

# Database Monitoring Index for Adaptive Soft Sensors and the Application to Industrial Process

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.14260

Published online October 21, 2013 in Wiley Online Library (wileyonlinelibrary.com)

*In chemical plants, soft sensors are used to predict difficult-to-measure process variables. Soft sensor models must adapt to process changes by using new measured data. However, when a model is reconstructed with data that have low variation, the model cannot predict abrupt changes of process characteristics. The predictive performance of adaptive models depends on databases. We therefore propose an index to monitor database, that is, database monitoring index (DMI), and a database monitoring method using the DMI. The DMI is based on similarity between two data. The more similar two data are the smaller value the DMI has. New data are stored when the minimum DMI-value of the data exceeds a threshold. Through the analysis of simulation data and real industrial data, we confirmed that databases can be appropriately managed and the predictive accuracy of adaptive soft sensor models increased by using the proposed method. © 2013 American Institute of Chemical Engineers AICHE J, 60: 160–169, 2014*

**Keywords:** process control, soft sensor, degradation, adaptive model, database, monitoring

## Introduction

Soft sensors are widely used to predict process variables that are difficult to measure online.<sup>1</sup> An inferential model is constructed between the variables that are easy to measure online and those that are not, and an objective variable,  $y$ , is then predicted using that model.

Their use, however, is accompanied by some practical difficulties. One of these difficulties is the degradation of the soft sensor models. The predictive accuracy of soft sensors tends to decrease gradually for several reasons, including changes in the state of the chemical plant, catalyzing performance loss, and sensor and process drift. This is called as the degradation of soft sensor models.

It is strongly desired to solve the degradation of a soft sensor model. To reduce the degradation, the model is reconstructed with newest data. A moving window (MW) model<sup>2,3</sup> and a recursive model<sup>4</sup> are categorized as a sequentially updating type and a distance-based just-in-time (JIT) model,<sup>5</sup> a correlation-based JIT model,<sup>6</sup> and a locally weighted partial least-square model<sup>7</sup> are categorized as a JIT type. For example, an MW model is constructed with data that are measured most recently, and a distance-based JIT model is constructed with data whose similarity with prediction data are higher than those of other data. The indexes such as the Euclidian distance and correlation are applied to the similarity. For the reconstruction of the model in MW models and JIT models, new data are stored in the database and the data regarded as old are deleted from the database.

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.<sup>8–10</sup> 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 handled by using a TD model without reconstruction of the model. The models such as MW, JIT, and TD models that can predict  $y$ -values while adapting to states of a plant are called adaptive models.<sup>11</sup> In addition, when data distributions are multimodal, multiple modeling approaches<sup>12,13</sup> can be combined with adaptive models.

There are no adaptive models having high predictive ability in all process states, and the prediction accuracy of each adaptive model depends on a process state.<sup>14</sup> Kaneko et al. categorized the degradation of a soft sensor model and discussed characteristics of adaptive models, such as MW, JIT, and TD models, based on the classification results, and confirmed the discussion results through the numerical simulation data and real industrial data analyses.<sup>15</sup> The predictive accuracy of TD models was high when the shift of  $X$ -values or  $y$ -values occurred and this is true regardless of the rapidity of the degradation. Meanwhile, when the slope of  $X$  and  $y$  changes rapidly, the predictive model can be constructed by the JIT method if there is the shift of  $X$ -values. However, if there is no shift of  $X$ -values, JIT models cannot adapt to the degradation. On the other hand, the MW models based on the support vector regression (SVR) method, which is one of the nonlinear regression methods, and the time variable can adapt to rapid changes of the slope of  $X$  and  $y$  even without shift of  $X$ -values (Kaneko and Funatsu. Adaptive soft sensor model using online support vector regression and the time variable. *AIChE J.* submitted).

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

While the appropriate use of the MW and JIT models enables soft sensors to adapt to the changes of the relationship between  $\mathbf{X}$  and  $\mathbf{y}$ , there remain some problems for the introduction of soft sensors into practice. One of the problems is that reconstructed models have a high tendency to specialize in predictions over a narrow data range. Subsequently, when variations in the process variables occur, these models cannot predict the resulting variations in data with a high degree of accuracy. However, if the model is not reconstructed frequently, the predictive ability of the model decreases due to the slow change of process states such as process and sensor drifts.

Therefore, in this study, to construct adaptive models with high predictive accuracy for wide data range, we monitor database appropriately. When the number of data in a database is too large, it takes much time to construct MW models and JIT models. Hence, it is required to judge whether new measured data should be stored in a database or not. Data having much information should be included in the database, but data with little information are not required in the database. Although the selection of new measurement data to store in the database in JIT modeling has been applied based on the prediction error in order to reduce the amount of the database,<sup>16,17</sup> to the best of our knowledge, there is as yet no research that manages database based on the amount of information of data.

We are developing the database monitoring index (DMI) and the database managing method with the DMI.<sup>18</sup> The DMI is an index based on similarity between two data, and is defined as the ratio of absolute difference of  $\mathbf{y}$  divided by similarity of  $\mathbf{X}$ . The more similar two data are, the smaller value the DMI has.

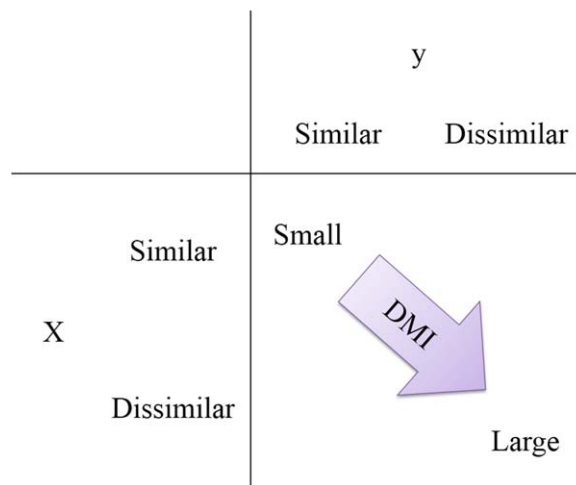
When new data are measured, the DMI-values are calculated between the new data and all data in a database. The data whose minimum value of the DMI-values is large has much information. Therefore, by storing only such data, the amount of information can increase while controlling the number of data in the database. The DMI enables a database to be managed with the consideration of not only the shifts of  $\mathbf{X}$  and  $\mathbf{y}$  but also the change of the slope between  $\mathbf{X}$  and  $\mathbf{y}$ . In addition, the proposed data managing method can be combined with not only the adaptive soft sensor models<sup>19,20</sup> but also the process monitoring models<sup>21,22</sup> that are updated or reconstructed with database including new measurement data.

