Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models

Hiromasa Kaneko and Kimito Funatsu
Dept. of Chemical System Engineering, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

DOI 10.1002/aic.14006
Published online January 14, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Soft sensors are used widely to estimate a process variable which is difficult to measure online. One of the crucial difficulties of soft sensors is that predictive accuracy drops due to changes of state of chemical plants. It is called as the degradation of soft sensor models. In this study, we attempted to classify this degradation of models in terms of changes in an explanatory variable and an objective variable, and the rapidity of the changes. Moreover, we discussed characteristics of adaptive soft sensor models, based on the classification results. By analyzing simulated data sets and a real industrial data set, we could obtain knowledge and information on appropriate adaptive models for each type of the degradation. © 2013 American Institute of Chemical Engineers *AIChE J*, 59: 2339–2347, 2013
Keywords: process control, soft sensor, degradation, adaptive model, predictive ability

Introduction

In industrial plants, soft sensors have been widely used to estimate process variables that are difficult to measure online.[1,2] An inferential model is constructed between those variables that are easy to measure online and those that are not, and an objective variable, **y**, is then estimated using that model. Through the use of soft sensors, the values of **y** can be estimated with a high degree of accuracy.

Their use, however, involves some practical difficulties. One crucial difficulty is that their predictive accuracy decreases due to changes in the state of chemical plants, catalyzing performance loss, sensor, and process drift, and the like. This is called as the degradation of soft sensor models. If the degradation is not solved, it is difficult to identify reasons of abnormal situations. On the site of plants, when a prediction error of **y** is above a threshold, it is recognized as an abnormal situation. There is no effective method to judge whether the reason of the abnormal situation is the trouble of **y**-analyzer or the degradation of a soft sensor model under present circumstances.

It is, therefore, strongly desired to solve the degradation. To reduce the degradation of a soft sensor model, the model is reconstructed with newest data. A moving window (MW) model[3–5] and a recursive model[6] are categorized as a sequentially updating type and a distance-based just-in-time (JIT) model[7] and a correlation-based JIT model[8] are categorized as a JIT type. For example, a MW model is constructed with data that are measured most recently and a distance-based JIT model is constructed with data whose distances to prediction data are smaller than those of other data. Many excellent results have been reported based on the use of MW models and JIT models.

Meanwhile, problems of reconstructing a model such as the incorporation of abnormal data with training data and an increase of maintenance costs were discussed, and then, a model based on the time difference (TD) of **y** and that of explanatory variables, **X**, was proposed.[9,10] This model is referred to as a TD model. The effects of deterioration with age such as the drift and gradual changes in the state of a plant can be accounted for by using a TD model without reconstruction of the model.

Kaneko et al.[9] analyzed data obtained from the operation of a distillation column and compared a MW model and a TD model. The results suggested that the predictive ability of each adaptive model changed depending on states of the plant. For instance, the TD model could predict **y** with high accuracy soon after and before process variations and the MW model had high predictive ability in process variations. We can say that the understanding of the state of a plant and the use of an appropriate model for each state are needed for highly accurate prediction.

In this study, therefore, we classify the degradation of soft sensor models in terms of the types of changes in **X** and **y** and the rapidity of the changes, and then, discuss adaptive models such as MW, JIT, and TD models, which is based on the classification results to construct and select models that can handle each degradation. Finally, the results of the discussion are confirmed by using simulation data sets and real industrial data.

Method

Classification of the degradation

Figure 1 shows basic concepts of the degradation of a linear soft sensor model constructed between an explanatory variable, **x**, and **y**. Figures 1a, b represent shifts of **y**-values

Correspondence concerning this article should be addressed to K. Funatsu at funatsu@chemsys.t.u-tokyo.ac.jp.

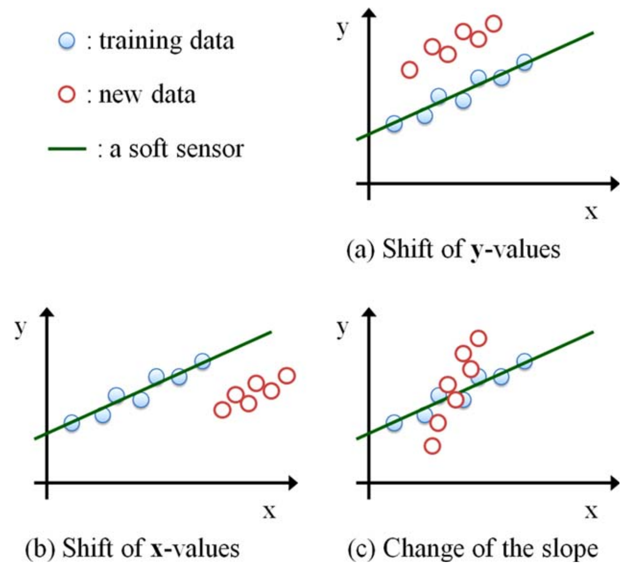


Figure 1. Classification of the degradation of a linear soft sensor model.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

and **x**-values, respectively. These are corresponding to sensor and process drift, scale deposition on a pipe, a change of operating condition such as the amount of raw material, and so on. The slope does not change between training data and new data, but values of a **y**-variable or an **x**-variable shift.

Figure 1c represents a change of the slope of **x** and **y**. This is corresponding to catalyzing performance loss, a change of operating condition such as a concentration in raw material, and so on. Of course, shifts of **y**-values and **x**-values and a change of the slope will occur simultaneously.

When we focus on the rate of the degradation, each shift or change happens gradually, rapidly, or instantly. For example, catalyzing performance loss, process, and sensor drift, a change of external temperature, and scale deposition on a pipe occur gradually; a sharp change in raw material occurs rapidly; and correction of drift, regular repair of a plant, and a stoppage of a pipe occur instantly. Of course, this rapidity is continuous in fact.

It is said that four types of sensor faults are considered: bias, complete failure, drifting, and precision degradation.[11] Bias and drifting can be included in our classification, that is, bias means instant shifts of **y**-values and **x**-values and drifting means gradual shifts of **y**-values and **x**-values. However, all types of sensor faults must be detected as early as possible and abnormal sensors should recover. After the recovery or the elimination of abnormal variables, soft sensor models should be reconstructed and used for prediction. We consider sensor faults as not the degradation of soft sensor models but problems that are detected quickly.

Characteristics of adaptive models

Table 1 shows the characteristics of TD, MW, and JIT models. A TD model can adapt shifts of both **y**-values and **x**-values because it achieves the same effect as a bias update. Even when the shifts happen gradually, rapidly, and instantly, a TD model can follow the shifts appropriately. On the one hand, it is difficult for a MW model to adapt a rapid or instant shift because old data before the shift remain in training data.

