Adaptive Database Management Based on the Database Monitoring Index for Long-term Use of Adaptive Soft Sensors

Hiromasa Kaneko and Kimito Funatsu

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Corresponding author: Kimito Funatsu

Tel: +81-3-5841-7751

Fax: +81-3-5841-7771

E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

Abstract

Soft sensors are an essential tool for controlling chemical and industrial plants. To adapt to new process states, these soft sensors require adaptation mechanisms. As the performance of adaptive soft sensors depends on the quality of its database and the size of which will have some upper limit, a database monitoring index (DMI) has been developed. Additional information is only added to the database if it has a sufficiently high DMI value. In this study, we propose a DMI optimization method and adaptive database management scheme. DMI hyperparameters are set automatically using an initial database, and the main database adapts to new process states by storing data with large prediction errors. Case studies using simulated and real industrial datasets confirm that the proposed scheme enables soft sensors to operate with high predictive accuracy over the long term.

Keywords: Process control; Soft sensor; Model degradation; Adaptive model; Database management.

1 Introduction

Soft sensors [1,2] are commonly used for the rapid and efficient control of industrial and chemical plants. A soft sensor, which is a numerical model, can estimate the value of a difficult-to-measure variable (the y-variable) using values of easy-to-measure variables (X-variables). There are three approaches to constructing a soft sensor model: first-principles approaches, data-based techniques, and their hybrid. Because of the volume of operational data, data-based or hybrid soft sensors are predominant. Soft sensors are applied to many industrial processes [3-5].

One of the crucial problems with soft sensors is model degradation [6], which leads to a deterioration in performance. This often occurs after a change of process states in a plant, such as catalyst deterioration, changes in raw materials, fouling on pipes, and equipment degradation. Model degradation prevents the practical use of soft sensors by engineers and operators.

Adaptive soft sensors [7] can reduce model degradation by using new measurement data. The three main approaches to constructing adaptive soft sensors are the moving window (MW), just-in-time (JIT), and time difference (TD) approaches. An MW soft sensor is constructed using the most recent dataset from the plant. The number of data used in model construction corresponds to the window size. A linear model can be updated recursively [8]. To deal with nonlinearity between X and y, a nonlinear model can be used [9] and multiple local linear models are combined [10,11]. A final predicted y-value is obtained by weighting y-values predicted with nonlinear MW models, according to process states, to adapt to time-varying and multiple states [12]. Variable selection improves the prediction accuracy of an MW model [13]. MW models can be applied to fault detection [14] and process control [15]. A JIT soft sensor estimates the value of the y-variable by selecting only data close to that of the X-variables and constructing a new model [16] or by weighting data close to that of the X-variables, which is locally-weighted regression [17] and is applied to pharmaceutical process [18]. A similarity is based on not only distance but also correlation [19] and Gaussian mixture model [20]. A local model can be constructed with a nonlinear regression method such as kernel partial least squares regression [21] and kernel principal component regression [22]. JIT modelling can be applied to process monitoring [23]. A TD soft sensor uses the temporal differences between y-variable values and those of the corresponding X-variables, and adds a known y-value to the TD of the y-variable to form a prediction [24]. Ensemble learning was applied to TD models and the improvement of prediction accuracy achieved [25]. As no all-round

adaptive soft sensors exist, the selective use of adaptive soft sensors based on process states is essential [26].

Generally soft sensor performance depends on a database; that is, an uninformative dataset cannot provide predictive soft sensors. Especially for MW and JIT approaches in which a model is reconstructed online, we have to consider database management (DBM) since a database is updated online [27]. For example, an MW model cannot adapt to rapid changes in process states when its database includes only small variations in data, and a predictive JIT model cannot be constructed if there are no similar data to the input datum of the X-variables. Not all data can be stored in a database, as there are limits to the storage capacity, and old data often have a negative effect on the performance of soft sensors [6]. Therefore, DBM is important for the maintenance of predictive soft sensors.

