



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2023, 2023:49

<https://doi.org/10.28919/cmbn/7941>

ISSN: 2052-2541

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO USING BIRESPONSE MULTIPREDICTOR LOCAL POLYNOMIAL NONPARAMETRIC REGRESSION

MILLATUL ULYA^{1,2}, NUR CHAMIDAH^{3,4,*}, TOHA SAIFUDIN^{3,4}

¹Doctoral Study Program of Mathematics and Natural Sciences, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

²Department of Agroindustrial Technology, Faculty of Agriculture, Universitas Trunojoyo Madura, Bangkalan 69162, Indonesia

³Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

⁴Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology, Airlangga University, Surabaya 60115, Indonesia

Copyright © 2023 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Mango's internal quality can be determined based on its acidity and sweetness in the form of pH and total soluble solids (TSS) content. Research on fruit internal quality prediction based on near-infrared spectroscopy generally uses parametric regression modeling such as linear and partial least square regression. The study proposed biresponse multipredictor local polynomial nonparametric regression to determine mango's internal quality. The study aims to apply the theory of biresponse multipredictor local polynomial nonparametric regression for predicting the mango's internal quality in the form of pH and TSS value. We created R code for estimating nonparametric

*Corresponding author

E-mail address: nur-c@fst.unair.ac.id

Received March 04, 2023

regression model of the biresponse multipredictor local polynomial. The predicted performance is determined by Mean Absolute Percentage Error (MAPE) value. The MAPE value of mango's pH and TSS prediction is 4.473%. This means that the proposed method was highly accurate for predicting the mango maturity because the MAPE value is less than 10%.

Keywords: prediction, local polynomial; nonparametric regression; mango; quality.

2020 AMS Subject Classification: 62G05, 62G08, 62P95.

1. INTRODUCTION

Generally, the internal quality of fruit can be evaluated using two approaches: destructive and non-destructive. Near-infrared spectroscopy (NIRS) has become one of the most promising and widely utilized non-destructive methods of investigation in many fields, including agriculture [1]. It has many advantages, such as rapid method, simple sample preparation, and environmental friendliness because no chemicals are employed, and minimal waste is produced [2]. Significantly, it can evaluate many quality attributes simultaneously [3]. One of the uses of NIRS in agriculture is to predict the internal quality of fruit and vegetables.

Mango is a climacteric fruit that is usually consumed when it is mature. Because of its appearance, taste, flavour, and overall nutritional benefits, mango (*Mangifera indica* L.) is one of the most significant and popular tropical fruits commercialized globally [4]. At maturity, most mango varieties are green and have low firmness, but external color and firmness alone are insufficient to suggest maturity [5]. Mature mangoes have low acidity and a high degree of sweetness. The pH value represents the acidity level of the fruit, while the Total Soluble Solids (TSS) content value represents the level of fruit sweetness. These two variables are indicators of the internal quality of mango. According to recent studies, pH and TSS content are valid parameters for mango maturity prediction. The pH and TSS value of fruit can valid evaluated using destructive analysis in the laboratory, which is complex sample preparation, time-consuming, laborious, and produces waste. Therefore, it is better to predict pH and TSS values non-destructively based on NIRS.

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

Supply chain stakeholders need to manage fruit ripening to give customers the most excellent quality fruit [6]. Over the last few decades, NIRS has been commonly applied as a reliable and speedy non-destructive method for predicting fruit quality [7]. Regression analysis is a statistical tool for examining the functional relationship between response and predictor variables and it is used in the NIRS-based prediction of fruit internal quality. There are two methods for analyzing regression: parametric regression and nonparametric regression. The majority of previous mango quality prediction research focused on parametric regression techniques such as simple linear regression [8–9], multiple linear regression [9–12], partial least square regression (PLSR) [10–13], principal component regression [10–11] and Kernel PLSR [14–15]. Ulya et al. [16–17] performed a nonparametric regression prediction of mango fruit quality based on an uniresponse multipredictor local polynomial estimator.

In general, the nonparametric regression can be applied to multiresponse and multipredictor cases. Estimation of the nonparametric regression models for these cases is equivalent to estimation of their regression functions. To estimate these nonparametric regression functions, there are some estimators which are often used, for examples local linear estimators [18–21], local polynomial estimators [22–26], kernel estimators [27–30], and spline estimators [31–37]. One of the spline estimators in nonparametric regression model approach which has superior in smoothing capabilities for analyzing biresponse data is local polynomial estimator. Also, the nonparametric regression can be applied to biresponse and multipredictor matters. The biresponse local polynomial nonparametric regression can be implemented to various areas, including estimating the growth curve of children for two years using a biresponse local linear estimator [38], and modeling of children's weight in East Java [39]. Some cases of biresponse and multipredictor include predicting the internal quality of mango based on absorbance spectral data of NIRS [40].

The theory of estimating a nonparametric regression model based on a biresponse multipredictor local polynomial estimator has been developed by Ulya et al. [40]. Its results show that, theoretically, a nonparametric regression approach based on a biresponse

multipredictor local polynomial estimator can predict mangoes' internal quality. However, its method only applies to simulation studies using generated data. It has not been implemented directly in actual cases. Therefore, further research is needed to implement the estimation theory of a nonparametric regression model based on a biresponse multipredictor local polynomial estimator to predict mangoes' pH and TSS values.