Meanwhile, a TD model cannot adapt a change of the slope.[10] A MW model should be used to follow a gradual change of the slope by adding new data to training data. However, it is difficult for a MW model to handle a rapid or instant change because the model is affected by old data before the change.

In case of a JIT model, which is constructed with data sets close to test data in the space of **x**, appropriate selection of data sets will be performed if a shift of **x**-values happens. However, besides that, data sets after a shift of **y**-values or a change of the slope cannot be selected because there is no change in the space of **x** as shown in Figures 1a, c. When a shift of **x**-values and a change of slope happen simultaneously, a JIT model can adapt these changes appropriately if **x**-values change clearly and adequate amount of data in the new situation is stored in database. This is because appropriate selection of data sets can be performed owing to a shift of **x**-values.

As shown in Table 1, the use of a TD model is recommended for shifts of **y**-values and **x**-values because a TD model is not reconstructed and, therefore, is not affected by online abnormal data. In addition, maintenance cost of a TD model is low. On the other hand, MW and JIT models are affected by online abnormal data and maintenance cost is high because these models are reconstructed online. However, if a change of the slope occurs, a TD model cannot adapt the change. A MW model or a JIT model should be used depending on the situation of process.

Results and Discussion

To test predictive ability of each adaptive model when each type of the degradation occurs, we used simulated data sets and a real industrial data set.

Modeling of the simulation data

The number of **X**-variables was set as two. First, **X** of uniform pseudorandom numbers whose range was from 0 to 10 was prepared. Then, **y** was set as follows

$$\mathbf{y}=\mathbf{X}\mathbf{b}+\mathrm{UOD}+N(0,0.1) \tag{1}$$

where **b** means the magnitude of contribution of **X** to **y**; UOD means unobserved disturbances; and $N(0,\ 0.1)$ is

Table 1. Characteristics of TD, MW, and JIT Models

Degradation				
Type	Rapidity	TD Model	MW Model	JIT Model
Shift of **y**-value	Gradual	○[*]	○	×[†]
	Rapid	○	Δ[‡]	×
	Instant	○	×	×
Shift of **x**-value	Gradual	○	○	○
	Rapid	○	Δ	○
	Instant	○	×	○
Change of the slope	Gradual	×	○	×
	Rapid	×	Δ	×
	Instant	×	×	×
Shift of **x**-value and change of the slope	Gradual	×	○	○×[#]
	Rapid	×	Δ	○×
	Instant	×	×	○×
		TD Model	MW Model	JIT Model
Online abnormal data		Not affected	Affected	Affected
Maintenance cost		Low	High	High

*The model can handle the degradation well.
†The model can handle the degradation to some extent.
‡The model cannot handle the degradation.
#It depends on a situation whether the model can handle the degradation or not.

Table 2. r_{pred}^2 Values of Each Model for Each UOD

UOD	TD Model	MW Model	JIT Model
(2)	1.000	0.999	0.965
(3)	0.998	0.782	0.747
(4)	0.977	0.816	0.781

random numbers from normal distribution given a standard deviation of 0.1 and a mean of 0. Shifts of **y**-values and **x**-values, and changes of the slope (Figure 1) can be represented by changing UOD, **X**, and **b**, respectively. Additionally, we can consider gradual, rapid, or instant shifts and changes. Last, $N(0, 0.1)$ was added also to **X**.

The number of data was 200. The first 100 data were used for training, and the next 100 data were the test data. We used a partial least squares (PLS) method[12] as a regression approach and set the number of data for constructing a MW model and a JIT model as 20. In this article, not a correlation-based JIT model[8] but a Euclidian distance-based JIT model was used because there is little correlation between **X**-variables. The first and second **X**-variables are represented as $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively.

Shift of y-Values. First, we explain the prediction results where three types of UOD were used as follows

$$\mathrm{UOD} = 0.01t \tag{2}$$

$$\mathrm{UOD} = 3\sin(0.04t) \tag{3}$$

$$\mathrm{UOD} = \begin{cases} 0 \ (1 \le t \le 20, 101 \le t \le 120) \\ 5 \ (21 \le t \le 100, 121 \le t \le 200) \end{cases} \tag{4}$$

where t means time. Equations 2–4 represent gradual, rapid, and instant changes of UOD, respectively.

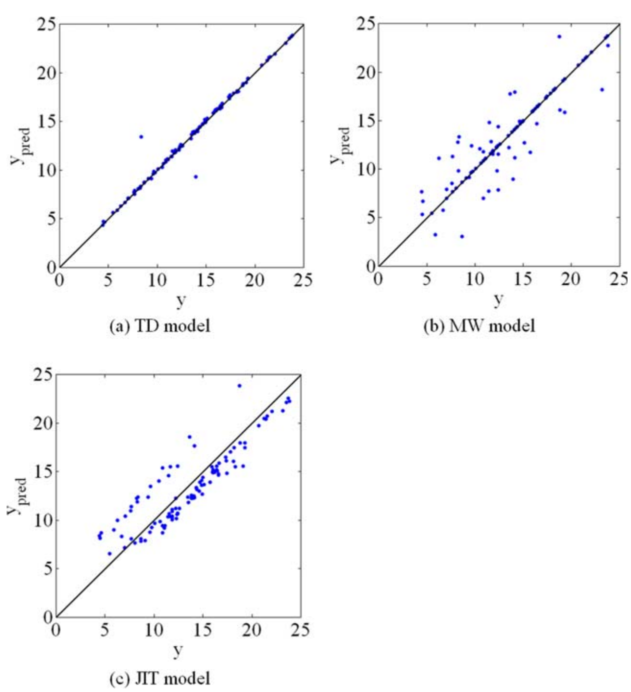


Figure 2. Relationships between y and predicted y of test data for each model when Eq. 4 was used as UOD.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Table 2 shows the r_{pred}^2 values of each model for each UOD. The r_{pred}^2 is the determination coefficient, r^2, for test data and the detail is given in Appendix. As we discussed, the TD models had high predictive ability for all types of UOD. Though the MW model could follow the gradual change of UOD, the r_{pred}^2 values got low when UOD changed rapidly and instantly. The r_{pred}^2 value of the JIT model was the lowest in all cases, indicating appropriate selection of training data could not be performed.

Figure 2 shows the relationships between **y** and predicted **y** of test data for each model when Eq. 4 was used as UOD. The plot of Figure 2a shows a much tighter cluster of predicted values along the diagonal for the TD model though the MW model also had high predictive accuracy. There seem two outliers in Figure 2a, which are corresponding to the times when the value of UOD changes from 0 to 5 and from 5 to 0. Practically, these changes mean, for example, the instant shift of **y**-values just after drift correlation. The occurrence of this event can be noticed in advance and, therefore, there is no problem with the outliers in practice. In addition, the same outliers are included also in training data. The accuracy of the TD model can be improved by eliminating the outliers.

