Prediction of Glucose Concentration Hydrolysed from Oil Palm Trunks Using a PLSR-Based Model

Wan Sieng Yeo[1], Mieow Kee Chan[2(✉)], and Nurul Adela Bukhari[3]

[1] Department of Chemical Engineering, Faculty of Engineering and Sciences, Curtin University Malaysia, CDT 250, 98000 Miri, Sarawak, Malaysia
[2] Centre for Bioprocess Engineering, Faculty of Engineering and the Built Environment, SEGi University, Jalan Teknologi, Kota Damansara, 47810 Petaling Jaya, Selangor Darul Ehsan, Malaysia
mkchan@segi.edu.my
[3] Energy and Environment Unit, Engineering and Processing Research Division, Malaysian Palm Oil Board (MPOB), 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia

Abstract. Oil palm trunks are biomass and it contains starch that can be used to produce higher value-added glucose for bioethanol, lactic acid, food and beverage productions. An immobilised enzymes hydrolysis that does not require high temperature, strong acids, and an additional separation process is preferable for the conversion of starch to glucose as compared to acid hydrolysis involving hydrochloric acid or sulphuric acid. Notice that a limited study focuses on utilisation of least-square regression models to predict the glucose concentration from an immobilised enzymes hydrolysis. Hence, this study developed a least square model, namely a locally weighted kernel partial least square regression (LW-KPLSR) model to forecast the glucose concentration produced from the immobilised enzymes hydrolysis of the oil palm trunks. Its predictive performance results were determined, evaluated, and compared with its counterparts. LW-KPLSR has a more accurate glucose concentrations prediction than others since its Ea value is 103% to 195% lower.

Keywords: Oil palm trunks · Glucose concentration · Hydrolysis process · Prediction · Partial least square regression model

1 Introduction

Malaysia is the second-largest palm oil producer in the world after Indonesia [1] with 19.47 million tonnes of palm oil was produced in the year 2020 [2, 3]. This implies the generation of a large amount of biomass due to replantation and milling activities. Biomass such as oil palm trunk (OPT) consists of a high amount of starch [4] which can be converted to glucose. Acid hydrolysis is the commonly used method to convert starch to glucose with the help of hydrochloric acid and sulphuric acid [5, 6]. However, an additional separation process is required to purify glucose from the by-products, for instance, furans, before glucose can be used as the substrate for the fermentation

P. Vasant et al. (Eds.): ICO 2021, LNNS 371, pp. 927–937, 2022.
https://doi.org/10.1007/978-3-030-93247-3_88

process. Enzymatic hydrolysis produces a high yield of glucose from starch at a mild process condition due to the selectivity of the enzyme. It could be done by adding α-amylase and glucoamylase into the starch mixture at the same time, to produce glucose.

Attempts have been done to optimise glucose production via enzymatic hydrolysis, to reduce the operating cost and obtain high quality yield [7–9]. Studies also focus on modeling, where the enzyme kinetic parameters were calculated [10] to describe the reaction mechanism [11]. Prediction of yield could be obtained when these parameters/constants are available. However, these constant values are experiment dependent and in depth knowledge of biochemistry and microbiology is required. On another hand, machine learning makes use of massive data to develop models via a mathematical approach. It recognises the pattern of data distribution and can perform prediction without explicit, rule-based programming [12].

Machine learning algorithms including partial least square regression (PLSR) based models that involve mathematical approaches have been widely used in a variety of applications [13–15]. PLSR based models are famous since they are dimension reduction methods, simple, and can cope with collinearity between variables [16]. Recently, a PLSR-based model, namely locally weighted Kernel partial least square regression (LW-KPLSR) has been developed by Yeo, Saptoro and Kumar [17] for nonlinear processes. Its predictive capability has been investigated using different case studies from the literature [17] and the experimental data for the bleaching of fabric cotton. However, it has not been applied to the experimental data for a hydrolysis process.