The study aims to apply a theory of nonparametric regression estimation of biresponse multipredictor based on local polynomial estimators for predicting the pH and TSS value of mango. Next, develop an algorithm for the estimation process and run the R code to estimate the predicted value of pH and TSS content. The study reveals that the nonparametric regression based on biresponse multipredictor local polynomial estimator is very accurate in predicting the pH and TSS values of mangoes. It is expected that the algorithm and R code developed in this study would serve as the basis for designing an instrument to detect mango maturity.

2. MATERIALS AND METHODS

Estimating local polynomial nonparametric regression model for biresponse multipredictor cases has been studied by [40]. The model can be used to predict the mango's internal quality, such as pH and total soluble solids that represent the acidity and sweetness of mango. The local polynomial nonparametric regression model for biresponse multipredictor is presented in the following Theorem and Lemma [40].

Theorem. *Given the data in pairs $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i^{(r)})$, $i = 1, 2, \dots, n$ where i is the number of observations, p is the number of predictor variables, and r is the number of response variables. They meet the biresponse multipredictor nonparametric regression model as follows:*

$$(1) \quad y_i^{(r)} = \sum_{j=1}^p f_r(x_{ij}) + \varepsilon_i^{(r)}$$

where $\underline{y}_i = (y_i^{(1)}, y_i^{(2)})^T$, $\underline{f}(x_{ij}) = (f_1(x_{ij}), f_2(x_{ij}))^T$ is a function of unknown shape and $\varepsilon_i = (\varepsilon_i^{(1)}, \varepsilon_i^{(2)})^T$ is the measurement error with mean is $\mathbf{0}$, and variance is Σ_i , so local polynomial estimator for

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

estimating nonparametric regression function of biresponse multipredictor at x_j around x_{0j} is:

$$(2) \quad \hat{f}(x_j) = x^*(x_0) \left((\mathbf{X}^T(x_0) \mathbf{K}_h(x_0) \mathbf{V}^{-1} \mathbf{X}(x_0))^{-1} \mathbf{X}^T(x_0) \mathbf{K}_h(x_0) \mathbf{V}^{-1} y \right)$$

where, $x^*(x_0) = \begin{pmatrix} x_1^*(x_0) & 0 \\ 0 & x_2^*(x_0) \end{pmatrix}$ and $x^*(x_0)$ is a vector of size $2n \times (2+d_1+d_2)$, $\beta(x_0)$ is a vector

of size $(2+d_1+d_2) \times 1$ where $d_1 = \sum_{j=1}^p d_j^{(1)}$ and $d_2 = \sum_{j=1}^p d_j^{(2)}$. ε is a vector of size $2n \times 1$. d_1 and d_2 are

the polynomial degree of the first and second response, and d_j is the polynomial degree of the j^{th} predictor where $j = 1, 2, \dots, p$.

Lemma. In biresponse multipredictor local polynomial nonparametric regression, there is a correlation between the responses, enabling model estimation using the local polynomial estimator. In addition to kernel weighting, we apply the following weighting matrix for the variance and covariance (V) of the vector ε :

$$(3) \quad \mathbf{V} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\Sigma_{rr} = \text{diag}(\sigma_1^{2(r)}, \sigma_2^{2(r)}, \dots, \sigma_n^{2(r)})$, $\Sigma_{rs} = \Sigma_{sr} = \text{diag}(\sigma_1^{(r)(s)}, \sigma_2^{(r)(s)}, \dots, \sigma_n^{(r)(s)})$. V is a weighting matrix of size $2n \times 2n$. Σ_{rr} For $r = 1, 2$ is a diagonal matrix containing the k^{th} response error variance.

The study used experimental data from 186 mango samples tested using NIRS to obtain spectral data and validated by destructive pH and TSS analysis in the laboratory. The type of mango in this study was *Gadung Klonal 21*, harvested from Sukorejo District, Pasuruan Regency, Indonesia. The predictor variables were the absorbance spectral data of NIRS that used wavelength of 900–1650 nm. In contrast, the response variables were the acidity and sweetness of mango in the form of pH and TSS content.

The study consists of several stages as follows: (1). The data collection experimentally using 186 samples of mango *Gadung Klonal 21*. The samples were analyzed the spectral data using

NIR spectrometer then validated the pH and TSS values using destructive analysis in the laboratory. (2). Spectral data pre-processing to minimize unwanted effects of spectral data. (3). The dimension reduction using Principal Component Analysis (PCA). (4). Outlier analysis using Hotelling's T^2 test. (5). Statistical descriptive analysis. (6). Predicting pH and TSS value of mango using biresponse multipredictor local polynomial nonparametric regression model. (7). Calculate the predictive performance criteria.

The study used Mean Absolute Percentage Error (MAPE) as the criteria of predictive performance. The formulation of MAPE is as follows [41]:

$$(4) \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|\hat{y}_i - y_i|}{y_i} \right) \times 100\%$$

$$(5) \quad \text{Overall } MAPE = \frac{((MAPE_{tr} \times n_{tr}) + (MAPE_{ts} \times n_{ts}))}{(n_{tr} + n_{ts})}$$

where n is the sample size of observations, \hat{y}_i is the predicted value of the i^{th} response variable, a y_i is the measured value of the response variable, tr the training data; and ts is the testing data.

The best prediction model has the lowest MAPE. Moreno *et al.* [41] classified the MAPE value interpretation into four classes, namely MAPE value less than 10% is a highly accurate prediction, 10–20% is an accurate prediction, 20–50% MAPE is a reasonable prediction and MAPE value more than 50% is inaccurate prediction.