The distribution of the plot in Figure 2c is divided in two. This indicates that the training data for predicting test data could not be selected appropriately in the construction of the JIT models when UOD was 0 or 5. We concluded that TD models should be used for these kinds of the degradation of a soft sensor model.

Change of the Slope. Second, we set the magnitude of contribution of $\mathbf{x}_1$ to **y**, b_1, as 1 and that of $\mathbf{x}_2$ to **y**, b_2, as follows

$$b_2 = 3\sin(0.01\pi t) + 1 \tag{5}$$

$$b_2 = 3\sin(0.02\pi t) + 1 \tag{6}$$

Equations 5 and 6 represent gradual and rapid changes of b_2, respectively. Then, we compared the prediction results of the MW, JIT, and TD models.

The r_{pred}^2 values and the root-mean-squared error (RMSE) values of test data ($\mathrm{RMSE_P}$) of each model for each b_2 are shown in Tables 3 and 4, respectively. The detail of $\mathrm{RMSE_P}$ is given in Appendix. In both cases where Eqs. 5 and 6 were used, the MW models had the highest predictive accuracy in three adaptive models as discussed in "Characteristics of adaptive models." However, the MW models could not adapt the latest relationship between **X** and **y** because the number of data for the construction of the MW models was set as 20, and then, the MW models were constructed with data where the relationship between **X** and **y** was old. Hence, the r_{pred}^2 values of the MW models were not so high.

In addition, as shown in Table 4, the $\mathrm{RMSE_P}$ values of the MW models increased as the rate of the change of the slope increased, reflecting that it is difficult for a MW model to handle a rapid or instant change of the slope because the model is affected by old data as mentioned in

Table 3. r_{pred}^2 Values of Each Model for Each b_2

Equation for b_2	TD Model	MW Model	JIT Model
(5)	−6.077	0.546	−5.932
(6)	0.377	0.574	0.009

Table 4. RMSE$_P$ Values of Each Model for Each b_2

Equation for b_2	TD Model	MW Model	JIT Model
(5)	15.21	3.85	15.05
(6)	8.85	7.32	11.17

"Characteristics of adaptive models." But, the MW models could predict test data more accurately than the JIT and TD models did.

Figure 3 shows time plots of regression coefficients of the MW model when Eq. 5 was used as b_2. In this case study, there is no correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$. Though regression coefficients of $\mathbf{x}_1$ did not change so much those of $\mathbf{x}_2$ represented the sine curve of Eq. 5. However, the change of regression coefficients of $\mathbf{x}_2$ lags behind actual b_2. This is because each regression model is constructed with the latest 20 data for the MW model, and the model is affected by old data even if newest data are included in training data.

Time plots of regression coefficients of the JIT model are shown in Figure 4 when Eq. 5 was used as b_2. Both regression coefficients of $\mathbf{x}_1$ and $\mathbf{x}_2$ did not change meaningfully though the contribution of $\mathbf{x}_2$ to $\mathbf{y}$ changes in fact. There was no change of data in the space of $\mathbf{X}$ and, thus, the appropriate selection of data sets failed for the JIT model.

Shift of x-Values and Change of the Slope. Last, we changed the second **X**-variable, $\mathbf{x}_2$, to $\mathbf{x}_{2,\ \mathrm{new}}$ below

$$\mathbf{x}_{2,\mathrm{new}} = \mathbf{x}_2 + 0.01t \tag{7}$$

$$\mathbf{x}_{2,\mathrm{new}} = \mathbf{x}_2 + \begin{cases} 0 \ (1 \le t \le 20, 101 \le t \le 120) \\ 20 \ (21 \le t \le 100, 121 \le t \le 200) \end{cases} \tag{8}$$

where $\mathbf{x}_{2,\ \mathrm{new}}$ means the shifted values of $\mathbf{x}_2$. Equations 7 and 8 represent gradual and instant changes of $\mathbf{x}_{2,\ \mathrm{new}}$, respectively. We also set b_2 as Eq. 5 or as follows

$$b_2 = \begin{cases} 1 \ (1 \le t \le 20, 101 \le t \le 120) \\ 21 \ (21 \le t \le 100, 121 \le t \le 200) \end{cases} \tag{9}$$

b_1 is 1 as is the case in Change of the Slope.

Table 5 shows the r_{pred}^2 values of each model for each $\mathbf{x}_{2,\ \mathrm{new}}$ and each b_2. When $\mathbf{x}_{2,\ \mathrm{new}}$ changed gradually and b_2 was constant, the MW model and the TD model could predict test data with high accuracy as they could in "Shift of y-Values." The MW model could follow the shift of **x**-values, and the TD model was not affected by the shift. Meanwhile, appropriate selection of training data could not be performed in constructing the JIT model, and then, the r_{pred}^2 value was lower than the other models and the plot of **y**-values and predicted **y**-values was not close to diagonal as is the case in Figure 2c. Both data before the shift of **x**-values and data after that were selected for training data because the shift occurred slowly.

However, when the shift of **x**-values happened instantly as Eq. 8, data for the construction of the JIT model were selected successfully and the JIT model had more predictive accuracy than the MW model and the TD model did. We confirmed that the adequate shift of **x**-values is needed for the selection of data in the space of **X**. It should be noted that the accuracy and the predictive ability of the TD model can be improved if the outliers in training data and test data are removed as mentioned in "Shift of y-Values."

In cases where both $\mathbf{x}_{2,\ \mathrm{new}}$ and b_2 change gradually, the TD model could not follow the change of the slope and the appropriate JIT model could not be constructed as shown on the low values of r_{pred}^2, -6.07 and -5.68, in Table 5. The MW model had the highest predictive accuracy in three adaptive models though the r_{pred}^2 value was not so high. The MW model could adapt the change of b_2 given in Eq. 5 to some extent. However, when $\mathbf{x}_{1,\ \mathrm{new}}$ shifted instantly as given in Eq. 8, the r_{pred}^2 value was 0.213 and very low, reflecting that the MW model could hardly follow the instant shift of **x**-values.