Besides, it is found that a minimal study is considering regression models including PLSR based model to estimate the glucose concentration from the hydrolysis processes. Hence, this study aims to develop LW-KPLSR using the experimental data for immobilized enzymes hydrolysis of the oil palm trunks to predict the glucose concentration. Then, the predictive performance of the LW-PLSR was evaluated using the root mean square error (*RMSE*), the error of approximation (E_a), and the coefficient of determination (R^2). And these predictive results were also compared with other existing models such as locally weighted partial least square (LW-PLSR), PLSR, and principal component regression (PCR). The following sections are the research methodology, results, and discussions, as well as the conclusions.

2 Research Methodology

In this section, hydrolysis of OPT was described, followed by the descriptions of the LW-KPLSR model development, the data splitting and parameters setting, as well as the evaluation of the predictive performance of the regression models. Lastly, the computer configurations and software used in this study are illustrated.

2.1 Hydrolysis of OPT

Starch was extracted from OPT by heating method [18]. The immobilised enzymes, namely α-amylase and glucoamylase, were dispersed into the extract after it was cold to room temperature. The hydrolysis experiment was conducted at varied stirring speeds

(150 to 300 rpm), the mass of OPT (5 to 20 g), and hydrolysis time (8 to 24 h). The concentration of glucose was measured by using the One Touch Select Simple Glucometer [19].

2.2 Regression Model Development

Generally, the LW-KPLSR model is an improved model from LW-PLSR [17, 20] and the LW-PLSR is extended from PLSR [21]. The LW-KPLSR that was developed by Yeo, Saptoro and Kumar [17] for highly nonlinear processes are utilised in this study. The similarity measurement used in this LW-KPLSR model is the Euclidean distance-based similarity index, ω_n which is obtained based on the distance, d_n between a query, $x_{q,}$ and the historical input data, x_n. The ω_n and a similarity matrix, Ω can be determined using Eqs. 1 and 2, respectively [22].

$$\omega_n = exp(-\frac{d_n}{\phi\sigma_n}) \tag{1}$$

$$d_n = \sqrt{(x_n - x_q)^T (x_n - x_q)} \tag{2}$$

where ϕ is a localisation parameter, σ_d is the standard deviation of d_n (n = 1, 2, …,N).

In the LW-KPLS model, the input and output variables, x and y for n number of the sample can be denoted as Eqs. 3 and 4 where M and L are numbers of x and y, respectively.

$$x_n = [x_{n1}, x_{n2}, ..., x_{nM}]^T \tag{3}$$

$$y_n = [y_{n1}, y_{n2}, ..., y_{nL}]^T \tag{4}$$

To obtain the predicted output, $\hat{y}_q$ from the LW-KPLSR model, the following steps have to be followed and conducted [17].

1. Obtain both Kernel matrices for input variables, V and query, V_q in which the input and output variables in Eqs. 3 and 4 are mapped into a higher dimensional feature space utilising the polynomial Kernel function as shown in Eq. 5.

$$k(\mathrm{x},\mathrm{y}) = (x^T y + 1)^b \tag{5}$$

2. Perform mean centering on these obtained Kernel matrices, V and V_q using the following Eqs. 6 and 7.

$$\tilde{V} = \left(I - \frac{1}{n} 1_n 1_n^T\right) V \left(I - \frac{1}{n} 1_n 1_n^T\right) \tag{6}$$

$$\tilde{V}_q = (V_q - \frac{1}{n} 1_{N_2} 1_n^T V)(I - \frac{1}{n} 1_n 1_n^T) \tag{7}$$

3. Obtain a dual representation of a scaled version of projection direction, B via a dual kernel partial least square discrimination using Eq. 8.
   $B = YY^{'} V\beta$ with the normalisation,

$$\beta = \frac{\beta}{\|\beta\|} \tag{8}$$

4. Calculate the re-scaled query, and input variable matrices, V_q, and V using Eqs. 9 and 10.

$$X_q = V_q B \tag{9}$$

$$X = VB \tag{10}$$

5. Figure out the number of latent variables K and set k = 1.
6. Obtain a similarity matrix Ω using Eqs. 1, 2 and 11.

$$\Omega = dig\{\omega_1, \omega_2, ..., \omega_N\} \tag{11}$$

7. Determine X_k, Y_k, and $X_{q,k}$ using Eqs. 12–16.

$$X_k = X - 1_N \left[\overline{X_1}, \overline{X_2}, ..., \overline{X_M}\right] \tag{12}$$

$$Y_k = Y - 1_N \left[\overline{Y_1}, \overline{Y_2}, ..., \overline{Y_L}\right] \tag{13}$$

$$X_{q,k} = X_q - 1_N \left[\overline{X_1}, \overline{X_2}, ..., \overline{X_M}\right]^T \tag{14}$$

$$\overline{X_m} = \frac{\sum_{n=1}^{N} \omega_n X_{nm}}{\sum_{n=1}^{N} \omega_n} \tag{15}$$

$$\overline{Y_l} = \frac{\sum_{n=1}^{N} \omega_n Y_{nl}}{\sum_{n=1}^{N} \omega_n} \tag{16}$$

8. Let $\hat{y_q} = [\overline{y_1}, \overline{y_2}, ..., \overline{y_L}]^T$.
9. Get the kth latent variable of X_k using Eqs. 17 and 18.

$$t_k = X_k w_k \tag{17}$$

$$w_k = \frac{X_k^T \Omega Y_k}{\|X_k^T \Omega Y_k\|} \tag{18}$$

10. Attain the k[th] loading vector of X_k and the k[th] regression coefficient vector using Eqs. 19 and 20.

$$p_k = \frac{X_k^T \Omega t_k}{t_k^T \Omega t_k} \tag{19}$$

$$q_k = \frac{Y_k^T \Omega t_k}{t_k^T \Omega t_k} \tag{20}$$

11. Obtain the k[th] latent variable of X_q using Eq. 21.

$$t_{q,k} = X_{q,k}^T w_k \tag{21}$$

12. Substitute $\hat{y}_q$ with $\hat{y}_q + t_{q,j} q_j$ where $t_{q,k}$ is the k[th] latent variable of X_q.
13. If k = K, then complete the prediction using LW-KPLSR model. Otherwise, place Eqs. 22–24.

$$X_{k+1} = X_k - t_k p_k^T \tag{22}$$

$$Y_{k+1} = Y_k - t_k q_k^T \tag{23}$$

$$X_{q,k+1} = X_{q,k} - t_{q,k} p_k \tag{24}$$

14. Let $k = k + 1$ and go back to Step 9.

2.3 Data Splitting and Parameters Set for the Least Square Regression Models

In this study, a total number, N of 18 datasets were collected from the immobilised enzymes hydrolysis of the oil palm trunks. They were saved in a CSV file and were divided into a ratio of 75:25 where the number of training data, N_1 is 14, and the number of testing data, N_2 is 4. These datasets consist of the input or observed variables, namely stirring speed (rpm), the mass of OPT (g) and hydrolysis time (h). Meanwhile, the output or target variable is glucose concentration in mmol per litre. Both training and testing data involving the input and output variables were executed in MATLAB software using LW-KPLSR, LW-PLSR, PLSR, and PCR models. The number of latent variables for all these least-square regression models is set as 1.

Besides, for LW-KPLSR and LW-PLSR models, their φ is 0.1 as it gives the best results [17]. Meanwhile, since LW-KPLSR consists of a kernel function, its kernel parameter, b has to be tuned. According to Mongillo [23], and Orr [24], the value of b within the range of 0.01 to 10 provides the lower error. Usually, b equals 1 gives the lowest predicted error [20], hence b is fixed as 1 in this study. The parameters used in the least square regression models for this study are tabulated in Table 1.

Table 1. Parameters used for the LW-KPLSR, LW-PLSR, PLSR, and PCR models.

Parameters	N	N_1	N_2	LV	φ	b
Values	18	14	4	1	0.1	1

2.4 Evaluation of the Predictive Performance of the Regression Models

This study utilised *RMSE*, E_a, and R^2 to evaluate the predictive performance of the LW-KPLSR, LW-PLSR, PLSR, and PCR models. *RMSE* is a goodness-of-fit indicator that shows the differences in observed and target values [25, 26]. The lower the *RMSE*, the better the predictive performance of the model. *RMSE* can be calculated using Eq. 25 as shown below [25]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{25}$$

where y_i and $\hat{y}_i$ are the actual and predicted output values, respectively.

However, there could be some possible cases that its *RMSE* for the training dataset is the lowest and testing datasets have the highest *RMSE* [20, 27]. Then, it will be a trouble to evaluate the overall predictive performance of this regression model. E_a that is shown in Eq. 26 was adopted from Saptoro, Vuthaluru and Tadé [28] and Yeo, Saptoro, Kumar and Research [20] to address this problem.

$$E_a = \left(\frac{N_1}{N}\right) RMSE_1 + \left(\frac{N_2}{N}\right) RMSE_2 + |RMSE_1 - RMSE_2| \tag{26}$$

Moreover, R^2 indicates the comparison between the total of the squared errors to the total of the squared deviations about its mean. In this study, the R^2 that is shown in Eq. 27 [25] was employed to measure the goodness of fit between the actual and the predicted values. The closer the R^2 to 1 the better the predictive performance of the model [29].

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{27}$$

Besides, the percentage error (*PE*) which is displayed in Eq. 28 [25, 30] was also adopted in this study. *PE* was used to understand the differences between the *RMSE*, E_a, and R^2 between two regression models.

$$PE = \left| \frac{\widehat{y}_i - y_i}{y_i} \right| \times 100\% \tag{28}$$

2.5 Computer Configurations and Software

This study was performed using an Acer Swift 5 Thin and Light Laptop Intel Core i7 11th gen to perform the simulation works. The hardware and software computer configuration specifications of this laptop are Windows 10 Home 64-bit, up to 4.2 GHz Intel Core i7, 16.0 GB random-access memory, 512 GB solid-state drive storage, and MATLAB version R2021a.

3 Results and Discussions

Figure 1 presented the combination effect of stirring speed, the mass of OPT, and hydrolysis time on glucose production. The size of bubbles indicated the amount of glucose produced in mmol/L. Big bubbles were observed in Fig. 1(a) at the region 200–250 rpm and >20 g of OPT. This showed that high stirring speed and high amount of OPT produced were desired as a high amount of glucose was produced. This is because a high stirring speed increased the mass transfer, while the high mass of OPT provided more substrate for the enzyme to catalyse the hydrolysis process. In terms of hydrolysis time, 15–25 h was desired, as illustrated in Fig. 1(b). Under the optimised condition, the experiment was repeated by increasing the mass of OPT to 40 g. The result showed that the highest concentration of glucose, 30.1 mmol/L, was produced by using 30 g of OPT, 225 rpm for 16 h at 60 °C.

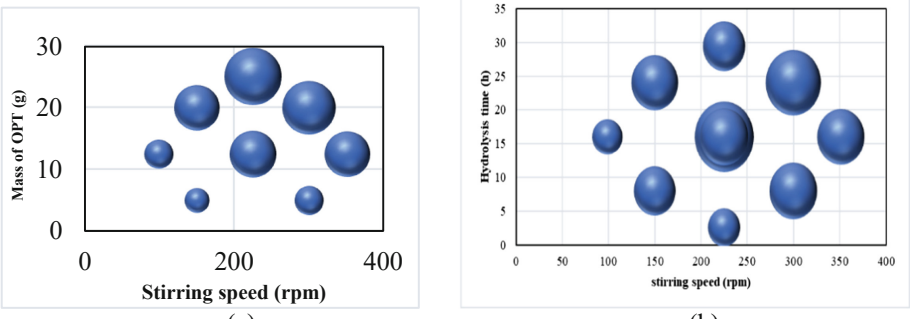

Fig. 1. The effect of stirring speed and (a) mass of OPT (g), (b) hydrolysis time on glucose production.

In this study, the experimental data for the immobilised enzymes hydrolysis of the oil palm trunks were utilised to build LW-KPLSR, LW-PLSR, PLSR, and PCR models. To evaluate the predictive performance of these regression models, *RMSE*, and R^2 for both training and testing data as well as E_a were calculated and tabulated in Table 2. $RMSE_1$ and $RMSE_2$ represent the *RMSE* for training and testing data, respectively. Meanwhile, the ${R_1}^2$ and ${R_2}^2$ are the R^2 for training and testing data, respectively. From Table 2, it can be seen that LW-KPLSR has the lowest value of *Ea* while both of its R^2 are more than 0.87. From Table 2, notice that the E_a for LW-KPLSR is 195%, 120%, and 103% lower than LW-PLSR, PLSR, and PCR, respectively.

Table 2. Predictive performance results from LW-KPLSR, LW-PLSR, PLSR and PCR models.

Models	LW-KPLSR	LW-PLSR	PE (%)	PLSR	PE (%)	PCR	PE (%)
$RMSE_1$	2.1555	1.9713	9	3.8111	77	4.4279	48
${R_1}^2$	0.9055	0.9256	2	0.6509	28	0.4631	121
$RMSE_2$	2.2505	5.8456	160	2.2891	2	2.1876	101
${R_2}^2$	0.8784	0.6932	21	0.8252	6	0.7334	97
E_a	2.2716	6.7065	195	4.9949	120	6.1703	103

Although LW-PLSR has slightly better $RMSE_1$ and ${R_1}^2$ than LW-KPLSR, its $RMSE_2$ and ${R_2}^2$ are 160% and 21% higher than LW-KPLSR. It is due to the help of polynomial Kernel in LW-KPLSR which ables to map the datasets into a higher dimension to obtain a better prediction [20]. Also, the polynomial Kernel function is one of the famous Kernel functions in machinery learning applications [31]. On the other hand, LW-KPLSR and LW-PLSR also provide better results for their training data as compared to PLSR and PCR in which its $RMSE_1$ is lower and its ${R_1}^2$ is higher. These results could due to the presence of a locally weighted model both LW-KPLSR and LW-PLSR which utilises a weighted Euclidean distance-based approach to choose the more relevant historical data for better prediction [32]. The absence of polynomial Kernel function in the LW-PLSR causes its poorer prediction for the testing data than PLSR and PCR.

Additionally, PLSR produces better predictive results than PCR where its *Ea* and $RMSE_1$ are lower and its ${R_1}^2$ and ${R_2}^2$ are higher than PCR. This is because PLSR includes both input and output variables in its model development while PCR is only involved the input variables [33]. Hence, the PLSR transformation objective can be done by finding the maxima between the input and output variables to describe better variance for prediction but PCR can only maximise the block coverage within the input variables [34, 35]. Nevertheless, the LW-KPLSR model is still the best predictive model to estimate the glucose concentration for the immobilised enzymes hydrolysis of the oil palm trunks. As can be seen from Fig. 2(a) and Fig. 2(b), the predicted glucose concentrations from LW-KPLSR are closer to the actual glucose concentration as compared to LW-PLSR, PLSR, and PCR. In conclusion, LW-KPLSR is more appropriate to be used for the prediction of glucose concentration from the immobilised enzymes hydrolysis of the oil palm trunks.

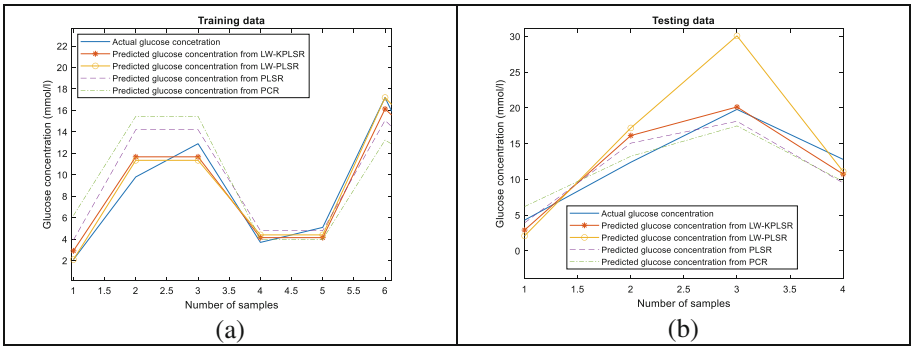

Fig. 2. Comparisons between the actual and predicted glucose concentrations from different least square regression models (a) for training data, (b) for testing data.

4 Conclusions

The experimental work indicated that stirring speed, hydrolysis time, and mass of OPT contributed significant impacts to the enzymatic process for glucose production. The result showed that the highest concentration of glucose, 30.1 mmol/L, was produced by using 30 g of OPT, 225 rpm for 16 h at 60 °C. In this study, an LW-KPLSR model was developed to predict the glucose concentration from the immobilised enzymes hydrolysis of OPT. For the overall predictive performance, the LW-KPLSR model predicted more accurate glucose concentrations than LW-PLSR, PLSR, and PCR models since its E_a value is 103% to 195% lower. Moreover, its R^2 values are more than 0.8 and also closer to 1 in which results indicate that the deviation between the actual and predicted glucose concentrations is not big. Hence, it can be concluded that LW-KPLSR is suitable to be used to estimate the glucose concentration for this hydrolysis process. It is suggested to further study more different experimental data of glucose concentration hydrolysed from oil palm trunks, then more data can be used to develop a PLSR-based model which can lead to more accurate predictive performance.

Acknowledgments. The authors would like to thank the supports from Curtin University Malaysia, SEGi University, and Malaysian Palm Oil Board (MPOB). Also, the authors declare no conflict of interest.

References

1. Analytics, P.O.: Palm Oil Analytics. Singapore (2017)
2. Ooi, L.C.L., et al.: SureSawitTM True-To-Type—A high throughput universal single nucleotide polymorphism panel for DNA fingerprinting, purity testing and origin verification in oil palm. J. Oil Palm Res. **31**, 561–571 (2019)
3. Kushairi, A., et al.: Oil palm economic performance in Malaysia and R&D progress in 2018. J. Oil Palm Res. **31**(2), 165–194 (2019)

4. Sulaiman, O., Salim, N., Nordin, N.A., Hashim, R., Ibrahim, M., Sato, M.: The potential of oil palm trunk biomass as an alternative source for compressed wood. BioResources **7**(2), 2688–2706 (2012)
5. Azmi, A., Malek, M., Puad, N.: A review on acid and enzymatic hydrolyses of sago starch. Int. Food Res. J. **24**(12), 265–273 (2017)
6. Bukhari, N.A., Loh, S.K., Bakar, N.A., Ismail, M.: Hydrolysis of residual starch from sago pith residue and its fermentation to bioethanol. Biores. Technol. **46**(8), 1269–1278 (2017)
7. Ude, M.U., Oluka, I., Eze, P.C.: Optimization and kinetics of glucose production via enzymatic hydrolysis of mixed peels. J. Biosour. Bioprod. **5**(4), 283–290 (2020)
8. Samaranayake, M.D., De Silv, A.B.: Optimization of liquefaction and sacch production of glucose syrup from Cas optimized conditi. J. Chem. Res. **7**(7), 16–25 (2017)
9. Acosta-Pavas, J.C., Alzate-Blandon, L., Ruiz-Colorado, A.A.: Enzymatic hydrolysis of wheat starch for glucose syrup production. Dyna **87**(214), 173–182 (2020)
10. Choi, B., Rempala, G.A., Kim, J.K.: Beyond the Michaelis-Menten equation: accurate and efficient estimation of enzyme kinetic parameters. Sci. Rep. **7**(1), 1–11 (2017)
11. Boeckx, J., Hertog, M., Geeraerd, A., Nicolai, B.: Kinetic modelling: an integrated approach to analyze enzyme activity assays. Plant Methods **13**(1), 1–12 (2017)
12. Dobbelaere, M.R., Plehiers, P.P., Van de Vijver, R., Stevens, C.V., Van Geem, K.M.: Machine learning in chemical engineering: strengths, weaknesses, opportunities, and threats. Engineering (2021)
13. Purwanto, A., Research, M.: Partial least squares structural squation modeling (PLS-SEM) analysis for social and management research: a literature review. J. Ind. Eng. Manag. Res. **2** (4), 114–123 (2021)
14. Khatri, P., Gupta, K.K., Gupta, R.K.: Environment: a review of partial least squares modeling (PLSM) for water quality analysis. Model. Earth Syst. Environ. **7**(2), 703–714 (2021)
15. Martínez, J.L., Leiva, V., Saulo, H., Liu, S.: Estimating the covariance matrix of the coefficient estimator in multivariate partial least squares regression with chemical applications. Chemom. Intell. Lab. Syst. **214**, 104328 (2021)
16. Cammnitiello, I., Lombardo, R., Durand, J.-F.J.Q.: Quantity: robust partial least squares regression for the evaluation of justice court delay. **51**(2), 813–827 (2017)
17. Yeo, W.S., Saptoro, A., Kumar, P.: Development of adaptive soft sensor using locally weighted Kernel partial least square model. Chem. Prod. Process. Model. **12**(4), 1–13 (2017)
18. Eom, I.-Y., Yu, J.-H., Jung, C.-D., Hong, K.-S.: Efficient ethanol production from dried oil palm trunk treated by hydrothermolysis and subsequent enzymatic hydrolysis. Biotechnol. Biofuels **8**(1), 1–11 (2015)
19. Philis-Tsimikas, A., Chang, A., Miller, L.J.J.O.D.S.: Technology: precision, accuracy, and user acceptance of the OneTouch SelectSimple blood glucose monitoring system. **5**(6), 1602–1609 (2011)
20. Yeo, W.S., Saptoro, A., Kumar, P.J.I., Research, E.C.: Adaptive soft sensor development for non-Gaussian and nonlinear processes. **58**(45), 20680–20691 (2019)
21. Ren, M., Song, Y., Chu, W.: An improved locally weighted PLS based on particle swarm optimization for industrial soft sensor modeling. Sensors **19**(19), 4099 (2019)
22. Ma, M., Khatibisepehr, S., Huang, B.: AB ayesian framework for real-time identification of locally weighted partial least squares. AIChE J. **61**(2), 518–529 (2015)
23. Mongillo, M.: Choosing basis functions and shape parameters for radial basis function methods. SIAM Undergraduate Research Online **4**(190–209), 2–6 (2011)
24. Orr, M.J.: Technical Report, Center for Cognitive Science. University of Edinburgh (1996)
25. Yeo, W.S., Lau, W.J.: Predicting the whiteness index of cotton fabric with a least squares model. Cellulose **28**(13), 8841–8854 (2021). https://doi.org/10.1007/s10570-021-04096-y

26. Harmel, R.D., Smith, P.K., Migliaccio, K.W.: Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. Trans. ASABE **53**(1), 55–63 (2010)
27. Dou, Y., Sun, Y., Ren, Y., Ren, Y.: Artificial neural network for simultaneous determination of two components of compound paracetamol and diphenhydramine hydrochloride powder on NIR spectroscopy. Anal. Chim. Acta **528**(1), 55–61 (2005)
28. Saptoro, A., Vuthaluru, H., Tadé, M.: Presented at the Proceedings of the International Conference on Modeling and Simulation (2006)
29. Yeo, W.S., Saptoro, A., Kumar, P.: Missing data treatment for locally weighted partial least square-based modelling: a comparative study. Asia-Pac. J. Chem. Eng. **15**(2), e2422 (2020)
30. Guang, W., Baraldo, M., Furlanut, M.: Calculating percentage prediction error: a user's note. Pharmacol. Res. **32**(4), 241–248 (1995)
31. Zhou, J., Zeng, S., Zhang, B.: Kernel nonnegative representation-based classifier. applied Intelligence. 1-21 (2021)
32. Kano, M., Fujiwara, K.: Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications. J. Chem. Eng. Japan. 12we167 (2012)
33. Thien, T.F., Yeo, W.S.: A comparative study between PCR, PLSR, and LW-PLS on the predictive performance at different data splitting ratios. Chem. Eng. Commun. 1–18 (2021)
34. Cramer, R.D.: Design: partial least squares (PLS): its strengths and limitations. Perspect. Drug Discovery Des. **1**(2), 269–278 (1993)
35. Carrascal, L.M., Galván, I., Gordo, O.: Partial least squares regression as an alternative to current regression methods used in ecology. Oikos **118**(5), 681–690 (2009)