To verify the effectiveness of the proposed method, we analyze simulation data where the relationship between  $\mathbf{X}$  and  $\mathbf{y}$  is nonlinear and data variation is small for a constant time. The drift of  $\mathbf{y}$  is also considered. Then, the proposed method is applied to real industrial data of a distillation column. The management of databases using the DMI makes it possible for adaptive soft sensor models to adapt to rapid changes of process characteristics after long states of small variations in  $\mathbf{X}$  and  $\mathbf{y}$ . In this study, it is assumed that there are no abnormal data and no outliers in new measurement data. In practice, therefore, abnormal data and outliers must be detected first by using multivariate statistical process control methods.<sup>23</sup>

## Method

### DMI

The DMI proposed in this article for managing databases is defined between two data,  $(\mathbf{x}_i, y_i)$  and  $(\mathbf{x}_j, y_j)$ , as follows



**Figure 1. The basic concept of the DMI.**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

$$\text{DMI} = \frac{|y_i - y_j|^a}{\text{sim}(\mathbf{x}_i, \mathbf{x}_j)} \quad (1)$$

where  $\text{sim}(\mathbf{x}_i, \mathbf{x}_j)$  is similarity between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , and  $a$  is a constant. For example, inverse of Euclidian distance and Mahalanobis distance, correlation, Gaussian kernel (GK) are used as similarity. The inverse of Euclidian distance (iED) is given as follows

$$\text{sim}(\mathbf{x}_i, \mathbf{x}_j) = \text{iED}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|^2}} \quad (2)$$

and GK is given as follows

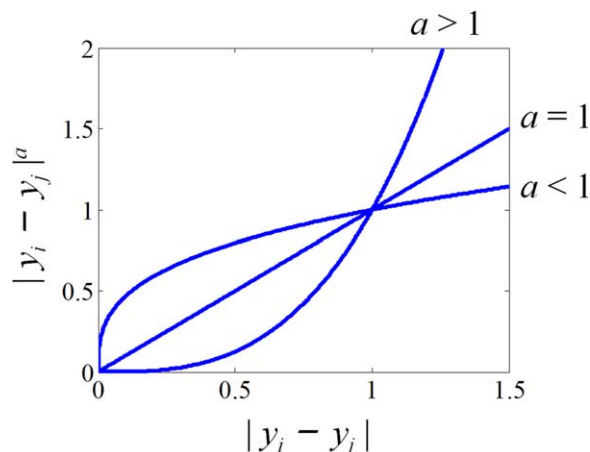
$$\text{sim}(\mathbf{x}_i, \mathbf{x}_j) = \text{GK}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (3)$$

where  $\gamma$  is a tuning parameter controlling the width of the kernel function. In our application, the similarity index is GK. The DMI used in this study is given as follows

$$\text{DMI} = \frac{|y_i - y_j|^a}{\exp\left(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\right)} \quad (4)$$

The basic concept of the DMI is shown in Figure 1. The DMI-values are large when two data of  $\mathbf{X}$  and  $\mathbf{y}$  are dissimilar, whereas the DMI-values are small when two data of  $\mathbf{X}$  and  $\mathbf{y}$  is similar. Although the dissimilarity used in process monitoring<sup>24,25</sup> is calculated between a data set and another data set, the proposed DMI is calculated between a datum and another datum. Therefore, the DMI can quantify data similarity even for nonlinear and non-Gaussian processes.

The weight of  $\mathbf{y}$  for  $\mathbf{X}$  can be changed by varying the  $a$ -value in Eqs. 1 and 4. Figure 2 shows the relationship between  $|y_i - y_j|$  and  $|y_i - y_j|^a$ . The smaller  $a$ -values compared with 1 mean the larger values of  $|y_i - y_j|^a$  and DMI-values for small variations of  $\mathbf{y}$  such as process and sensor drifts. However, much care should be taken because the DMI is sensitive to noise, which is one of the small variations. On the other hand, when the  $a$ -values are larger than 1, the DMI-values are small for small variations of  $\mathbf{y}$ . Although the DMI has small values in the drift of  $\mathbf{y}$ , the DMI is not much affected by noise and is robust over noise.



**Figure 2.** The relationships between  $|y_i - y_j|$  and  $|y_i - y_j|^a$ .<sup>18</sup>

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

The DMI can be discussed by changing  $a$ -values to be appropriate for process states and suitable to objective of the database management.

The GK used in this study is one of the similarity indexes and other similarity indexes can be applied to the DMI. Of course, the curse of dimensionality must be handled sometimes.

#### Database management with DMI

The flow of the proposed database management is shown in Figure 3. When new data is given, the DMI-values are calculated between the new data and all data included in the database. If the minimum value of the DMI-values exceeds the threshold  $P_{DMI}$ , the new data are stored in the database. If not, the new data are not stored in the database or the new data replaces the data in the database that has the minimum DMI-value. Then, if the size of the database exceeds the upper limit, data are deleted from the database. If the current database has the necessary and sufficient data, it is reasonable that the oldest data are eliminated as the minimum DMI values for all the data in the database exceed the  $P_{DMI}$  at least and the information loss is not so different in the elimination of each data. In addition, the process state where the oldest data were measured would be different from the process state where the new data were measured. For example, accurate JIT models cannot be constructed with databases where both the data before the  $y$ -shift and the data after  $y$ -shift exist.<sup>15</sup> The data before the  $y$ -shift should be immediately removed for JIT modeling. Although this elimination of the data before the  $y$ -shift is not covered in this study, those data will not exist so much by deleting the oldest data from the database.

The proposed DMI enables the database used for the construction of not only MW models but also JIT models to be managed by using the proposed DMI. In JIT modeling, some data must be selected from the enormous numbers of data or the weights of the enormous numbers of data must be set by using an index such as distance and correlation. Highly accurate JIT models can be constructed with less time by appropriately managing the database and keeping the amount of information.

In the field of structure activity relationship, structure-activity landscape index (SALI)<sup>26</sup> is famous as an index using similarity of  $\mathbf{X}$  and that of  $\mathbf{y}$  like Eq. 1. The SALI is defined between two molecules  $p$  and  $q$  as follows

$$\text{SALI} = \frac{|A_p - A_q|}{1 - \text{sim}(p, q)} \quad (5)$$

where  $A_p$  and  $A_q$  are the activity of  $p$  and  $q$ , respectively, and  $\text{sim}(p, q)$  is similarity of chemical structures between  $p$  and  $q$ . The different activity and the high similarity lead to large SALI values. The SALI has achieved significant results<sup>27</sup> such as the detection of the activity cliff, which is combination of molecules whose structures are similar and activities vary greatly.

## Results and Discussion

To verify the effectiveness of the proposed method, we analyzed simulation data and industrial distillation column data. The relationships between  $\mathbf{X}$  and  $\mathbf{y}$  have strong nonlinearity for the simulation data set.