To reduce the size of a database, JIT models select new measurement data to be stored based on the prediction errors of the y-variable [28] and are applied to adaptive predictive control [29]. Jin et al. proposed that new data should be replaced with the most similar data in the database [30]. However, the information in the database and that contained in new measurement data was not considered in the data selection. Kaneko et al. proposed a database monitoring index (DMI) for DBM that examines the amount of information contained in a new measurement datum [31]. Only data with high DMI values are stored in the database, meaning that the total amount of information can increase while the number of data is maintained. This DBM with DMI can be applied to soft sensor analyses as well as process monitoring [32], where models are updated or reconstructed with a database that includes new measurement data.

To operate DMI, we must set the threshold that judges whether a new measurement datum is stored in the database. This threshold plays an important role in maintaining a high-quality database. An automatic determination method for the threshold has been proposed [33]. However, some DMI hyperparameters must also be set, i.e., the weight of the y-variable. In addition, it is difficult to adapt DMI-based DBM to new process states. Therefore, we modify the offline analysis of DMI, introducing an adaptive mechanism to DBM by monitoring the predictive performance of a current soft sensor and proposing a DBM method for long-term database use. Of course, both MW- and JIT-based soft sensors can be used with the proposed DBM, as we will demonstrate through case studies.

To verify the effectiveness of the proposed DBM system, we analyze numerical simulation data and real industrial data measured in a distillation column. Then, adaptive soft sensors are used to predict a transition concentration in a formalin

production process.

2 Method

2.1 Database monitoring index (DMI) [27,31]

The DMI for managing databases is defined between two data $(\mathbf{x}^{(i)}, y^{(i)})$ and $(\mathbf{x}^{(j)}, y^{(j)})$ as follows:

$$\text{DMI} = \frac{|y^{(i)} - y^{(j)}|^a}{\text{sim}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})}, \quad (1)$$

where $\text{sim}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is the similarity between $\mathbf{x}^{(i)} \in R^{1 \times m}$ (m is the number of X-variables), which cannot be zero, and $\mathbf{x}^{(j)} \in R^{1 \times m}$, and a is a constant. Various similarity indexes can be used, such as the Gaussian kernel:

$$\text{sim}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp\left(-\gamma \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2\right), \quad (2)$$

where γ is a tuning parameter controlling the width of the kernel function. In this study, we use the inverse Euclidian distance:

$$\text{sim}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \frac{1}{\sqrt{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}}. \quad (3)$$

Assuming that there are no completely the same data, the denominator of eq. (3) is not zero. Using eqs. (1) and (3), the DMI used in this study is calculated as follows:

$$\text{DMI} = |y^{(i)} - y^{(j)}|^a \sqrt{\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}. \quad (4)$$

The DMI is large when two X and y data are dissimilar, and vice-versa.

If the minimum DMI value between a new datum and data in the database exceeds the threshold P_{DMI} , the new datum contains sufficient new information and is stored in the database.

2.2 Determination of the threshold P_{DMI} and optimization of a

To manage the database, we first optimize P_{DMI} and a . Assuming there exist some data that are similar to other data and are not indispensable in training data, we can discuss P_{DMI} while repeating the deletion of one of such similar data and the check of the predictive ability of the regression model constructed with the remaining training data.

The process for optimizing P_{DMI} and a is illustrated in Figure 1. After selecting a candidate value for a , a regression model is constructed with the current training data, and the prediction accuracy of the model is checked using the root-mean-square error of cross-validation and external validation ($\text{RMSE}_{\text{CVEX}}$) [33]:

$$\text{RMSE}_{\text{CVEX}} = \sqrt{\frac{\sum_{i=1}^n (y^{(i)} - y_{\text{pred}}^{(i)})^2}{n}}, \quad (5)$$

where n is the number of training data divided into the remaining training data and the deleted training data that are used as external data; $y^{(i)}$ is the y -value of the i^{th} data point; and $y_{\text{pred}}^{(i)}$ is the y -value of the i^{th} data point predicted with cross-validation or external validation. For the remaining training data, y -values are predicted with cross-validation. For the deleted training data, y -values are predicted with the regression model constructed with the remaining training data, which is external validation. $\text{RMSE}_{\text{CVEX}}$ is the RMSE calculated between actual y -values and predicted y -values for both the remaining training data and the deleted training data. At the first iteration of Figure 1, $\text{RMSE}_{\text{CVEX}}$ is the same as the RMSE of only cross-validation since there are no deleted training data. The DMI values are then calculated for all pairs of training data. The k pairs of data with the lowest DMI values are selected, and one of each pair is deleted from the training data; i.e., k data are removed from the training data. If enough training data to construct a regression model remain, we repeat the process of model construction, $\text{RMSE}_{\text{CVEX}}$ calculation, and the deletion of data, while saving the DMI values of the removed data. If not, we change the value of a and check the relationship between $\text{RMSE}_{\text{CVEX}}$ and the DMI values of the removed data for each loop. If there is no other candidate value for a , we examine the relationship between $\text{RMSE}_{\text{CVEX}}$, the DMI values of the k deleted data, and a . P_{DMI} is set as the maximum DMI value for as long as $\text{RMSE}_{\text{CVEX}}$ does not exceed the required RMSE. The corresponding value of a

