Application of Online Support Vector Regression for Soft Sensors

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.14299
Published online December 19, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Soft sensors have been widely used in chemical plants to estimate process variables that are difficult to measure online. One of the crucial difficulties of soft sensors is that predictive accuracy drops due to changes in state of chemical plants. Characteristics of adaptive soft sensor models such as moving window models, just-in-time models and time difference models were previously discussed. The predictive accuracy of any traditional models decreases when sudden changes in processes occur. Therefore, a new soft sensor method based on online support vector regression (SVR) and the time variable was developed for constructing soft sensor models adaptive to rapid changes of relationships among process variables. A nonlinear SVR model with the time variable is updated with the most recent data. The proposed method was applied to simulation data and real industrial data, and achieved higher predictive accuracy than traditional ones even when time-varying changes in process characteristics happen. © 2013 American Institute of Chemical Engineers *AIChE J*, 60: 600–612, 2014

Keywords: process control, soft sensor, degradation, online support vector machine, time variable

Introduction

Soft sensors are widely used to predict process variables that are difficult to measure online.[1] 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. Through the use of soft sensors, the values of **y** can be predicted with a high degree of accuracy.

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. 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 of a soft sensor model. To reduce the degradation, the model is reconstructed with newest data. A moving window (MW) model[2,3] and a recursive model[4] are categorized as a sequentially updating type and a distance-based just-in-time (JIT) model,[5] a correlation-based JIT model[6] and a locally-weighted partial least square model[7] 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.

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 of **y** and that of explanatory variables **X**, was proposed.[8–10] This model is referred to as a time difference (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, whereas adapting to states of a plant are called adaptive models.[11] In addition, when data distributions are multimodal, multiple modeling approaches[12,13] 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.[14] 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.[15] Table 1 shows the compiled results. 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. The MW models are suitable for gradual changes of the slope of **X** and **y**. Meanwhile, when the slope of **X** and **y**

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

Table 1. Characteristics of TD, MW, and JIT Model[15]

Degradation				
Type	Rapidity	TD model	MW model	JIT model
Shift of **y**-value	Gradual	○[a]	○	×[c]
	Rapid	○	△[b]	×
	Instant	○	×	×
Shift of **x**-value	Gradual	○	○	○
	Rapid	○	△	○
	Instant	○	×	○
Change of the slope	Gradual	×	○	×
	Rapid	×	△	×
	Instant	×	×	×
Shift of **x**-value and change of the slope	Gradual	×	○	○×[d]
	Rapid	×	△	○×
	Instant	×	×	○×
		TD model	MW model	JIT model
Online abnormal data		Not affected	Affected	Affected
Maintenance cost[e]		Low	High	High

[a]The model can handle the degradation well.
[b]The model can handle the degradation to some extent.
[c]The model cannot handle the degradation.
[d]It depends on a situation whether the model can handle the degradation or not. If there is the appropriate shift of **X** values, the model can handle the change of the slope.
[e]This includes the model construction, abnormal data handling and the database management.

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. The predictive ability of current TD and MW models is not entirely sufficient when rapid changes of the slope, i.e., time-varying changes in a process, occur, and, thus, novel techniques are required to solve this problem. In addition, to the best of our knowledge, there is no description of the degradation in the presence of nonlinearity between **X** and **y** in any references.

The objectives in this study are, therefore, as follows:

1. The improvement of predictive ability of soft sensor models in the degradation that the changes in processes, or the slope changes, are rapid and time-dependent.
2. The handling of any nonlinear relationships between **X** and **y.**

When both (1) and (2) are achieved, we can cope with abrupt changes in process characteristics.

We adopt a sequentially updating approach since not TD and JIT models but a traditional MW model, which is a linear model, can adapt to gradual changes of the slope even when there is no shift of **X** values. By updating a nonlinear regression model, both (1) and (2) will be able to be solved because the change of the slope is equivalent to the nonlinear change of the slope. However, sequential updates of a nonlinear regression model take a lot of time. We, therefore, apply the online support vector regression (OSVR) method[16] that efficiently update a support vector regression (SVR)[17] model, which is a nonlinear regression model, to soft sensor modeling. Nevertheless, it will be difficult for a nonlinear regression model to adapt to time-varying processes even when the model is updated, and, thus, we propose to add a variable representing time in **X** variables.

To verify the effectiveness of the proposed method, we analyze various types of simulation data where the relationships between **X** and **y** change from moment to moment and have strong nonlinearity. The performance of the proposed models is compared with that of other traditional adaptive models. Then, the proposed method is applied to industrial polymer process data. Of course, the number of data for model construction in MW modeling and in JIT modeling is important for the high prediction accuracy. The discussion on the number of data is included in another article.[18]

Method

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. As soft sensor analyses, the OSVR method is applied to cell concentration prediction in a simulated bioreactor,[19] short-term traffic flow prediction[20], forecasting respiratory motion[21] and feedwater flow rate prediction in a nuclear power plant.[22]

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

Minimize

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

where $y_{i,}$ and $\mathbf{x}_i$ are training data, f is a SVR model, $\mathbf{w}$ is a weight vector, ε 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. 1 is the ε-insensitive loss function and given as follows

$$|y_i-f(\mathbf{x}_i)|_\varepsilon=\max\left(0,|y_i-f(\mathbf{x}_i)|-\varepsilon\right) \tag{2}$$

Through the minimization of Eq. 1, 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 **x** is represented as follows

$$f(\mathbf{x})=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)K(\mathbf{x}_i,\mathbf{x})+b \tag{3}$$

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

$$K(\mathbf{x}_i,\mathbf{x})=\exp\left(-\gamma\|\mathbf{x}_i-\mathbf{x}\|^2\right) \tag{4}$$

where γ is a tuning parameter controlling the width of the kernel function. From Eqs. 1 and 2, α_i and α_i^* in Eq. 3 are obtained by minimizing the equation given as

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}K_{ij}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)-\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right)+\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right) \tag{5}$$

subject to

$$0\le\alpha_i,\alpha_i^*\le C\; i=1,2,\ldots,N \tag{6}$$

$$\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0 \tag{7}$$

K_{ij} in Eq. 5 is represented as follows

$$K_{ij}=K\left(\mathbf{x}_i,\mathbf{x}_j\right) \tag{8}$$

Now, we define θ_i as follows

$$\theta_i=\alpha_i-\alpha_i^* \tag{9}$$

From Eqs. 3, 4, and 8, a predicted $\mathbf{y}$ value of data $\mathbf{x}_i$ is given as

$$f(\mathbf{x}_i)=\sum_{j=1}^{N}K_{ij}\theta_j+b \tag{10}$$

where θ_i meets the following equation

$$\sum_{i=1}^{N}\theta_i=0 \tag{11}$$

The error function h is defined as

$$\begin{aligned}h(\mathbf{x}_i)&\equiv f(\mathbf{x}_i)-y_i\\&=\sum_{j=1}^{N}K_{ij}\theta_j+b-y_i\end{aligned} \tag{12}$$

Then the KKT conditions can be summarized as follows

$$h(\mathbf{x}_i)>\varepsilon,\;\theta_i=-C \tag{13}$$

$$h(\mathbf{x}_i)=\varepsilon,\;-C<\theta_i<0 \tag{14}$$

$$-\varepsilon<h(\mathbf{x}_i)<\varepsilon,\;\theta_i=0 \tag{15}$$

$$h(\mathbf{x}_i)=-\varepsilon,\;0<\theta_i<C \tag{16}$$

$$h(\mathbf{x}_i)<-\varepsilon,\;\theta_i=C \tag{17}$$

Each training data must meet one of Eqs. 13–17. All training data can be divided into the following sets: error support vectors E, which meet Eqs. 13 or 17, margin support vectors S, which meet Eqs. 14 or 16 and remaining vectors R, which meet Eq. 15. Figure 1 shows the data regions of E, S and R.

When new data $\mathbf{x}_c$, y_c is added, there is no need to update the SVR model θ_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 θ_c that is θ_i corresponding to $\mathbf{x}_c$ is set as 0, and θ_c, θ_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. However, assuming no such movements, variations of $h(\mathbf{x}_i)$, θ_c, θ_i and b, $\Delta h(\mathbf{x}_i)$, $\Delta\theta_c$, $\Delta\theta_i$ and Δb, respectively, can be represented from Eqs. 11 and 12 as follows

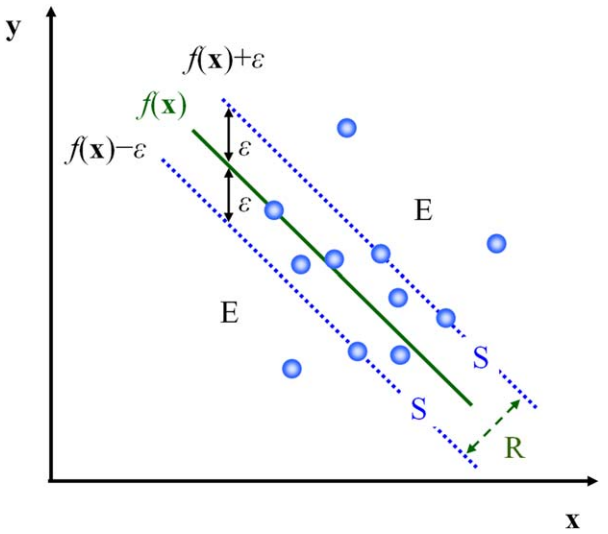

Figure 1. Data Regions for OSVR.

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

$$\Delta h(\mathbf{x}_i)=K_{ic}\Delta\theta_c+\sum_{j=1}^{N}K_{ij}\Delta\theta_j+\Delta b \tag{18}$$

$$\Delta\theta_c+\sum_{j=1}^{N}\Delta\theta_j=0 \tag{19}$$

The θ_i values of the training data belonging to E and R did not change because of Eqs. 13, 15, and 17, and, thus, Eq. 18 can be transformed as

$$\Delta h(\mathbf{x}_i)=K_{ic}\Delta\theta_c+\sum_{j\in S}K_{ij}\Delta\theta_j+\Delta b \tag{20}$$

The $h(\mathbf{x}_i)$ values of the training data belonging to S are settled due to Eqs. 14 and 16. Thus, Eqs. 19 and 20 can change to

$$\sum_{j\in S}K_{ij}\Delta\theta_j+\Delta b=-K_{ic}\Delta\theta_c\;\forall i\in S \tag{21}$$

$$\sum_{j\in S}\Delta\theta_j=-\Delta\theta_c \tag{22}$$

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

$$\Delta b=\delta\Delta\theta_c \tag{23}$$

$$\Delta\theta_i=\delta_i\Delta\theta_c\;\forall i\in S \tag{24}$$

where

$$\begin{bmatrix}\delta\\ \delta_{S_1}\\ \vdots\\ \delta_{S_M}\end{bmatrix}=-\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & K_{S_1S_1} & \cdots & K_{S_1S_M}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & K_{S_MS_1} & \cdots & K_{S_MS_M}\end{bmatrix}^{-1}\begin{bmatrix}1\\ K_{S_1c}\\ \vdots\\ K_{S_Mc}\end{bmatrix} \tag{25}$$

$$\delta_i=0\;\forall i\notin S \tag{26}$$

Here M is the number of the training data that belong to S. From Eqs. 20, 23, and 24, $h(\mathbf{x}_i)$ for the training data belonging to E and R can be transformed as

$$\begin{aligned}\Delta h(\mathbf{x}_i) &= K_{ic}\Delta\theta_c + \sum_{j\in \mathrm{S}} K_{ij}\Delta\theta_j + \Delta b \\ &= K_{ic}\Delta\theta_c + \sum_{j\in \mathrm{S}} K_{ij}\delta_j\Delta\theta_c + \delta\Delta\theta_c \\ &= \left(K_{ic} + \sum_{j\in \mathrm{S}} K_{ij}\delta_j + \delta\right)\Delta\theta_c \\ &= \gamma\Delta\theta_c\end{aligned} \tag{27}$$

where

$$\gamma = K_{ic} + \sum_{j=1}^{N} K_{ij}\delta_j + \delta \tag{28}$$

From Eqs. 24 and 27, $\Delta\theta_c$ for the movement of each training data is represented as

$$\Delta\theta_c = {\delta_i}^{-1}\Delta\theta_i \;\forall i \in \mathrm{S} \tag{29}$$

$$\Delta\theta_c = \gamma^{-1}\Delta h(\mathbf{x}_i) \;\forall i \notin \mathrm{S} \tag{30}$$

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. 29 and 30. 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. 13–17. When one data is deleted from training data, the same iterative calculation is performed until all the data meet the KKT conditions.

In this study, the already developed OSVR programs[23] were used as the machine learning software.

Time variable

By updating a nonlinear regression model, the model will be able to deal with new nonlinear relationships between **X** and **y**. However, even the updated model cannot adapt to time-varying process changes. We, therefore, propose to add the variable representing process time to **X**-variables. By modeling the changes of process characteristics with this time variable, the model will be able to appropriately predict following **y** values. In addition, even when the changes in process are nonlinear in terms of time, an updated nonlinear model with the time variable can be adaptive to the changes.

The time variable is an interval scale and the zero point can be set arbitrarily.

Results and Discussion

To verify the effectiveness of the proposed method, we analyzed various types of simulation data and industrial polymer process data. The relationships between **X** and **y** change from moment to moment for one simulation data set and the relationships have strong nonlinearity for the other simulation data set. The compared models are as follows:

- PLS: Non-adaptive model constructed with the PLS method;
- SVR: Non-adaptive model constructed with the SVR method;
- TDPLS: TD model constructed with the PLS method[24];
- TDSVR: Nonlinear TD model constructed with the SVR method[10];
- JITPLS: PLS model constructed with data whose Euclidian distances to new data are smaller than those of other data;
- LWPLS: PLS model constructed by weighting data with a similarity matrix[25];
- MWPLS: PLS model constructed with data that are measured most recently;
- OSVR: Updated SVR model.

Using the TDPLS and TDSVR models, the time difference from newly obtained data of **y** was predicted. In JIT modeling, JITPLS and LWPLS update the model with the database in which newly obtained data was stored. MWPLS and OSVR models were updated with newly obtained data.

Each model including the time variable was also used in the case studies. The hyperparameters for the PLS, SVR and LWPLS models were selected with the five-fold cross-validation. When JITPLS and LWPLS models were used, the time variable was also considered in selecting the data and weighting data with a similarity matrix, respectively.

Data in which the relationship between X and y is time-varying

The number of **X** variables was set as two. First, $\mathbf{x}_1$ and $\mathbf{x}_2$ of uniform pseudorandom numbers whose ranges were from 0 to 10 was prepared. Then, **y** was set as follows

$$\mathbf{y} = [\,\mathbf{x}_1 \quad \mathbf{x}_2\,][\,1 \quad b_2\,]^{\mathrm{T}} + \mathrm{N}(0, 0.1) \tag{31}$$

where b_2 means the magnitude of contribution of $\mathbf{x}_2$ to **y** and $N(0, 0.1)$ is random numbers from normal distribution given a standard deviation of 0.1 and a mean of 0. We set b_2 as follows

$$b_2 = 3\sin(0.01\pi t) + 1 \tag{32}$$

$$b_2 = 3\sin(0.02\pi t) + 1 \tag{33}$$

Equations 32 and 33 represent gradual and rapid changes of b_2, respectively, and t was set as 1, 2,., 200, and the number of data was 200. The first 100 data were used for training and the next 100 data were the test data. These data set is the same as that of "*Change of the Slope*" in the reference[15] and used to discuss the number of data for the model construction in the reference.[18] We assumed that new data was obtainded at the next step of the measurement and only data including **y** measurement were collected because the **y** variable has time delay of measurement and is not measured for every sampling time of **X** variable in real plants.

The window size, which means the number of data for the model construction, was set as 20 and 50 for the JITPLS, MWPLS and OSVR models. For the OSVR models, we decided the hyperparameters to be the best prediction performance on the last 50 or 80 training data. When the window size was 20, the number of the last training data was 80, and when the window size was 50, the number of the last training data was 50.

Table 2 shows the prediction results for test data when only $\mathbf{x}_1$ and $\mathbf{x}_2$ were used and the time variable was not used as **X** variables. The ${r_{\mathrm{pred}}}^2$ is the determination coefficient for test data and represents the prediction accuracy of each model. The higher the ${r_{\mathrm{pred}}}^2$ value of a model is, the more predictive accuracy of the model has. The $RMSE_{\mathrm{P}}$ is the root-mean-square error for test data. The lower $RMSE_{\mathrm{P}}$

Table 2. The Prediction Results without the Time Variable using the Simulation Data where the Relationship between x and y is Time-Varying

	b_2: Eq. 32		b_2: Eq. 33	
Model	r_p^2	$RMSE_P$	r_p^2	$RMSE_P$
PLS	−10.6	19	0.092	11
SVR	−9.9	19	0.067	11
TDPLS	−6.1	15	0.376	9.0
TDSVR	−9.9	19	−0.19	12
JITPLS, 20[a]	−6.0	15	0.009	11
JITPLS, 50[a]	−6.0	15	0.053	11
LWPLS	−5.9	15	0.045	11
MWPLS, 20[a]	0.546	3.8	0.574	7.3
MWPLS, 50[a]	−1.3	8.7	−0.44	13
OSVR, 20[a]	0.498	4.1	0.536	7.6
OSVR, 50[a]	−1.1	8.3	−0.20	12

[a]The number of data for the model construction.

values of models mean the higher prediction accuracy of the models. In both cases of b_2 of Eqs. 32 and 33, the prediction ability of the PLS, SVR, TDLS, TDSVR, JITPLS, and LWPLS was low because the r_{pred}^2 value was low and the $RMSE_P$ value was high for each model. Those models could not adapt to the time-dependent process characterictics.

The prediction results of the MWPLS models were better than those of other models when the window size is 20, but nevertheless, the r_{pred}^2 values were low and the $RMSE_P$ values were high for the MWPLS models. The increase of the window size confounded the predictive ability of the models. Even when the SVR models were updated by using the OSVR method, the positive results could not be obtained relative to those of the MWPLS models. Any models could not accurately adapt to the time-varying change of the relationships between **X** and **y**.

The prediction results are shown in Table 3 when t was added to **X** variables ($\mathbf{x}_1$ and $\mathbf{x}_2$) as the time variable. In PLS, SVR, TDPLS, TDSVR, JITPLS and LWPLS modeling, some of the prediction results improved, compared with the results without the time variable, but yet, the improvements were not enough. In both cases of b_2, the r_{pred}^2 values highly increased and the $RMSE_P$ values significantly decreased by using the MWPLS models of the window size 20 with the time variable. Furthermore, the significant improvement of the predictive ability for the OSVR models was confirmed by the addition of the time variable. The proposed models could appropriately deal with the time-varying change of the relationship between **X** and **y** by updating the SVR models and by adding the time variable to **X** variables.

For the MWPLS models, the predictive accuracy drastically decreased by increaseing the window size from 20 to 50. The nonlinear and time-dependent changes in process characteristics were linearly-approximated in MWPLS modeling. By increasing the window size, the difference between the actual nonlinear relationship and the linearly-approximated relationship would be large. On the other hand, there was little change in the prediction accuracy of the OSVR models when the window size varied. By adding the time variable and enabling to model the nonlinear change in process characteristics, the robust OSVR models could be constructed for the window size.

Figures 2 and 3 shows the relationships between measured and predicted **y** for test data in MWPLS and OSVR modeling with two variables ($\mathbf{x}_1$ and $\mathbf{x}_2$) and three variables ($\mathbf{x}_1$, $\mathbf{x}_2$ and t) when Eqs. 32 and 33 were used as b_2, respectively. The window sizes are 20 for Figure 2 and 50 for Figure 3. From the plots, we can say that the data came close to the diagonal and the prediction accuracy improved by adding the time variable as **X**-variables. Moreover, by using the OSVR modeling, the plots show tight clusters of predicted values along the diagonals in both cases of b_2, meaning that the prediction accuracy of the models was suitably high. It was confirmed that the proposed models could deal with the sudden and time-varying changes in the processes.

Data in which the relationship between X and y is nonlinear

To verify the performance of the proposed method when the relationship between **X** and **y** is nonlinear, analyses using two types of data were performed. The relationships between **X** and **y** in the data was given as follows

$$y = \sin(x_1)\cos(x_2) + 0.1x_1 \tag{34}$$

$$\exp(y) = \left(1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3{x_1}^2 - 14x_2 + 6x_1x_2 + 3{x_2}^2)\right) \times \left(30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12{x_1}^2 + 48x_2 - 36x_1x_2 + 27{x_2}^2)\right) \tag{35}$$

Equations 34 and 35 are described in the reference[26] as test problems, and on Eq. 35 $0.1 \times {}_1$ was added to the raw equation. The data of **X**-variables were generated to be randomly walked within ±3 for Eq. 34 and within ±2 for Eq. 35. Figure 4 shows the trajectories of the data of **X**-variables. The color bars represent **y** values given by Eqs. 34 and 35. Random numbers from the normal distribution with a mean of 0 and a standard deviation 0.01 were added to the **y** variables. In the reference,[18] the same data set was used to discuss the window size of adaptive soft sensor models. For Eqs. 34 and 35, the 1,100 data were generated, the first 100 data were used for training, and the next 1,000 data were the test data. We assumed that new data was obtained at the next step of the measurement and only data including **y** measurement were collected as was the case in "*Data in which the relationship between X and y is time-varying*". The window size was set as 20 and 50 for the JITPLS, MWPLS and OSVR models.

Table 4 shows the prediction results of test data when only $\mathbf{x}_1$ and $\mathbf{x}_2$ were used as **X**-variables. For the nonadaptive models, the prediction accuracy was low even when the SVR method, a nonlinear regression method, was used. However, the r_{pred}^2 value exceeded 0.8 and relatively good

Table 3. The Prediction Results with the Time Variable using the Simulation Data where the Relationship between x and y is Time-Varying

	b_2: Eq. 32		b_2: Eq. 33	
Model	r_p^2	$RMSE_P$	r_p^2	$RMSE_P$
PLS	−10.6	19	−6.4	31
SVR	−1.2	8.4	−5.9	20
TDPLS	−6.1	15	0.376	9.0
TDSVR	−19.2	26	−20.1	52
JITPLS, 20[a]	0.407	4.4	0.261	9.6
JITPLS, 50[a]	−0.25	6.4	−0.19	12
LWPLS	−0.30	6.5	0.265	9.6
MWPLS, 20[a]	0.848	2.2	0.848	4.4
MWPLS, 50[a]	0.147	5.3	0.239	9.8
OSVR, 20[a]	0.996	0.35	0.994	0.85
OSVR, 50[a]	0.995	0.42	0.993	0.93

[a]The number of data for the model construction.

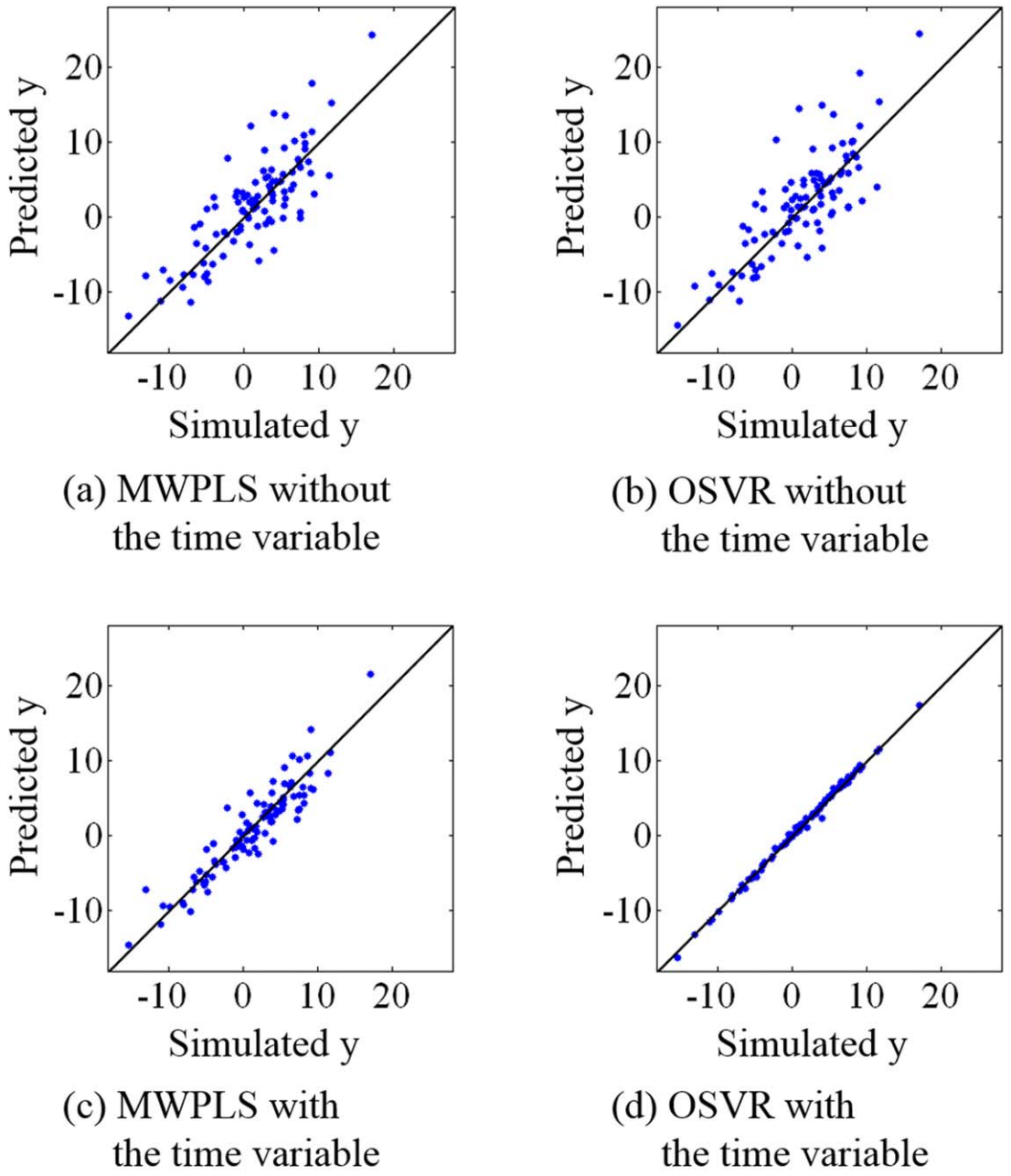

Figure 2. Relationships between y and predicted y of test data when Eq. 32 was used as b_2.

The numbers of data for the model construction are 20. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

prediction results could be obtained by using the adaptive models. The nonlinear relationships between **X** and **y** of test data are different from those of training data as shown in Figure 4, and, therefore, the predictive accuracy decreased for the nonadaptive SVR models.

For the LWPLS model, the r_{pred}^2 value was lower and the $RMSE_{\text{P}}$ value was higher than those of the other adaptive models when Eq. 35 was used. This was because there are no old data where the nonlinear relationship between **X** and **y** was the same as that of new data, and the LWPLS model could not be constructed appropriately. Even in this situation, the OSVR models indicated the same or high performance of prediction, compared to the other adaptive models.

In JITPLS and MWPLS modeling, the increase of the window size led to the decrease of the prediction accuracy as it was in the results of "*Data in which the relationship between X and y is time-varying*". The linear model is locally constructed and the nonlinear relationships between **X** and **y** is linearly-approximated in each modeling. The prediction accuracy would become low as the window size increased, and the linear approximation of the local model got bad. On the other hand, by using the OSVR models, when the window size increased from 20 to 50, the predictive ability did not largely change or slightly improved. It was confirmed that the predictive models can be constructed regardless of the window size by updating the nonlinear models.

Table 5 shows the prediction results when t was added to **X** variables as the time variable. There is little difference between the results of Table 4 and those of Table 5. We can say that the time variable did not contribute to the increase

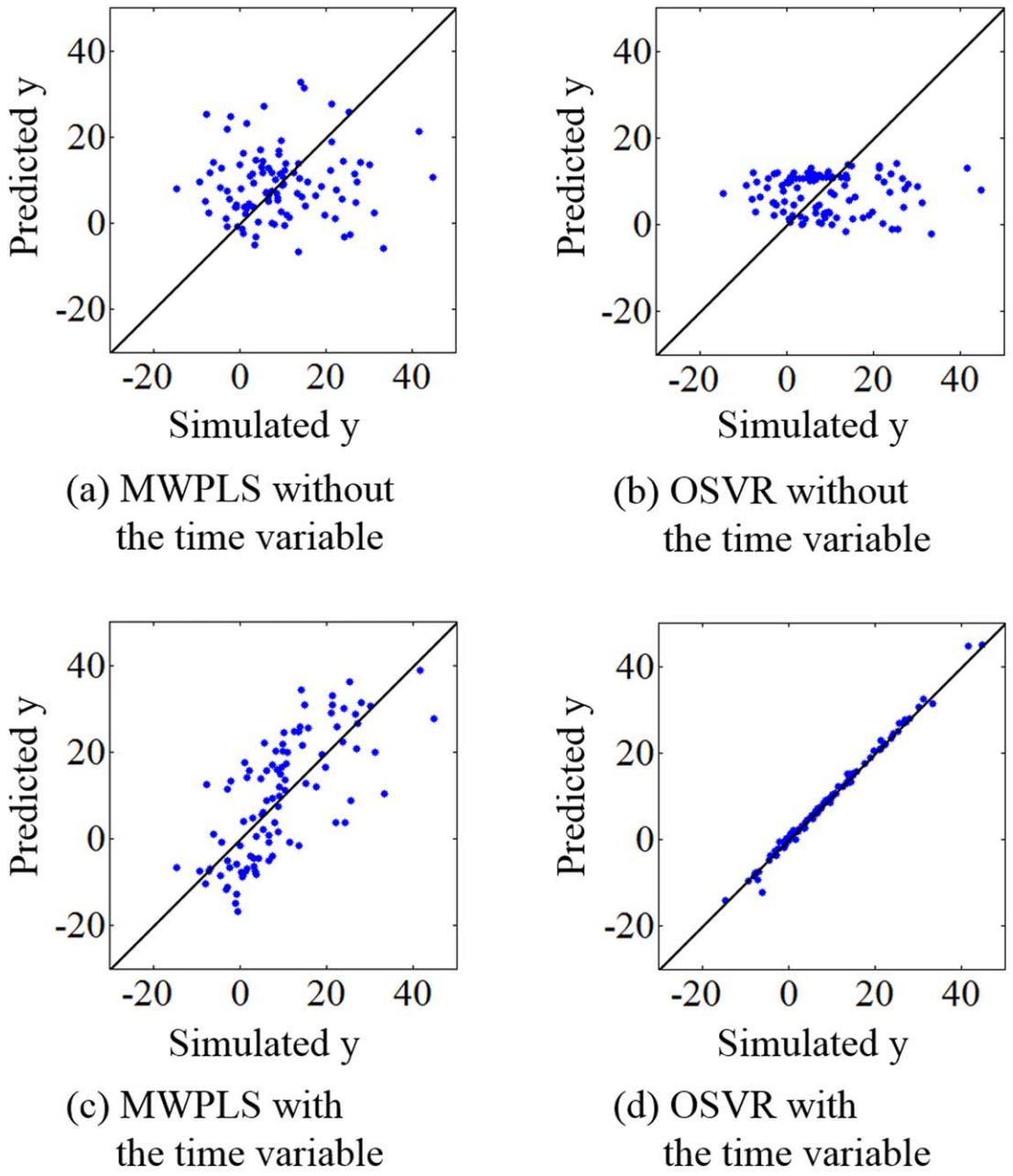

Figure 3. Relationships between y and predicted y of test data when Eq. 33 was used as b_2.

The numbers of data for the model construction are 50. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of the prediction accuracy when the relationship between **X** and **y** is only nonlinear.

Figures 5 and 6 shows the relationships between measured and predicted **y** for test data in JITPLS, LWPLS, MWPLS and OSVR modeling when Eqs. 34 and 35 were used, respectively. The window sizes of the JIT, MWPLS and OSVR models are 50. From Figure 5, the plot shows a much tighter clustering of predicted values along the diagonal in OSVR modeling, reflecting the higher prediction of **y**. The OSVR model could adequately adapt to the nonlinear relationship between **X** and **y**.

As shown in the plot of Figure 6b, relatively many data are far from the diagonal for the LWPLS model, indidating that the LWPLS model could not deal with new relationships between **X** and **y** that did not exist in the database. From the plots of Figure 6a and 6c, the positive bias of the prediction errors could be seen when the simulated **y** values are more than 9 for the JITPLS and MWPLS models. Meanwhile, the plot of Figure 6d shows no bias and a tight cluster of predicted values along the diagonal line for the OSVR model. This means that the OSVR had the good performance of prediction. However, when **y** values were small, some data were far from the diagonal and predictive accuracy was relatively low for the OSVR model as well as in JITPLS and MWPLS modeling. The optimal hyperparameters and the optimal window size of the OSVR model would differ according to the relationships between **X** and **y**. The predictive ability of the OSVR will improve by selecting OSVR parameters that are appropriate for each process situation.

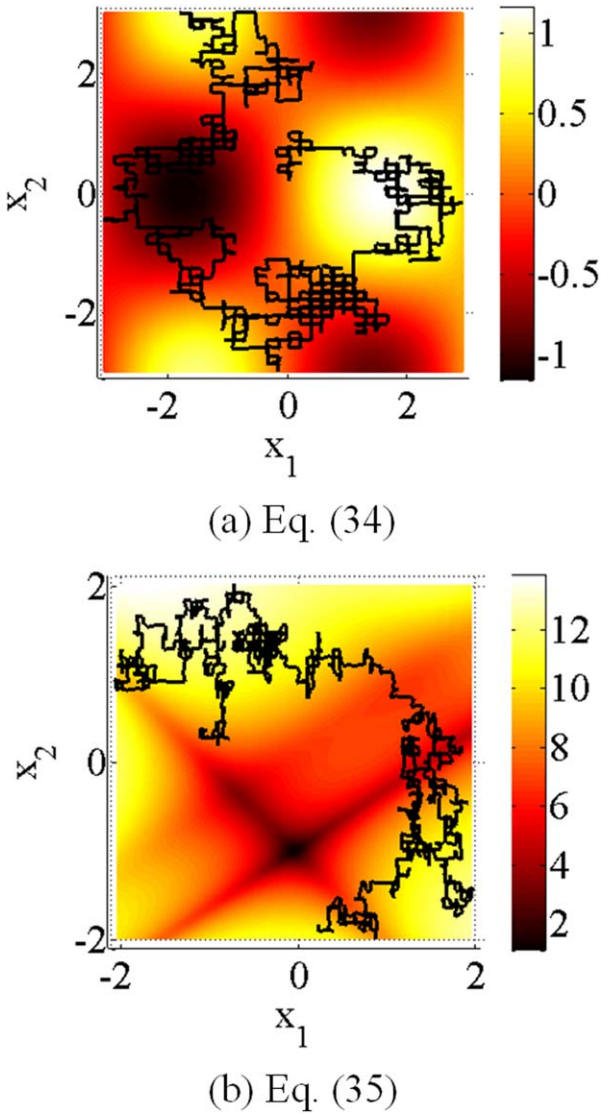

Figure 4. Trajectories of simulation data of X variables.[18]

The color bars represent y values. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Application to industrial polymer processes

We applied the proposed methods to actual industrial data obtained during an industrial polymer process at Mitsui Chemicals, Inc., to verify the prediction ability. Industrial polymer processes generally produce many grades of products. Therefore, when a polymer grade changes, it is important to reduce the quantity of off-grade material. Thus, an early and accurate judgment on whether the polymer quality is within the given specifications or not is made by using soft sensors because it is impossible to perform online measurements of a large number of polymer quality variables by using hard sensors. Constructing models with high-prediction performance is difficult under the present circumstances. One of the reasons is that there can be nonlinear relationship between a polymer quality variable **y**, and other process variable **X.**[18,27–31] In addition, impurities can be mixed in with raw materials and the relationships between **X** and **y** will depend on the amount of the impurities.

We attempted to construct nonlinear models between **X** and **y** with various kinds of polymer grades. The constructed models can be applied to data of any kinds of polymer grades if the relationship between **X** and **y** can be extrapolated to those data. Data measured in steady state of many grades were collected in this study.

Table 4. The Prediction Results without the Time Variable using the Simulation Data where the Relationship between x and y is Nonlinear

	Eq. 34		Eq. 35	
Model	r_p^2	$RMSE_P$	r_p^2	$RMSE_P$
PLS	−1.76	1.022	−1.82	3.704
SVR	0.068	0.593	−0.65	2.831
TDPLS	0.985	0.0747	0.969	0.389
TDSVR	0.954	0.1315	0.981	0.308
JITPLS, 20[a]	0.997	0.0342	0.985	0.275
JITPLS, 50[a]	0.978	0.0907	0.976	0.343
LWPLS	0.987	0.0700	0.855	0.840
MWPLS, 20[a]	0.994	0.0484	0.982	0.299
MWPLS, 50[a]	0.967	0.1117	0.954	0.472
OSVR, 20[a]	0.999	0.0167	0.992	0.197
OSVR, 50[a]	0.999	0.0165	0.989	0.233

[a]The number of data for the model construction.

The **y** variables represent melt flow rate (MFR) and density, whose measurement delay was considered, and the **X** variables represent 38 variables such as the temperature in the reactor and the pressure and concentration of the monomer, comonomer, and hydrogen. MFR was logarithmically-transformed. We used the data monitored from January 2005 to April 2007 as the training data, and that from May 2007 to May 2008 as the test data.

The prediction results for MFR and density are shown in Tables 6 and 7, respectively. In this study, the window size was set as 200 and 300 for the JITPLS, MWPLS and OSVR models due to the large number of **X** variables. SVR parameters optimized with a five-fold cross validation on training data were used as the parameters of an OSVR model.

As can be seen in Table 6, the r_{pred}^2 values of the SVR models are higher than those of the PLS models and the results are vice versa for the $RMSE_P$ values. The prediction accuracy of the SVR models was higher than that of the PLS models because the SVR models could represent the nonlinear relationship between **X** variables and MFR. The predictive ability decreased by using the TDPLS, JITPLS, LWPLS and MWPLS models, but the r_{pred}^2 values and the $RMSE_P$ values improved by using the TDSVR models. This is probably because the time difference could handle gradual changes in the process and the SVR model could deal with

Table 5. The Prediction Results with the Time Variable using the Simulation Data where the Relationship between x and y is Nonlinear

	Eq. 34		Eq. 35	
Model	r_p^2	$RMSE_P$	r_p^2	$RMSE_P$
PLS	−0.45	0.7393	−70.6	18.68
SVR	0.115	0.5786	0.001	2.256
TDPLS	0.985	0.0747	0.969	0.389
TDSVR	0.985	0.0741	0.981	0.307
JITPLS, 20[a]	0.994	0.0460	0.982	0.299
JITPLS, 50[a]	0.973	0.1002	0.987	0.360
LWPLS	0.988	0.0686	0.868	0.802
MWPLS, 20[a]	0.992	0.0552	0.981	0.307
MWPLS, 50[a]	0.963	0.1182	0.960	0.440
OSVR, 20[a]	0.999	0.0210	0.991	0.204
OSVR, 50[a]	0.999	0.0182	0.987	0.254

[a]The number of data for the model construction.

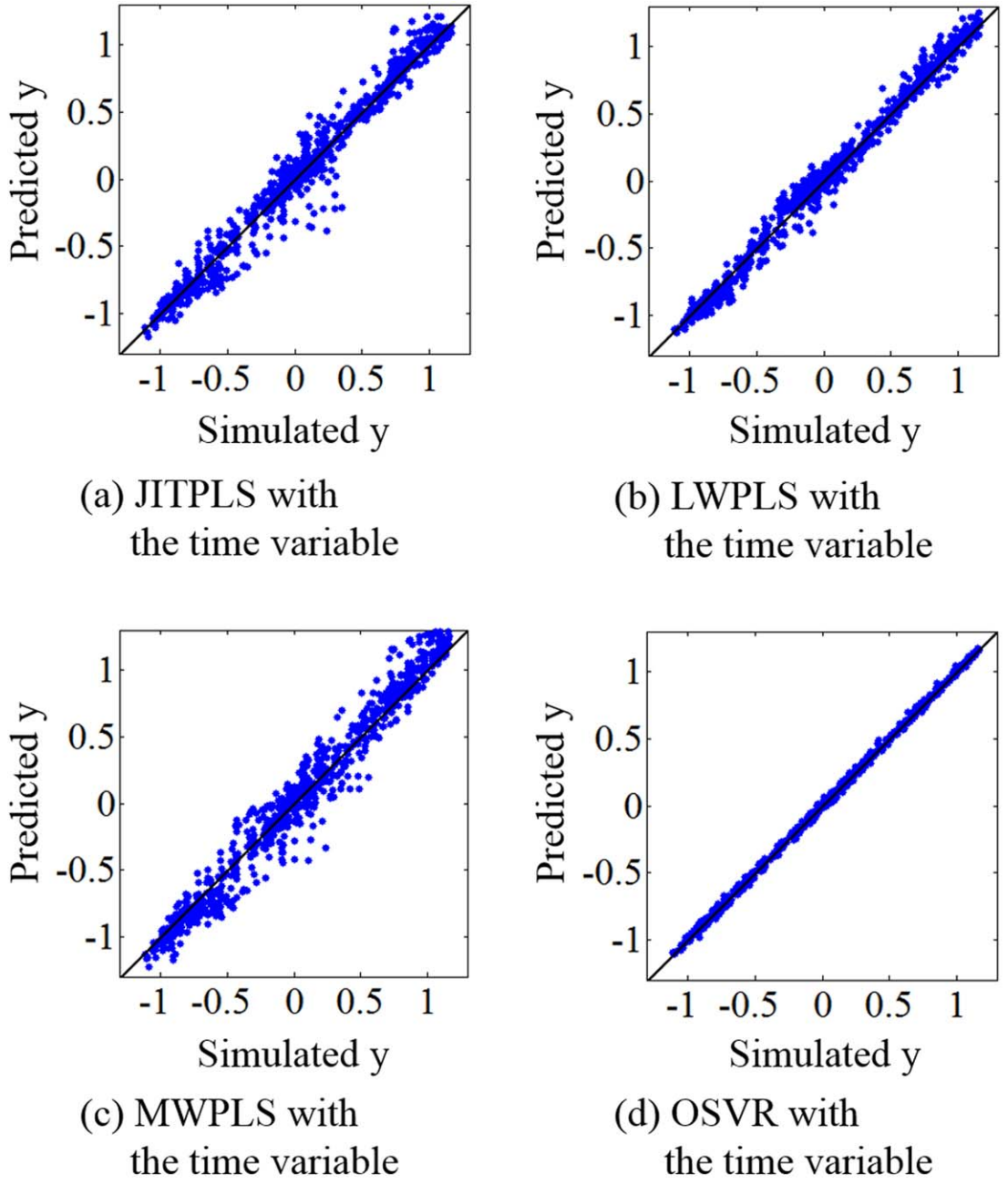

Figure 5. Relationships between y and predicted y of test data when simulation data were generated with Eq. 34.

The numbers of data for the model construction are 50 for the JITPLS, MWPLS and OSVR models. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the nonlinearity between **X**-variables and MFR. In addition, the OSVR models improved the r_{pred}^2 values and the $RMSE_{\text{P}}$ values. The prediction accuracy of the SVR models was higher than that of the other models. It can be said that the soft sensors could adapt the changes in the characteristic of the polymer process by updating the SVR models.

The increase of the window size from 200 to 300 increased the predictive ability of the OSVR models. On the one hand, for JITPLS and MWPLS models, which are linear regression models, linear approximation of the nonlinear relationship between **X** and **y** is not allowed when the window size increases. On the other hand, for an OSVR model, which is a nonlinear regression model, the nonlinearity between **X** and **y** can be handled and, thus, a stable model will be constructed by increasing the window size.

In the case of MFR, there was little difference between the prediction results without the time variable and those with the time variable. The time variable did not affect the r_{pred}^2 values and the $RMSE_{\text{P}}$ values of the OSVR models. From the results of the simulation data analyses, the addition of the time variable to the OSVR models improved the prediction accuracy when the relationship between **X** and **y** is time-varying, and the time variable did not affect the prediction results of the OSVR models when the relationship between **X** and **y** is nonlinear. The relationship between **X** variables and MFR would be nonlinear and would not be time-dependent.

Figure 7 shows the relationships between measured and predicted MFR with test data for some models. The window sizes are 300 for the MWPLS and OSVR models. For the LWPLS model, the relatively large variation of the

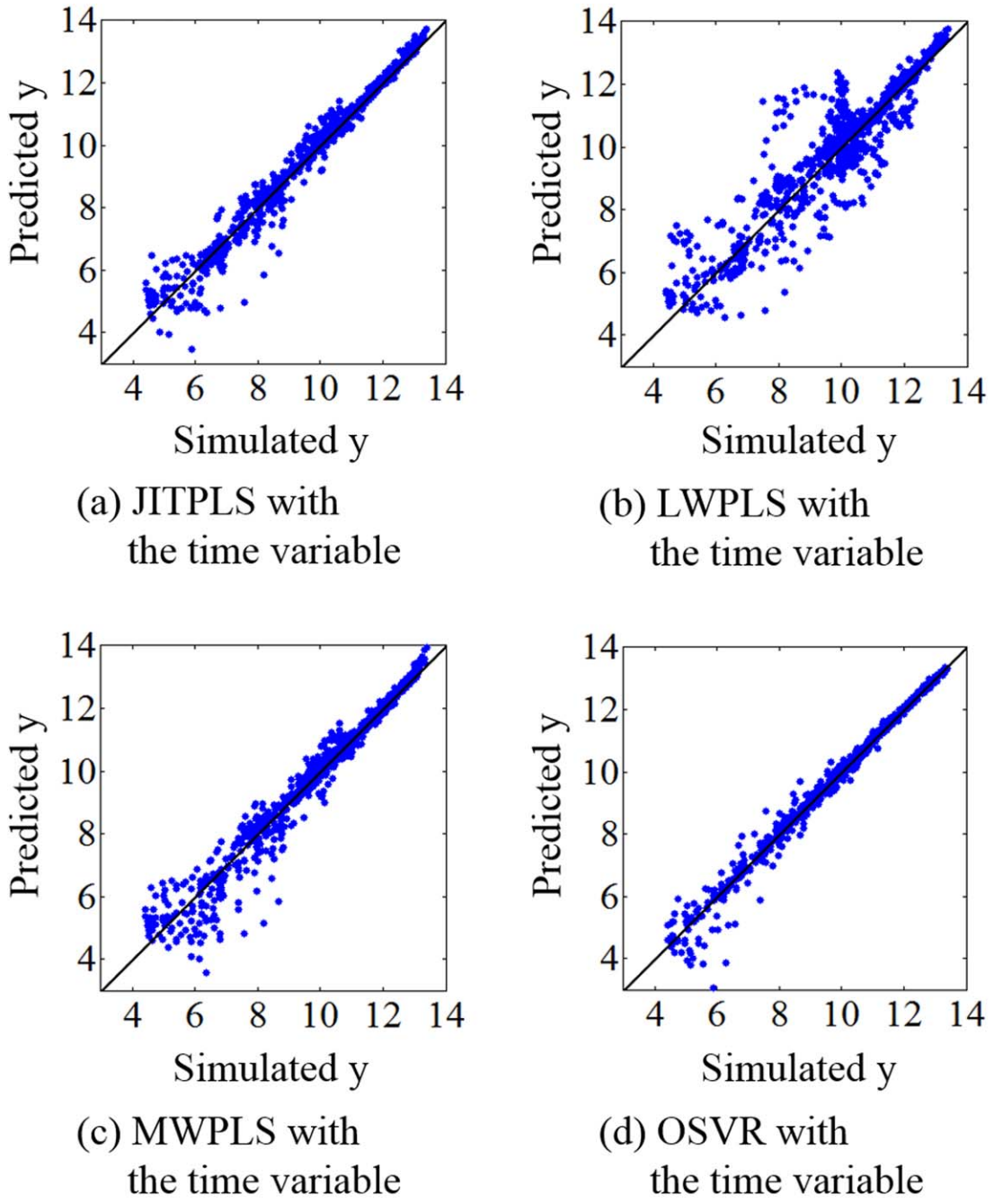

Figure 6. Relationships between y and predicted y of test data when simulation data were generated with Eq. 35.

The numbers of data for the model construction are 50 for the JITPLS, MWPLS and OSVR models. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

prediction errors of small MFR values could be seen from Figure 7b, and for the MWPLS model, the relatively large variation of the prediction errors of large MFR values could be seen from Figure 7c. Meanwhile, the plot of Figure 7d shows much tighter clusters of predicted values along the diagonal than those of the other models, reflecting the higher prediction of MFR by using the proposed model.

Table 7 is the prediction results for density. The prediction accuracy of the SVR models was higher than that of the PLS models as was the case with MFR. This will come from the nonlinearity between **X** variables and density. In the adaptive models, the r_{pred}^2 value was low and the $RMSE_{\text{P}}$ value was high for the linear TDPLS models, compared with those of the SVR models, but the other adaptive models could improve the predictive accuracy. Besides, the proposed OSVR model with the time variable achieved the highest performance of the density prediction. The results of the proposed model were superior to those of the OSVR models without the time variable, indicating that the relationship between **X** variables and density is time-varying from the results of the analyses with two types of the simulation data. The relationship between **X** variables and density may be time-varying and depend on the amount of impurities in raw materials.

Along with the case of MFR, on the one hand, the increase of the window sizes decreased the prediction

Table 6. The Prediction Results for MFR

Model	Without the time variable		With the time variable	
	r_p^2	$RMSE_P$	r_p^2	$RMSE_P$
PLS	0.968	0.176	0.964	0.186
SVR	0.989	0.104	0.988	0.109
TDPLS	0.941	0.238	0.941	0.238
TDSVR	0.991	0.095	0.990	0.097
JITPLS, 200[a]	0.987	0.113	0.985	0.119
JITPLS, 300[a]	0.980	0.139	0.979	0.141
LWPLS	0.987	0.111	0.987	0.109
MWPLS, 200[a]	0.980	0.137	0.981	0.135
MWPLS, 300[a]	0.979	0.143	0.978	0.143
OSVR, 200[a]	0.992	0.088	0.992	0.088
OSVR, 300[a]	0.993	0.083	0.993	0.083

[a]The number of data for the model construction.

Table 7. The Prediction Results for Density

Model	Without the time variable		With the time variable	
	r_p^2	$RMSE_P$ ($\times 10^{-3}$)	r_p^2	$RMSE_P$ ($\times 10^{-3}$)
PLS	0.923	3.35	0.926	3.29
SVR	0.953	2.61	0.931	3.19
TDPLS	0.933	3.12	0.933	3.12
TDSVR	0.960	2.43	0.954	2.58
JITPLS, 200[a]	0.971	2.05	0.969	2.11
JITPLS, 300[a]	0.960	2.41	0.960	2.44
LWPLS	0.968	2.15	0.962	2.37
MWPLS, 200[a]	0.967	2.19	0.963	2.32
MWPLS, 300[a]	0.960	2.42	0.959	2.45
OSVR, 200[a]	0.949	2.73	0.974	1.96
OSVR, 300[a]	0.956	2.54	0.976	1.88

[a]The number of data for the model construction.

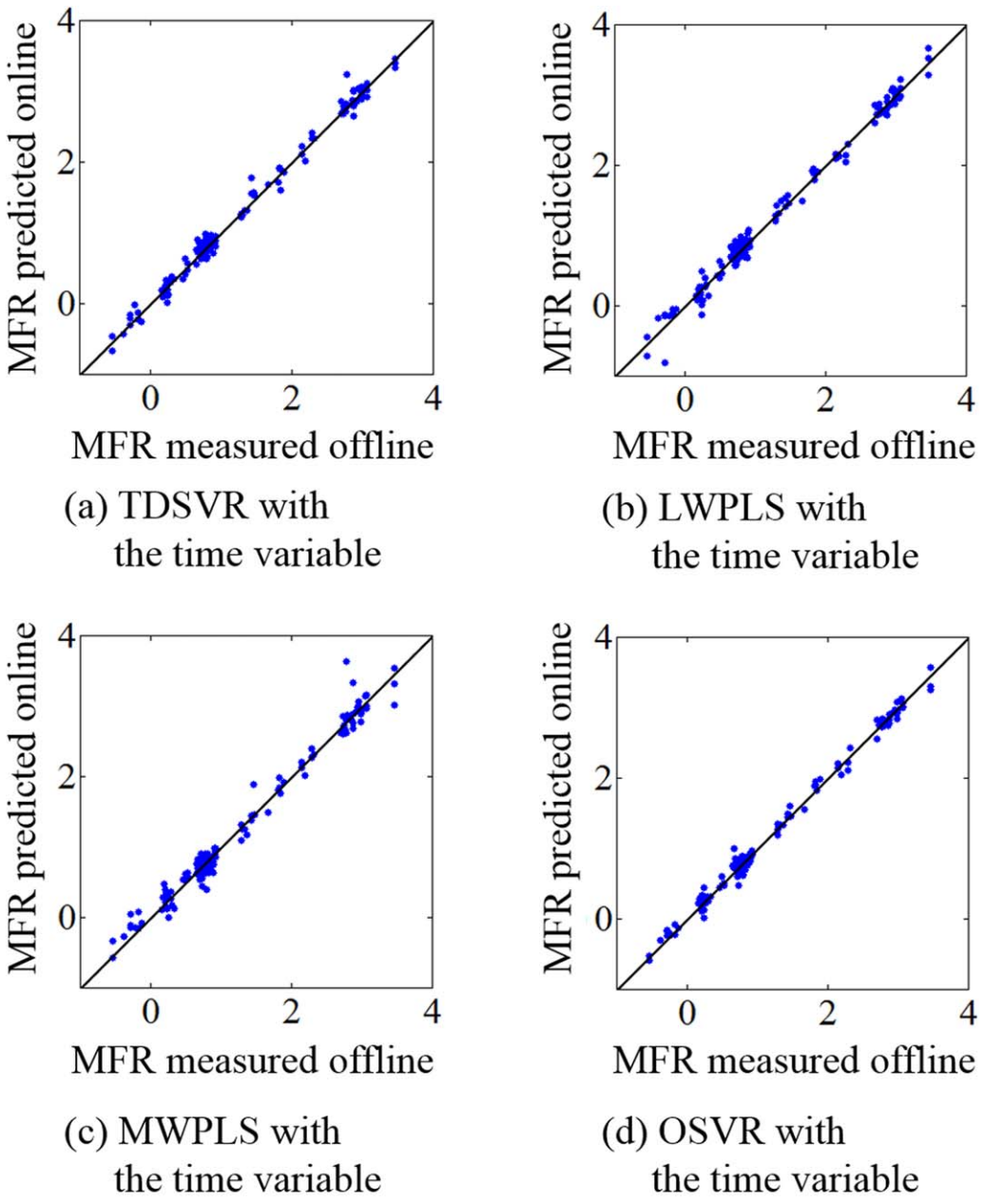

Figure 7. Relationships between measured and predicted MFR with test data.

The numbers of data for the model construction are 300 for the MWPLS and OSVR models. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

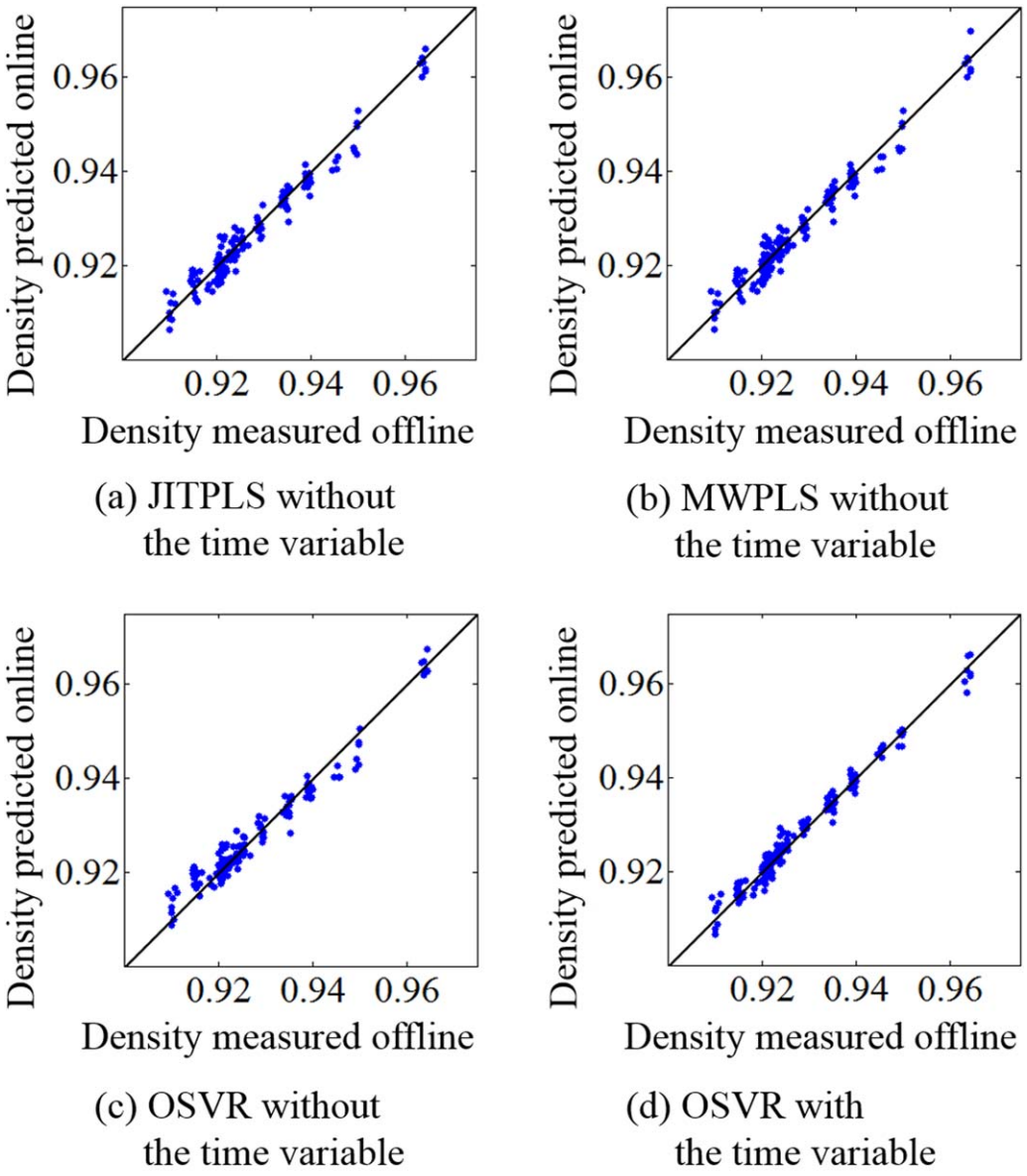

Figure 8. Relationships between measured and predicted density with test data.

The numbers of data for the model construction are 300. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

accuracy for the JITPLS and MWPLS models; on the other hand, the predictive ability improved with the increase of the window size for the OSVR model construction. The update of the SVR model and the time variable can deal with the nonlinear and time-varying relationship between process variables, and, therefore, the increase of the data can contribute to a predictive and stable model for the proposed method.

Figure 8 shows the relationships between measured and predicted density with test data. The window sizes are 300. From the plots of Figure 8a, b, and c, the negative bias of the prediction errors could be seen when the actual values of density are from 0.94 to 0.95 for the JITPLS, MWPLS and OSVR models without the time variable. By using the OSVR model with the time variable, the bias did not exist and much tighter clusters of predicted values along the diagonal could be shown, reflecting the higher prediction of density. It was confirmed that the proposed model can accurately adapt to the nonlinear relationships between $\mathbf{X}$ and $\mathbf{y}$ and the time-varying changes in the relationships, i.e., the changes in process characteristics.

The hyperparameters and the window size of the OSVR models were fixed in prediction in all the case studies. However, the appropriate parameters must differ depending on a state in a process. Therefore, the further improvement of the prediction accuracy in OSVR modeling will be achieved by selecting the adequate parameters in response to process characteristics.

Conclusion

In this article, to accurately predict $\mathbf{y}$ values even when the changes in the relationship between $\mathbf{X}$ and $\mathbf{y}$ are nonlinear and time-varying, we proposed a new adaptive soft sensor model combining the OSVR method and the time

variable. By adding the time variable to **X** variables and updating the SVR model efficiently, the proposed model can adapt to the abrupt change in process characteristics.

Through the analyses of various simulation data, it was confirmed that the OSVR models with the time variable had the good performance of prediction when the relationship between **X** and **y** changed from moment to moment and had strong nonlinearity. In addition, the proposed model could deal with the new nonlinear relationships between **X** and **y** that did not exist in the database. In addition, the superiority of the proposed method was confirmed through the real industrial data analyses. The appropriate selection of the parameters of an OSVR model depending on process states and the ensemble prediction using multiple OSVR models will be able to improve the prediction accuracy.

The window size and the way to determine to hyperparameters in MW and JIT modeling is important for high-prediction accuracy. The discussion on the window size, the hyperparamters and calculation time of adaptive soft sensor models is included in another article.[18] We set the optimization of the window size of the OSVR models as one of the future works.

The main issue of the proposed method is that an updated model specializes in predictions over a narrow data range, after there is little variation in a process over a long period as it is for sequentially updating models and JIT models. Appropriate control of the database is required for solving the issue.

We believe that by applying our proposed method to process control and achieving the adaptation to process characteristics and the highly accurate prediction, chemical plants will be operated effectively and stably.

Acknowledgment

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 Mitsui Chemicals, Inc., and the financial support of Mizuho Foundation for the Promotion of Sciences.

Literature Cited

1. Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Comput Chem Eng*. 2009;33:795–814.
2. 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.
3. Kadlec P, Gabrys B. Local learning-based adaptive soft sensor for catalyst activation prediction, *AIChE J*. 2010;57:1288–1301.
4. Qin SJ. Recursive PLS algorithms for adaptive data modeling. *Comput Chem Eng*. 1998;22:503–514.
5. Cheng C, Chiu MS. A new data-based methodology for nonlinear process modeling. *Chem Eng Sci*. 2004;59:2801–2810.
6. Fujiwara K, Kano M, Hasebe S, Takinami A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE J*. 2009; 55:1754–1765.
7. Schaal S, Atkeson CG, Vijayakumar S. Scalable techniques from onparametric statistics for real time robot learning. *Appl Intell*. 2002; 17:49–60.
8. Kaneko H, Funatsu K. Maintenance-free soft sensor models with time difference of process variables. *Chemom Intell Lab Syst*. 2011; 107:312–317.
9. 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.
10. 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.
11. Kadlec P, Grbic R, Gabrys B. Review of adaptation mechanisms for data-driven soft sensors. *Comput Chem Eng*. 2011;35:1–24.
12. 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.
13. 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.
14. 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.
15. Kaneko H, Funatsu K. Classification of the degradation of soft sensor models and discussion on adaptive models. *AIChE J*. 2013;59: 2339–2347.
16. Ma J, Theliler J, Perkins S. Accurate on-line support vector regression. *Neural Comput*. 2003;15:2683–2703.
17. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
18. Kaneko H, Funatsu K, Adaptive soft sensor model using online support vector regression with the time variable and discussion on appropriate hyperparameters and window size. *Comput Chem Eng*. 2013;58:288–297.
19. Iplikci, S. Online trained support vector machines-based generalized predictive control of non-linear systems. *Int J Adapt Control Signal Process*. 2006;20:599–621.
20. Neto MC, Jeong YS, Jeong MK, Han LD. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst Appl*. 2009;36:6164–6173.
21. Ernst F, Schweikard A. Forecasting respiratory motion with accurate online support vector regression. *Int J CARS*. 2009;4:439–447.
22. Omitaomu OA, Jeong MK, Badiru AB. Online support vector regression with varying parameters for time-dependent data. *IEEE Trans Syst Man Cybern A Syst Humans*. 2011;41:191–197.
23. Parrella F. Online Support Vector Regression. Software available at http://onlinesvr.altervista.org/ (accessed December 14, 2013).
24. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst*. 2001;58:109–130.
25. Kim S, Kano M, Nakagawa H, Hasebe S. Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. *Int J Pharm*. 2011;421: 269–274.
26. Li G, Aute V, Azarm S. An accumulative error based adaptive design of experiments for offline metamodeling. *Struct Multidiscip O*. 2010;40:137–155.
27. McAuley KB, MacGregor JF. On-line inference of polymer properties in an industrial polyethylene reactor. *AIChE J*. 1991;37:825–835.
28. Ohshima M, Tanigaki M. Quality control of polymer production processes. *J Process Control*. 2000;10:135–148.
29. Lee EH, Kim TY, Yeo, YK. Prediction of the melt index in a high-density polyethylene process. *J Chem Eng Jpn*. 2007;40:840–846.
30. Oh SJ, Lee J, Park S. Prediction of pellet properties for an industrial bimodal high-density polyethylene process with Ziegler−Natta catalysts. *Ind Eng Chem Res*. 2005;44:8–20.
31. Kaneko H, Arakawa M, Funatsu K. Novel soft sensor method for detecting completion of transition in industrial polymer processes. *Comput Chem Eng*. 2011;35:1135–1142.

Manuscript received Feb. 14, 2013, and revision received Sept. 20, 2013.