When the slope of **x** and **y** changed instantly, that is, Eq. 9 was used as b_2, the r_{pred}^2 value of the JIT model was very high if **x**-values also changed instantly. Figure 5 shows the relationships between **y** and predicted **y** of test data for each model when Eq. 8 was used as $\mathbf{x}_{1,\ \mathrm{new}}$ and Eq. 9 was used as b_2. On the one hand, many data are far from the diagonal in the plots of Figures 5a, b; on the other hand, the plot of Figure 5c shows a much tighter cluster of predicted values along the diagonal for the JIT model. But, if **x**-values changed gradually, the predictive accuracy of the MW, JIT, and TD models was low totally. We can say that the data selection of a JIT model succeeds and a predictive JIT model

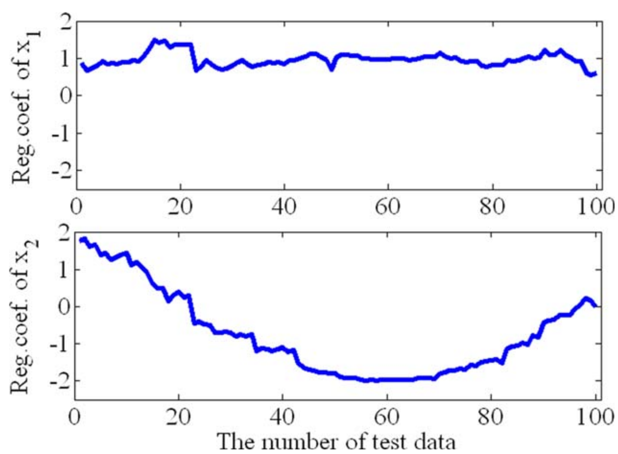


Figure 3. Time plots of regression coefficients of the MW model when Eq. 5 was used as b_2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

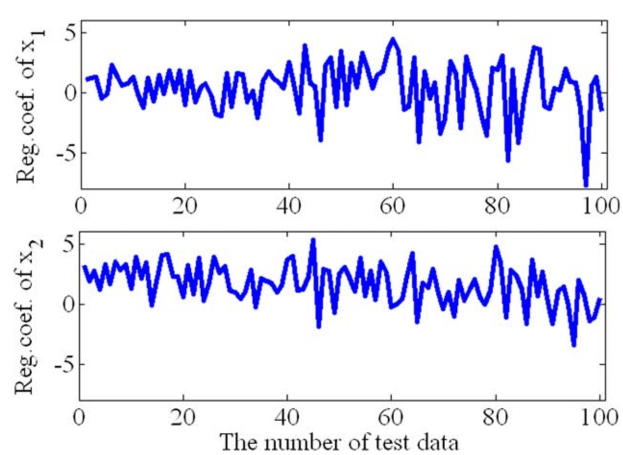


Figure 4. Time plots of regression coefficients of the JIT model when Eq. 5 was used as b_2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

Table 5. r_{pred}^2 Values of Each Model for Each $\mathbf{x}_{1,\ new}$ and Each b_2

$\mathbf{x}_{1,\ new}$	Equation for b_2	TD Model	MW Model	JIT Model
(7)	–	1.000	0.999	0.967
(8)	–	0.696	0.264	0.911
(7)	(5)	−6.07	0.539	−5.68
(8)	(5)	−0.680	0.213	−6.78
(7)	(9)	0.576	0.527	0.498
(8)	(9)	0.897	0.807	0.990

can be constructed if the change of the slope and the shift of $\mathbf{x}$-values occur simultaneously.

Figure 6 shows the time plots of regression coefficients of the JIT model when Eq. 8 was used as $\mathbf{x}_{2,\ new}$ and Eq. 9 was used as b_2. When the number of test data was 20 and $\mathbf{x}_{2,\ new}$ changed from around 0 to around 20, the regression coefficient of $\mathbf{x}_{2,\ new}$ appropriately changed from around 1 to 21, which was the actual b_2 as shown in Eq. 9. For the JIT model, the selection of a data set succeeded and appropriate regression models could be constructed. Additionally, the regression coefficients in Figure 6 did not lag behind actual b_2 unlike Figure 3. This is because data selection was performed with database for every test data and the constructed model did not affected by stored data that are close in terms of time.

However, when the number of data close to prediction data is small, data far from prediction data are also selected for the JIT model. As shown in Figure 6, regression coefficients of $\mathbf{x}_{1,\ new}$ were sometimes large when the number of test data is from 1 to 20. In these cases, the number of data where b_2 equals 1 was only 20 and then not all those data were selected for the construction of regression models in fact. Thus, careful attention is required to build relationships between a stored data set and prediction data.

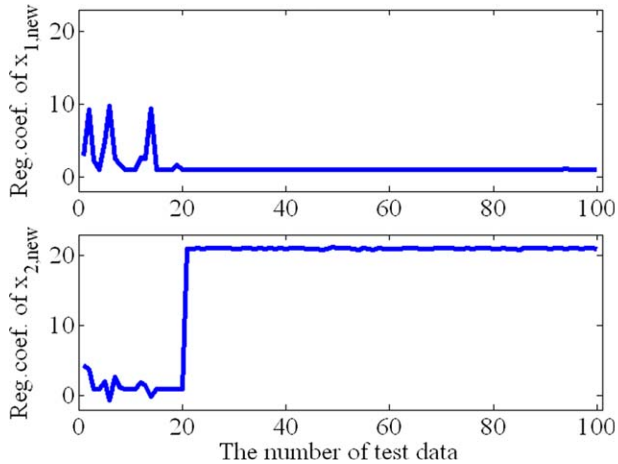


Figure 6. Time plots of regression coefficients of the JIT model when Eq. 8 was used as $\mathbf{x}_{2,\ new}$ and Eq. 9 was used as b_2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

The time plots of regression coefficients of the MW model are shown in Figure 7 when Eq. 8 was used as $\mathbf{x}_{2,\ new}$ and Eq. 9 was used as b_2. Until the number of test data is 20, each data where b_2 equals 1 is added to data where b_2 equals 21 for the MW model. After that, each data where b_2 equals 21 is added to data where b_2 equals 1. In these time spans, therefore, the appropriate MW model could not be constructed, reflecting that regression coefficients of $\mathbf{x}_{1,\ new}$ and $\mathbf{x}_{2,\ new}$ changed improperly.

Analysis of real industrial data

We analyzed the data obtained from an operation of a distillation column at Mizushima works, Mitsubishi Chemical Corporation to compare the performance of each adaptive model when the degradation of soft sensors occurs. It should be careful that the objective of this case study is not the construction of high predictive model but the comparison of each adaptive model.

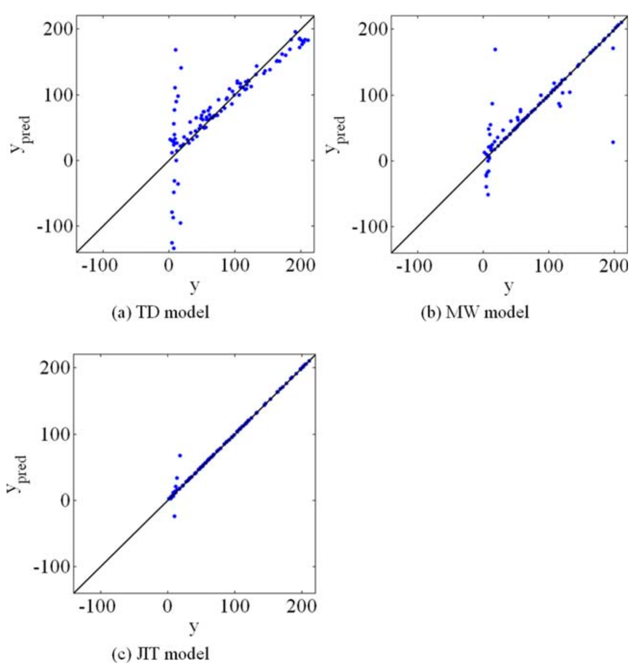


Figure 5. Relationships between y and predicted y of test data for each model when Eq. 8 was used as $\mathbf{x}_{2,\ new}$ and Eq. 9 was used as b_2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

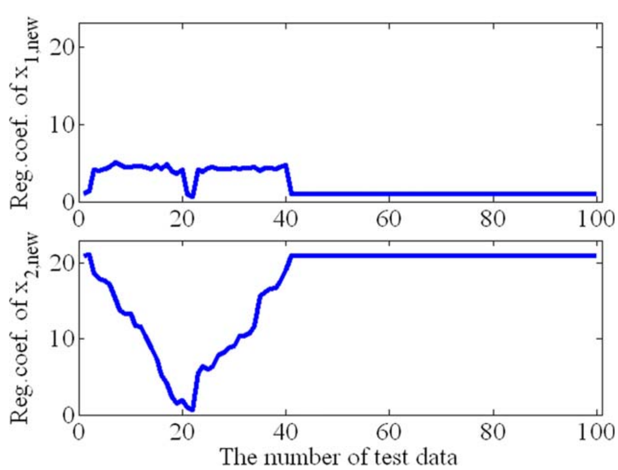


Figure 7. Time plots of regression coefficients of the MW model when Eq. 8 was used as $\mathbf{x}_{2,\ new}$ and Eq. 9 was used as b_2.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

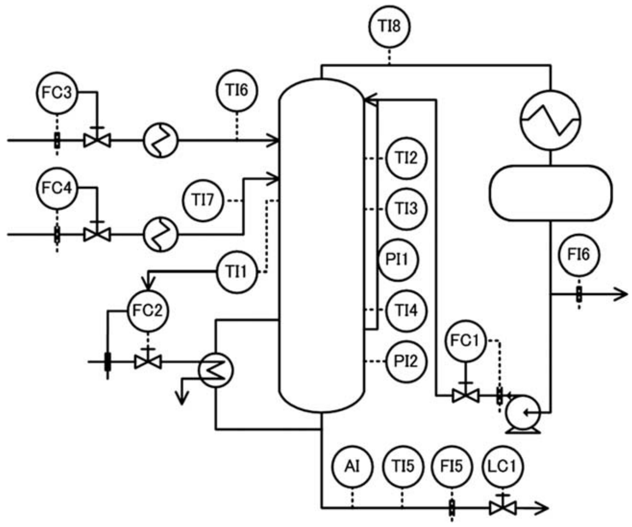


Figure 8. A schematic representation of the distillation column.

Figure 8 shows a schematic representation of the distillation column, and Table 6 shows the process variables. A **y**-variable is the concentration of the bottom product having a lower boiling point; and **X**-variables are that represent 19 variables such as temperature and pressure. The input variables are F3 and F4, and the operational variables are F1 and F2.

The measurement interval of **y** is 30 min. For training data, we used data from monitoring that took place from January to March 2003, and for test data, we used data from April 2003 to December 2006. Data that reflects variations caused by **y**-analyzer fault were eliminated in advance.

The TD was calculated between the present values and those that were 30 min before the present time because the measurement interval of **y** was 30 min, and then, the smallest interval was also 30 min. If the time interval is large, the model will be affected by the disturbance during the interval. We used a PLS method as a regression approach and set the number of data for constructing a MW model and a JIT model as 100.

Table 6. Process Variables

No.	Symbol	Objective Variable
	AI	Bottom Product Concentration
No.	Symbol	Explanatory Variables
1	FC1	Reflux flow
2	FC2	Reboiler flow
3	FC3	Feed 1 flow
4	FC4	Feed 2 flow
5	FI5	Bottom flow
6	FI6	Top flow
7	LC1	Liquid level
8	PI1	Pressure 1
9	PI2	Pressure 2
10	TI1	Temperature 1
11	TI2	Temperature 2
12	TI3	Temperature 3
13	TI4	Temperature 4
14	TI5	Bottom temperature
15	TI6	Feed 1 temperature
16	TI7	Feed 2 temperature
17	TI8	Top temperature
18	F4/F3=R	Reflux ratio
19	F1/F6=F	Feed flow ratio

To incorporate the dynamics of process variables into soft sensor models, **X** included each explanatory variable that was delayed for durations ranging from 0 to 60 min in steps of 10 min, that is, **X** consists of seven time points (0, 10, 20, 30, 40, 50, and 60 min) times 19 variables ($7 \times 19=133$). The normal PLS model was constructed and the r^2 and q^2 values of the model were 0.974 and 0.973, respectively. The details of r^2 and q^2 are explained in Appendix. Meanwhile, the TD model was constructed by using the PLS method and the r^2 and q^2 values of the model were also 0.985 and 0.984, respectively.

Additionally, to consider UOD, only the first and second score vectors of the normal PLS model were used as **X**-variables and then the PLS model was constructed again. The r^2 and q^2 values of the PLS model of two components were both 0.961. Meanwhile, the r^2 and q^2 values of the TD model were both 0.977, respectively.

Table 7 shows $RMSE_P$ values for each model in the distillation column. In the results of two components, the $RMSE_P$ value of the MW model was the smallest and that of the JIT model was largest. The slope of **X** and **y** must change in this plant and the MW model could adapt this change. On the other hand, when 133 variables were used as **X**-variables, the predictive accuracy of the MW and the JIT models was high and that of the TD model was lowest. For the JIT model, when only two components were used and information on **X** lacked, the training data sets were not selected appropriately because of unobserved disturbance of **X**. But, by increasing **X**-variables and information on **X**, the appropriate selection of the training data sets could be performed according to the change of **X** and, therefore, the predictive ability of the JIT model increased.

On the contrary, the MW model decreased the predictive accuracy, increasing the information on **X** from two components to 133 variables. When 133 variables were used as **X**-variables, the model would overfit to training data because the number of training data was 100 and smaller than that of **X**-variables. We confirmed the calculation of the MW model where the number of training data was 500 and 1000, and then, the $RMSE_P$ values were larger than that of 100 training data. If the number of data was large, the MW model was more affected by old data in training data and, thus, the predictive accuracy of the MW model decreased.

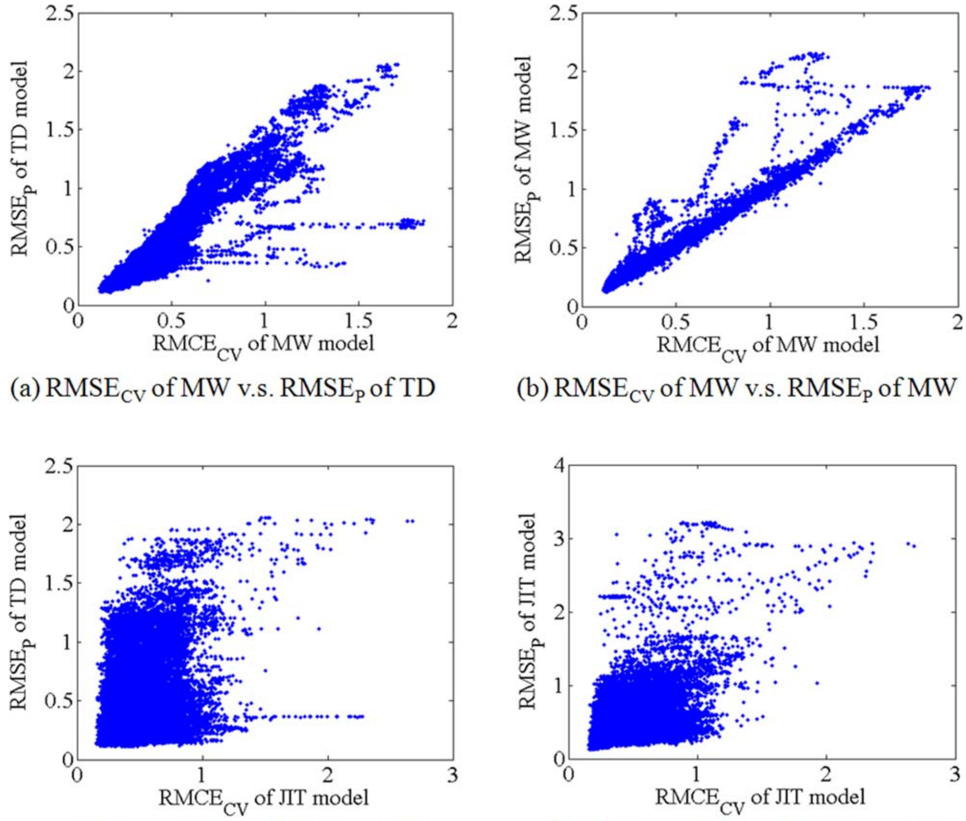
The $RMSE_P$ value of the TM model with two components did not change much from that of 133 variables, reflecting that the degree of unobserved disturbance did not affect the predictive ability of a TD model.

To investigate the relationship between a state of the plant and prediction accuracy of each adaptive model, we checked the relationships between the $RMSE_{CV}$ values of the MW and JIT models, and the $RMSE_P$ values calculated with 100 data where the each adaptive model predicted the **y**-values. Figures 9 and 10 show the relationships between $RMSE_{CV}$ and $RMSE_P$ when the two components of the

Table 7. $RMSE_P$ Values for Each Model in the Distillation Column

X	TD Model	MW Model	JIT Model
Two components of the PLS model*	0.492	0.435	0.524
133 variables	0.478	0.451	0.455

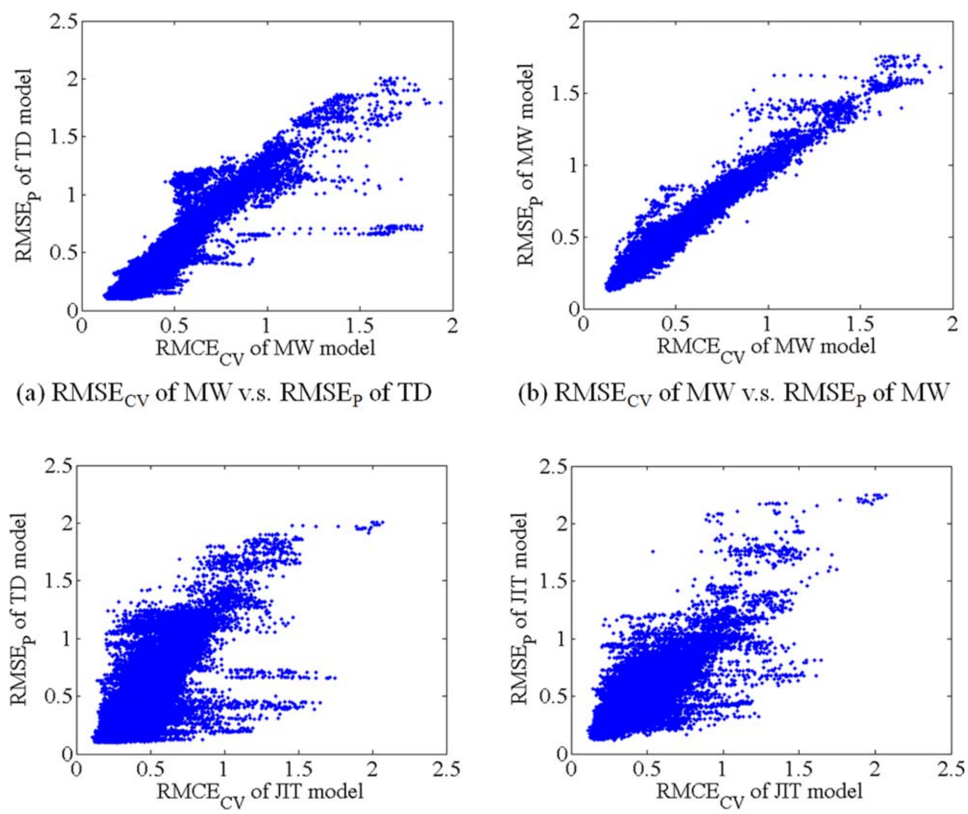
*The PLS model was constructed with 133 variables.



