

Title	Development of soft-sensor using locally weighted PLS with adaptive similarity measure
Author(s)	Kim, Sanghong; Okajima, Ryota; Kano, Manabu; Hasebe, Shinji
Citation	Chemometrics and Intelligent Laboratory Systems (2013), 124: 43-49
Issue Date	2013-05
URL	http://hdl.handle.net/2433/174103
Right	© 2013 Elsevier B.V.
Type	Journal Article
Textversion	author

Development of Soft-sensor Using Locally Weighted PLS with Adaptive Similarity Measure

Sanghong Kim^a, Ryota Okajima^{a,1}, Manabu Kano^{b,*}, Shinji Hasebe^a

^a*Dept. of Chemical Engineering, Kyoto University, Kyoto 6158510, Japan*

^b*Dept. of Systems Science, Kyoto University, Kyoto 6068501, Japan*

Abstract

Recently, just-in-time (JIT) modeling, such as locally weighted partial least squares (LW-PLS), has attracted much attention because it can cope with changes in process characteristics as well as nonlinearity. Since JIT modeling derives a local model from past samples similar to a query sample, it is crucial to appropriately define the similarity between samples. In this work, a new similarity measure based on the weighted Euclidean distance is proposed in order to cope with nonlinearity and to enhance estimation accuracy of LW-PLS. The proposed method can adaptively determine the similarity according to the strength of the nonlinearity between each input variable and an output variable around a query sample. The usefulness of the proposed method is demonstrated through numerical examples and a case study of a real cracked gasoline fractionator of an ethylene production process.

Keywords: Soft-sensor, Just-in-time model, Locally weighted partial least squares, Locally weighted regression, Distillation process

*Corresponding author. Tel.: +81-(0)75-753-3367; Fax: +81-(0)75-753-3371.

Email address: manabu@human.sys.i.kyoto-u.ac.jp (Manabu Kano)

¹Present address: Nippon Steel Corporation, Tokyo 1008071, Japan

1. Introduction

In various industrial processes, it is necessary to measure and control product quality to produce high-quality, competitive products. However, online measurement is not always available due to unacceptable expenses of analytical instruments or long measurement/analysis delay. To solve this problem, inferential models using online measured variables as predictor variables have been adopted in many fields such as the chemical, bioprocess, steel, and pharmaceutical [1, 2]. According to the recent questionnaire survey of process control in the chemical industry in Japan [3], 90% of the inferential models are constructed by using linear regression methods such as multiple regression analysis (MRA) and partial least squares (PLS). This fact shows that linear models are practically useful. In some cases, nonlinear models are required to achieve high estimation accuracy for processes having strong nonlinearity. Thus, nonlinear modeling methods such as neural networks [4–7], support vector regression [8–10] and polynomial functions [11–13] have been used to construct nonlinear inferential models.

The above-mentioned questionnaire survey revealed that the most important problem of current inferential models is how to cope with changes in process characteristics and keep high estimation accuracy for a long period of time, i.e., model maintenance [3]. The importance of this problem was also pointed out in [1, 14]. To cope with changes in process characteristics, many kinds of recursive modeling methods, which update models by prioritizing newer samples, have been developed [15]. When process characteristics change gradually, the prioritized samples are supposed to be similar to a query sample, for which an output estimation is required. For such a case, recursive methods can cope with gradual changes in process characteristics. However, they cannot cope with an abrupt

26 change in process characteristics caused by replacement of a catalyst, cleaning
27 of equipment, etc., because a query sampled just after an abrupt change becomes
28 significantly different from the prioritized samples.

29 Locally weighted regression (LWR) [16], which is also called just in time
30 learning, lazy learning or model-on-demand, constructs a local model by pri-
31 oritizing samples in a database according to the similarity between them and a
32 query sample. Hence, LWR can cope with abrupt changes as well as gradual
33 ones in contrast to the recursive methods introduced in [15]. Furthermore, it can
34 cope with nonlinearity since it builds a local model repeatedly. To build an ac-
35 curate model with LWR, the similarity needs to be properly defined. In general,
36 similarity is defined on the basis of the Euclidean distance or the Mahalanobis
37 distance [10, 17–22]. Other similarity measures proposed so far include the an-
38 gle [14, 24], the distance between an output estimate for a query sample derived
39 by a global model and output measurements for samples in a database [23, 25],
40 the correlation [26, 27] and the weighted Euclidean distance [28–30]. In addi-
41 tion to define the similarity properly, it is crucial to update a database when new
42 data become available in order to cope with changes in process characteristics.
43 More detailed explanation and review of the problem of the changes in process
44 characteristics and LWR can be found in [31].

45 This study focuses on the problem of nonlinearity and the definition of the sim-
46 ilarity, and does not deal with the problem of changes in process characteristics.
47 The similarity based on the weighted Euclidean distance is further investigated for
48 its simplicity. PLS is adopted for local modeling since it can cope with collinear-
49 ity and has been widely accepted in various fields. The main contribution of this
50 paper is to discuss how the weight of each input should be determined and to pro-

51 pose a method for deriving appropriate weights from operation data stored in a
 52 database.

53 The rest of this paper is organized as follows. In Section 2, the algorithm of
 54 locally weighted PLS (LW-PLS) is explained. Section 3 discusses how to deter-
 55 mine the weight of each input, and a method for deriving the appropriate weights
 56 from operation data is proposed. Section 4 shows the effectiveness of the pro-
 57 posed method through numerical examples. In Section 5, an application result of
 58 the proposed method to an industrial distillation process is reported. Finally, this
 59 research is concluded in Section 6.

60 2. Locally Weighted Partial Least Squares

61 The n th sample ($n = 1, 2, \dots, N$) of input and output variables is denoted by

$$\mathbf{x}_n = [x_{n1}, x_{n2}, \dots, x_{nM}]^T \quad (1)$$

$$\mathbf{y}_n = [y_{n1}, y_{n2}, \dots, y_{nL}]^T \quad (2)$$

62 where M is the number of input variables, L is the number of output variables and
 63 the superscript T denotes the transpose of a vector or matrix. $\mathbf{X} \in \Re^{N \times M}$ and
 64 $\mathbf{Y} \in \Re^{N \times L}$ are input and output variable matrices whose n th rows are \mathbf{x}_n^T and \mathbf{y}_n^T ,
 65 respectively. N is the number of samples.

66 LW-PLS is a just-in-time (JIT) modeling method. \mathbf{X} and \mathbf{Y} are stored in a
 67 database in order to construct a local PLS model. When an output estimation is
 68 required for a query sample \mathbf{x}_q , the similarity ω_n between \mathbf{x}_q and \mathbf{x}_n is calculated,
 69 and a local PLS model is constructed by weighting samples with a similarity ma-
 70 trix $\mathbf{\Omega} \in \Re^{N \times N}$ defined by

$$\mathbf{\Omega} = \text{diag}(\omega_1, \omega_2, \dots, \omega_N). \quad (3)$$

71 In general, the output estimate $\hat{\mathbf{y}}_q \in \mathfrak{R}^L$ is calculated through the following
72 procedure.

73 1. Determine the number of latent variables R and set $r = 1$.

74 2. Calculate the similarity matrix Ω .

75 3. Calculate \mathbf{X}_r , \mathbf{Y}_r and $\mathbf{x}_{q,r}$

$$\mathbf{X}_r = \mathbf{X} - \mathbf{1}_N[\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M] \quad (4)$$

$$\mathbf{Y}_r = \mathbf{Y} - \mathbf{1}_N[\bar{y}_1, \bar{y}_2, \dots, \bar{y}_L] \quad (5)$$

$$\mathbf{x}_{q,r} = \mathbf{x}_q - [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M]^T \quad (6)$$

$$\bar{x}_m = \frac{\sum_{n=1}^N \omega_n x_{nm}}{\sum_{n=1}^N \omega_n} \quad (7)$$

$$\bar{y}_l = \frac{\sum_{n=1}^N \omega_n y_{nl}}{\sum_{n=1}^N \omega_n} \quad (8)$$

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

77 4. Derive the r th latent variable of \mathbf{X}

$$\mathbf{t}_r = \mathbf{X}_r \mathbf{w}_r \quad (9)$$

78 where \mathbf{w}_r is the eigenvector of $\mathbf{X}_r^T \Omega \mathbf{Y}_r \mathbf{Y}_r^T \Omega \mathbf{X}_r$ which corresponds to the
79 maximum eigen value.

80 5. Derive the r th loading vector of \mathbf{X}

$$\mathbf{p}_r = \frac{\mathbf{X}_r^T \Omega \mathbf{t}_r}{\mathbf{t}_r^T \Omega \mathbf{t}_r} \quad (10)$$

81 and the regression coefficient vector

$$\mathbf{q}_r = \frac{\mathbf{Y}_r^T \Omega \mathbf{t}_r}{\mathbf{t}_r^T \Omega \mathbf{t}_r}. \quad (11)$$

82 6. Derive the r th latent variable of \mathbf{x}_q

$$t_{q,r} = \mathbf{x}_{q,r}^T \mathbf{w}_r . \quad (12)$$

83 7. If $r = R$, calculate the output estimate

$$\hat{\mathbf{y}}_q = [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_L]^T + \sum_{r=1}^R t_{q,r} \mathbf{q}_r \quad (13)$$

84 and finish estimation. Otherwise, set

$$\mathbf{X}_{r+1} = \mathbf{X}_r - \mathbf{t}_r \mathbf{p}_r^T \quad (14)$$

$$\mathbf{Y}_{r+1} = \mathbf{Y}_r - \mathbf{t}_r \mathbf{q}_r^T \quad (15)$$

$$\mathbf{x}_{q,r+1} = \mathbf{x}_{q,r} - t_{q,r} \mathbf{p}_r . \quad (16)$$

85 8. Set $r = r + 1$ and go to step 4.

86 When the similarity matrix Ω is an identity matrix, LW-PLS becomes the same
 87 as linear PLS. At step 3, the weighted mean of each variable is subtracted from
 88 each column of \mathbf{X} , \mathbf{Y} and \mathbf{x}_q^T to make the query sample near to the origin of
 89 the multidimensional space. At steps 4-8, the latent variable \mathbf{t} , the loading vector
 90 \mathbf{p} and the regression coefficient vector \mathbf{q} are derived iteratively, and the output
 91 estimate $\hat{\mathbf{y}}_q$ is calculated when $r = R$.

92 The definition of the similarity affects the estimation performance of LW-PLS
 93 significantly. In the original algorithm of LW-PLS [32], the similarity ω_n is de-
 94 fined on the basis of the Euclidean distance. The similarity based on the Euclidean
 95 distance or the Mahalanobis distance is used most frequently to construct a local
 96 regression model [10, 17–21, 32]. In addition, the estimation accuracy can be im-
 97 proved by using the similarity based on the weighted Euclidean distance [28–30].

98 In these researches, weights were defined on the basis of regression coefficients
 99 of a global MRA model or an LW-PLS model where all the weights were one.
 100 However, the conventional weighted Euclidean distance may deteriorate the esti-
 101 mation performance as described below. Hence, this research discusses how the
 102 weight of each input should be determined in Section 3.1 and proposes a method
 103 for deriving appropriate weights from operation data in Section 3.2.

104 **3. New Similarity Measure**

105 *3.1. How Should Weights Be Determined?*

106 In the present work, it is assumed that the number of output variables is one,
 107 and the following form of the similarity ω is investigated:

$$\omega_n = \exp\left(-\frac{d_n}{\sigma_d \varphi}\right) \quad (17)$$

$$d_n = \sqrt{(\mathbf{x}_n - \mathbf{x}_q)^T \Theta (\mathbf{x}_n - \mathbf{x}_q)} \quad (18)$$

$$\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_M) \quad (19)$$

108 where σ_d is a standard deviation of d_n ($n = 1, 2, \dots, N$) and φ is a localization
 109 parameter; the similarity decreases steeply when φ is small and gradually when φ
 110 is large. In addition, $\Theta \in \Re^{M \times M}$ is a weighting matrix and θ_m is a weight of the
 111 m th input variable.

112 Figure 1 shows simple examples, in which a relationship between a local linear
 113 model and weights θ_m ($m = 1, 2, 3$) is illustrated by using very small number of
 114 samples. In each figure, it is assumed that the values and the weights of the other
 115 inputs, and φ are constant. Relationship between input 1 and the output variable
 116 is linear as shown in Figure 1 (top). Large θ_1 causes overfitting by prioritizing
 117 samples 4 and 5; therefore, θ_1 should be small. On the contrary, the relationship

118 between input 2 and the output variable is nonlinear as shown in Figure 1 (middle).
119 Thus, θ_2 should be large to cope with nonlinearity between input 2 and the output
120 variable. In addition, the strength of nonlinearity around a query sample may
121 change depending on the value of the input variable as shown in Figure 1 (bottom).
122 In this case, θ_3 should be large for query 1 and small for query 2.

123 The weights proposed in [28–30] do not necessarily correspond to the strength
124 of nonlinearity around a query sample. For example, a regression coefficient of
125 an input can be large even when the input-output relationship is linear. In such a
126 case, the large weight might cause a deterioration of the estimation performance
127 as shown in Figure 1 (top).

128 3.2. Proposed Procedure for Calculating Similarity

129 In Section 3.1, it was revealed that weights of inputs should correspond to
130 strength of nonlinearity between the inputs and an output around a query sample.
131 In addition, a regression coefficient, i.e. slope in Figure 1, significantly changes
132 around a query sample when the nonlinearity around it is strong. Although the re-
133 gression coefficient is constant when input-output relationship is linear, it changes
134 depending on the strength of nonlinearity. To evaluate the change of the regres-
135 sion coefficient of an input around a query sample and to determine the weights,
136 the weighted variance of each input's regression coefficients of LW-PLS models
137 is utilized. The similarities between a query sample and samples in a database are
138 utilized as the weights when the weighted variance is calculated. Since similarity
139 depends on the weight θ , iterative calculation is conducted to derive similarity and
140 θ .

141 Offline and online calculation procedures of the weights are as follows.

142 · Offline part

- 143 1. Determine the number of latent variables R , the localization parameter φ
 144 and the maximum iteration number I .
- 145 2. Set $i = 1$ and $\theta_{m,i-1} = 1$ for all m .
- 146 3. Regard each of N samples in the database as a query sample and construct
 147 N LW-PLS models by using $\theta_{m,i-1}$.
- 148 4. Calculate the variance $V_{m,i}$ of N regression coefficients of the m th input
 149 variable $a_{nm,i}$, then set $\theta_{m,i} = (V_{m,i})^\alpha$. Here, α is a tuning parameter.
- 150 5. If $i = I$ or the following equation is satisfied for all m , finish the offline
 151 calculation. Here, ε_1 is a tolerance.

$$\left| \frac{\theta_{m,i} - \theta_{m,i-1}}{\theta_{m,i-1}} \right| \leq \varepsilon_1 \quad (20)$$

- 152 6. Set $i = i + 1$ and go to step 3.

153 In the offline part, θ_m is first set to 1, then θ_m is updated to $(V_{m,i})^\alpha$; a_{nm} and V_m
 154 are calculated repeatedly until θ_m converges.

155 · Online part

- 156 1. Determine the maximum iteration number J , and set $j = 1$ and $\theta_{m,j-1} = \theta_m$
 157 obtained in the offline part.
- 158 2. Calculate the similarity $\omega_{n,j-1}$ by using $\theta_{m,j-1}$.
- 159 3. Calculate the weighted variance $V_{m,j}$ of a_{nm} obtained in the offline part.

$$V_{m,j} = \sum_{n=1}^N \omega_{n,j-1} (a_{nm} - \bar{a}_m)^2 / \sum_{n=1}^N \omega_{n,j-1} \quad (21)$$

$$\bar{a}_m = \sum_{n=1}^N \omega_{n,j-1} a_{nm} / \sum_{n=1}^N \omega_{n,j-1} \quad (22)$$

- 160 4. Set $\theta_{m,j} = [(V_{m,j})^\alpha + \theta_{m,j-1}] / 2$.

161 5. If $j = J$ or the following equation is satisfied for all m , finish the online
 162 calculation. Here, ε_2 is a tolerance.

$$\left| \frac{\theta_{m,j} - \theta_{m,j-1}}{\theta_{m,j-1}} \right| \leq \varepsilon_2 \quad (23)$$

163 6. Set $j = j + 1$ and go to step 2.

164 In the online part, θ_m is updated by using the weighted variance V_m of a_{nm} ob-
 165 tained in the offline part and the similarity ω_n between a query sample to evaluate
 166 the strength of nonlinearity around a query sample.

167 This procedure contains seven parameters to be determined: the number of
 168 latent variables R , the localization parameter φ , the tuning parameter α , the max-
 169 imum iteration number in the offline part I and in the online part J , and the tol-
 170 erance in the offline part ε_1 and in the online part ε_2 . R , φ and α can be de-
 171 termined by applying cross validation to all data or by building and validating
 172 models with different datasets, i.e., model construction data and parameter tun-
 173 ing data. The proposed method includes the conventional LW-PLS, which uses
 174 normal Euclidean distance since the proposed method becomes the same as the
 175 conventional one when $\alpha = 0$. Thus, the estimation accuracy of the proposed
 176 LW-PLS model is the same as or better than that of the conventional LW-PLS
 177 model when α is tuned properly.

178 4. Numerical Example

179 In this section, the proposed method is compared with the conventional meth-
 180 ods in two numerical examples. The following four methods are compared.

181 LW-PLS 1) LW-PLS with $\theta_m = 1$.

182 LW-PLS 2) LW-PLS with θ_m defined as the absolute value of the m th variable's
 183 regression coefficient of a global MRA model [28].

184 LW-PLS 3) LW-PLS with θ_m defined as the absolute value of the m th variable's
 185 regression coefficient of a LW-PLS model constructed by LW-PLS 1 [29].

186 LW-PLS 4) LW-PLS with θ_m defined by the proposed method.

187 4.1. Problem Settings

188 The following two cases are investigated; in each case, x_m and y are inputs
 189 and an output, respectively.

190 · Case 1

$$w_m \sim N(0, 0.02^2) \quad (m = 0, 1, 2, 3) \quad (24)$$

$$s_m \sim \text{rand}(-5, 5) \quad (m = 1, 2, 3) \quad (25)$$

$$x_m = s_m + w_m \quad (m = 1, 2, 3) \quad (26)$$

$$y = 10s_1 + 5s_2^2 + \exp(s_3) + w_0 \quad (27)$$

191 · Case 2

$$w_m \sim N(0, 0.02^2) \quad (m = 0, 1, \dots, 6) \quad (28)$$

$$s_m \sim \text{rand}(-5, 5) \quad (m = 1, 2, \dots, 6) \quad (29)$$

$$x_m = s_m + w_m \quad (m = 1, 2, \dots, 6) \quad (30)$$

$$x_7 = s_6 + w_6 \quad (31)$$

$$y = s_2^3 + 3s_3 + s_4^2 + \exp(s_5) + 3s_6 + w_0 \quad (32)$$

192 Here, $\text{rand}(a, b)$ denotes the uniform random distribution in closed interval $[a, b]$,
 193 and $N(\mu, \sigma^2)$ denotes the normal distribution whose mean is μ and standard de-
 194 viation is σ . In both cases, 3000 samples were generated and divided into three
 195 groups: samples for model construction (1000 samples), parameter tuning (1000
 196 samples) and model validation (1000 samples). Models were constructed with
 197 different values of localization parameter φ , the number of latent variables R , and
 198 α , by using samples for model construction. Then, the estimation errors were
 199 calculated by using samples for parameter tuning, and the set of parameters that
 200 minimized the estimation error was selected. The search range of φ , R and α is
 201 $[0.01, 0.03, \dots, 0.09]$, $[1, 2, 3]$ and $[0.01, 0.03, \dots, 0.09]$, respectively. The ap-
 202 propriate search range of the parameters depends on the situation; therefore, it is
 203 recommended to make the search range wide enough in order to get the optimal
 204 parameters. In LW-PLS 4, tolerances ε_1 and ε_2 are 0.01. Both of the maximum
 205 iteration numbers I and J are 30.

206 4.2. Results and Discussions

207 Table 1 shows the selected parameters and root mean square error for valida-
 208 tion samples (RMSE 1). The proposed method achieved the minimum RMSE 1
 209 in both cases and was considerably superior to the conventional methods. Figure 2
 210 shows the relationship between RMSE for parameter tuning samples (RMSE 2).
 211 and φ when the proposed method is applied to case 1 ($R = 3$). RMSE 2 was large
 212 when φ was too small or too large. Overfitting occurred when φ was too small,
 213 and models were unable to cope with nonlinearity between input and output vari-
 214 ables when φ was too large. Tables 2 and 3 show θ_m when $\mathbf{x}_q = [0, 0, 3]^T$ and
 215 $\mathbf{x}_q = [0, 0, -3]^T$ in case 1, respectively. Figure 3 shows the transition of θ_m in the
 216 online part of the weights calculation procedure. Here, θ_m is normalized so that

Table 1: Selected parameters and RMSE for validation samples in numerical examples.

Case	Method	R	φ	α	RMSE
1	LW-PLS 1	3	0.05	-	3.84
	LW-PLS 2	3	0.05	-	5.29
	LW-PLS 3	3	0.05	-	4.69
	LW-PLS 4	3	0.03	0.8	1.59
2	LW-PLS 1	6	0.21	-	18.93
	LW-PLS 2	6	0.09	-	16.53
	LW-PLS 3	5	0.21	-	21.17
	LW-PLS 4	6	0.06	0.8	5.31

217 the sum of θ_m be 1 in LW-PLS 2, 3 and 4. In case 1, where the relationship be-
 218 tween x_1 and the output is linear, therefore, θ_1 should be 0. When $\mathbf{x}_q = [0, 0, 3]^T$,
 219 θ_3 should be larger than θ_2 because x_3 has stronger nonlinearity around the query
 220 sample than x_2 , i.e.

$$\text{abs}\left(\frac{\partial^2 y}{\partial x_2^2} \Big|_{\mathbf{x}=[0,0,3]^T}\right) < \text{abs}\left(\frac{\partial^2 y}{\partial x_3^2} \Big|_{\mathbf{x}=[0,0,3]^T}\right) \quad (33)$$

221 where $\text{abs}(a)$ denotes the absolute value of a . On the other hand, when $\mathbf{x}_q =$
 222 $[0, 0, -3]^T$, θ_2 should be larger than θ_3 because

$$\text{abs}\left(\frac{\partial^2 y}{\partial x_2^2} \Big|_{\mathbf{x}=[0,0,-3]^T}\right) > \text{abs}\left(\frac{\partial^2 y}{\partial x_3^2} \Big|_{\mathbf{x}=[0,0,-3]^T}\right) \quad (34)$$

223 . The proposed method derived appropriate θ for both query samples while the
 224 other methods could not. This is the reason why the proposed method could
 225 achieve the best performance in the four methods.

Table 2: Derived weights of $x_q = [0, 0, 3]^T$ in case 1

Method	θ_1	θ_2	θ_3
LW-PLS 1	1.00	1.00	1.00
LW-PLS 2	0.53	0.08	0.39
LW-PLS 3	0.32	0.02	0.66
LW-PLS 4	0.00	0.32	0.68

Table 3: Derived weights of $x_q = [0, 0, -3]^T$ in case 1

Method	θ_1	θ_2	θ_3
LW-PLS 1	1.00	1.00	1.00
LW-PLS 2	0.53	0.08	0.39
LW-PLS 3	0.57	0.34	0.09
LW-PLS 4	0.01	0.81	0.18

226 5. Application to an Industrial Distillation Process

227 In this section, an application result of the proposed method to an industrial
 228 distillation process is reported. A soft-sensor for estimating the aroma concen-
 229 tration was constructed in order to realize highly efficient operation of a cracked
 230 gasoline (CGL) fractionator of an ethylene production process at the Showa Denko
 231 K.K. (SDK) Oita plant in Japan. Aroma denotes the generic name for benzene,
 232 toluene, xylene and styrene, etc. In this case study, linear PLS, LW-PLS 1, 2, 3
 233 and 4 were compared. The search range of φ , R and α is $[0.2, 0.4, \dots, 2, 2.5, 3.0,$
 234 $\dots, 10]$, $[1, 2, \dots, 9]$, $[0.2, 0.4, \dots, 2.0]$, respectively.

235 5.1. CGL Fractionator

236 A schematic diagram of the CGL fractionator of the ethylene production pro-
237 cess is shown in Figure 4. The CGL fractionator is controlled with multivariable
238 model predictive control (MPC) with an optimizer, and the aroma concentration
239 in the CGL is used as one of the constraints in the optimizer. Although the opera-
240 tion data of the CGL fractionator are stored in the database every hour, the aroma
241 concentration is analyzed in a laboratory usually once a day because of its long
242 analysis time. For safety, the process was operated at a condition that has a wide
243 margin and is far from the constraints. Thus, real-time accurate estimation of the
244 aroma concentration is crucial in order to make the operating condition closer to
245 the constraint and reduce the energy consumption.

246 5.2. Operation Data

247 Although 19 variables are measured in the CGL fractionator, only eight vari-
248 ables were selected as the input variables of the soft-sensor on the basis of the
249 process knowledge. In addition, the coil outlet temperature of a cracking furnace
250 measured four hours before was used together with the selected input variables,
251 since the product composition is affected by the operating condition of the crack-
252 ing furnace which is located in the upstream of the CGL fractionator, and it takes
253 about four hours for materials to reach the CGL fractionator from the cracking
254 furnace. Hence, the total number of input variables is nine. The selected input
255 variables of the soft-sensor are listed in Table 4 and Figure 4. The operation data
256 obtained from January 1, 2010 to August 4, 2011 were stored in the database.
257 Then, the tuning parameters were determined using these data, and the aroma
258 concentration was estimated for the operation data obtained from August 6, 2011

Table 4: Input variables of the soft-sensor for the CGL fractionator.

No.	Variable name
1	Outlet cracked gasoline density
2	Reboiler flow rate
3	Outlet cracked kerosene flow rate
4	Reflux volume
5	Outlet cracked gasoline flow rate
6	Tray #4 differential pressure
7	Tower top temperature
8	Feed flow rate
9	Cracked furnace coil outlet temperature

259 to December 31, 2011. Here, all variables were mean-centered and scaled in order
 260 to make each variable's standard deviation one.

261 5.3. Results and Discussions

262 Table 5 shows the selected parameters and RMSE 1. In LW-PLS 4, tolerances
 263 ε_1 and ε_2 are 0.01. The maximum iteration numbers I and J are 20 and 30,
 264 respectively. The average calculation time of output estimation for each query
 265 was 4.8 msec when Intel® Core™ i7-2620M (2.7 GHz×2) and 8 GB RAM were
 266 used.

267 In this process, the output variable (aroma concentration) is measured to one
 268 place of decimal, thus, the differences of RMSEs between Linear PLS and LW-
 269 PLS 1, and between LW-PLS 2, 3 and 4 are not significant. The reason why
 270 LW-PLS 2, 3 and 4 derived the better result than the other methods might be that
 271 the strength of nonlinear effect of each input on the output is different. Table 6

Table 5: Selected parameters and RMSE for validation samples in a case study of the CGL fractionator.

Method	R	φ	α	RMSE
Linear PLS	2	-	-	1.20
LW-PLS 1	2	6.5	-	1.15
LW-PLS 2	2	1.0	-	0.99
LW-PLS 3	2	1.0	-	0.98
LW-PLS 4	2	1.4	1.2	1.03

Table 6: Changes of weights in a case study of the CGL fractionator when LW-PLS 4 is applied.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9
Maximum value	0.32	0.12	0.13	0.09	0.12	0.10	0.09	0.10	0.12
Mean value	0.26	0.09	0.11	0.08	0.09	0.09	0.09	0.08	0.11
Minimum value	0.21	0.07	0.10	0.07	0.08	0.08	0.08	0.08	0.09
Standard deviation	0.03	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01

272 shows the maximum, mean and minimum values, and standard deviation of the
273 m th weight θ_m when LW-PLS 4 is applied. Here, $\theta_1, \theta_2, \dots, \theta_8$ for each query
274 are normalized so that their sum becomes 1. θ_1 is the largest and the nonlinear
275 effect of input 1 on the output is expected to be strong. In addition, the strength of
276 nonlinear effect of each input on the output does not seem to depend on the value
277 of since the standard deviations of the weights are small. This could be the reason
278 why RMSEs of LW-PLS 2, 3 and 4 are similar.

279 **6. Conclusion**

280 To construct highly accurate locally weighted partial least squares (LW-PLS)
281 models, an adaptive similarity measure was proposed. In the proposed method,
282 weights of input variables are determined through iterative calculation by using
283 the weighted variance of the regression coefficients. The results of the case studies
284 showed that the proposed method could adaptively derive the appropriate weights
285 and more accurate models than the conventional methods in numerical examples.
286 Furthermore, root mean square error was improved by 11.3 % by using the pro-
287 posed method compared to LW-PLS in which conventional similarity based on the
288 Euclidean distance without weights is used. These results clearly demonstrate the
289 usefulness of the proposed method, which uses newly defined similarity based on
290 the weighted Euclidean distance.

291 **7. Acknowledgments**

292 This work was partially supported by the Japan Society for the Promotion of
293 Science (JSPS), Grant-in-Aid for Scientific Research (C) 21560793.

294 **References**

- 295 [1] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process
296 industry, *Comput. and Chem. Eng.* 33 (2009) 795–814.
- 297 [2] T. Rajalahti, O. M. Kvalheim, Multivariate data analysis in pharmaceuticals:
298 A tutorial review, *Int. J. Pharm.* 417 (2011) 280–290.
- 299 [3] M. Kano, M. Ogawa, The state of the art in chemical process control in

- 300 japan: Good practice and questionnaire survey, *J. Proc. Cont.* 20 (2010)
301 969–982.
- 302 [4] G. Andersson, P. Kaufmann, L. Renberg, Non-linear modelling with a cou-
303 pled neural network - PLS regression system, *J. Chemom.* 10 (1996) 605–
304 614.
- 305 [5] G. Baffi, E. B. Martin, A. J. Morris, Non-linear projection to latent struc-
306 tures revisited (the neural network PLS algorithm), *Comput. Chem. Eng.* 23
307 (1999) 1293–1307.
- 308 [6] S. J. Qin, T. J. McAvovy, Nonlinear PLS modeling using neural networks,
309 *Comput. Chem. Eng.* 16 (1992) 379–391.
- 310 [7] D. Pérez-Marin, A. Garrido-Varo, J. E. Guerrero, J. C. Gutiérrez-Estrada,
311 Use of artificial neural networks in near-infrared reflectance spectroscopy
312 calibrations for predicting the inclusion percentages of wheat and sunflower
313 meal in compound feedingstuffs, *Appl. Spectrosc.* 60 (2006) 1062–1069.
- 314 [8] V. R. Nadadoor, S. L. S. H. Siegler, A. B.-Z. W. C. McCaffrey, Online
315 sensor for monitoring a microalgal bioreactor system using support vector
316 regression vector regression, *Chemom. Intell. Lab. Syst.* 44 (2012) 2101–
317 2105.
- 318 [9] I. Barman, C. R. Kong, N. C. Dingari, R. R. Dasari, M. S. Feld, Development
319 of robust calibration models using support vector machines for spectroscopic
320 monitoring of blood glucose, *Chemom. Intell. Lab. Syst.* 82 (2012) 9719–
321 9726.

- 322 [10] D. E. Lee, J. H. Song, S. O. Song, E. S. Yoon, Weighted support vector ma-
323 chine for quality estimation in the polymerization process, *Ind. Eng. Chem.*
324 *Res.* 44 (2005) 2101–2105.
- 325 [11] S. Wold, N. K. Wold, B. Skagerberg, Nonlinear PLS modeling, *Chemom.*
326 *Intell. Lab. Syst.* 7 (1989) 53–65.
- 327 [12] G. Robertsson, Contributions to the problem of approximation of non-linear
328 data with linear pls in an absorption spectroscopic context, *Chemom. Intell.*
329 *Lab. Syst.* 47 (1999) 99–106.
- 330 [13] A. I. Abdel-Rahmana, G. J. Lim, A nonlinear partial least squares algorithm
331 using quadratic fuzzy inference system, *J. Chemom.* 23 (2009) 530–537.
- 332 [14] S. Y. Chang, E. H. Baughman, B. C. McIntosh, Implementation of locally
333 weighted regression to maintain calibrations on FT-NIR analyzers for indus-
334 trial processes, *Appl. Spectrosc.* 55 (2001) 1199–1206.
- 335 [15] P. Kadlec, R. Grbić, B. Gabrys, Review of adaptation mechanisms for data-
336 driven soft sensors, *Comput. and Chem. Eng.* 35 (2011) 1–24.
- 337 [16] W. S. Cleveland, Robust locally weighted regression and smoothing scatter-
338 plots, *J. Am. Stat. Assoc.* 74 (1979) 829–836.
- 339 [17] A. T. Walden, P. Prescott, Identification of trends in annual maximum sea
340 levels using robust locally weighted regression, *Estuar. Coast. Shelf. S.* 16
341 (1983) 17–26.
- 342 [18] W. S. Cleveland, S. J. Devlin, Locally weighted regression: An approach to
343 regression analysis by local fitting, *J. Am. Stat. Assoc.* 83 (1988) 596–610.

- 344 [19] T. Naes, T. Isaksson, Locally weighted regression and scatter correction for
345 near-infrared reflectance data, *Anal. Chem.* 62 (1990) 664–673.
- 346 [20] V. Centner, D. L. Massart, Optimization in locally weighted regression,
347 *Anal. Chem.* 70 (1998) 4206–4211.
- 348 [21] H. H. Leung, Y. S. Huang, C. X. Cao, Locally weighted regression for desul-
349 phurisation intelligent decision system modeling, *Simulat. Model. Pract.*
350 *Theor.* 12 (2004) 413–423.
- 351 [22] Z. Ge, Z. Song, A comparative study of just-in-time-learning based methods
352 for online soft sensor modeling, *Chemom. Intell. Lab. Syst.* 104 (2010) 306–
353 317.
- 354 [23] C. Cheng, M. S. Chiu, A new data-based methodology for nonlinear process
355 modeling, *Chem. Eng. Sci.* 59 (2004) 2801–2810.
- 356 [24] Z. Ge, Z. Song, Online monitoring of nonlinear multiple mode processes
357 based on adaptive local model approach, *Control Eng. Pract.* 16 (2008)
358 1427–1437.
- 359 [25] Z. Y. Wang, T. Isaksson, B. R. Kowalski, New approach for distance mea-
360 surement in locally weighted regression, *Anal. Chem.* 66 (1994) 249–260.
- 361 [26] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami, Soft-sensor development
362 using correlation-based just-in-time modeling, *AIChE J.* 55 (2009) 1754–
363 1765.
- 364 [27] K. Fujiwara, M. Kano, S. Hasebe, Development of correlation-based clus-

- 365 tering method and its application to software sensing, *Chemom. Intell. Lab.*
366 *Syst.* 101 (2010) 130–138.
- 367 [28] H. Shigemori, M. Kano, S. Hasebe, Optimum quality design system for
368 steel products through locally weighted regression model, *J. Proc. Cont.* 21
369 (2011) 293–301.
- 370 [29] S. Kim, M. Kano, H. Nakagawa, S. Hasebe, Estimation of active pharma-
371 ceutical ingredients content using locally weighted partial least squares and
372 statistical wavelength selection, *Int. J. Pharm.* 421 (2011) 269–274.
- 373 [30] H. Nakagawa, T. Tajima, M. Kano, S. Kim, S. Hasebe, T. Suzuki, H. Nak-
374 agami, Evaluation of infrared-reflection absorption spectroscopy measure-
375 ment and locally weighted partial least-squares for rapid analysis of residual
376 drug substances in cleaning processes, *Anal. Chem.* 84 (2012) 3820–3826.
- 377 [31] M. Kano, K. Fujiwara, Virtual sensing technology in process industries:
378 trends and challenges revealed by recent industrial applications, *J. Chem.*
379 *Eng. Jpn.* (2012) doi: 10.1252/jcej.12we167.
- 380 [32] S. Schaal, C. G. Atkeson, S. Vijayakumar, Scalable techniques from non-
381 parametric statistics for real time robot learning, *Appl. Intell.* 17 (2002)
382 49–60.

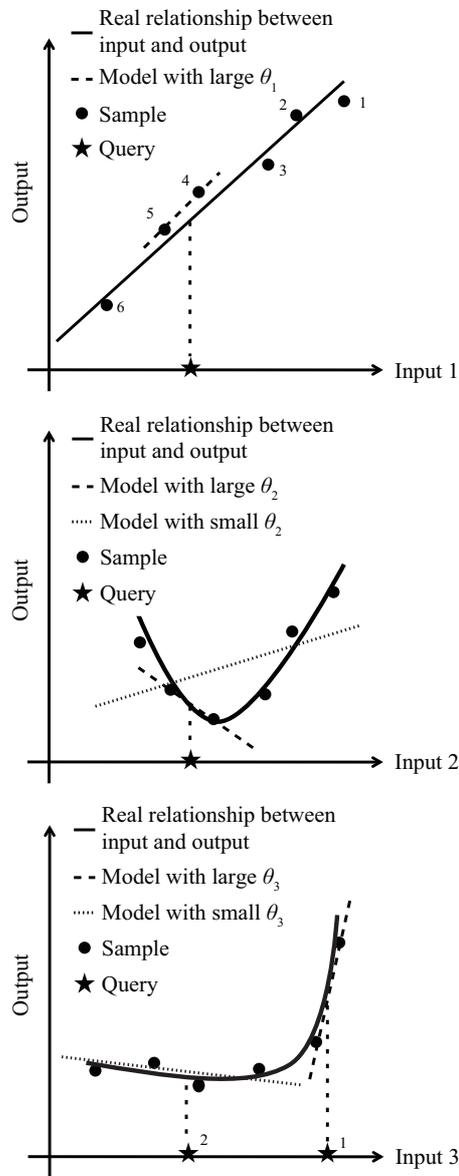


Figure 1: The relationship between a local model and weights θ_m ($m = 1, 2, 3$). (top) a case where relationship between an input and an output is linear. (middle) a case where relationship between an input and an output is nonlinear. (bottom) a case where the strength of nonlinearity changes depending on the value of an input variable.

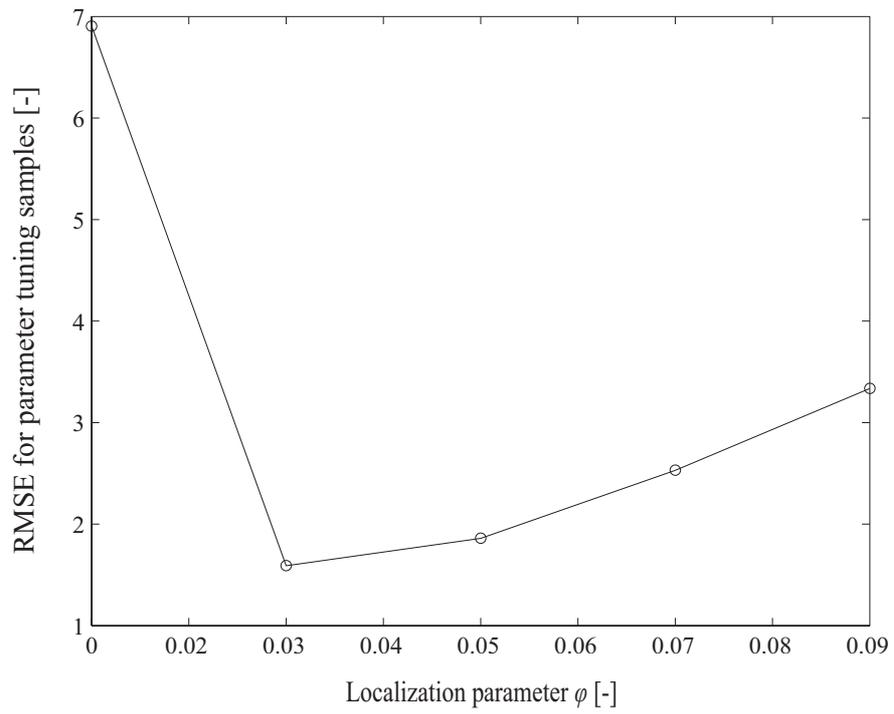


Figure 2: The relationship between RMSE for parameter tuning samples and the localization parameter φ when the proposed method is applied in case 1 ($R = 3$).

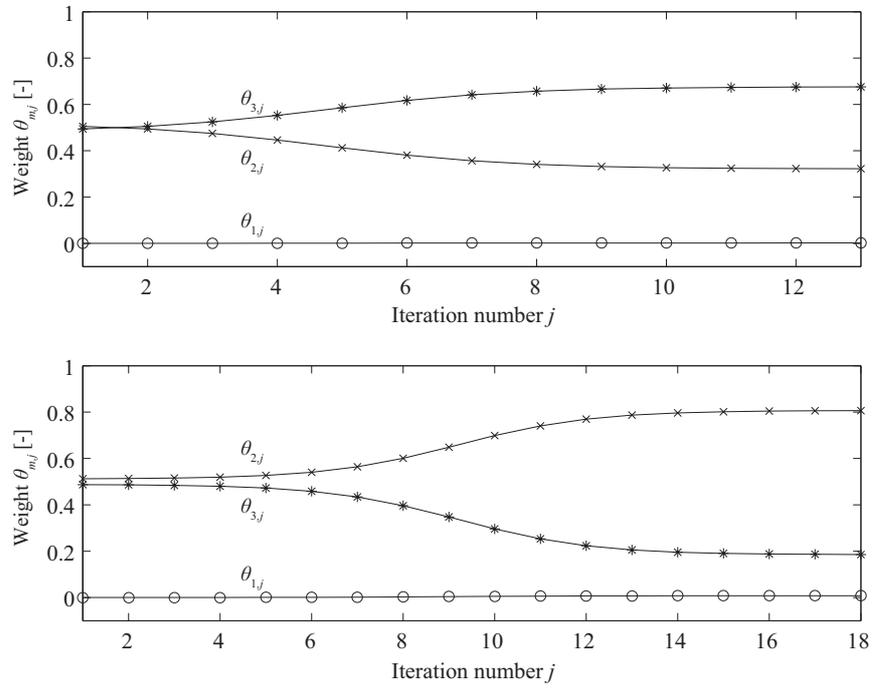


Figure 3: The transition of θ_m in the online part of the weights calculation procedure in case 1. (top) $\mathbf{x}_q = [0, 0, 3]^T$. (bottom) $\mathbf{x}_q = [0, 0, -3]^T$.

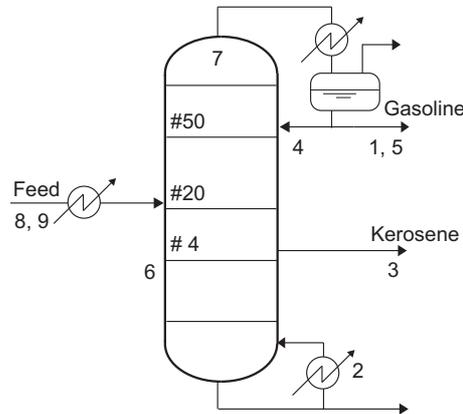


Figure 4: Schematic diagram of the CGL fractionator of the ethylene production process at the Showa Denko K.K. (SDK) Oita plant.