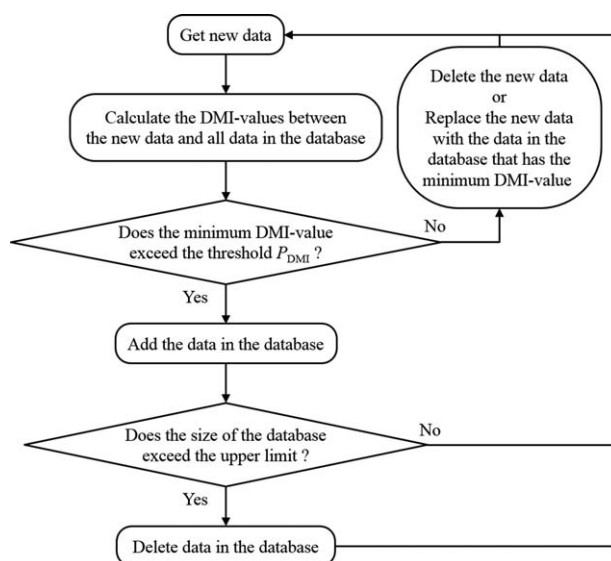
#### Data in which the relationship between $\mathbf{X}$ and $\mathbf{y}$ is nonlinear

The analysis using data, where the relationship between  $\mathbf{X}$  and  $\mathbf{y}$  is nonlinear, was performed to verify the performance of the proposed method. The relationship between  $\mathbf{X}$  and  $\mathbf{y}$  in the data was given as follows

$$y = \sin(x_1)\cos(x_2) + 0.1x_1 \quad (6)$$

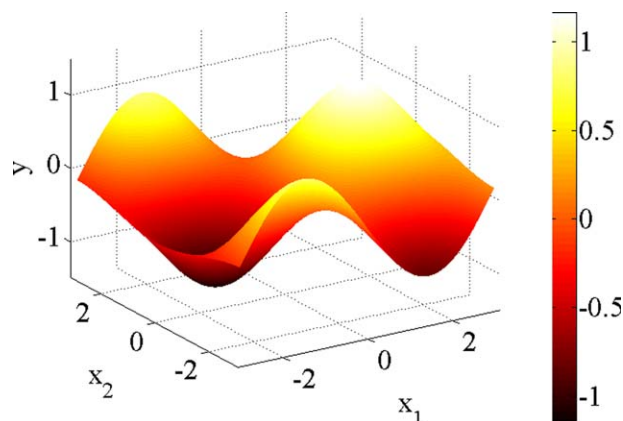
Equation 6 is described in the reference<sup>28</sup> as a test problem and  $0.1x_1$  was added to the raw equation. Figure 4 shows the relationship between  $x_1$ ,  $x_2$ , and  $y$  in Eq. 6. The colorbar represents  $y$ -values.

The data of  $\mathbf{X}$ -variables were generated to be randomly walked within  $\pm 3$ . Random numbers from the normal distribution with a mean of 0 and a standard deviation 0.01 were added to the  $y$ -variable. The time plots of  $x_1$ ,  $x_2$ , and  $y$  without the drift of  $y$  for the simulation data are shown in Figure 5. From time 786 to 986, the data variation comes from only noise and this situation seems to be in Figure 5. The first



**Figure 3.** The flow of monitoring database with the DMI.





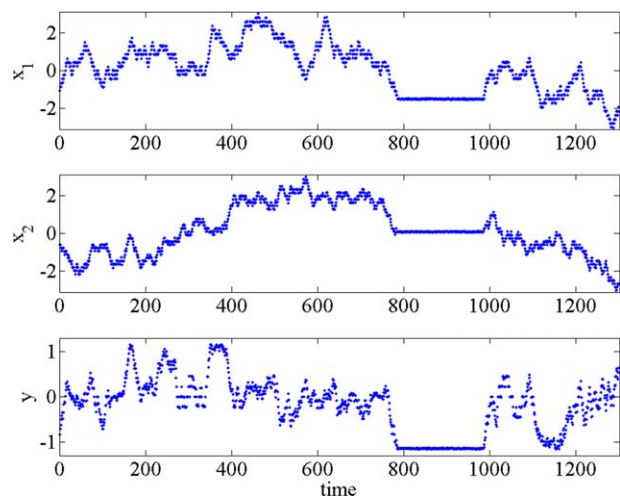
**Figure 4.** The relationship among  $x_1$ ,  $x_2$ , and  $y$  for the simulation data.<sup>18</sup>

The colorbar represents  $y$ -values. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

100 data were used for training and the next 1200 data were the test data.

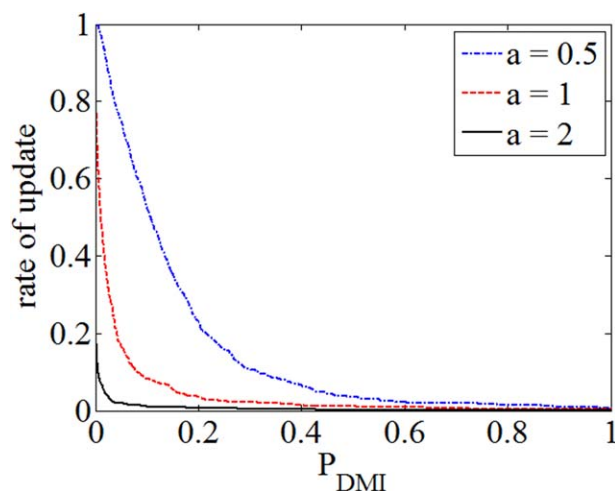
In this study, two cases were assumed; No drift (Case 1) and Linear time-varying drift (Case 2). The weight of the drift is 0.01 times of the standard deviation of  $y$  and this drift was added to  $y$  in Case 2.

**CASE 1: NO DRIFT.** The  $\gamma$ -value in Eq. 4 was optimized with 5-fold cross validation using SVR<sup>29</sup> of GK. The DMI-values were calculated with the  $\gamma$ -value of  $2^{-5}$  and Eq. 4, and the database was updated according to the flow of Figure 3. The upper limit of the number of data in the database was set as 50 and the old data was deleted automatically. We set the  $a$ -value as 0.5, 1, and 2 while changing the  $P_{\text{DMI}}$ -value from 0 to 1 in steps of 0.0001, and then, calculated the ratio of the update of the database. Figure 6 shows the results. As the  $P_{\text{DMI}}$ -values became large, the ratio of the database update was low. Additionally, the change of the  $a$ -values caused the change of the relationships between the  $P_{\text{DMI}}$ -values and the ratio of the database update.



**Figure 5.** The time plots of  $x_1$ ,  $x_2$ , and  $y$  without the drift of  $y$  for the simulation data.<sup>18</sup>

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



**Figure 6.** The relationship between  $P_{\text{DMI}}$  and rate of update when no drift of  $y$  existed for the simulation data.<sup>18</sup>

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

Then we predicted  $y$ -values with the online support vector regression (OSVR) model (Kaneko and Funatsu. Adaptive soft sensor model using online support vector regression and the time variable. *AIChE J.* submitted)<sup>30</sup> which is one of the adaptive models. The details of OSVR are shown in Appendix A. To consider measurement time of  $y$ , it was assumed that  $y$ -values can be obtained after gaining time of 5, although  $X$ -values can be given in real time. When a  $y$ -value is obtained, the DMI judges whether the new data should be stored in the database or not (Figure 3).