is then assigned, and the database can be managed so as not to exceed the required RMSE.

Every soft sensor has required predictive accuracy to perform some objective. The required RMSE depends on objective of soft sensors. In this study, we assume that required predictive accuracy can be represented as required RMSE in actual situations. However, RMSE for $\text{RMSE}_{\text{CVEX}}$ and required RMSE in this paper can be altered to other indexes such as coefficient of determination r^2 (r^2_{CVEX} and required r^2) as necessary.

2.3 Database management (DBM)

The process of DBM is shown in Figure 2. When the prediction error for a new datum ($\mathbf{x}^{(i)}, y^{(i)}$) is large, the current database or soft sensor model does not adapt to the current process state, even for small DMI values. Thus, new data with large prediction errors may be added to the database. The error threshold is three times the standard deviation of y-errors, with cross-validation in reference to the three-sigma limit. Otherwise, the DMI values between a new datum and all data included in the database are calculated. If the minimum DMI value exceeds P_{DMI} , the new datum is stored in the database. If not, the new datum is not stored, or replaces the datum in the database with the minimum DMI value. Then, if the size of the database exceeds some upper limit, the oldest data are deleted. Databases can be managed flexibly and adaptively to handle new process states by considering both prediction errors and DMI values.

The oldest datum is discarded from the database even when the datum is very informative, which will happen because there is a limit to the storage capacity for the database and other data in the database are also very informative in this case. The DBM can delay the deletion of informative data, compared with no DBM.

There are two types of over-adaptation in adaptive soft sensors; over-adaptation of a regression model and over-adaptation of a database. Over-adaptation of a regression model is outside the scope of this paper and selection of values of hyperparameters in regression models is important to prevent the over-adaptation. On the other hand, when regression models are updated only with steady state data, i.e. uninformative data, for a long time, the models specialize in predictions over a narrow data range, which is an example of over-adaptation of a database. By using the proposed DBM system, uninformative data are not stored and only informative data are stored in a database. We therefore can prevent over-adaptation of a database.

3 Results and Discussion

To verify the effectiveness of the proposed method, we analyzed numerical simulation data and industrial distillation column data. The simulation dataset exhibits a nonlinear relationship between its X-variables. The proposed DBM was then applied to a formalin production plant.

The two types of adaptive soft sensor that most require DBM are MW and JIT. We used a moving window support vector regression (SVR) model [10] as an MW approach (for SVR, see the reference [34]), and a locally-weighted partial least squares (LWPLS) model [18,19] as a JIT approach. The hyperparameters in each model were optimized with 5-fold cross-validation. The SVR hyperparameters could be determined very quickly [36].

The set of candidate values for a was $\{-10, -9, -8, \dots, 13, 14, 15\}$, k was set to 5, and the required RMSE was set to 1.05 times the value of $\text{RMSE}_{\text{CVEX}}$ calculated from the first batch of training data. Five-fold cross-validation was used to calculate $\text{RMSE}_{\text{CVEX}}$.

3.1 Numerical simulation data where the relationship between X and y is nonlinear

We considered two X-variables related according to:

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

The basic form of eq. (6) has been used as a test problem [37], and we have added the term $0.1x_1$. Data for the X-variables were generated using a random walk procedure. Random numbers from the normal distribution with a mean of 0 and standard deviation of 0.01 were added to the y-variable. The time plots of x_1 , x_2 , and y for the simulation data are shown in Figure 3. Some of the variation in the simulation data is solely due to noise. The first 500 data were used for training, and the next 1,000 data formed the test dataset. The y-values were assumed to be obtained after 20 units of time.