(a) RMSE$_{CV}$ of MW v.s. RMSE$_P$ of TD
(b) RMSE$_{CV}$ of MW v.s. RMSE$_P$ of MW
(c) RMSE$_{CV}$ of JIT v.s. RMSE$_P$ of TD
(d) RMSE$_{CV}$ of JIT v.s. RMSE$_P$ of JIT

Figure 9. Relationships between RMSE$_{CV}$ and RMSE$_P$ when the two components the PLS model were used as X-variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]



(a) RMSE$_{CV}$ of MW v.s. RMSE$_P$ of TD
(b) RMSE$_{CV}$ of MW v.s. RMSE$_P$ of MW
(c) RMSE$_{CV}$ of JIT v.s. RMSE$_P$ of TD
(d) RMSE$_{CV}$ of JIT v.s. RMSE$_P$ of JIT

Figure 10. Relationships between RMSE$_{CV}$ and RMSE$_P$ when the 133 variables were used as X-variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

PLS model and the 133 variables were used as **X**-variables, respectively. The each point represents the relationship between the each $\mathrm{RMSE}_{\mathrm{CV}}$ value calculated at a time and the each $\mathrm{RMSE}_{\mathrm{P}}$ value calculated with 100 data backward from the same time. In Figures 9a and 10a, the $\mathrm{RMSE}_{\mathrm{CV}}$ values of the MW models have high correlations with the $\mathrm{RMSE}_{\mathrm{P}}$ values of the TD models. When the slope of **X** and **y** changes, the predictive accuracy of a TD model decreases as discussed in "Classification of the degradation." The degradation of the predictive ability of a MW model means the change of the slope of **X** and **y** because predictive models cannot be constructed when a state of process is shifting and accordingly the relationship between **X** and **y** is changing.

Meanwhile, the predictive accuracy of the TD models was sometimes high even though the $\mathrm{RMSE}_{\mathrm{CV}}$ values of the MW models were large as shown in Figures 9a and 10a. In fact, the **y**-values shifted in those cases. For the MW models, the $\mathrm{RMSE}_{\mathrm{CV}}$ value got large and the predictive accuracy decreased because a regression model was constructed with a data set including data before the shift of **y**-values and data after that. However, the TD models could adapt the shift of **y**-values and predict the **y**-values with high accuracy, and then, the $\mathrm{RMSE}_{\mathrm{P}}$ values were small. We, therefore, confirmed that the predictive accuracy of a TD model decreases when the slope of **X** and **y** changes and that a TD model can adapt the shift of **y**-values.

In Figures 9b and 10b, there are correlations between the $\mathrm{RMSE}_{\mathrm{CV}}$ and the $\mathrm{RMSE}_{\mathrm{P}}$ of the MW models. When a state of the plant changes, that is, the shift of **X**-values or **y**-values, or the change of the slope of **X** and **y** happens, the construction of appropriate regression models is difficult and the $\mathrm{RMSE}_{\mathrm{CV}}$ values are large. In this situation, the prediction accuracy of the next data will decrease. Therefore, the correlation between the $\mathrm{RMSE}_{\mathrm{CV}}$ and the $\mathrm{RMSE}_{\mathrm{P}}$ of a MW model is reasonable.

On the other hand, the $\mathrm{RMSE}_{\mathrm{P}}$ values were sometimes large in the cases of the small values of the $\mathrm{RMSE}_{\mathrm{CV}}$. When a state of the plant changes rapidly, the $\mathrm{RMSE}_{\mathrm{CV}}$ values will increase, but the predictive accuracy of new data will further decrease. Of course, it can be said that the prediction accuracy of test data is lower than the validation accuracy of training data from a statistical perspective.

When only two components were used as **X**-variables, the $\mathrm{RMSE}_{\mathrm{CV}}$ values of the JIT model were poorly correlated with the $\mathrm{RMSE}_{\mathrm{P}}$ values of the TD model and the JIT model from Figures 9c, d. The appropriate selection of a data set for the JIT model could not performed probably because only two components could not represent the changes of **X**-variables fully. As a result, the predictive ability of the JIT model changed independently of a state of the plant and it is reasonable that it did not affect the predictive accuracy of the TD model and the JIT model. Meanwhile, as shown in Figures 10c, d, correlations between the $\mathrm{RMSE}_{\mathrm{CV}}$ values of the JIT model and the $\mathrm{RMSE}_{\mathrm{P}}$ values of the TD model and the JIT model were confirmed by increasing the information on **X**-variables. The 133 variables could represent the change of **X**; the data selection was appropriately performed to some extent and, thus, the $\mathrm{RMSE}_{\mathrm{CV}}$ of the JIT model could reflect a state of the plant.

Figure 11 shows the time plots of $\mathrm{RMSE}_{\mathrm{P}}$, $\mathrm{RMSE}_{\mathrm{CV}}$ of the MW model and average of **y** in March and April 2005. The $\mathrm{RMSE}_{\mathrm{P}}$ values of the all models, especially the TD model, were high around 250 h. The $\mathrm{RMSE}_{\mathrm{CV}}$ values were relatively high as shown in Figure 11b and the shift of **y**-values did not occur as shown in Figure 11c around this time. It, therefore, can be said that the slope of **X** and **y** changed. As discussed in "Characteristics of adaptive models," the TD model could not adapt this situation. After that, however, the $\mathrm{RMSE}_{\mathrm{P}}$ values of the TD model decreased and the predictive accuracy of the TD model was higher than that of the MW model. This will be because the data during variation of the plant were included in training data of the MW model. In this situation, the TD model is desirable for a soft sensor model.

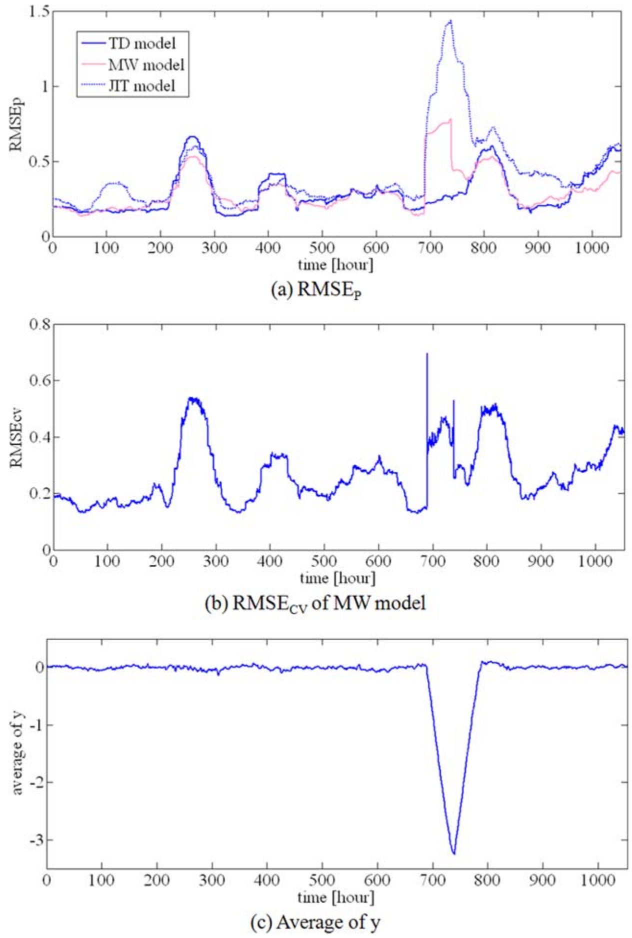


Figure 11. Time plots of $\mathrm{RMSE}_{\mathrm{P}}$ and $\mathrm{RMSE}_{\mathrm{CV}}$ of MW model and average of y in March and April 2005.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

After 700 h, the $\mathrm{RMSE}_{\mathrm{P}}$ values of the MW model and the JIT model increased though those of the TD model did not change in Figure 11a. The rapid shift of **y**-values happened in this situation as shown in Figure 11c. The TD model could adapt the rapid shift of **y**-values and predict **y**-values appropriately as discussed in "Characteristics of adaptive models." The $\mathrm{RMSE}_{\mathrm{P}}$ values of all models increased after the **y**-shift. In this time, the $\mathrm{RMSE}_{\mathrm{CV}}$ values were high and the shift of **y**-values did not occur and, therefore, this situation is the same as that around 250 h.

As stated above, we confirmed that the discussion in "Characteristics of adaptive models" is applicable to real industrial data.

Conclusions

In this article, we categorized the degradation of a soft sensor model and discussed characteristics of MW, JIT, and TD models, based on the classification results. Then, we confirmed the discussion results with simulated data sets. The results suggested that there exist appropriate models for each type of the degradation. When the shift of **y**-values or **x**-values occurs, a TD model is suitable. But, we should give attention to data just after the instant shift. A MW model is appropriate for the gradual change of the slope of **x** and **y**.

When the slope of **x** and **y** changes instantly, a JIT model should be used if the shift of **x**-values also happens and the amount of training data that are close to new data is enough, but if not, a MW model will have higher predictive ability than a JIT model and a TD model do. However, the predictive accuracy of a MW model is not sufficient and, therefore, other methods are needed for predictive soft sensors.

Then, the each adaptive model was compared by using the real industrial data of a distillation column. The predictive accuracy of the TD model was high when the shift of **y**-values occurred and that of the MW model was high when the slope of **X** and **y** changed. Meanwhile, if the effect of unobserved disturbance was large by design, the shift of **x**-values was small and then the data selection in the JIT model was not performed well. However, when the shift of **x**-values happens, the predictive model could be constructed by the JIT method even if there is nonlinearity between **X** and **y**.

By using the knowledge we obtained in this study, the predictive accuracy of soft sensors will improve, and then, chemical plants can be operated stably.

Acknowledgments

H. Kaneko is grateful for financial support of the Japan Society for the Promotion of Science (JSPS) through a Grant-in-Aid for Young Scientists (B) (No. 24760629). The authors acknowledge the support of Mizushima Works, Mitsubishi Chemical Co., and the financial support of Mizuho Foundation for the Promotion of Sciences and the Toyota Physical and Chemical Research Institute.

Literature Cited

1. Kano M, Nakagawa Y. Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry. *Comput Chem Eng*. 2008;32:12–24.
2. Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Comput Chem Eng*. 2009;33:795–814.
3. Kaneko H, Arakawa M, Funatsu K. Development of a new soft sensor method using independent component analysis and partial least squares. *AIChE J*. 2009;55:87–98.
4. Kaneko H, Arakawa M, Funatsu K. Applicability domains and accuracy of prediction of soft sensor models. *AIChE J*. 2011;57:1506–1513.
5. Kadlec P, Gabrys B. Local learning-based adaptive soft sensor for catalyst activation prediction, *AIChE J*. 2010;57:1288–1301.
6. Qin SJ. Recursive PLS algorithms for adaptive data modeling. *Comput Chem Eng*. 1998;22:503–514.
7. Cheng C, Chiu MS. A new data-based methodology for nonlinear process modeling. *Chem Eng Sci*. 2004;59:2801–2810.
8. Fujiwara K, Kano M, Hasebe S, Takinami A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE J*. 2009;55:1754–1765.
9. Kaneko H, Funatsu K. Maintenance-free soft sensor models with time difference of process variables. *Chemom Intell Lab Syst*. 2011;107:312–317.
10. Kaneko H, Funatsu K. A soft sensor method based on values predicted from multiple intervals of time difference for improvement and estimation of prediction accuracy. *Chemom Intell Lab Syst*. 2011;109:197–206.
11. Dunia R, Qin SJ, Edgar TF, McAvoy TJ. Identification of faulty sensors using principal component analysis. *AIChE J*. 1996;42:2797–2812.
12. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Sys*. 2001;58:109–130.

Appendix

In this article, r^2, q^2, root-mean-square error (RMSE), and $\mathrm{RMSE_{CV}}$ were used as measures of accuracy and the predictive ability of regression models, and were defined as

$$r^2 = 1 - \frac{\sum \left(y_{\mathrm{obs}} - y_{\mathrm{calc}}\right)^2}{\sum \left(y_{\mathrm{obs}} - \bar{y}\right)^2} \tag{A1}$$

$$q^2 = 1 - \frac{\sum \left(y_{\mathrm{obs}} - y_{\mathrm{pred}}\right)^2}{\sum \left(y_{\mathrm{obs}} - \bar{y}\right)^2} \tag{A2}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum \left(y_{\mathrm{obs}} - y_{\mathrm{calc}}\right)^2}{n}} \tag{A3}$$

$$\mathrm{RMSE}_{\mathrm{CV}} = \sqrt{\frac{\sum \left(y_{\mathrm{obs}} - y_{\mathrm{pred}}\right)^2}{n}} \tag{A4}$$

where y_{obs} is the measured y value; y_{calc} is the calculated y value; y_{pred} is the predicted y value in the cross-validation procedure; and n is the number of data points. In this study, the leave-one-out cross-validation method was used in the calculation of y_{pred}. In the above equations, r^2 represents the fitting accuracy of the constructed models, and q^2 represents the predictive accuracy of the constructed models. Values close to unity for both r^2 and q^2 are favorable. Comparison of the r^2 and q^2 values requires the use of models constructed with the same objective variable data. The lower the RMSE and $\mathrm{RMSE_{CV}}$ values, the greater the predictive accuracy achieved with the constructed model.

${r_{\mathrm{pred}}}^2$ and $\mathrm{RMSE_P}$ are values of r^2 and RMSE calculated from the test data, respectively.

Manuscript received Mar. 26, 2012, and revision received Dec. 7, 2012.