Table 1 shows the prediction results of the OSVR models, changing  $P_{\text{DMI}}$  values when  $a$  equals 1. The  $r_p^2$  is the determination coefficient  $r^2$  for test data and  $\text{RMSEP}$  is the root-mean-square error (RMSE) for test data. The larger values of  $r_p^2$  and the smaller values of  $\text{RMSEP}$  mean the more predictive accuracy of the model. From Table 1, the models with less update frequency had more predictive ability than the model updating each time ( $P_{\text{DMI}}$  is 0). However, when the  $P_{\text{DMI}}$ -values increased more than 0.07,  $r_p^2$  decreased and  $\text{RMSEP}$  increased. When  $P_{\text{DMI}}$  was not so large, only new data dissimilar to the data in the database were stored whereas new data similar to the data in the database were not stored, and hence, informative data in the database were not removed. Therefore, the variety of the database increased and the OSVR model constructed with that database could predict wide data range of  $y$  with high accuracy. On the other hand, when  $P_{\text{DMI}}$  was too large, the number of new data stored in the database was so small that the OSVR model could not adapt to the new process state. We can say that the appropriate  $P_{\text{DMI}}$ -values exist for the database management with the DMI.

The result of  $P_{\text{DMI}}$  of 0.07 where the  $r_p^2$ -value was maximum and the  $\text{RMSEP}$ -value was minimum showed that high prediction accuracy was achieved with only 12.0% model update by appropriately selecting data that should be stored in the database. Data required to represent the nonlinear relationship between  $X$  and  $y$  could be properly selected by using the proposed DMI.

Figure 7 shows the time plots of simulated and predicted  $y$  from 950 to 1050 when no drift of  $y$  existed. When the

**Table 1. The Prediction Results for Each  $P_{\text{DMI}}$  when No Drift of  $y$  Existed for the Simulation Data when  $a$  Equals 1<sup>18</sup>**

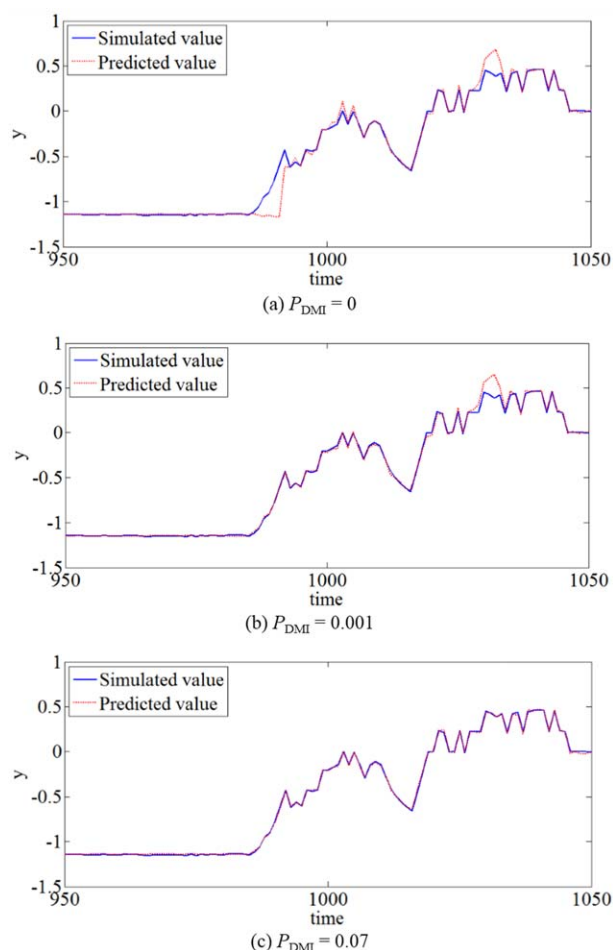
$P_{\text{DMI}}$	Rate of update	$r_p^2$	$\text{RMSE}_P$
0	1.0000	0.9962	0.0378
0.0001	0.9775	0.9964	0.0367
0.0002	0.9517	0.9961	0.0381
0.0005	0.8975	0.9972	0.0323
0.0007	0.8675	0.9978	0.0288
0.001	0.8317	0.9976	0.0300
0.002	0.7667	0.9975	0.0307
0.005	0.6500	0.9971	0.0330
0.007	0.5900	0.9976	0.0301
0.01	0.5200	0.9970	0.0333
0.02	0.3733	0.9975	0.0305
0.05	0.1675	0.9968	0.0343
0.07	0.1200	0.9980	0.0270
0.1	0.0825	0.9930	0.0508
0.2	0.0358	0.9902	0.0603
0.5	0.0108	0.8789	0.2121
0.7	0.0058	0.7869	0.2814
5	0.0000	-0.1407	0.6510

model was updated each time, the prediction errors were large in around time 990 when the abrupt process change happened (Figure 7a). This is because the model was constructed with only data including the small variation, and specialized in that state. Thus, the model could not adapt to the next abrupt change. Meanwhile, when the  $P_{\text{DMI}}$ -value was 0.001 and the model update with data in the small variation was avoided, the model could adapt to the abrupt variation in around time 990 (Figure 7b). Besides, from Figure 7c, the high prediction accuracy was achieved also in the case of the model update frequency of 12.0% ( $P_{\text{DMI}}$  is 0.07). We could confirmed that the appropriate selection of data required to the database with the proposed DMI enables to predict wide range of  $y$  with high accuracy.

**CASE 2: EXISTING DRIFT.** The drift was added to only  $y$ -variable and we conducted another case study as we did in Case 1. The relationship between  $P_{\text{DMI}}$  and rate of update is shown in Figure 8. The increase of  $P_{\text{DMI}}$  and the change of  $a$  affected the decrease of update and the change of relationship between  $P_{\text{DMI}}$  and the rate of update, respectively. Figure 8 is very similar to Figure 8 because the difference in the data of Case 1 and Case 2 is only the drift added to  $y$ -variable.

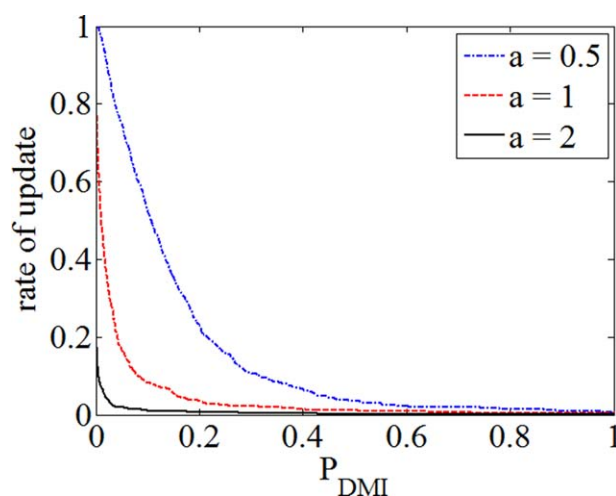
Table 2 shows the prediction results of the OSVR models. Even when the drift of  $y$  existed, the  $r_p^2$ -value increased and the  $\text{RMSE}_P$ -value decreased by selecting the threshold  $P_{\text{DMI}}$  appropriately. When the  $r_p^2$ -value was maximum and the  $\text{RMSE}_P$ -value was minimum, the  $P_{\text{DMI}}$ -value was 0.002 and the rate of update was 79.9%. The rate increased compared with that of no drift (Case 1) since the OSVR model had to adapt frequently to the change of the relationship between  $X$  and  $y$  due to the drift.

Figure 9 shows the time plots of simulated and predicted  $y$  from 950 to 1050 when the  $P_{\text{DMI}}$ -values are 0, 0.002, and 0.07. In the case of updating model each time, the model adapted to the data with the small variation and could not accurately predict the  $y$ -values in the rapid variation (Figure 9a). Meanwhile, from Figures 9b and 9c, the updated model could predict  $y$ -values in not only the small variation but also in the rapid variation after that with high accuracy by using the DMI and selecting the data for the model update. Even with the drift of  $y$ , the appropriate database management could be achieved using the proposed DMI.



**Figure 7. The time plots of simulated and predicted  $y$  from 950 to 1050 when no drift of  $y$  existed for the simulation data.<sup>18</sup>**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



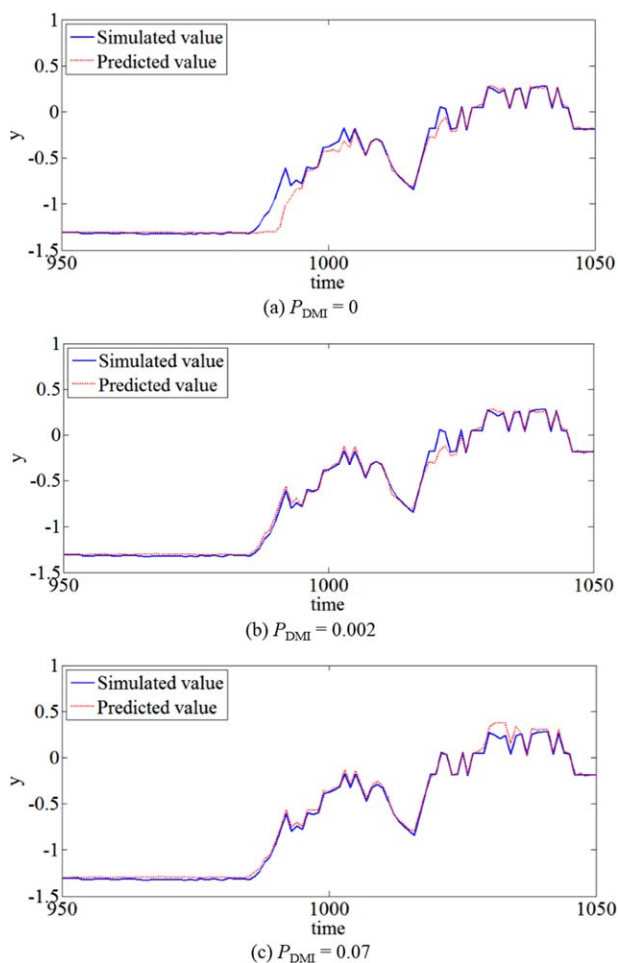
**Figure 8. The relationship between  $P_{\text{DMI}}$  and rate of update when drift of  $y$  existed for the simulation data.<sup>18</sup>**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

**Table 2. The Prediction Results for Each  $P_{DMI}$  when Drift of  $y$  Existed for the Simulation Data when  $a$  Equals 1<sup>18</sup>**

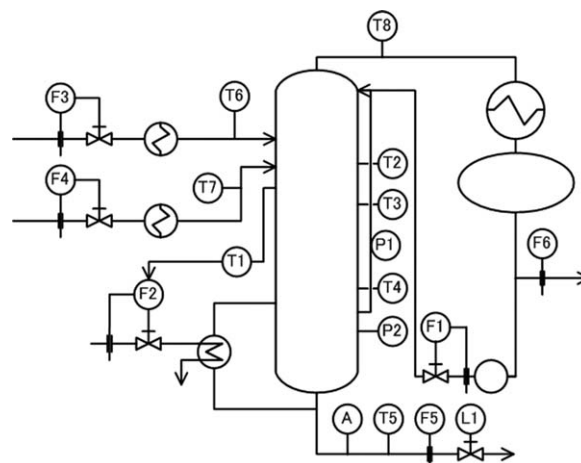
$P_{DMI}$	Rate of update	$r_p^2$	$RMSEP$
0	1.0000	0.9958	0.0416
0.0001	0.9850	0.9957	0.0424
0.0002	0.9642	0.9956	0.0427
0.0005	0.9233	0.9955	0.0432
0.0007	0.9025	0.9959	0.0413
0.001	0.8675	0.9966	0.0374
0.002	0.7992	0.9971	0.0344
0.005	0.6817	0.9971	0.0344
0.007	0.6208	0.9969	0.0359
0.01	0.5633	0.9964	0.0385
0.02	0.4233	0.9965	0.0382
0.05	0.2208	0.9958	0.0416
0.07	0.1725	0.9959	0.0413
0.1	0.1258	0.9954	0.0434
0.2	0.0675	0.9898	0.0649
0.5	0.0267	0.9708	0.1099
0.7	0.0183	0.9717	0.1083
5	0.0000	0.8773	0.2253

In these case studies, we changed the  $P_{DMI}$ -values and checked the predictive accuracy of the updated model, but in practice, the adequate  $P_{DMI}$ -value must be set beforehand in managing database with the DMI.



**Figure 9. The time plots of simulated and predicted  $y$  from 950 to 1050 when drift of  $y$  existed for the simulation data.<sup>18</sup>**

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



**Figure 10. Schematic representation of the distillation column.**

### Distillation column data

We analyzed data obtained from the operation of a distillation column at the Mizushima Works, Mitsubishi Chemical Corporation. Figure 10 shows schematic representation of the distillation column, and Table 3 shows the process variables. The  $y$ -variable is the concentration of the bottom product with the lowest boiling point, and the  $X$ -variables are the 19 variables given in Table 3. The measurement interval of  $y$  was 30 min and  $X$ -variables are measured every minute. The OSVR method was used as the regression method.

We collected data of the 42 days from January 1, 2003. Figure 11 shows the time plot of  $y$ . The data of the first week are training data and the data after that are test data.

The relationships between  $P_{DMI}$  and rate of update are shown in Figure 12. The relationships changed with the change of  $a$ . Although Figures 6 and 8 are very similar, Figure 12 is different from Figures 6 and 8 because the data sets are completely different and also the numbers of  $X$ -variables are different. Table 4 is the prediction results of the OSVR model when  $a$  equals 1. The  $r_p^2$ -values tended to increase and the  $RMSEP$ -values tended to decrease with the increase of the  $P_{DMI}$ -values from 0. However, when the

**Table 3. Process Variables Measured in the Distillation Column**

	Symbol	Process variables
$y$	A	Bottom product concentration
$x_1$	F1	Reflux flow
$x_2$	F2	Reboiler flow
$x_3$	F3	Feed 1 flow
$x_4$	F4	Feed 2 flow
$x_5$	F5	Bottom flow
$x_6$	F6	Top flow
$x_7$	L1	Liquid level
$x_8$	P1	Pressure 1
$x_9$	P2	Pressure 2
$x_{10}$	T1	Temperature 1
$x_{11}$	T2	Temperature 2
$x_{12}$	T3	Temperature 3
$x_{13}$	T4	Temperature 4
$x_{14}$	T5	Bottom temperature
$x_{15}$	T6	Feed 1 temperature
$x_{16}$	T7	Feed 2 temperature
$x_{17}$	T8	Top temperature
$x_{18}$	F4/F3=R	Reflux ratio
$x_{19}$	F1/F6=F	Feed flow ratio

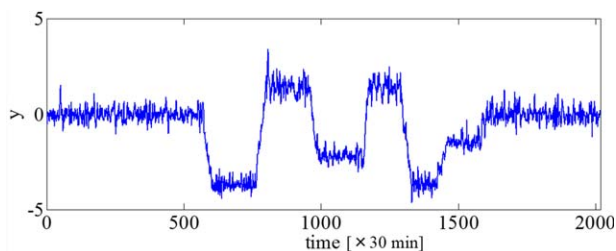


Figure 11. The time plot of  $y$  for the distillation column data.