4. MAIN RESULTS

This section describes the implementation of nonparametric regression models based on the biresponse and multipredictor local polynomial estimator for predicting the mango's pH and TSS value.

4.1. Spectral Data Pre-processing

Before generating the prediction models, some pre-treatment approaches can minimize unwanted effects such as random noise, high-frequency noise, light scattering, and any other external

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

effects induced by environmental or instrumental variables. Smoothing also successfully decreases high-frequency noise. Savitzky-Golay (SG) smoothing is one of the area's most extensively utilized smoothing algorithms [42]. The Unscrambler X 10.4 software was used to perform spectral pre-treatment based on SG smoothing. Figure 1 shows the raw and pre-treated (SG smoothing) spectral data.

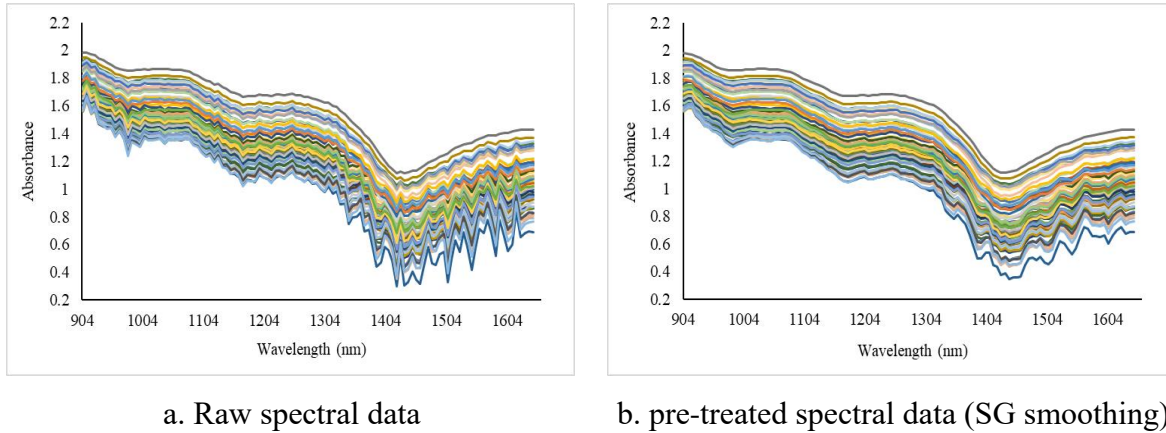


Figure 1. Absorbance Spectral Data

After pre-treatment of the absorbance spectral data, PCA was used to decrease the dimension of the spectral data using the singular value decomposition algorithm. Two latent variables were selected to explain 99.75% of the variation. This study used Hotelling's T^2 ellipse analysis to identify spectral outliers. There were 21 outliers detected in this analysis. These outliers were removed because they had the potential to affect the model. Outliers in a sample can give useful information, but they can also be non-representative samples contributing to model errors [43]. The final sample included 165 observations separated into two parts: 80% training and 20% testing datasets (132 training data samples and 33 testing data samples).

4.2. The Dimension Reduction and Outlier Analysis

Before dimension reduction and outlier analysis, matrixes were created from the pH, TSS, and spectral data with the size of 186 x 114. The 186 samples are represented by the matrix rows,

while the 114 columns represent the predictors (absorbance spectral data) and response variables (y_1 and y_2). The wavelengths of 112 NIR spectra from each mango sample were used as predictor variables (P_1, P_2, \dots, P_{112}). The response variables in the first and second columns detailed the measured pH and TSS values associated with each sample. After dimension reduction and outlier analysis, the matrixes size changed to 165 x 4. The 165 samples are represented the sample size. In contrast, the 4 columns represent the pH and TSS values, and the last two columns were the new predictor variables (x_1 and x_2) after dimension reduction using PCA. The biresponse multipredictor local polynomial nonparametric regression approach was carried out by using the open-source software R.

4.3. The Algorithm of Biresponse Multipredictor Local Polynomial Nonparametric Regression Model for Predicting pH and TSS Values of Mango.

The algorithm in this study refers to previous study by [40], with the two response variables, including pH and TSS value. While the predictor variables are two variables derived from spectral data dimension reduction. The optimal bandwidth is determined by finding the minimum Generalized Cross Validation (GCV) value in each response and predictor.

Algorithm for determining the optimal bandwidth value without a weighting matrix (\mathbf{V}^{-1}) is as follows:

Step 1: Defining the response variable $y^{(r)}$, $r=1,2$ where the first response variable is pH value and the second one is TSS value, and the predictor variables x_j , $j=1,2$

Step 2: Determine the kernel function utilized, based on the Gaussian Kernel, using the following equation:

$$(6) \quad \mathbf{K}_h(x) = \frac{1}{h} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x}{h}\right)^2\right), -\infty < x < \infty$$

Step 3: Defining the matrix $\mathbf{X}(x_0)$ according to Equation (2)

Step 4: Determine the bandwidth values partially on the 1 to d polynomial order on each

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

predictor for each response (h_c) where ll is the lower limit and ul is the upper limit of the bandwidth value, and δ is the addition of the bandwidth value for each iteration.

