the correlation between the latent variables in the PLS model. Partial model adaptation is based on the industrial observation that the covariance structure for minor changes in the process operating conditions is similar to its original structure (Hwang *et.al,* 1999). When the partially updated model does not work well, all of the model parameters, including the PLS regression coefficients, are updated using the previous PLS model parameters, in a recursive manner.

Partial model adaptation The update target in the partial model adaptation is determined according to the result of the model performance assessment. Both the mean and standard deviations of the predictor variables and response variables often change in the case of a small shift in the process operation, which causes the performance of the PLS model to be degraded. Therefore, updating both the mean and standard deviation can improve the model's performance. In the case where the correlation structure among the variables is violated, the computational load required for updating all of the PLS parameters is high. The inner relation of the score vectors in the reduced space can be refreshed. The update of the regression coefficients between $\mathbf{T}$ and $\mathbf{U}$ can take the place of the insignificant computation of all the PLS model parameters.

If the number of samples in the old modeling set is N_0 and the number of samples in the new updating block is N_1, then the adjustable update values $\bar{x}_{j,1}$ and $\sigma_{x_{j,1}}$ of the mean $\bar{x}_{j,0}$ and standard deviation $\sigma_{x_{j,0}}$, respectively, for the jth variable of the $\boldsymbol{x}$ data are given by

$$\bar{x}_{j,1} = \frac{1}{N_0 + N_1}\left(N_0 \bar{x}_{j,0} + \sum_{i=1}^{N_1} x_{ij,1} \right) \tag{5}$$

$$\sigma_{x_{j,1}} = \left\{ \frac{1}{N_0 + N_1 - 1}\left(\sum_{i=1}^{N_0} x_{ij,0}^2 + \sum_{i=1}^{N_1} x_{ij,1}^2 - (N_0 + N_1)\bar{x}_{j,1}^2 \right) \right\}^{1/2} \tag{6}$$

The second term in Eq. (5) can be replaced by the median of $\boldsymbol{x}_j$, in order for it to be robust to outliers. For the y data, the formula is the same as that described above, with y being substituted for x. The effect of new updating blocks can be weighted more heavily, by increasing the number of samples and the block size, N_1, and vice versa. Also, the adjustable update values, $b_{a,1}$, of the inner regression coefficient, $\mathrm{b}_{a,0}$, calculated from the $\mathbf{t}_{a,0}$ and $\mathbf{u}_{a,0}$ score vectors for the ath latent variable, can be represented by the following equation:

$$b_{a,1} = b_{a,0} + \left(\sum_{i=1}^{N_0} t_{ia,0}^2 + \sum_{i=1}^{N_1} t_{ia,1}^2 \right)^{-1} \left(\sum_{i=1}^{N_1} t_{ia,1}(u_{ia,1} - b_{a,0} t_{ia,1}) \right) \tag{7}$$

where $t_{ia,1}$ and $u_{ia,1}$ are the elements of the $\mathbf{t}_{a,1}$ and $\mathbf{u}_{a,1}$ score vectors obtained from the newly gathered data, respectively.

Complete model adaptation based on block-wise recursive PLS algorithm Partial model adaptation is more effective in the case of a small shift from the operating boundary of the original model. However, the operating region moves gradually outside of the original region, in the case of significant process changes, such as catalyst deactivation, equipment aging, sensor and process drifting and cleaning. Hence, complete remodeling is required in such cases. Using the previous model parameters, the block-wise recursive PLS algorithm is introduced (Qin, 1998). Updating the PLS model involves performing PLS on the existing model and the new sub-model. Let $\{\mathbf{T}, \mathbf{W}, \mathbf{P}, \mathbf{B_D}, \mathbf{Q}\}$ be the PLS results for data $\{\mathbf{X}, \mathbf{Y}\}$, where $\mathbf{T}$ is the score matrix, $\mathbf{W}$ is the weighting matrix used in PLS, $\mathbf{P}$ and $\mathbf{Q}$ are loading matrices for $\mathbf{X}$ and $\mathbf{Y}$, and $\mathbf{B_D}$ is a diagonal matrix of inner model coefficients. From the theorem proposed by Qin, assuming two PLS results, $\{\mathbf{T}_0, \mathbf{W}_0, \mathbf{P}_0, \mathbf{B}_{\mathbf{D}0}, \mathbf{Q}_0\}$ for data $\{\mathbf{X}_0, \mathbf{Y}_0\}$ and $\{\mathbf{T}_n, \mathbf{W}_n, \mathbf{P}_n, \mathbf{B}_{\mathbf{D}n}, \mathbf{Q}_n\}$ of the new block data $\{\mathbf{X}_n, \mathbf{Y}_n\}$, performing PLS regression on $\begin{bmatrix} \mathbf{P}_0^{\mathrm{T}} \\ \mathbf{P}_n^{\mathrm{T}} \end{bmatrix}, \begin{bmatrix} \mathbf{B}_{\mathbf{D}0}\mathbf{Q}_0^{\mathrm{T}} \\ \mathbf{B}_{\mathbf{D}n}\mathbf{Q}_n^{\mathrm{T}} \end{bmatrix}$ results in the same regression model as that obtained by performing PLS regression on the data pair $\begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_n \end{bmatrix}, \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_n \end{bmatrix}$. The new PLS result for the data pair taking into consideration the loading coefficients and inner regression coefficient is denoted as $\{\mathbf{T}_1, \mathbf{W}_1, \mathbf{P}_1, \mathbf{B}_{D1}, \mathbf{Q}_1\}$. If the model performance is still degraded after updating, the PLS model is successively updated with the next new sub-models.

4. Results and discussion

4.1. Performance assessment of the NO_X emission model

Before the proposed approach is applied, an initial PLS regression model should be established. A dataset of 4020 samples was used to build an initial PLS model. The parameters for the performance assessment and the adaptive modeling can be obtained from the initial model and historical long-term data. The window size was set to 720 sampling points, and the block size was selected as one half of the window size, i.e. 360 sampling points. Starting from the initial PLS model with the predefined parameters, the models were assessed using industrial data recorded over a period of three months. The median value of every 360 samples was marked onto the specific region of the CUSUM chart. Figure 2 illustrates a multiple CUSUM chart for two performance measures during a period of approximately one month. It demonstrates the extrapolation feature of the model, which means that the model has good prediction power, but that the diagnostic performance obtained using the previous model can be degraded, due to the mean shift of the predictor variables resulting from the change in the heater conditions. To diagnose the detailed causes, the contribution plot analysis of T_R^2 for a point indicated with an arrow in Figure 2 was performed. The contribution plot in Figure 3 shows that the fifth variable, namely the flow rate of the crude oil entering the heater, has a high value, which implies that there is an increase in the heating load. This caused the means of the process variables to shift out of the reference conditions, however the covariance structure between the heater operating conditions and NO_X emissions remained consistently unchanged, because the APE_R values in Figure 2 were below the decision limit.

4.2. Adaptive NO_X emission modeling

The predictive NO_X emissions model was updated during a three month period whenever the operating mode changed significantly. Figure 4 shows the result of the performance monitoring for the predictive model. Most of the points for the two measures are under the decision limits. This indicates that the model was well adapted to a process shift of over 1σ. Numerous CUSUM peaks are shown, due to the process shifts followed by the adaptation actions of the model. The proposed method was compared with the block-wise recursive PLS algorithm in a moving window approach (MW-RPLS). The prediction results of MW-RPLS are depicted in Figure 5, along with the results of the proposed method.

Except for a couple of points, the prediction results are almost identical. The 338th sample, which is marked with an arrow, shows a point for which the prediction accuracy is worse than that of the proposed method. The concentration of NO_X emitted from the stack should be below 200 ppm according to the prevailing environmental regulations. The MW-RPLS approach is susceptible to producing false alarms, however the proposed method can avoid such false alarms, owing to its updating only the scaling parameters for each new block, which enables the model to adapt more rapidly than the MW-RPLS method. In the successive change of operating conditions, the MW-RPLS approach may not take account of recent status changes, because of its updating based on old long-term data with the window size used for the model update. Also, the proposed method requires a much lower update frequency. In this case study, the update frequency used for the scaling parameter is 30, and that used for all other model parameters is 12. It is shown that the proposed approach is much more efficient than the previous approach, which is updated 107 times.

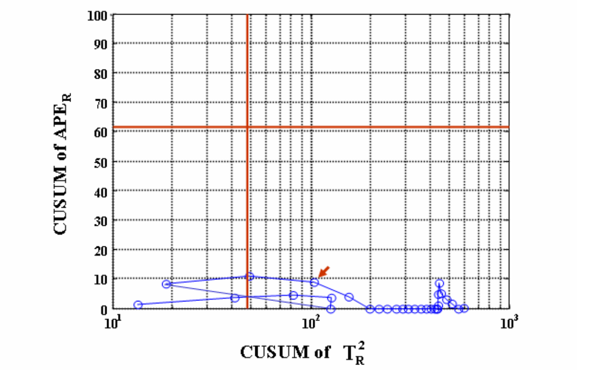

Figure 3. Multiple CUSUM chart for the two performance measures during a specific period.

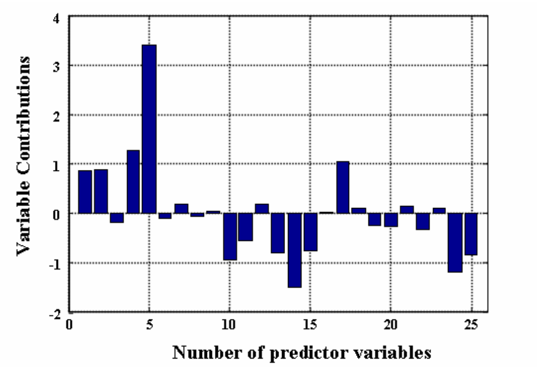

Figure 4. Variable contributions for the specific point indicated with an arrow in Figure 3.

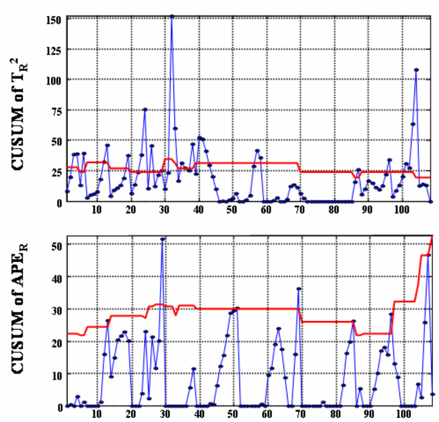

Figure 5. Results of monitoring the model performance for the proposed adaptive modeling scheme

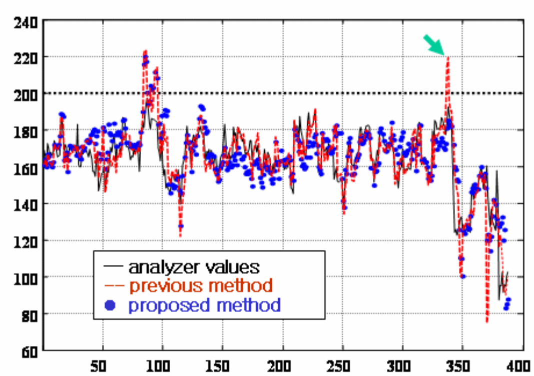

Figure 6. Comparison of prediction results of the proposed method with those of the block-wise recursive PLS modeling scheme

Acknowledgements

The authors gratefully acknowledge the partial financial support of the Korea Science and Engineering Foundation provided through the Advanced Environmental Biotechnology Research Center (R11-2003-006) at Pohang University of Science and Technology, the IMT2000 project (ID: 00015993) fund of the Ministry of Information Communication, and the Brain Korea 21 program initiated by the Ministry of Education, Korea.

References

Baines, G., Haynes, R., and Stabell, J. (1997), "Predicting boiler emissions with neural networks," *TAPPI Journal*, 80(5), 57-61.

Collins, M., and Terhune, K. (1994), "A model solution for tracking pollution," *Control Engineering*, June.

Hwang, D.-H., Han, C., (1999), "Real-time monitoring for a process with multiple operating modes" Cont. Eng. Prac. 7, 891-902.

Kocijan, J. (1997), "An approach to multivariate combustion control design," *J. of Process Control*, 7(4), 291-301.

Qin, S. J. (1998), "Recursive PLS for adaptive monitoring," *Comput. Chem. Engng*, 22, 503-514.

Quinn, S. L. (2002), "Improving the desulfurization process using adaptive multivariate statistical modeling," *AISE Steel Technology*, October 37-41.

Wang, X., Kruger, U., and Lennox, B. (2003), "Recursive partial least squares algorithms for monitoring complex industrial processes," *Control Engineering Practice*, 11, 613-632.