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

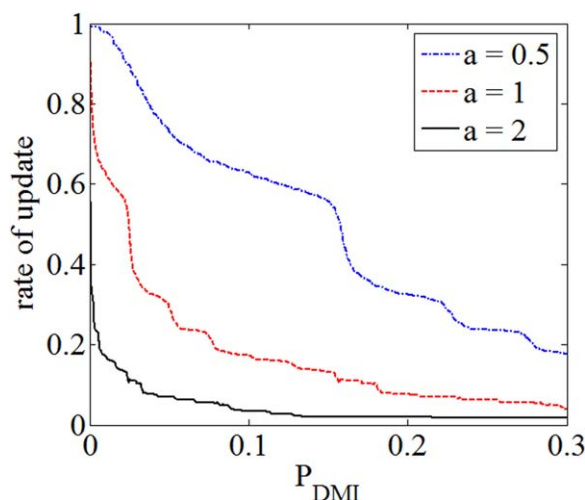


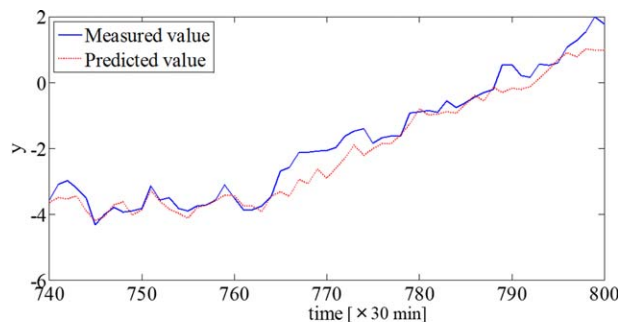
Figure 12. The relationship between  $P_{DMI}$  and rate of update for the distillation column data.

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

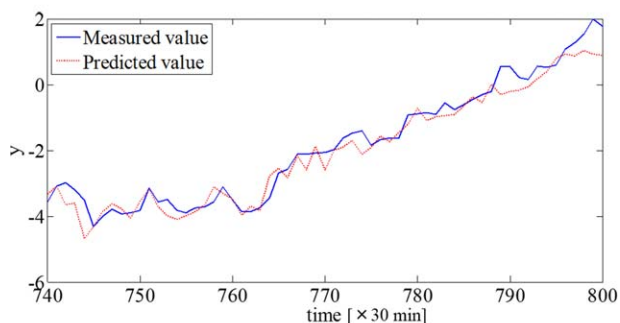
$P_{DMI}$ -values exceeded 0.005, the  $r_p^2$ -values decreased and the  $RMSEP$ -values increased. In the small  $P_{DMI}$ -value, the variety of the database and the prediction accuracy of the OSVR model increased by storing only informative new data in the database. But, when the  $P_{DMI}$ -value was too large, few new data were stored, and hence, the appropriate model update could not be performed. From the results of  $P_{DMI}$  of 0.005 where the  $r_p^2$ -value was maximum and the  $RMSEP$ -value was minimum, it was confirmed that the model updated only about 67% of new data could predict  $y$ -values with the highest accuracy.

Table 4. The Prediction Results for Each  $P_{DMI}$  for the Distillation Column Data when  $a$  Equals 1

$P_{DMI}$	Rate of update	$r_p^2$	$RMSEP$
0	1.0000	0.9572	0.3505
0.0001	0.9780	0.9573	0.3504
0.0005	0.9036	0.9574	0.3498
0.001	0.8405	0.9561	0.3551
0.005	0.6690	0.9580	0.3476
0.01	0.6274	0.9568	0.3524
0.05	0.2946	0.9538	0.3644
0.1	0.1720	0.9297	0.4493
0.5	0.0179	0.9401	0.4149



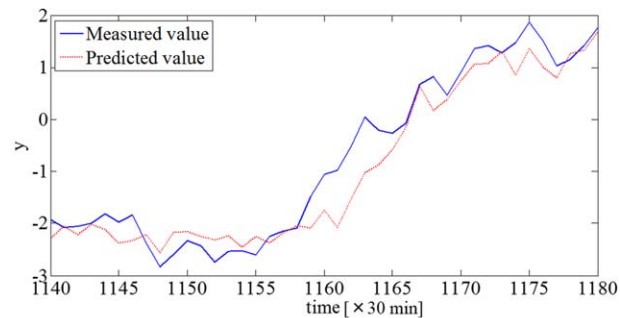
(a)  $P_{DMI} = 0$



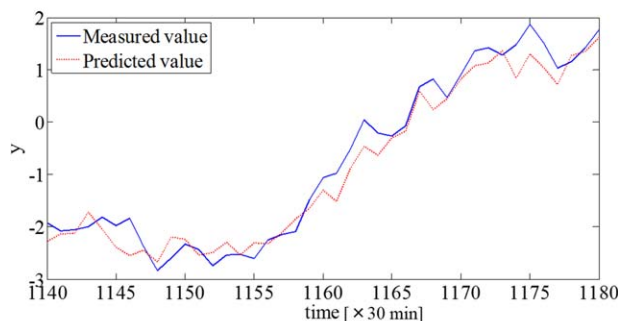
(b)  $P_{DMI} = 0.005$

Figure 13. The time plots of measured and predicted  $y$  from 740 to 800 for the distillation column data.

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



(a)  $P_{DMI} = 0$



(b)  $P_{DMI} = 0.005$

Figure 14. The time plots of measured and predicted  $y$  from 1140 to 1180 for the distillation column data.

[Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



Figures 13 and 14 show the time plots of measured and predicted  $y$ . When the OSVR model was updated each time, the prediction errors were large in the relatively large variations after the small variations (see around time 770 in Figure 13a and around time 1160 in Figure 14a). The model that specialized in the small variations could not adapt to the rapid time-varying variations occurring subsequently. Meanwhile, by managing the database with the DMI and the flow of Figure 3, the  $y$ -values were accurately predicted even in the rapid variations after the stable states (Figures 13b and 14b). It was confirmed that the database can be appropriately managed with the DMI and enables soft sensor models to predict  $y$ -values for wide data range with high accuracy.

## Conclusion

We proposed the DMI that is an index to monitor databases and the method that manages databases for adaptive soft sensors with the DMI. Decision whether new measured data are stored in the database or not is made, based on the DMI-values of the new data. By storing only informative data, the amount of information increases whereas the number of the data is controlled. Through the analysis of the simulation data and the real industrial data, we confirmed that the database can be appropriately managed with the DMI and accordingly the predictive accuracy of the soft sensor models increases and also the data range where the  $y$ -values can be accurately predicted increases.

Although only MW models were used as adaptive soft sensor models in this article, the proper management of databases is required for JIT models.<sup>31</sup> In JIT modeling, the data similar to the prediction data are selected or the data are weighted to be larger weights for the data that are more similar to the prediction data. JIT models cannot adapt to the latest state of a plant without storing new data in databases. However when the number of the data in databases is too large, it takes too much time to construct JIT models. JIT models can be appropriately operated by selecting the data that should be stored in database with the DMI (Figure 3).

In the case studies, we set the  $\alpha$ -value as one, changed the  $P_{\text{DMI}}$ -value and checked the performance of the soft sensor models. In fact, the appropriate  $\alpha$ -value and  $P_{\text{DMI}}$ -value must be set beforehand, the way of which is one of the future works. Additionally, the sensitivity of those parameters to the predictive ability of a soft sensor model should be investigated. In our case studies, the first training data are totally stored in the database, but the DMI-values among the training data can be calculated and then the data having little information can be eliminated from the training data. Here-with compact database will be able to be maintained, including much amount of information.

In this article, we assumed that there are no abnormal data and no outliers in new measurement data, however, the abnormal data and outliers deteriorate the predictive ability of adaptive models. In practice, the fault detection<sup>23</sup> is one of the essential tasks.

We believe that by managing the database and increasing the predictive accuracy of adaptive soft sensor models, chemical plants will be operated effectively and 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 Corporation, and the financial support of Mizuho Foundation for the Promotion of Sciences.

## Literature Cited

- Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Comput Chem Eng.* 2009;33:795–814.
- 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.
- Kadlec P, Gabrys B. Local learning-based adaptive soft sensor for catalyst activation prediction. *AIChE J.* 2010;57:1288–1301.
- Qin SJ. Recursive PLS algorithms for adaptive data modeling. *Comput Chem Eng.* 1998;22:503–514.
- Cheng C, Chiu MS. A new data-based methodology for nonlinear process modeling. *Chem Eng Sci.* 2004;59:2801–2810.
- Fujiwara K, Kano M, Hasebe S, Takinami A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE J.* 2009; 55:1754–1765.
- Schaal S, Atkeson CG, Vijayakumar S. Scalable techniques from onparametric statistics for real time robot learning. *Appl Intell.* 2002; 17:49–60.
- Kaneko H, Funatsu K. Maintenance-free soft sensor models with time difference of process variables. *Chemom Intell Lab Syst.* 2011; 107:312–317.
- 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.
- Kaneko H, Funatsu K. Development of soft sensor models based on time difference of process variables with accounting for nonlinear relationship. *Ind Eng Chem Res.* 2011;50:10643–10651.
- Kadlec P, Grbic R, Gabrys B. Review of adaptation mechanisms for data-driven soft sensors. *Comput Chem Eng.* 2011;35:1–24.
- Yu J. Multiway gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes. *Ind Eng Chem Res.* 2012;51:13227–13237.
- Yu J. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Comput Chem Eng.* 2012;41:134–144.
- Okada T, Kaneko H, Funatsu K. Development of a model selection method based on reliability of a soft sensor model. *Songklanakarin J Sci Technol.* 2012;34:217–222.
- Kaneko H, Funatsu K. Classification of the degradation of soft sensor models and discussion on adaptive models. *AIChE J.* 2013;52:1322–1334.
- Kalmukale AG, Chiu MS, Wang QG. Partitioned model-based IMC design using JITL modeling technique. *J Process Control.* 2007;17: 757–769.
- Kansha Y, Chiu MS. Adaptive generalized predictive control based on JITL technique. *J Process Control.* 2009;17:1067–1072.
- Kaneko H, Funatsu K. Development of a New Index to Monitor Database for Soft Sensors. *J Comput-Aided Chem.* 2013;14:11–22.
- Yu J. Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. *Chem Eng Sci.* 2012;82:22–30.
- Yu J, Chen K, Rashid MM. A Bayesian model averaging based multi-kernel Gaussian process regression framework for nonlinear state estimation and quality prediction of multiphase batch processes with transient dynamics and uncertainty. *Chem Eng Sci.* 2013;93:96–109.
- Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J.* 2008;54:1811–1829.
- Yu J. A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis. *J Process Control.* 2012;22:778–788.
- Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring. *Ind Eng Chem Res.* 2013;52:3543–3562.
- Rashid MM, Yu J. A new dissimilarity method integrating multidimensional mutual information and independent component analysis for non-Gaussian dynamic process monitoring. *Chemom Intell Lab Syst.* 2012;115:44–58.
- Rashid MM, Yu J. Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity method. *Ind Eng Chem Res.* 2012;51:10910–10920.



26. Guha R, Drie JHV. Structure-activity landscape index: Identifying and quantifying activity cliffs. *J Chem Inf Model*. 2008;48:646–658.
27. Stumpfe D, Bajorath J. Methods for SAR visualization, *RSC Adv*. 2012;2:369–378.
28. Li G, Aute V, Azarm S. An accumulative error based adaptive design of experiments for offline metamodeling. *Struct Multidiscip Optim*. 2010;40:137–155.
29. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer, 2006.
30. Ma J, Theliler J, Perkins S. Accurate on-line support vector regression. *Neural Comput*. 2003;15:2683–2703.
31. Kano M, Fujiwara K. Virtual sensing technology in process industries: Trends and challenges revealed by recent industrial applications. *J Chem Eng Jpn*. 2013;46:1–17.

## Appendix A

### OSVR

The SVR method applies a support vector machine (SVM) to regression analysis and can be used to construct nonlinear models by applying a kernel trick as well as the SVM. The OSVR method is a method efficiently updating a SVR model to meet the Karush–Kuhn Tucker (KKT) conditions that the SVR model must fulfill when training data are added or deleted.

The primal form of SVR can be shown to be the following optimization problem.

Minimize

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i |y_i - f(\mathbf{x}_i)|_\varepsilon \quad (\text{A1})$$

where  $y_i$  and  $\mathbf{x}_i$  are training data,  $f$  is a SVR model,  $\mathbf{w}$  is a weight vector,  $\varepsilon$  is a threshold, and  $C$  is a penalizing factor that controls the trade-off between model complexity and training errors. The second term of Eq. A1 is the  $\varepsilon$ -insensitive loss function and given as follows

$$|y_i - f(\mathbf{x}_i)|_\varepsilon = \max(0, |y_i - f(\mathbf{x}_i)| - \varepsilon) \quad (\text{A2})$$

Through the minimization of Eq. A1, we can construct a regression model that has good balance between generalization capabilities and the ability to adapt to the training data. A  $y$ -value predicted by inputting data  $\mathbf{x}$  is represented as follows

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (\text{A3})$$

where  $N$  is the number of training data,  $b$  is a constant term, and  $K$  is a kernel function. The kernel function in our application is a radial basis function

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\right) \quad (\text{A4})$$

where  $\gamma$  is a tuning parameter controlling the width of the kernel function. From Eqs. A1 and A2,  $\alpha_i$  and  $\alpha_i^*$  in Eq. A3 are obtained by minimizing the equation given as

$$\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N K_{ij} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \quad (\text{A5})$$

subject to

$$0 \leq \alpha_i, \alpha_i^* \leq C \quad i = 1, 2, \dots, N \quad (\text{A6})$$

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \quad (\text{A7})$$

$K_{ij}$  in Eq. A5 is represented as follows

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) \quad (\text{A8})$$

Now, we define  $\theta_i$  as follows

$$\theta_i = \alpha_i - \alpha_i^* \quad (\text{A9})$$

From Eqs. A3, A4, and A8, a predicted  $y$ -value of data  $\mathbf{x}_i$  is given as

$$f(\mathbf{x}_i) = \sum_{j=1}^N K_{ij} \theta_j + b \quad (\text{A10})$$

where  $\theta_i$  meets the following equation

$$\sum_{i=1}^N \theta_i = 0 \quad (\text{A11})$$

The error function  $h$  is defined as

$$h(\mathbf{x}_i) \equiv f(\mathbf{x}_i) - y_i$$

$$= \sum_{j=1}^N K_{ij} \theta_j + b - y_i \quad (\text{A12})$$

Then the KKT conditions can be summarized as follows

$$h(\mathbf{x}_i) > \varepsilon, \quad \theta_i = -C \quad (\text{A13})$$

$$h(\mathbf{x}_i) = \varepsilon, \quad -C < \theta_i < 0 \quad (\text{A14})$$

$$-\varepsilon < h(\mathbf{x}_i) < \varepsilon, \quad \theta_i = 0 \quad (\text{A15})$$

$$h(\mathbf{x}_i) = -\varepsilon, \quad 0 < \theta_i < C \quad (\text{A16})$$

$$h(\mathbf{x}_i) < -\varepsilon, \quad \theta_i = C \quad (\text{A17})$$

Each training data must meet one of Eqs. A13–A17. All training data can be divided into the following sets: error support vectors, E, which meet Eq. A13 or A17, margin support vectors, S, which meet Eq. A14 or A16, and remaining vectors, R, which meet Eq. A15.

When new data  $\mathbf{x}_c$ ,  $y_c$  are added, there is no need to update the SVR model  $\theta_i$ ,  $b$  if  $\mathbf{x}_c$  belongs to R. On the other hands, if  $\mathbf{x}_c$  belongs to E or S, the initial value of  $\theta_c$  that is  $\theta_i$  corresponding to  $\mathbf{x}_c$  is set as 0, and  $\theta_c$ ,  $\theta_i$ , and  $b$  are gradually changed to meet the KKT conditions. There are possibilities that each training data moves to another region due to the changes. But, assuming no such movements, variations of  $h(\mathbf{x}_i)$ ,  $\theta_c$ ,  $\theta_i$ , and  $b$ ,  $\Delta h(\mathbf{x}_i)$ ,  $\Delta \theta_c$ ,  $\Delta \theta_i$ , and  $\Delta b$ , respectively, can be represented from Eqs. A11 and A12 as follows

$$\Delta h(\mathbf{x}_i) = K_{ic} \Delta \theta_c + \sum_{j=1}^N K_{ij} \Delta \theta_j + \Delta b \quad (\text{A18})$$

$$\Delta \theta_c + \sum_{j=1}^N \Delta \theta_j = 0 \quad (\text{A19})$$

The  $\theta_i$ -values of the training data belonging to E and R did not change because of Eqs. A13, A15, and A17, and thus, Eq. A18 can be transformed as

$$\Delta h(\mathbf{x}_i) = K_{ic} \Delta \theta_c + \sum_{j \in S} K_{ij} \Delta \theta_j + \Delta b \quad (\text{A20})$$

The  $h(\mathbf{x}_i)$ -values of the training data belonging to S are settled due to Eqs. A14 and A16. Thus Eqs. A19 and A20 can change to

$$\sum_{j \in S} K_{ij} \Delta \theta_j + \Delta b = -K_{ic} \Delta \theta_c \quad \forall i \in S \quad (\text{A21})$$

$$\sum_{j \in S} \Delta \theta_j = -\Delta \theta_c \quad (\text{A22})$$

Then,  $\Delta \theta_c$ ,  $\Delta \theta_i$  and  $\Delta b$  can be represented as

$$\Delta b = \delta \Delta \theta_c \quad (\text{A23})$$

$$\Delta \theta_i = \delta_i \Delta \theta_c \quad \forall i \in S \quad (\text{A24})$$

where

$$\begin{bmatrix} \delta \\ \delta_{S_1} \\ \vdots \\ \delta_{S_M} \end{bmatrix} = - \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K_{S_1 S_1} & \cdots & K_{S_1 S_M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K_{S_M S_1} & \cdots & K_{S_M S_M} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ K_{S_1 c} \\ \vdots \\ K_{S_M c} \end{bmatrix} \quad (\text{A25})$$

$$\delta_i = 0 \quad \forall i \notin S \quad (\text{A26})$$

Here  $M$  is the number of the training data that belong to  $S$ . From Eqs. A20, A23, and A24,  $h(\mathbf{x}_i)$  for the training data belonging to  $E$  and  $R$  can be transformed as

$$\begin{aligned} \Delta h(\mathbf{x}_i) &= K_{ic} \Delta \theta_c + \sum_{j \in S} K_{ij} \Delta \theta_j + \Delta b \\ &= K_{ic} \Delta \theta_c + \sum_{j \in S} K_{ij} \delta_j \Delta \theta_c + \delta \Delta \theta_c \\ &= \left( K_{ic} + \sum_{j \in S} K_{ij} \delta_j + \delta \right) \Delta \theta_c \\ &= \gamma \Delta \theta_c \end{aligned} \quad (\text{A27})$$

where

$$\gamma = K_{ic} + \sum_{j=1}^N K_{ij} \delta_j + \delta \quad (\text{A28})$$

From Eqs. A24 and A27,  $\Delta \theta_c$  for the movement of each training data is represented as

$$\Delta \theta_c = \delta_i^{-1} \Delta \theta_i \quad \forall i \in S \quad (\text{A29})$$

$$\Delta \theta_c = \gamma^{-1} \Delta h(\mathbf{x}_i) \quad \forall i \notin S \quad (\text{A30})$$

The absolute  $\Delta \theta_i$ -values for each training data to move from the current region to another region, i.e. from  $E$  to  $S$ , from  $S$  to  $E$  or  $R$  and from  $R$  to  $S$ , are calculated by using Eqs. A29 and A30. The minimum value of the absolute  $\Delta \theta_i$ -values calculated with all training data is selected, and the data having the minimum  $\Delta \theta_i$ -value is actually moved to a new region. The calculation of the absolute  $\Delta \theta_i$ -values and the movement of the data having the minimum value of the absolute  $\Delta \theta_i$ -values are repeated until each of all the training data meets the KKT conditions, namely, one of Eqs. A13–A17. When one data are deleted from training data, the same iterative calculation is performed until all the data meet the KKT conditions.

*Manuscript received May 20, 2013, and revision received Aug. 18, 2013.*