Step 5: Determine the diagonal matrix ($\mathbf{K}_h(x)$) for the response variable $y_c^{(r)}, r=1,2$

Step 6: Calculate the GCV value for each $h_c \in [seq(ll, ul, \delta)]$ on the order of 1 to d using the following Equation:

$$(7) \text{ GCV}(h) = \frac{n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \hat{f}(x_{ij}))^2}{(1 - \text{tr}[\mathbf{I} - \mathbf{A}(h_j)] / n)^2} \text{ where } n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \hat{f}(x_{ij}))^2 \text{ is the Mean}$$

Square Error (MSE)

Step 7: The set of bandwidth values that produces the minimum GCV value is the optimal bandwidth value on a predetermined order

In addition to the algorithm above, this study also uses an algorithm for establishing a weighting matrix (\mathbf{V}^{-1}) and an algorithm for model estimation using a weighting matrix (\mathbf{V}^{-1}). The use of the two algorithms in this study refers to [40].

4.4. Statistical Descriptive of the pH and TSS Values

Before performing nonparametric regression analysis, it is necessary to analyze descriptively about the data. Table 1 presents a statistical description of the pH and TSS values of 186 mango samples. While Table 2 shows a statistical description of the pH and TSS values of 165 samples and the two predictor variables. The two predictor variables (x_1 and x_2) were the new variables that resulted from the dimension reduction of absorbance spectral data on 112 wavelengths (P_1, P_2, \dots, P_{112}).

Table 1. Statistical Descriptive of pH and TSS Values of 186 Mango Samples.

Description	pH	TSS (°Brix)
Mean	4.29	5.60
Median	4.12	5.70
Deviation standard	0.61	1.29
Minimum	2.80	2.10
Maximum	5.79	9.00

Table 2. Statistical Descriptive of pH, TSS, x_1 and x_2 Values of 165 Mango Samples.

Description	pH	TSS (°Brix)	X_1	X_2
Mean	4.27	5.55	-0.07	-0.03
Median	4.12	5.40	-0.12	-0.01
Deviation standard	0.47	1.02	1.37	0.20
Minimum	3.03	3.50	-2.91	-0.49
Maximum	5.66	8.40	3.47	0.42

The next stage was made a scatter plot between the pH value with x_1 and x_2 and the TSS value with x_1 and x_2 . The scatter plot can indicate whether there is a specific trend in the relationship between the response and predictor variable, as well as indications of cases of heteroscedasticity. Figures 2 and 3 show the scatter plot of pH and TSS values vs. the predictor variables.

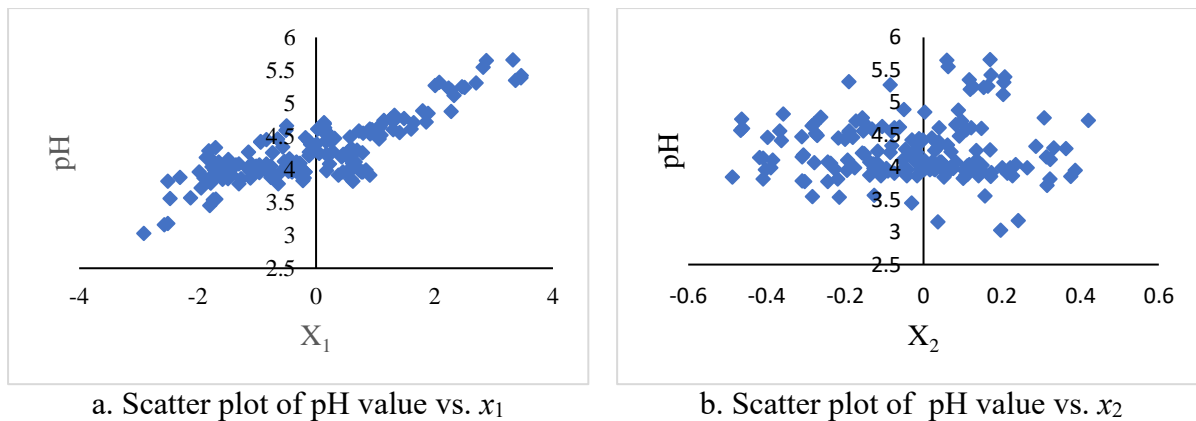
**Figure 2.** Scatter Plot of pH Value vs. The Predictor Variables

Figure 2 shows the scatter plot between pH value against x_1 and x_2 , which shows a linear trend with a steep slope (Fig. 2a) and gentle slope, and the data is spread out (Fig. 2b). It was reinforced by the results of the linearity test which showed that the significance values for deviation from linearity were 0.591 (pH vs. X_1) and 0.068 (pH vs. X_2), which means that there is a linear relationship between pH and x_1 and x_2 .

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

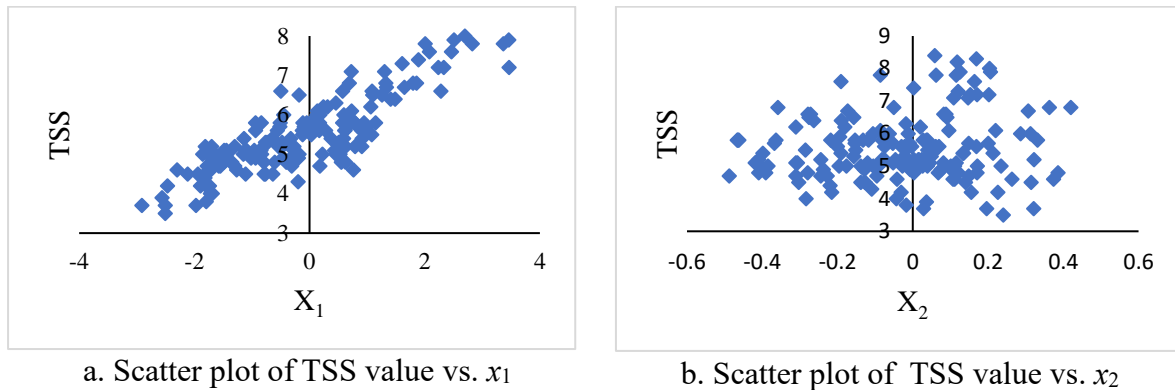


Figure 3. Scatter Plot of TSS Value vs. The Predictor Variables

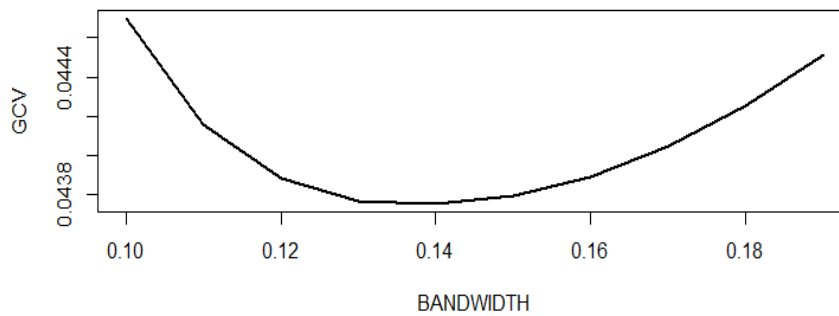
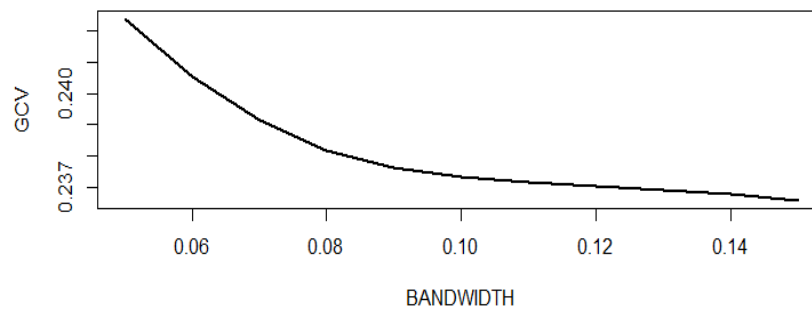
Figure 3 provides the scatter plot between the TSS value against x_1 and x_2 . Figure 3a shows a linear trend with a steep slope, and Figure 2b shows no trend or data spread. It was proved by the results of the linearity test, which showed that the significance values for deviation from linearity were 0.053 (TSS vs. x_1), meaning there is a strong linear relationship between TSS and x_1 . While the significance values for deviation from linearity of TSS vs. x_1 were 0.001, there is no linear relationship between the TSS value and x_2 .

4.5. Determination of the Optimal Bandwidth

The next step is determining the optimal bandwidth value without weighting matrix (\mathbf{V}^{-1}) in the training data. The bandwidth determination is carried out partially in each response and predictor. Table 3 shows the optimal bandwidth determination based on the minimum GCV. Based on Table 3, the best order polynomial is first order that resulted minimum GCV. Besides, the linearity test also shows a linear relationship between the response variables, especially with x_1 . So, the nonparametric regression modeling process will use a first-order polynomial. This step is expected to produce a regression model that fits the observation data.

Table 3. Optimal Bandwidth Determination Partially Based on Minimum GCV.

Response	Predictor	Polynomial Order	Optimal bandwidth	Minimum GCV
1	X ₁	1	0.14	0.04375*
		2	0.22	0.04379
		3	0.23	0.04376
	X ₂	1	0.15	0.23362*
		2	0.14	0.2327665
		3	0.16	0.23862
2	X ₁	1	0.14	0.04375*
		2	0.22	0.23204
		3	0.22	0.23317
	X ₂	1	0.09	1.06505
		2	0.11	1.07053
		3	0.13	1.07462

**Figure 4.** The Plot of Bandwidth vs. GCV (y_1, x_1)**Figure 5.** The Plot of Bandwidth vs. GCV (y_1, x_2)

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

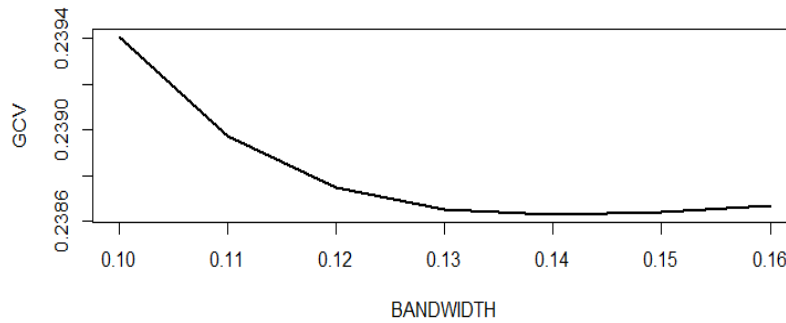


Figure 6. The Plot of Bandwidth vs. GCV (y_2, x_1)

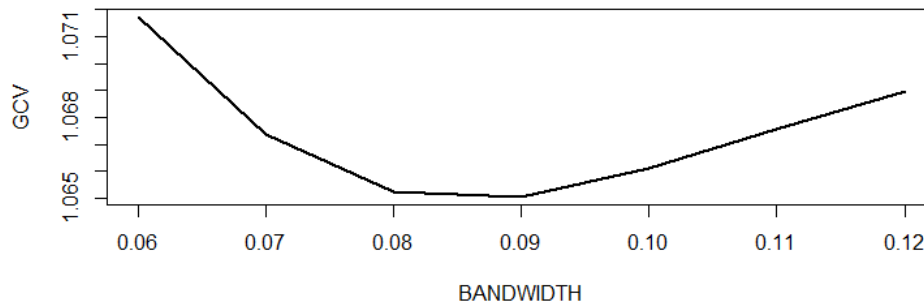


Figure 7. The Plot of Bandwidth vs. GCV (y_2, x_2)

Based on Table 3, the best order polynomial is first order that resulted minimum GCV on each predictor for each response. So, the best combination bandwidths are shown in Table 4.

Table 4. The Optimal Bandwidth Combination with Minimum GCV Value

The predictor variables	Polynomial order	The bandwidth value of the response variable	
		pH	TSS
X_1	1	0.14	0.15
X_2	1	0.14	0.09

Furthermore, the bandwidth values in Table 4 are used for estimating the predicted value of pH and TSS, so the first and the second response error 2 values will be obtained, which will be used to create the (\mathbf{V}^{-1}) weighting matrix. The next stage estimates the biresponse multipredictor nonparametric regression model based on a local polynomial estimator using a weighting matrix (\mathbf{V}^{-1}).

A summary of the estimation results for the best combination can be seen in Table 5, which shows the MAPE and MSE values for the training, testing, and overall data. The observation plots and the results of the estimation of the two responses on training data can be seen in Figures 8 and 9. While the plots of testing data can be seen in Figures 10 and 11.

Table 5. The MAPE and MSE Values of Training, Testing, and Overall Data.

The criteria	Data		
	Training	Testing	Overall
MAPE	3.729 %	7.466 %	4.476 %
MSE	1,185	0.219	0.992

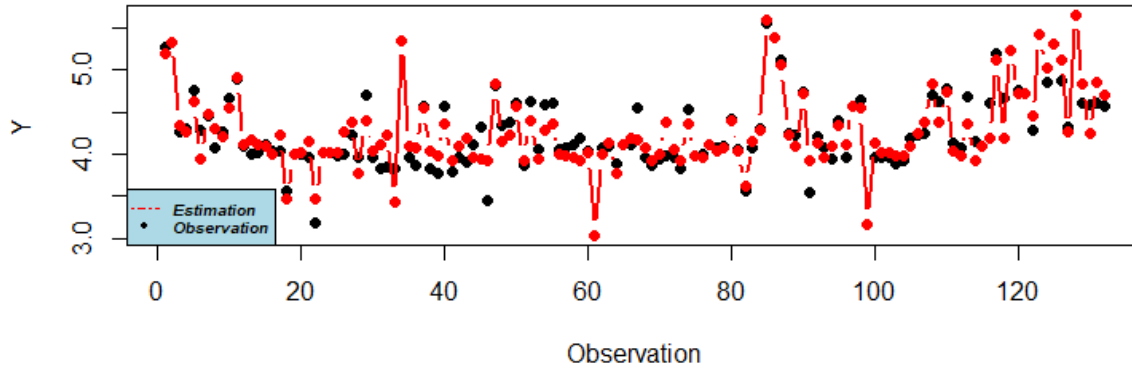


Figure 8. The Plot of Observation vs. Estimation of pH Value in Training Data.

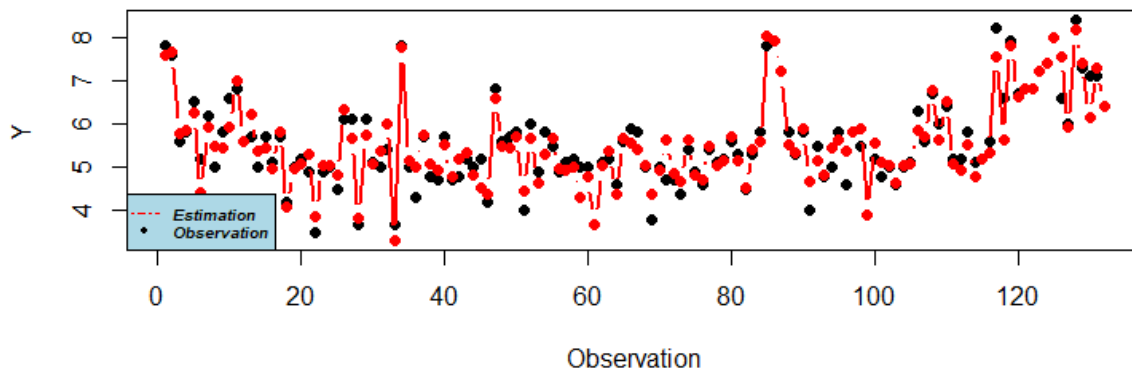


Figure 9. The Plot of Observation vs. Estimation of TSS Value in Training Data.

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

Figures 8 and 9 show that the predicted value is close to the observed value. It was 89% of training data that has absolute percentage error less than 10%. It illustrated that the model on sample data that was formed is very good. The results of estimating the pH and TSS values using biresponse multipredictor nonparametric regression based on the local polynomial estimator on training data yield a MAPE value of 3.729%. The MAPE value is still less than 10% which is included in the category of highly accurate prediction [41].

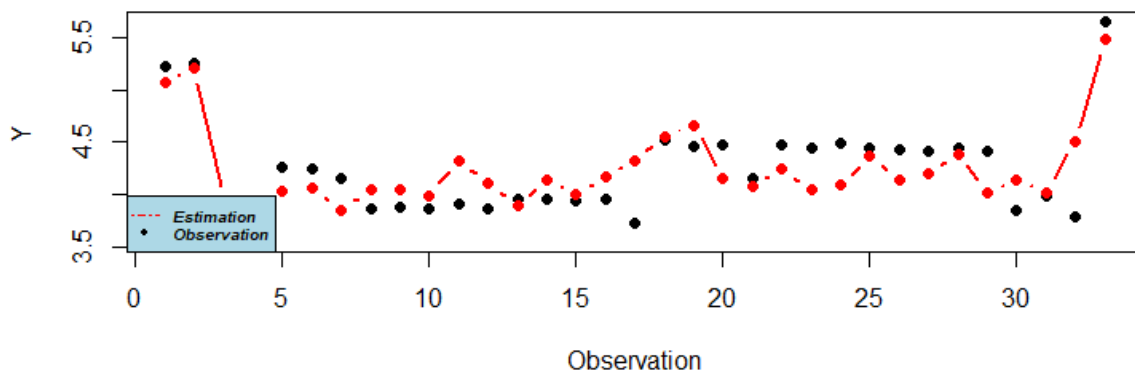


Figure 10. The Plot of Observation vs. Estimation of pH Value in Testing Data.

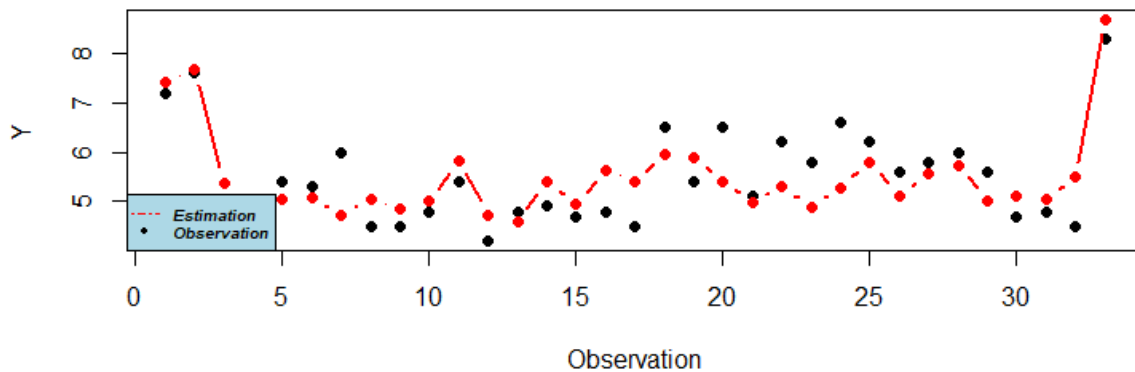


Figure 11. The Plot of Observation vs. Estimation of TSS Value in Testing Data.

Figures 10 and 11 show that that the predicted value is quite close to the observed value on testing data. The results of estimating the pH and TSS values using biresponse multipredictor nonparametric regression based on the local polynomial estimator on testing data yield a MAPE value of 7.466%. The nonparametric regression model is highly accurate in predicting mango's pH and TSS value on testing

data with less than 10% MAPE value [41].

The MAPE value of overall data in this study is 4.473%. The study proves that the bi-responses multi-predictors local polynomial nonparametric regression model is highly accurate for predicting the mango's pH and TSS values. It is in line with [16–17] that local polynomial nonparametric regression is highly accurate for predicting the sweetness and acidity of mangoes. In fact, the MAPE values of local polynomial nonparametric regression is lower than the MAPE value of parametric regression on the case of predicting mango's pH and TSS values [16–17].

4. CONCLUSIONS

The biresponse multipredictor local polynomial nonparametric regression model estimation can be implemented to solve the prediction of mango's internal quality. The study shows that the biresponse multipredictor local polynomial nonparametric regression is highly accurate in predicting the pH and TSS of mango with the MAPE value of 4.473%. The algorithm and R code developed in this study are expected to be useful for designing an instrument to detect mango maturity.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

- [1] C. Pasquini, Near infrared spectroscopy: A mature analytical technique with new perspectives – A review, *Analytica Chimica Acta*. 1026 (2018), 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>.
- [2] Yusmanizar, I.S. Setiasih, S. Nurjanah, et al. Fast and non-destructive prediction of moisture content and chologenic acid of intact coffee beans using near infrared reflectance spectroscopy, *IOP Conf. Ser.: Mater. Sci. Eng.* 506 (2019), 012033. <https://doi.org/10.1088/1757-899x/506/1/012033>.
- [3] A.A. Munawar, D. von Hörsten, J.K. Wegener, et al. Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics, *Eng. Agric. Environ. Food*. 9 (2016), 208–215. <https://doi.org/10.1016/j.eaef.2015.12.004>.

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

- [4] S.N. Jha, K. Narsaiah, P. Jaiswal, et al. Nondestructive prediction of maturity of mango using near infrared spectroscopy, *J. Food Eng.* 124 (2014), 152–157. <https://doi.org/10.1016/j.jfoodeng.2013.10.012>.
- [5] S. Sohaib Ali Shah, A. Zeb, W.S. Qureshi, Mango maturity classification instead of maturity index estimation: A new approach towards handheld NIR spectroscopy, *Infrared Phys. Technol.* 115 (2021), 103639. <https://doi.org/10.1016/j.infrared.2021.103639>.
- [6] T. Nordey, J. Joas, F. Davrieux, et al. Robust NIRS models for non-destructive prediction of mango internal quality, *Sci. Horticult.* 216 (2017), 51–57.
- [7] B.M. Nicolai, T. Defraeye, B. De Ketelaere, et al. Nondestructive measurement of fruit and vegetable quality, *Annu. Rev. Food Sci. Technol.* 5 (2014), 285–312. <https://doi.org/10.1146/annurev-food-030713-092410>.
- [8] C. Watanawan, T. Wasusri, V. Srilaong, et al. Near infrared spectroscopic evaluation of fruit maturity and quality of export Thai mango (*Mangifera indica* L. var. Namdokmai), *Int. Food Res. J.* 21 (2014), 1073–1078.
- [9] K. Schulze, M. Nagle, W. Spreer, et al. Development and assessment of different modeling approaches for size-mass estimation of mango fruits (*Mangifera indica* L., cv. ‘Nam Dokmai’), *Computers Electronics Agric.* 114 (2015), 269–276. <https://doi.org/10.1016/j.compag.2015.04.013>.
- [10] S.N. Jha, A.R.P. Kingsly, S. Chopra, Non-destructive determination of firmness and yellowness of mango during growth and storage using visual spectroscopy, *Biosyst. Eng.* 94 (2006), 397–402. <https://doi.org/10.1016/j.biosystemseng.2006.03.009>.
- [11] S.N. Jha, S. Chopra, A.R.P. Kingsly, Modeling of color values for nondestructive evaluation of maturity of mango, *J. Food Eng.* 78 (2007), 22–26. <https://doi.org/10.1016/j.jfoodeng.2005.08.048>.
- [12] S.N. Jha, P. Jaiswal, K. Narsaiah, et al. Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy, *Sci. Horticult.* 138 (2012), 171–175. <https://doi.org/10.1016/j.scienta.2012.02.031>.
- [13] P. Rungpichayapichet, B. Mahayothee, M. Nagle, et al. Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango, *Postharv. Biol. Technol.* 111 (2016), 31–40. <https://doi.org/10.1016/j.postharvbio.2015.07.006>.
- [14] B.M. Nicolai, K.I. Theron, J. Lammertyn, Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple, *Chemometrics Intell. Lab. Syst.* 85 (2007), 243–252. <https://doi.org/10.1016/j.chemolab.2006.07.001>.

- [15] Y. Bao, J. Liu, Y. Zhong, et al. Kernel partial least squares model for pectin content in peach using near - infrared spectroscopy, *Int. J. Food Sci. Technol.* 56 (2020), 1877-1885. <https://doi.org/10.1111/ijfs.14817>.
- [16] M. Ulya, N. Chamidah, Multi-predictor local polynomial regression for predicting the acidity level of avomango (Gadung Klonal 21), *AIP Conf. Proc.* 2329 (2021), 060024. <https://doi.org/10.1063/5.0042290>.
- [17] M. Ulya, N. Chamidah, T. Saifudin, Predicting the sweetness level of avomango (Gadung Klonal 21) using multi-predictor local polynomial regression, *IOP Conf. Ser.: Earth Environ. Sci.* 733 (2021), 012009. <https://doi.org/10.1088/1755-1315/733/1/012009>.
- [18] N. Chamidah, E. Tjahjono, A.R. Fadilah, et al. Standard growth charts for weight of children in East Java using local linear estimator, *J. Phys.: Conf. Ser.* 1097 (2018), 012092. <https://doi.org/10.1088/1742-6596/1097/1/012092>.
- [19] E. Ana, N. Chamidah, P. Andriani, et al. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression, *J. Phys.: Conf. Ser.* 1397 (2019), 012067. <https://doi.org/10.1088/1742-6596/1397/1/012067>.
- [20] N. Chamidah, Y.S. Yonani, E. Ana, et al. Identification the number of Mycobacterium tuberculosis based on sputum image using local linear estimator, *Bulletin EEI.* 9 (2020), 2109–2116. <https://doi.org/10.11591/eei.v9i5.2021>.
- [21] N. Chamidah, B. Zaman, L. Muniroh, et al. Designing local standard growth charts of children in East Java Province using a local linear estimator, *Int. J. Innov. Creat. Change.* 13 (2020), 45–67.
- [22] N. Chamidah, K.H. Gusti, E. Tjahyono, et al. Improving of classification accuracy of cyst and tumor using local polynomial estimator, *TELKOMNIKA.* 17 (2019), 1492–1500. <http://doi.org/10.12928/telkomnika.v17i3.12240>.
- [23] N. Chamidah, I.N. Budiantara, S. Sunaryo, et al. Designing of child growth chart based on multi-response local polynomial modeling, *J. Math. