CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Kyoto University Research Information Repository
KURENAI
Kyoto University Research Information Repository
KYOTO UNIVERSITY

Title	Covariance-based locally weighted partial least squares for high-performance adaptive modeling
Author(s)	Hazama, Koji; Kano, Manabu
Citation	Chemometrics and Intelligent Laboratory Systems (2015), 146: 55-62
Issue Date	2015-08-15
URL	http://hdl.handle.net/2433/200670
Right	© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Type	Journal Article
Textversion	publisher

Kyoto University

Chemometrics and Intelligent Laboratory Systems 146 (2015) 55–62

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Covariance-based locally weighted partial least squares for high-performance adaptive modeling

Koji Hazama [1], Manabu Kano *

Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan

ARTICLE INFO

Article history:
Received 19 March 2015
Received in revised form 4 May 2015
Accepted 5 May 2015
Available online 13 May 2015

Keywords:
Just-in-time modeling
Locally weighted partial least squares
Soft-sensor
Process analytical technology
Calibration

ABSTRACT

Locally weighted partial least squares (LW-PLS) is one of Just-in-Time (JIT) modeling methods; PLS is used to build a local linear regression model every time when output variables need to be estimated. The prediction accuracy of local models strongly depends on the definition of similarity between a newly obtained sample and past samples stored in a database. To calculate the similarity, the Euclidean distance and the Mahalanobis distance have been widely used, but they do not take account of the relationship between input and output variables. This fact limits the achievable performance of LW-PLS and other locally weight regression methods. Thus, in the present work, covariance-based locally weighted PLS (CbLW-PLS) is proposed by integrating LW-PLS and a new similarity index based on the covariance between input and output variables. CbLW-PLS was applied to two industrial problems: soft-sensor design for estimating unreacted NaOH concentration in an alkali washing tower in a petrochemical process, and process analytical technology (PAT) for estimating concentration of a residual drug substance in a pharmaceutical process. The proposed similarity index was compared with six conventional indexes based on distances, correlations, or regression coefficients. The results have demonstrated that CbLW-PLS achieved the best prediction performance of all in both case studies.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Real-time monitoring and control of product quality are difficult in most manufacturing processes because product quality is not always measured in real time. On the other hand, the number of measured variables and the amount of data stored in databases are rapidly increasing. Such a situation has motivated us to predict the difficult-to-measure product quality from easily measurable process variables and to use predicted values instead of measurements for real-time monitoring and control. In other words, virtual sensing technology is crucial in predicting product quality or other important variables when online analyzers are not available [1]. Virtual sensing technology has been successfully applied to various processes in various industries. It is known as soft-sensors in the refinery/petrochemical industry, process analytical technology (PAT) in the pharmaceutical industry, and virtual metrology (VM) in the semiconductor industry.

In recent years, Just-in-Time (JIT) modeling has attracted a lot of attention in order to prevent deterioration of prediction accuracy due to changes in process characteristics and operating conditions. In fact, Kano and Ogawa reported in 2009 that the maintenance of models is the most critical issue concerning soft-sensors on the basis of the questionnaire survey of process control applications [2]. More than 30% of the engineers pointed out the necessity to cope with changes in process characteristics and operating conditions in order to keep the prediction performance of soft-sensors.

To cope with changes in process characteristics and operating conditions, various recursive methods have been proposed and their applications have been reported. A review of adaptation techniques was given by Kadlec et al. [3]. The concept drift theory was exploited to classify the algorithms into three different types: 1) moving windows techniques, 2) recursive adaptation techniques, and 3) ensemble-based methods. Recursive methods can adapt models to new operating conditions gradually, but the model may adapt excessively and not function in a sufficiently wide range of operating conditions when a process is operated within a narrow range for a certain period of time. An approach to prevent excessive recursive PLS update is minimizing the number of recursive PLS update runs while maintaining the model [4]. A more serious drawback of recursive methods is that they cannot cope with abrupt changes in process characteristics.

In such situations, JIT modeling is desirable. JIT modeling technique constructs a model every time when prediction is required so that it can adapt the model to time-varying process characteristics and operating conditions. It constructs a local model by weighting samples in a database according to the similarity between a newly obtained sample

* Corresponding author.
E-mail address: manabu@human.sys.i.kyoto-u.ac.jp (M. Kano).
[1] Present affiliation: A. T. Kearney, Tokyo, Japan.

http://dx.doi.org/10.1016/j.chemolab.2015.05.007
0169-7439/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(query) and past samples. The JIT modeling concept has been integrated with linear and nonlinear regression methods such as multiple regression analysis (MRA) [5] and support vector regression (SVR) [6,7]. JIT modeling and its industrial applications were recently surveyed by Kano and Fujiwara [1].

In particular, JIT modeling integrated with partial least squares (PLS) is called locally weighted PLS (LW-PLS) [8], which has been successfully applied to various industrial processes. In the pharmaceutical industry, for example, LW-PLS has been applied to estimation of active pharmaceutical ingredients (API) content with near infrared (NIR) spectroscopy [8], estimation of the amount of residual drug substances in cleaning processes with infrared-reflection absorption spectroscopy (IR-RAS) [9], and NIR-based real-time monitoring of ingredient concentration during blending [10]. Other applications of LW-PLS include inferential control of product quality in the petrochemical industry [11], maize hardness characterization in the food industry [12], VM in the semiconductor industry [13], and determination of clinical parameters in human serum samples with Fourier transform infrared (FTIR) spectroscopy [14]. In addition, several updating strategies including LW-PLS were compared in the prediction accuracy by using an NIR dataset of gasoline [15]. Furthermore, LW-PLS algorithm was extended to improve the estimation performance or to cope with different problems. Such extension includes locally weighted partial least squares-discriminant analysis (LW-PLS-DA) for non-linear classification [16] and a Bayesian framework providing a systematic way for real-time parameterization of the similarity function, selection of the local PLS model structure, and estimation of the corresponding model parameters [17].

The definition of similarity plays a crucial role in improving the prediction accuracy of JIT modeling technique including LW-PLS. Similarity indexes are usually defined on the basis of the Euclidean distance or the Mahalanobis distance [8,18]. Other similarity indexes take account of the angles between a query and samples in a database [6,19]. In addition, the prediction accuracy can be significantly improved by using the similarity index based on the weighted distance, whose weights are derived from physical properties of target material [9]. There have been various works that calculate the weighted distance based on the regression coefficients of MRA, PLS, and LW-PLS [5,20]. However, these methods require constructing a regression model in advance to calculate the similarity index, therefore the computational load is heavy.

In the present work, focusing on LW-PLS, we propose a new similarity index that takes account of the relationships both among input variables and among input and output variables with suppressing an increase in computational load. The proposed method is referred to as covariance-based LW-PLS (CbLW-PLS). Case studies are conducted through two different operation data in real plants to compare the proposed similarity index with other similarity indexes in the prediction performance of LW-PLS.

This paper is organized as follows: LW-PLS is described in Section 2, and the new similarity index is proposed in Section 3 The distribution of each similarity index is visualized through a numerical experiment in Section 4. The case studies are shown in Section 5 to demonstrate the effectiveness of the proposed method. Finally, the conclusion is given in Section 6.

2. Locally weighted partial least squares (LW-PLS)

In this section, PLS and LW-PLS are briefly explained.

2.1. Partial least squares (PLS)

In general, PLS is preferable to multiple regression or ordinary least squares (OLS) when a linear regression model is built from process data, because PLS can deal with multicollinearity that prevents from obtaining a reliable model by using OLS. Multicollinearity appears in a situation where input variables are nearly or completely linearly dependent; such a situation is common in process data analysis. To address this issue, PLS derives latent variables as linear combinations of input variables and uses them to predict output variables.

Suppose data of input variables and an output variable are given as $\mathbf{X} \in \Re^{N \times M}$ and $\mathbf{y} \in \Re^{N}$. These variables are mean-centered and properly scaled, e.g. normalized. A PLS model with K latent variables is expressed as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

$$\mathbf{y} = \mathbf{T}\mathbf{q} + \mathbf{f} \tag{2}$$

where $\mathbf{T} \in \Re^{N \times K}$ is a score matrix consisting of latent variables $\mathbf{t}_k \in \Re^{N}$ ($k = 1, 2, \ldots, K$), $\mathbf{P} \in \Re^{M \times K}$ consisting of $\mathbf{p}_k \in \Re^{M}$ is a loading matrix of $\mathbf{X}$, $\mathbf{q} \in \Re^{K}$ is a regression coefficient vector from latent variables to the output variable, and $\mathbf{E}$ and $\mathbf{f}$ are residuals.

In PLS, the model is constructed in an iterative manner through the NIPALS algorithm [21]. After $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{y}_1 = \mathbf{y}$ are set, the variable matrices at the k th iteration ($k \geq 2$) are written as

$$\mathbf{X}_k = \mathbf{X}_{k-1} - \mathbf{t}_{k-1}\mathbf{p}_{k-1}^{\mathrm{T}} \tag{3}$$

$$\mathbf{y}_k = \mathbf{y}_{k-1} - \mathbf{t}_{k-1}q_{k-1} \ . \tag{4}$$

The k th latent variable $\mathbf{t}_k$ is expressed as

$$\mathbf{t}_k = \mathbf{X}_k\mathbf{w}_k \tag{5}$$

where the k th weighting vector $\mathbf{w}_k$, the kth column of the weighting matrix $\mathbf{W}$, is determined so that the inner product between $\mathbf{t}_k$ and $\mathbf{y}_k$ is maximized under the constraint $\|\mathbf{w}_k\| = 1$. The Lagrange multiplier method enables us to derive $\mathbf{w}_k$, $\mathbf{p}_k$, and q_k as follows.

$$\mathbf{w}_k = \frac{\mathbf{X}_k^{\mathrm{T}}\mathbf{y}_k}{\left\|\mathbf{X}_k^{\mathrm{T}}\mathbf{y}_k\right\|} \tag{6}$$

$$\mathbf{p}_k = \frac{\mathbf{X}_k^{\mathrm{T}}\mathbf{t}_k}{\mathbf{t}_k^{\mathrm{T}}\mathbf{t}_k} \tag{7}$$

$$q_k = \frac{\mathbf{y}_k^{\mathrm{T}}\mathbf{t}_k}{\mathbf{t}_k^{\mathrm{T}}\mathbf{t}_k}. \tag{8}$$

This procedure is repeated until k reaches the number of adopted latent variables K. This PLS algorithm is knows as PLS1 because the number of output variables is one; PLS2 is available when multiple output variables need to be predicted simultaneously.

2.2. Locally weighted partial least squares (LW-PLS)

LW-PLS is a JIT modeling method that constructs a local regression model according to the similarity between a query (target sample) and past samples stored in a database [8]. It has attracted much attention as a tool for virtual sensing since it can cope with changes in operating conditions and process characteristics.

Here the algorithm of LW-PLS is explained. $\{x_{nm}\}$ and $\{y_{nl}\}$ ($n = 1, 2, \ldots, N$; $m = 1, 2, \ldots, M$; $l = 1, 2, \ldots, L$) are preprocessed measurements of input and output variables, where M and L are the numbers of input and output variables, respectively. As the preprocess, an adequate scaling is necessary to achieve high prediction performance. The same preprocess should be applied both to samples in the database and to the query. The n th sample is expressed as

$$\mathbf{x}_n = [x_{n1}, x_{n2}, \ldots, x_{nM}]^{\mathrm{T}} \tag{9}$$

$$\mathbf{y}_n = [y_{n1}, y_{n2}, \ldots, y_{nL}]^{\mathrm{T}} \ . \tag{10}$$

The input and output variable matrices $\mathbf{X}$ and $\mathbf{Y}$ consist of these vectors. In LW-PLS, $\mathbf{X}$ and $\mathbf{Y}$ are stored in a database, and the similarity index ω_n between a query $\mathbf{x}_q$ and the nth sample $\mathbf{x}_n$ is calculated to construct a local PLS model when an output prediction is required for $\mathbf{x}_q$. The predicted output $\hat{\mathbf{y}}_q$ is calculated through the following procedure.

1. Determine the number of latent variables K and set $k = 1$.
2. Calculate a similarity matrix $\mathbf{\Omega}$.

$$\mathbf{\Omega} = \mathrm{diag}\{\omega_1, \omega_2, \ldots, \omega_N\} \tag{11}$$

where diag{} denotes a diagonal matrix. The details are mentioned below.

3. Calculate $\mathbf{X}_k$, $\mathbf{Y}_k$, and $\mathbf{x}_{q,k}$

$$\mathbf{X}_k = \mathbf{X} - \mathbf{1}_N[\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M] \tag{12}$$

$$\mathbf{Y}_k = \mathbf{Y} - \mathbf{1}_N[\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_L] \tag{13}$$

$$\mathbf{x}_{q,k} = \mathbf{x}_q - [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M]^{\mathrm{T}} \tag{14}$$

$$\bar{x}_m = \frac{\sum_{n=1}^{N} \omega_n x_{\mathrm{nm}}}{\sum_{n=1}^{N} \omega_n} \tag{15}$$

$$\bar{y}_l = \frac{\sum_{n=1}^{N} \omega_n y_{\mathrm{nl}}}{\sum_{n=1}^{N} \omega_n} \tag{16}$$

where $\mathbf{1}_N \in \Re^N$ is a vector of ones.

4. Set $\hat{\mathbf{y}}_q = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_L]^{\mathrm{T}}$.
5. Derive the kth latent variable of $\mathbf{X}_k$:

$$\mathbf{t}_k = \mathbf{X}_k \mathbf{w}_k \tag{17}$$

where $\mathbf{w}_k$ is the eigenvector of $\mathbf{X}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{Y}_k\mathbf{Y}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{X}_k$, which corresponds to the maximum eigenvalue, and it is derived by

$$\mathbf{w}_k = \frac{\mathbf{X}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{Y}_k}{\left\|\mathbf{X}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{Y}_k\right\|} \quad . \tag{18}$$

6. Derive the kth loading vector of $\mathbf{X}_k$ and the k th regression coefficient vector.

$$\mathbf{p}_k = \frac{\mathbf{X}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{t}_k}{\mathbf{t}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{t}_k} \tag{19}$$

$$\mathbf{q}_k = \frac{\mathbf{Y}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{t}_k}{\mathbf{t}_k^{\mathrm{T}}\mathbf{\Omega}\mathbf{t}_k} \tag{20}$$

7. Derive the kth latent variable of $\mathbf{x}_q$.

$$t_{q,k} = \mathbf{x}_{q,k}^{\mathrm{T}}\mathbf{w}_k \tag{21}$$

8. Replace $\hat{\mathbf{y}}_q$ with $\hat{\mathbf{y}}_q + t_{q,k}\mathbf{q}_k$.
9. If $k = K$, then finish prediction. Otherwise, set

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{t}_k\mathbf{p}_k^{\mathrm{T}} \tag{22}$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k - \mathbf{t}_k\mathbf{q}_k^{\mathrm{T}} \tag{23}$$

$$\mathbf{x}_{q,k+1} = \mathbf{x}_{q,k} - t_{q,k}\mathbf{p}_k \quad . \tag{24}$$

10. Set $k = k + 1$ and go to step 5.

The definition of similarity significantly affects the prediction performance of LW-PLS. The similarity index ω_n is defined as

$$\omega_n = \exp\left(-\frac{\varphi d_n}{\sigma_d}\right) \tag{25}$$

$$d_n = \sqrt{(\mathbf{x}_n - \mathbf{x}_q)^{\mathrm{T}}\mathbf{\Theta}(\mathbf{x}_n - \mathbf{x}_q)} \tag{26}$$

$$\mathbf{\Theta} = \mathrm{diag}\{\theta_1, \theta_2, \ldots, \theta_M\} \tag{27}$$

where φ is a localization parameter, σ_d is the standard deviation of $\{d_n\}$, $\mathbf{\Theta}$ is a weighting matrix, and θ_m is a weight of the m th input variable. The weighting matrix $\mathbf{\Theta}$ does not have to be diagonal, but it is diagonal in most literatures. It is clear from Eq. (25) that LW-PLS is equivalent to PLS when $\phi = 0$ and $\omega_n = 1$. Thus, selecting the localization parameter properly can lead to constructing a LW-PLS model with higher accuracy than or at least the same accuracy as PLS. The localization parameter can be determined through cross-validation. The optimal value of ϕ is usually found in the range of 0 to 10.

Although an identity matrix is often used as $\mathbf{\Theta}$, the prediction performance of LW-PLS can be improved by determining $\mathbf{\Theta}$ in a way that the relationships among input variables and among input and output variables are taken into consideration. This fact motivates us to propose covariance-based LW-PLS (CbLW-PLS).

3. Covariance-based locally weighted partial least squares (CbLW-PLS)

In this section, a new similarity index, into which the relationships among input variables and among input and output variables are incorporated, is proposed to improve the prediction performance of LW-PLS. The proposed modeling method is called covariance-based LW-PLS (CbLW-PLS).

The new similarity index is calculated through Eqs. (25) and (26) so that the weighting matrix $\mathbf{\Theta}$ has the desired properties, that is, the relationships among input variables and among input and output variables are taken into account simultaneously. Given $\mathbf{\Theta} = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}$ with a matrix $\mathbf{\Gamma} \in \Re^{M \times \gamma}$ ($\gamma \le M$), Eq. (26) is rewritten as

$$d_n = \sqrt{(\mathbf{x}_n - \mathbf{x}_q)^{\mathrm{T}}\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}}(\mathbf{x}_n - \mathbf{x}_q)} \tag{28}$$

where the Euclidean distance is computed after linear transformation of samples $\{\mathbf{x}_n\}$ and $\mathbf{x}_q$ with $\mathbf{\Gamma}^{\mathrm{T}}$.

3.1. Conventional weighting

When $\mathbf{\Gamma}$ is a square matrix, i.e., $\gamma = M$, the similarity index is calculated on the M dimensional input variable space. When the M dimensional identity matrix $\mathbf{I}_M$ is employed as $\mathbf{\Gamma}$, $\mathbf{\Theta}$ is also an identity matrix; $\mathbf{\Theta} = \mathbf{I}_M$. In this case, d_n is the Euclidean distance on the original input variable space; in other words, the coordinates are not transformed.

In the case where an weight θ_m (≥ 0) is used for each variable, $\mathbf{\Gamma}$ is a diagonal matrix, whose diagonal elements are $\{\sqrt{\theta_m}\}$. This weighting is regarded as scaling of each input variable. Kim et al. investigated and compared various scaling methods in the prediction performance of LW-PLS [22]. For example, absolute values of regression coefficients of PLS or LW-PLS are used as $\{\theta_m\}$.

$$\mathbf{\Theta} = \mathrm{diag}\{|b_1|, |b_2|, \ldots, |b_M|\} \tag{29}$$

where $\mathbf{b} = [b_1, b_2, \ldots, b_M]^{\mathrm{T}}$ is a regression coefficient vector of PLS, LW-PLS, or another model. This method incorporates the relationships among input variables and among input and output variables into the similarity index. However, computational load is heavy, since a model has to be constructed in advance of similarity index calculation;

modeling requires to determine parameters such as the number of latent variables and a localization parameter.

Another well-known distance is the Mahalanobis distance, which takes account of the covariance of input variables by using the inverse of a covariance matrix of input variables as $\boldsymbol{\Theta}$.

$$\boldsymbol{\Theta} = \mathbf{S}_X^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^{\mathrm{T}} \tag{30}$$

Here the eigenvalue decomposition is applied to the covariance matrix $\mathbf{S}_X$:

$$\mathbf{S}_X = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}} \tag{31}$$

where $\mathbf{P}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix. The Mahalanobis distance is interpreted as the distance derived by linear transformation of a sample with $\boldsymbol{\Gamma}^{\mathrm{T}} = \boldsymbol{\Lambda}^{-1/2}\mathbf{P}^{\mathrm{T}}$. In other words, it is the distance computed after the coordinate rotation with the rotation matix $\mathbf{P}^{\mathrm{T}}$ and then the scaling with the diagonal matrix $\boldsymbol{\Lambda}^{-1/2}$. This procedure is associated with principal component analysis (PCA). $\mathbf{P}^{\mathrm{T}}$ represents linear transformation to perform rotation to the space spanned by principal components (PCs), and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the variances of PCs. Thus, the Mahalanobis distance is the distance normalized by using the standard deviations of PCs on the principal component space.

By employing a non-diagonal matrix such as $\mathbf{S}_X^{-1}$ as $\boldsymbol{\Theta}$, the relationship among input variables is incorporated into the similarity index. Another simple approach is to use a covariance matrix as a weighting matrix, i.e., $\boldsymbol{\Theta} = \mathbf{S}_X$. This weighting method is just the opposite of the Mahalanobis distance. In either case, the relationship between input variables and an output variable is not taken into account.

Correlation coefficients between input variables and an output variable can be used as weights $\{\theta_m\}$.

$$\boldsymbol{\Theta} = \mathrm{diag}\left\{\frac{|\mathbf{x}_1^{\mathrm{T}}\mathbf{y}|}{\|\mathbf{x}_1\|\|\mathbf{y}\|}, \frac{|\mathbf{x}_2^{\mathrm{T}}\mathbf{y}|}{\|\mathbf{x}_2\|\|\mathbf{y}\|}, \ldots, \frac{|\mathbf{x}_M^{\mathrm{T}}\mathbf{y}|}{\|\mathbf{x}_M\|\|\mathbf{y}\|}\right\} \tag{32}$$

where the mth input variable $\mathbf{x}_m$ and the output variable $\mathbf{y}$ are mean-centered. In this case, the relationship among input variables is not taken into account. The weighting matrices introduced here are summarized in Table 1.

3.2. New weighting

To construct a local regression model with high prediction accuracy, the similarity should be defined by evaluating both the relationship among input variables and the relationship between input and output variables. Moreover, the computational load should be reduced. From these viewpoints, the present work proposes a new similarity index, which takes account of both relationships and is easy to derive.

Table 1
Comparison of weighting matrices to calculate similarity indexes, which are integrated with LW-PLS. Weights are based on Euclidean distance, Mahalanobis distance, covariance among input variables S_X, correlation coefficients between input and output variables r_{xy}, regression coefficients of PLS and LW-PLS, and covariance among input and output variables.

Method	Weighting matrix $\boldsymbol{\Theta}$
Euclidean	$\mathbf{I}_M = \mathrm{diag}\{1, 1, \ldots, 1\}$
Mahalanobis	$\mathbf{S}_X^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^{\mathrm{T}}$
S_X	$\mathbf{S}_X = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}$
r_{xy}	$\mathrm{diag}\left\{\frac{\lvert\mathbf{x}_1^{\mathrm{T}}\mathbf{y}\rvert}{\|\mathbf{x}_1\|\|\mathbf{y}\|}, \frac{\lvert\mathbf{x}_2^{\mathrm{T}}\mathbf{y}\rvert}{\|\mathbf{x}_2\|\|\mathbf{y}\|}, \ldots, \frac{\lvert\mathbf{x}_M^{\mathrm{T}}\mathbf{y}\rvert}{\|\mathbf{x}_M\|\|\mathbf{y}\|}\right\}$
PLS	$\mathrm{diag}\{\lvert b_1\rvert, \lvert b_2\rvert, \ldots, \lvert b_M\rvert\}$
LW-PLS	$\mathrm{diag}\{\lvert b_1\rvert, \lvert b_2\rvert, \ldots, \lvert b_M\rvert\}$
Covariance	$\frac{\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{X}}{\|\mathbf{X}^{\mathrm{T}}\mathbf{y}\|^2}$

The proposed weighting method is based on the covariance between input variables and an output variable. Given $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]$ and $\mathbf{y}$, the weighting matrix $\boldsymbol{\Theta}$ is expressed as

$$\boldsymbol{\Gamma} = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{y}}{\left\|\mathbf{X}^{\mathrm{T}}\mathbf{y}\right\|} \tag{33}$$

$$\boldsymbol{\Theta} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{X}}{\left\|\mathbf{X}^{\mathrm{T}}\mathbf{y}\right\|^2} \tag{34}$$

where $\boldsymbol{\Gamma}$ is equivalent to the first weighting vector $\mathbf{w}_1$ of PLS1. Both the proposed weighting based on covariance and the conventional weighting based on correlation coefficients use the inner product $\mathbf{x}_m^{\mathrm{T}}\mathbf{y}$ to define the weighting matrix, but they are significantly different. The conventional correlation-based weighting matrix is diagonal; it scales each input variable separately. This means that the correlation-based weighting does not take account of the relationship among input variables explicitly. On the other hand, the covariance-based weighting matrix is non-diagonal; the covariance-based weighting takes account of the relationship among input variables as well as the relationship between input variables and the output variable. In addition, in the covariance-based weighting method, the distance is computed after projecting samples $\{\mathbf{x}_n\}$ and $\mathbf{x}_q$ onto one dimensional space with $\boldsymbol{\Gamma}^T$.

Another advantage of the covariance-based weighting method over conventional weighting methods is its light computational load. In conventional methods, to derive weights by taking account of both the relationship among input variables and the relationship between input and output variables, the weights are determined on the basis of regression coefficients of PLS or LW-PLS. These approaches are time-consuming, because they need to build one or more PLS models to derive the weights. On the other hand, the proposed CbLW-PLS does not require any model to calculate weights even though it takes account of both the relationship among input variables and the relationship between input and output variables. Among various weighting matrices listed in Table 1, the computational load of CbLW-PLS is similar to those of other methods, in which weights based on Euclidean distance, Mahalanobis distance, covariance among input variables S_X, or correlation coefficients between input and output variables r_{xy} are used. The average calculation time of output prediction for each query depends on the size of database, and it is usually less than 10 ms from the authors' experience.

3.3. CbLW-PLS

The proposed modeling method, LW-PLS with the covariance-based similarity index, is referred to as covariance-based locally weighted partial least squares (CbLW-PLS). The algorithm of CbLW-PLS is as follows:

1. Prepare an input variable matrix $\mathbf{X}$ and an output variable vector $\mathbf{y}$, where all variables are mean-centered and appropriately scaled.
2. Calculate the weighting matrix $\boldsymbol{\Theta}$ through Eq. (34).
3. Calculate the similarity index ω_n according to Eqs. (25) and (26).
4. Predict the output variable based on the LW-PLS algorithm.

4. Visualization of similarity

This section aims to visualize various similarity indexes listed in Table 1 through a numerical example. Given a query, the distribution of each similarity index between the query and samples in a database is displayed on a two dimensional space. 1000 samples of input variables $\{x_m\}$ $(m = 1, 2, 3, 4)$ were generated through the following equations and stored in the database.

$$x_1 = u_1 \tag{35}$$

$$x_2 = u_1 + 0.2u_2 \tag{36}$$

$$x_3 = 3(u_1 - 0.5)^3 + 0.2u_3 \tag{37}$$

$$x_4 = u_4 \tag{38}$$

$$u_m \sim \mathcal{U}(0,\ 1) \tag{39}$$

where $\mathcal{U}(a,b)$ is the Uniform distribution from a to b. The output variable y was related to the first principal component (PC1) t of $\{x_m + v_m\}$.

$$y = 3t^3 + v_0 \tag{40}$$

$$t = \text{PC1 of } (x_m + v_m) \tag{41}$$

$$v_0 \sim \mathcal{N}\left(0, 0.01^2\right) \tag{42}$$

$$v_m \sim \mathcal{N}\left(0, 0.02^2\right) \tag{43}$$

where $\mathcal{N}(\mu,\sigma^2)$ denotes the Normal distribution with mean μ and variance σ^2. 1000 samples of y were stored in the database together with those of $\{x_m\}$. The samples are visualized in Fig. 1, where blue small points are samples in the database and the red large point is the query.

The similarity indexes were calculated after all variables were normalized, i.e., their mean and variance are 0 and 1. The distributions of the similarity indexes are displayed on two dimensional spaces in Figs. 2 and 3.

In Fig. 2, the similarity index based on the Euclidean distance is distributed concentrically regardless of the relationship among input variables. The similarity index based on the Mahalanobis distance is adjusted according to the covariance of input variables. The similarity index based on the covariance of input variables S_X is distributed at right angles to that of the Mahalanobis distance.

The $x_1 - x_4$ plot at the upper right of Fig. 3 is representative of the similarity index based on the absolute values of the correlation coefficients among input and output variables r_{xy}. This plot shows that x_1 is more similar to y than x_4, because the similarity index depends more greatly on x_1 than x_4. In fact, x_1, which is correlated with x_2 and x_3, has stronger correlation with y than x_4. The similarity index based on the absolute values of the regression coefficients of PLS is similar to that of r_{xy}, because weighting is performed according to the contribution of each input variable to the output variable in both methods. The similarity index based on the absolute values of the regression coefficients of LW-PLS is different from that of PLS, because the LW-PLS model is local while the PLS model is global. Finally, the proposed covariance-based similarity index is shown at the bottom of Fig. 3. Since this similarity index is calculated on one dimensional space, it is distributed in one direction. The distribution is different from the others. This affects the prediction performance of LW-PLS as shown in the next section.

5. Industrial case studies

This section aims to compare a variety of LW-PLS with the similarity indexes introduced in Section 4. To validate the practicability of the proposed CbLW-PLS, the modeling methods were applied to two industrial case studies: 1) soft-sensor for estimating unreacted NaOH concentration in an alkali washing tower in a petrochemical process, and 2) process analytical technology (PAT) for estimating the concentration of a residual drug substance in a pharmaceutical process.

When the regression coefficients of PLS or LW-PLS are employed for calculating the similarity index, the parameters such as the number of latent variables or a localization parameter have to be determined in advance. In this work, optimal parameters were selected through leave-one-out cross validation (LOOCV), and the regression coefficients of constructed models were used for similarity index calculation.

5.1. Petrochemical process

The first case study focuses on soft-sensor design in the petrochemical process. The objective of the soft-sensor is to predict the concentration of unreacted NaOH in an alkali washing tower.

The alkali washing tower removes impurities, H_2S and CO_2, from cracked gas generated in the ethylene plant. The cracked gas is fed to the tower from the bottom and NaOH is fed from the top. Both H_2S and CO_2 in the cracked gas are reacted with NaOH through the following chemical reactions.

$$H_2S + 2NaOH \rightarrow Na_2S + 2H_2O$$

$$CO_2 + 2NaOH \rightarrow Na_2CO_3 + H_2O$$

The formulated salts, Na_2S and Na_2CO_3, and unreacted NaOH come out of the tower from the bottom, while the washed cracked gas comes out from the top. To maximize the productivity and the profitability of this process, the concentration of unreacted NaOH should be as low as possible. Since the concentration is measured at the laboratory by analysing samples, that is, it is not measured in real time, the soft-sensor plays a crucial role in achieving real-time monitoring and feedback control.

As a soft-sensor design method, LW-PLS is preferable because it can cope with changes in process characteristics and enables engineers to avoid frequent model maintenance. In addition, this company has already developed an in-house software to design soft-sensors based on LW-PLS. Thus, developing a new soft-sensor based on LW-PLS is faster and lower-cost than building another model including a first-principle model.

A total of 11 process variables such as temperature, flow rate, and pressure were selected and used for soft-sensor design. To take process dynamics into account, not only current measurements but also past measurements were used as input variables. This type of modeling has been widely used in the process industry [23]. The number of past measurements used as input variables was determined for each process variable according to the process knowledge. As a result, the number of

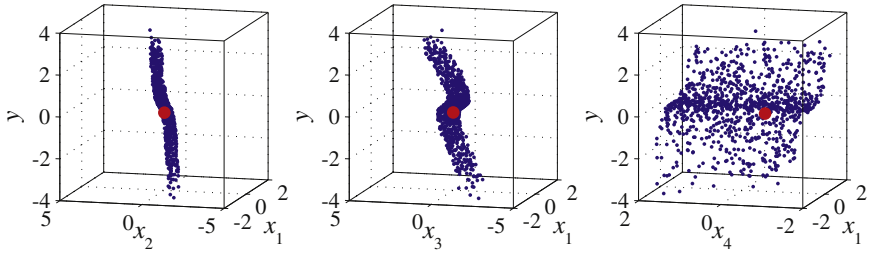

Fig. 1. Visualization of samples in the numerical example. One output variable y is nonlinearly correlated with four input variables x_1, x_2, x_3, and x_4.

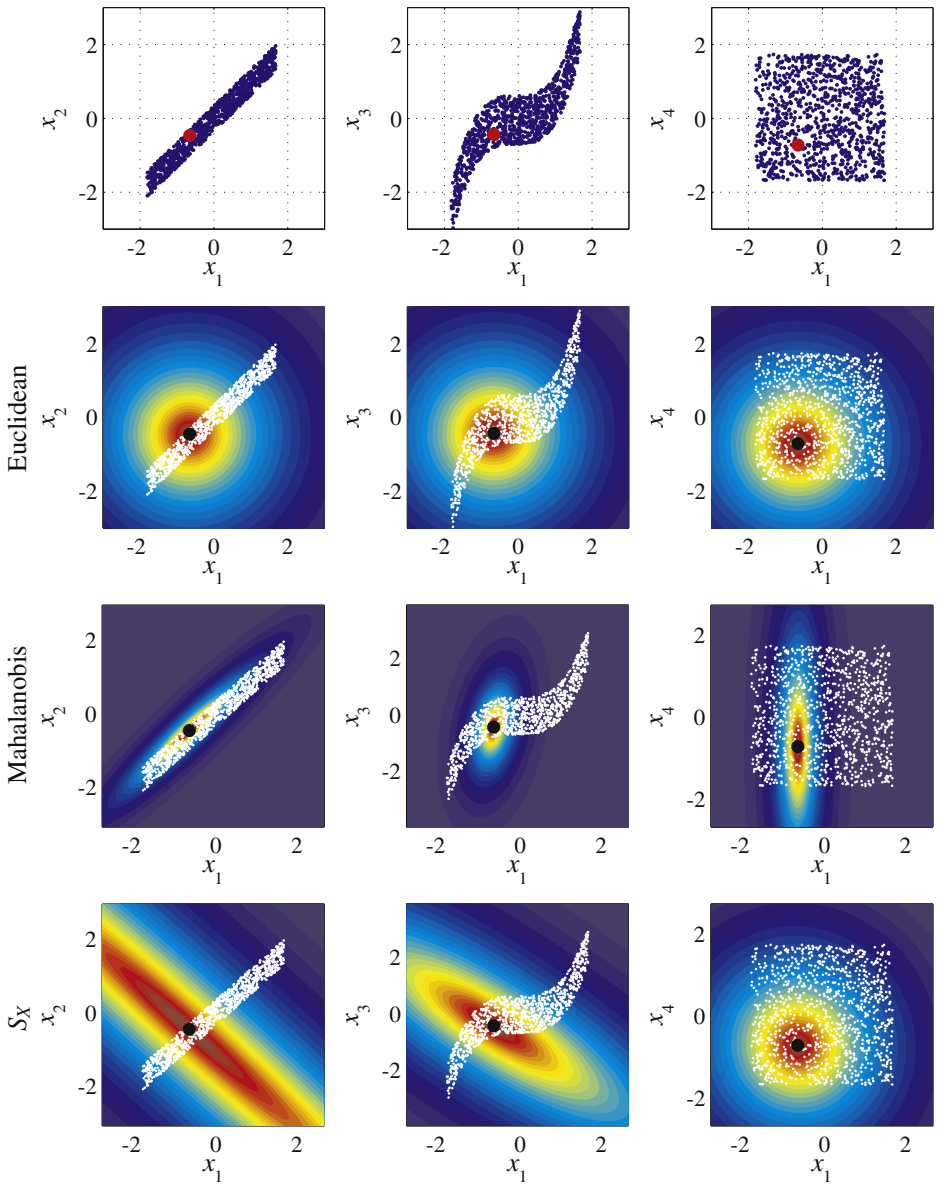

Fig. 2. Visualization of similarity indexes (1). The similarity index decreases as the color turns from red to blue. White points are samples in the database, and black ones are the query. (top) the relationships between input variables; blue small points are samples in the database, and red large points are the query. (upper middle) similarity index based on Euclidean distance. (lower middle) similarity index based on Mahalanobis distance. (bottom) similarity index based on covariance of input variables S_X.

input variables was determined to be 50 by experienced engineers and operators. The number of modeling samples was 211, and that of test samples was 141. Here, older samples were used as modeling samples, and newer ones were used as test samples.

The prediction results are summarized in Table 2, where #LV is the number of latent variables adopted in each LW-PLS model, ϕ is the localization parameter of LW-PLS, and the prediction accuracy is evaluated through the correlation coefficient between predicted values and reference values (R) and root mean square error of prediction (RMSEP). The results clearly show that the proposed CbLW-PLS, which combines LW-PLS with the covariance-based similarity index, achieved the best prediction performance of all. The selected parameters of LW-PLS, i.e., #LV and φ, of CbLW-PLS were close to those of LW-PLS with the similarity index based on Mahalanobis distance. Although Rs of both methods were the same, RMSEPs were different. In fact, RMSEP of CbLW-PLS was significantly smaller than the others.

In addition, the selected parameters and the prediction performance of LW-PLS with the similarity indexes based on the correlation coefficients r_{xy} and PLS regression coefficients were close to each other. The similarity indexes of these methods are based on the influence of each input variable on the output variable; therefore, the similar prediction accuracy was achieved. The similar results were observed in the other case study.

5.2. Pharmaceutical process

The second case study focuses on process analytical technology (PAT) in the pharmaceutical process. To minimize the risk of cross contamination, it is crucial to develop a rapid measurement method that enables continuous monitoring of the amount of residual drug substances in pharmaceutical manufacturing equipment after each cleaning. If it is detected with rapid measurement that the amount of residual drug substances is beyond the limit of acceptance criteria, the amount should be analyzed with a conventional method such as the swab method in more detail. Such a rapid measurement method will have the great advantage in mitigating the risk of the

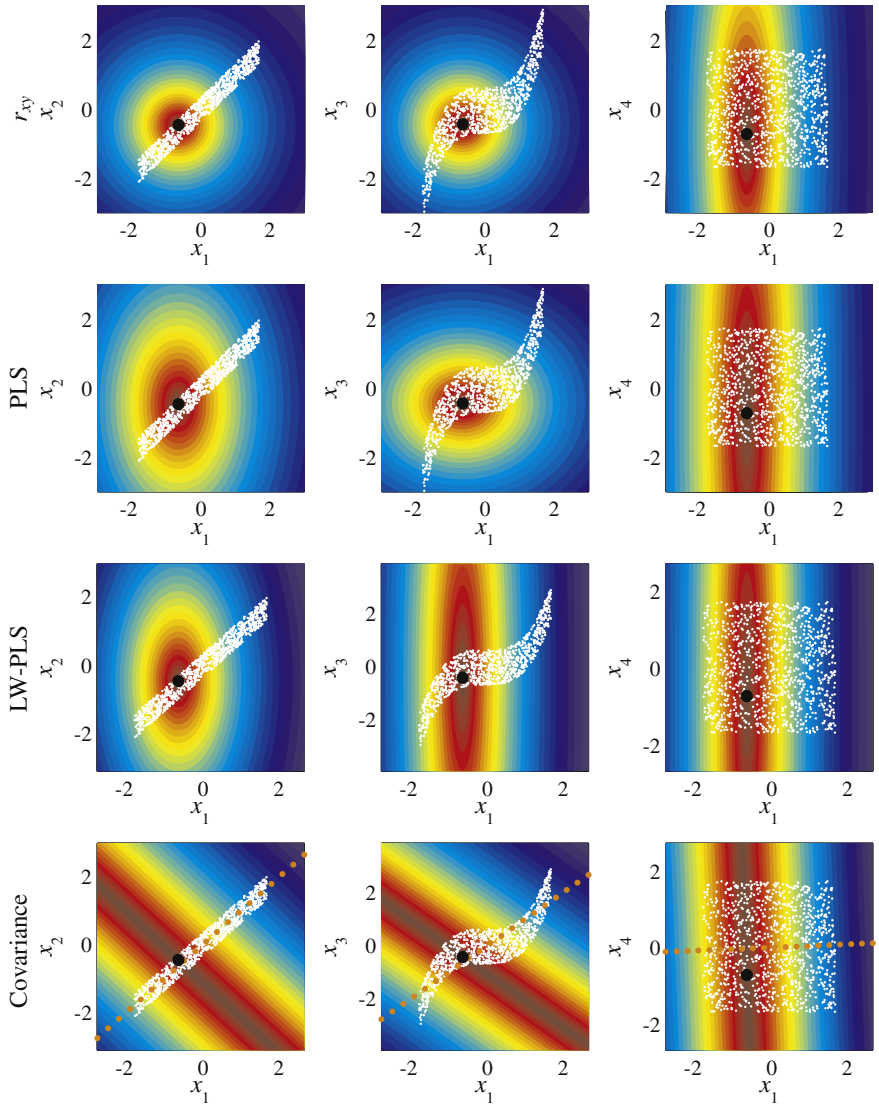

Fig. 3. Visualization of similarity indexes (2). The similarity index decreases as the color turns from red to blue. White points are samples in the database, and black ones are the query. (top) similarity index based on absolute values of correlation coefficients r_{xy}. (upper middle) similarity index based on absolute values of PLS regression coefficients. (lower middle) similarity index based on absolute values of LW-PLS regression coefficients. (bottom) the proposed covariance-based similarity index; the orange dotted line represents the first latent variable space.

cross contamination. In this way, rapid measurement methods of residual drug substances lead to paradigm shift of the cleaning process as PAT tools. To develop a rapid and accurate measurement method of residual drug substances without sampling, the use of infrared-reflection absorption spectroscopy (IR-RAS) and LW-PLS was evaluated [9].

The problem investigated in this case study is the same as that in [9]. The concentration of a residual drug substance, ibuprofen, was estimated by using the absorbance at 753 wavenumbers, which were measured by IR-RAS. The numbers of modeling, parameter determination, and test samples were 69, 53, and 63, respectively. The parameter determination samples were used to optimize the number of latent variables and the localization parameter. The test samples were used to evaluate the prediction accuracy of LW-PLS models constructed by using the modeling samples.

The prediction results summarized in Table 3 show that the calibration model constructed through CbLW-PLS achieved the best prediction

Table 2
Prediction results of LW-PLS with different similarity indexes (1): prediction of the unreacted NaOH concentration in an alkali washing tower in a petrochemical process.

Method	#LV	φ	R	RMSEP
Euclidean	6	1.5	0.856	1.003
Mahalanobis	9	0.8	0.870	0.940
S_X	6	1.6	0.866	0.931
r_{xy}	6	1.9	0.864	0.993
PLS	6	1.8	0.865	0.987
LW-PLS	6	1.5	0.862	1.002
Covariance	8	0.8	0.870	0.889

Table 3
Prediction results of LW-PLS with different similarity indexes (2): prediction of the concentration of a residual drug substance in a pharmaceutical process.

Method	#LV	φ	R	RMSEP
Euclidean	8	0.7	0.968	1.38
Mahalanobis	8	0.6	0.964	1.46
S_X	8	0.8	0.971	1.26
r_{xy}	8	0.7	0.971	1.29
PLS	8	0.9	0.974	1.24
LW-PLS	8	0.7	0.970	1.37
Covariance	8	0.7	0.973	1.17

performance of all. Although R of CbLW-PLS was slightly lower than that of LW-PLS with the similarity index based on PLS regression coefficients, CbLW-PLS outperformed all the others in RMSEP.

5.3. Discussion

In summary, the proposed method, CbLW-PLS, achieved the best performance in two case studies. In addition, CbLW-PLS is less time-consuming than the other methods, particularly the methods requiring model development in advance.

It should be also noticed that the selection of samples stored in the database have a huge effect on the prediction accuracy. Thus, appropriate selection of samples stored in the database is crucial in applications of LW-PLS. In general, samples should be stored in the database in descending order of their importance, which is evaluated from the three viewpoints: the newness, the (probability) density, and the nonlinearity. Newer samples are more important than older ones, samples become more important in the region where they are sparse, and more samples are needed to describe stronger nonlinearity. In the application of LW-PLS to a cracked gasoline (CGL) fractionator and a purification section of an acetyl plant [11], for example, dozens of newest samples were stored in the database to cope with recent changes in process characteristics, and hundreds of past samples were selected as core data, which were always stored in the database to prevent over-adaptation and cope with abrupt changes in process characteristics. It was reported that the use of core data was significantly useful to make LW-PLS soft-sensors robust. Further discussion on the database management is found in [11].

In addition, the treatment of missing data and outliers is important to achieve high and robust prediction performance regardless of modeling techniques. In conventional (non JIT) modeling methods, missing data are usually complemented by using other measurements on the basis of the correlation among variables. Such an approach is also available in JIT modeling. However, another easy-to-use approach may be adopted by taking account of the fact that a local model is built repeatedly in JIT modeling such as CbLW-PLS: to build a local model by excluding or ignoring the missing data (variables). Furthermore, it should be emphasized that LW-PLS including CbLW-PLS is robust against outliers because the weights on outliers usually become very small and the outliers are ignored when LW-PLS builds a local model, although any technique can be used for outlier detection.

6. Conclusions

To improve the prediction performance of LW-PLS, the new similarity index that takes account of the relationships among input variables as well as among input and output variables was proposed. The new LW-PLS based on the proposed similarity index is referred to as covariance-based locally weighted partial least squares (CbLW-PLS).

The advantages of CbLW-PLS over the other six methods were demonstrated through two industrial case studies: 1) soft-sensor design for estimating the unreacted NaOH concentration in the alkali washing tower in the petrochemical process, and 2) process analytical technology (PAT) for estimating the concentration of the residual drug substance in the pharmaceutical process. The similarity indexes compared with the proposed one were similarity indexes based on 1) Euclidean distance, 2) Mahalanobis distance, 3) weighted distance using covariance of input variables, 4) weighted distance using absolute values of correlation coefficients between input and output variables, 5) weighted distance using absolute values of regression coefficients of PLS, and 6) weighted distance using absolute values of regression coefficients of LW-PLS.

CbLW-PLS provided better prediction performance than the other methods in both case studies. Although it is inappropriate to conclude that CbLW-PLS is the best method in general, the present work succeeded to demonstrate the practicability of CbLW-PLS because it outperformed the other methods in different real problems in different industrial processes.

Acknowledgements

This study was partially supported by JSPS KAKENHI Grant Number 24560940. The authors greatly appreciate the support from Production Technology Department, Showa Denko K.K. and Pharmaceutical Technology Division, Daiichi Sankyo Co., Ltd.

References

[1] 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.
[2] M. Kano, M. Ogawa, The state of the art in chemical process control in Japan: good practice and questionnaire survey, J. Process Control 20 (2010) 969–982.
[3] P. Kadlec, R. Grbić, B. Gabrys, Review of adaptation mechanisms for data-driven soft sensors, Comput. Chem. Eng. 35 (2011) 1–24.
[4] F. Ahmed, S. Nazir, Y. Yeo, A new soft sensor based on recursive partial least squares for online melt index predictions in grade-changing HDPE operations, Chem. Prod. Process. Model. 4 (2009) 14–20.
[5] H. Shigemori, M. Kano, S. Hasebe, Optimum quality design system for steel products through locally weighted regression model, J. Process Control 21 (2011) 293–301.
[6] Z. Ge, Z. Song, Online monitoring of nonlinear multiple mode processes based on adaptive local model approach, Control. Eng. Pract. 16 (2008) 1427–1437.
[7] 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.
[8] 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.
[9] H. Nakagawa, T. Tajima, M. Kano, S. Kim, S. Hasebe, T. Suzuki, H. Nakagami, Evaluation of infrared-reflection absorption spectroscopy measurement and locally weighted partial least-squares for rapid analysis of residual drug substances in cleaning processes, Anal. Chem. 84 (2012) 3820–3826.
[10] H. Nakagawa, M. Kano, S. Hasebe, T. Miyano, T. Watanabe, N. Wakiyama, Verification of model development technique for NIR-based real-time monitoring of ingredient concentration during blending, Int. J. Pharm. 471 (2014) 264–275.
[11] 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.
[12] A. Guelpa, M. Bevilacqua, F. Marini, K. O'Kennedy, P. Geladi, M. Manley, Application of Rapid Visco Analyser (RVA) viscograms and chemometrics for maize hardness characterisation, Food Chem. 173 (2015) 1220–1227.
[13] T. Hirai, M. Kano, Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares, IEEE Trans. Semicond. Manuf. 28 (2015) 137–144.
[14] D. Perez-Guaita, J. Kuligowski, G. Quintás, S. Garrigues, M. Guardia, Modified locally weighted-partial least squares regression improving clinical predictions from infrared spectra of human serum samples, Talanta 107 (2015) 368–375.
[15] K. He, H. Cheng, W. Du, F. Qian, Online updating of NIR model and its industrial application via adaptive wavelength selection and local regression strategy, Chemom. Intell. Lab. Syst. 134 (2014) 79–88.
[16] M. Bevilacqua, F. Marini, Local classification: locally weighted-partial least squares-discriminant analysis (LW-PLS-DA), Anal. Chim. Acta 838 (2014) 20–30.
[17] M. Ma, S. Khatibisepehr, B. Huang, A Bayesian framework for real-time identification of locally weighted partial least squares, AIChE J. 61 (2015) 518–529.
[18] H. Leung, Y. Huang, C. Cao, Locally weighted regression for desulphurisation intelligent decision system modeling, Simul. Model. Pract. Theory 12 (2004) 413–423.
[19] C. Cheng, M.-S. Chiu, A new data-based methodology for nonlinear process modeling, Chem. Eng. Sci. 59 (2004) 2801–2810.
[20] S. Kim, R. Okajima, M. Kano, S. Hasebe, Development of soft-sensor using locally weighted PLS with adaptive similarity measure, Chemom. Intell. Lab. Syst. 124 (2013) 43–49.
[21] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[22] S. Kim, M. Kano, H. Nakagawa, S. Hasebe, Input variable scaling for statistical modeling, Comput. Chem. Eng. 74 (2015) 59–65.
[23] M. Kano, K. Miyazaki, S. Hasebe, I. Hashimoto, Inferential control system of distillation compositions using dynamic partial least squares regression, J. Process Control 10 (2000) 157–166.