First, P_{DMI} and a were automatically optimized in the SVR and LWPLS models using the proposed method. The optimized P_{DMI} and a , and the number of remaining data in the SVR and LWPLS models, are given in Table 1. The very small value of P_{DMI} for LWPLS comes from the large value of a . Using the proposed method, the number of training data was reduced efficiently, with the remaining data providing the predictive performance of the SVR and LWPLS models. The number of remaining data for SVR is less than that for LWPLS, indicating that SVR can reduce the number of training data

more effectively than LWPLS.

The upper limit of the database size was changed to 50, 100, 150, ..., 900, 950, and 1000. Figure 4 shows the values of $RMSE_p$, which is the RMSE for test data, for SVR and LWPLS with and without DBM. For both SVR and LWPLS, the value of $RMSE_p$ in the presence of DBM decreased, which means that the predictive accuracy increased, for almost all upper limits of the database. Overall, the $RMSE_p$ values for SVR were lower than those for LWPLS, because the predictions made using SVR were more stable and less affected by the influence of noise in this case study.

Figure 5 shows time plots of predicted and actual y -values for maximum database size of 100. When the database was continually updated and not being managed, the prediction errors from both SVR and LWPLS became large following the abrupt process change after 1000 iterations. The prediction errors reduced after several data including large variations had been stored (see Figure 5(a) and (c)). This is because the SVR and LWPLS models were constructed using data including only small variations, and were specialized in that state. Thus, the SVR and LWPLS models could not adapt to an abrupt change. In contrast, when the database was managed according to the method in Figure 2, and most data pertaining to small variations were not included in the database, the SVR and LWPLS models were able to adapt to the abrupt variation after 1000 iterations. The database update rates were 0.314 for SVR and 0.533 for LWPLS (see Figure 5(b) and (d)). This confirms that appropriate DBM with the proposed method enables a wide range of y -values to be predicted with a high degree of accuracy.

3.2 Real industrial data in a distillation column

We applied the proposed method to data obtained from the operation of a distillation column at the Mizushima Works, Mitsubishi Chemical Corporation. In this operation, the y -variable is the concentration of the bottom product, and the X -variables include 19 variables such as flow rate, temperature, and pressure. The measurement interval of the y -variable was 30 minutes, and the X -variables were measured every minute. Details of the dataset can be found in reference [24]. We collected data from the 42 days from 1 January 2003. Data from the first week were used as training data, with the remainder forming the test dataset. Thus, we had 336 training data and 1680 test data.

Table 2 presents the optimized values of P_{DMI} and a , as well as the number of remaining data for the SVR and LWPLS models. The small P_{DMI} value for SVR is a result of the large a value. The proposed method enabled the number of training data to

be reduced while maintaining the predictive performance of the SVR and LWPLS models. Specifically, the predictive SVR model could be constructed with only 16.7% of the training data, reflecting the fact that stable SVR models can be obtained with small datasets if appropriate data are selected.

Figure 6 shows the RMSE_p values of the test data for SVR and LWPLS with and without DBM for maximum database sizes of 50, 100, ..., 450, and 500. DBM clearly reduces the RMSE_p values for both SVR and LWPLS, especially when the upper limit of the database size is small. This shows that the proposed DBM increases the predictive accuracy of adaptive soft sensors.

Time plots of the predicted and actual y-values for SVR and LWPLS are shown in Figure 7 for a maximum database size of 100. The database update rates were 0.288 for SVR and 0.479 for LWPLS. Without DBM, the SVR and LWPLS models could not adapt to rapid changes from points 950 to 1000 and 1150 to 1200, and the predicted y-values took time to adjust back to the actual y-values. Using the proposed DBM, the y-values predicted by the SVR and LWPLS models were in good agreement with the actual y-values during both the stable period and the rapid changes from 950 to 1000 and 1150 to 1200. Thus, the proposed method achieved appropriate DBM for adaptive soft sensors with good predictive ability.

3.3 Application to a formalin production process

We applied the proposed method to a formalin production process at Mitsui Chemicals, Inc. The plant consists of a reactor, absorption tower, ion exchange tower, and demethanol tower. Methanol is a raw material used to produce formaldehyde through the following dehydrogenation reaction:

Formic acid can be produced by the further oxidation of formaldehyde. To monitor a reaction situation, the concentrations of methanol, formaldehyde, and formic acid are measured at the bottom of the absorption tower. Soft sensors monitor these concentrations in real time. In this paper, the y-variable is the concentration of formic acid, which is measured every few hours with a given delay, and the X-variables are 41 process variables such as temperature, pressure, and flow rate in each tower, which are measured online every minute. The y-variable is scaled from 0–1.

The 295 training data include the same variation as that in the test data

measured from April to June 2014. The number of test data is 133230, but only 210 data have y -measurements. The optimization results of P_{DMI} and a , and the number of remaining data, are given in Table 3. The small value of P_{DMI} is due to the large value of a . SVR produced a greater reduction in data volume than LWPLS, as in the previous case studies. This is because SVR's theoretical background of stable regression models requires fewer training data. An SVR model with high predictive accuracy can be constructed with fewer than half of the first batch of training data.

Figure 8 shows time plots of the predicted and actual y -values with and without DBM for an upper database size limit of 100. Without DBM, the SVR and LWPLS models could not adapt to the rapid increase in the y -variable at around 1350 h, and could not predict the formic acid concentration at all (see Figure 8(a) and (c)). However, the SVR and LWPLS models with the proposed DBM could predict the change in formic acid concentration beforehand (see Figure 8(b) and (d)), which is actually before the actual y -values adapt (due to the delay in measurement). If the threshold of the formic acid concentration is 0.3, for example, we can take action when a concentration value predicted by the SVR and LWPLS models with the proposed DBM exceeds this threshold. As the predicted values are obtained in real time, and much earlier than the measurements, we can achieve efficient process control using the proposed soft sensors.

4 Conclusions

We have proposed an overall DBM system based on DMI for adaptive (both MW- and JIT-based) soft sensors with long-term high predictive ability. Given an initial database, the hyperparameter a in DMI and the threshold P_{DMI} are automatically optimized. During the prediction phase, only informative data and data for new process states are stored in the database. The selection of which data should be stored is based on the DMI values and prediction errors of the new data. Case studies using simulated and real industrial data confirmed that highly predictive and adaptive soft sensors could be maintained over long periods with only a small number of data.

Our approach assumed that there were no outliers or abnormal data in either the database or the new data. Outliers and abnormal data have an adverse effect on the predictive ability of soft sensors, and should therefore be preprocessed (i.e., deleted) [38]. We believe that the use of our proposed method and appropriate DBM will improve the predictive accuracy of adaptive soft sensor models, allowing chemical plants to be operated more effectively and stably.

Acknowledgements

The authors acknowledge the support of the Japan Society for the Promotion of Science (JSPS) through a Grant-in-Aid for Young Scientists (B) (grant no. 24760629) and the Core Research for Evolutionary Science and Technology (CREST) project “Development of a knowledge-generating platform driven by big data in drug discovery through production processes” of the Japan Science and Technology Agency (JST). The authors also acknowledge the support of Mizushima Works, Mitsubishi Chemical Corporation, and Mitsui Chemicals Inc.

References

- [1] M. Kano, Y. Nakagawa, Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry, *Comput. Chem. Eng.* 32 (2008) 12–24.
- [2] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, *Comput. Chem. Eng.* 33 (2009) 795–814.
- [3] M. Kano, M. Ogawa, The state of the art in chemical process control in Japan: Good practice and questionnaire survey, *J. Process Contr.* 20 (2010) 969–982.
- [4] M. Kano, K. Fujiwara, Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications, *J. Chem. Eng. Jpn.* 46 (2013) 1–17.
- [5] S. Kim, M. Kano, S. Hasebe, A. Takinami, T. Seki, Long-term industrial applications of inferential control based on just-in-time soft-sensors, economical impact and challenges, *Ind. Eng. Chem. Res.* 52 (2013) 12346–12356.
- [6] H. Kaneko, K. Funatsu, Classification of the degradation of soft sensor models and discussion on adaptive models, *AIChE J.* 59 (2013) 2339–2347.
- [7] P. Kadlec, R. Grbic, B. Gabrys, Review of adaptation mechanisms for data-driven soft sensors, *Comput. Chem. Eng.* 35 (2011) 1–24.
- [8] S.J. Qin, Recursive PLS algorithms for adaptive data modeling, *Comput. Chem. Eng.* 22 (1998) 503–514.
- [9] H. Kaneko, K. Funatsu, Adaptive soft sensor model using online support vector regression with the time variable and discussion on appropriate hyperparameters and window size, *Comput Chem Eng.* 58 (2013) 288–297.
- [10] P. Kadlec, B. Gabrys, Local learning-based adaptive soft sensor for catalyst activation prediction. *AIChE J.* 57 (2010) 1288–1301.
- [11] W.M. Shao, X.M. Tian, Adaptive soft sensor for quality prediction of chemical

processes based on selective ensemble of local partial least squares models, *Chem. Eng. Res. Des.* 95 (2015) 113–132.

- [12] H. Kaneko, K. Funatsu, Adaptive soft sensor based on online support vector regression and bayesian ensemble learning for various states in chemical plants, *Chemom. Intell. Lab. Syst.* 137 (2014) 57–66.
- [13] Y.Q. Liu, Y.P. Pan, D.P. Huang, Development of a novel adaptive soft-sensor using variational Bayesian PLS with accounting for online identification of key variables, *Ind. Eng. Chem. Res.* 54 (2015) 338–350.
- [14] Z.Q. Ge, Z.H. Song, M. Kano, External analysis-based regression model for robust soft sensing of multimode chemical processes, *AIChE J.* 60 (2014) 136–147.
- [15] P. Wang, C.H. Yang, X.M. Tian, D.X. Huang, Adaptive nonlinear model predictive control using an on-line support vector regression updating strategy, *Chinese J. Chem. Eng.* 22 (2014) 774–781.
- [16] C. Cheng, M.S. Chiu, A new data-based methodology for nonlinear process Modeling, *Chem. Eng. Sci.* 59 (2004) 2801–2810.
- [17] S. Schaal, C.G. Atkeson, S. Vijayakumar, Scalable techniques from onparametric statistics for real time robot learning, *Appl. Intell.* 17 (2002) 49–60.
- [18] S. Kim, M. Kano, H. Nakagawa, S. Hasebe, Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection, *Int. J. Pharm.* 421 (2011) 269–274.
- [19] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami, Soft-sensor development using correlation-based just-in-time modeling, *AIChE J.* 55 (2009) 1754–1765.
- [20] M. Fan, Z.Q. Ge, Z.H. Song, Adaptive Gaussian mixture model-based relevant sample selection for JITL soft sensor development, 53 (2014) 19979–19986.
- [21] H.P. Jin, X.G. Chen, J.W. Yang, L. Wu, Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes, *Comput. Chem. Eng.* 71 (2014) 77–93.
- [22] X.F. Yuan, Z.Q. Ge, Z.H. Song, Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes, *Ind. Eng. Chem. Res.* 53 (2014) 13736–13749.
- [23] Q.C. Jiang, X.F. Yan, Just-in-time reorganized PCA integrated with SVDD for chemical process monitoring, *AIChE J.* 60 (2014) 949–965.
- [24] H. Kaneko, K. Funatsu, Maintenance-free soft sensor models with time difference of process variables, *Chemom. Intell. Lab. Syst.* 107 (2011) 312–317.
- [25] H. Kaneko, K. Funatsu, 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.* 109 (2011) 197–206.
- [26] H. Kaneko, T. Okada, K. Funatsu, The selective use of adaptive soft sensors based on process state, *Ind. Eng. Chem. Res.* 53 (2014) 15962–15968.
- [27] H. Kaneko, K. Funatsu, Development of a new index to monitor database for soft sensors. *J. Comput. Aided Chem.* 14 (2013) 11–22. (in Japanese)
- [28] A.G. Kalmukale, M.S. Chiu, Q.G. Wang, Partitioned model-based IMC design using JITL modeling technique, *J. Process Contr.* 17 (2007) 757–769.
- [29] Y. Kansha, M.S. Chiu, Adaptive generalized predictive control based on JITL technique, *J. Process Contr.* 17 (2009) 1067–1072.
- [30] H.P. Jin, X.G. Chen, J.W. Yang, L. Wu, Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes, *AIChE J.* 71 (2014) 77–93.
- [31] H. Kaneko, K. Funatsu, Database monitoring index for adaptive soft sensors and the application to industrial process, *AIChE J.* 60 (2014) 160–169.
- [32] Macgregor JF, Jaeckle C, Kiparissides C, Koutoudi. Process monitoring and diagnosis by multiblock PLS methods. *AIChE J.* 1994;5:826–838.
- [33] H. Kaneko, K. Funatsu, Automatic database monitoring for process control systems, *Modern Advances in Applied Intelligence Lecture Notes in Computer Science*, 8481 (2014) 410–419.
- [34] C.M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006.
- [35] J. Ma, J. Theliler, S. Perkins, Accurate on-line support vector regression, *Neural Comput.* 15 (2003) 2683–2703
- [36] H. Kaneko, K. Funatsu, Fast optimization of hyperparameters for support vector regression models with highly predictive ability, *Chemom. Intell. Lab. Syst.* 142 (2015) 64–69.
- [37] G. Li, V. Aute, S. Azarm, An accumulative error based adaptive design of experiments for offline metamodeling, *Struct. Multidiscip. O.* 40 (2010) 137–155.
- [38] M.S. Escobar, H. Kaneko, K. Funatsu, Combined generative topographic mapping and graph theory unsupervised approach for non-linear fault identification, *AIChE J.* 61 (2015) 1559–1571.

List of Figures captions

Figure 1. Procedure for optimizing P_{DMI} and α .

Figure 2. Process of DBM.

Figure 3. Time plots of x_1 , x_2 , and y for the simulation data.

Figure 4. RMSE_p values of test data for the simulation data, for various upper limits of the database size.

Figure 5. Time plots of predicted and actual y -values for the simulation data for maximum database size of 100. The update rates of the database are 0.314 for SVR and 0.533 for LWPLS.

Figure 6. RMSE_p values of test data for the industrial distillation column data, for various upper limits of the database size.

Figure 7. Time plots of predicted and actual y -values for the industrial distillation column data for a maximum database size of 100. The update rates of the database are 0.288 for SVR and 0.479 for LWPLS.

Figure 8. Time plots of predicted and actual y -values for the industrial formalin production process data for a maximum database size of 100. The update rates of the database are 0.0476 for SVR and 0.286 for LWPLS.

Table 1. Optimized P_{DMI} and a , and the number of remaining data (out of 500), for the numerical simulation data.

	P_{DMI}	a	# of remaining data
SVR	2.22×10^{-3}	1	160
LWPLS	1.22×10^{-15}	6	270

Table 2. Optimized P_{DMI} and a , and the number of remaining data (out of 336), for the industrial distillation column data.

	P_{DMI}	a	# of remaining data
SVR	2.03×10^{-10}	7	56
LWPLS	1.21	0	121

Table 3. Optimized P_{DMI} and a , and the number of remaining data (out of 295), for the formalin production process data.

	P_{DMI}	a	# of remaining data
SVR	2.85×10^{-11}	10	145
LWPLS	1.44	0	200

Figure 1. Procedure for optimizing P_{DMI} and α .

Figure 2. Process of DBM.

Figure 3. Time plots of x_1 , x_2 , and y for the simulation data.

Figure 4. $RMSE_P$ values of test data for the simulation data, for various upper limits of the database size.

Figure 5. Time plots of predicted and actual y -values for the simulation data for maximum database size of 100. The update rates of the database are 0.314 for SVR and 0.533 for LWPLS.

Figure 6. $RMSE_P$ values of test data for the industrial distillation column data, for various upper limits of the database size.

(a) SVR without DBM

(b) SVR with DBM

(c) LWPLS without DBM

(d) LWPLS with DBM

Figure 7. Time plots of predicted and actual y -values for the industrial distillation column data for a maximum database size of 100. The update rates of the database are 0.288 for SVR and 0.479 for LWPLS.

(a) SVR without DBM

(b) SVR with DBM

(c) LWPLS without DBM

(d) LWPLS with DBM

Figure 8. Time plots of predicted and actual y-values for the industrial distillation column data for a maximum database size of 100. The update rates of the database are 0.0476 for SVR and 0.286 for LWPLS.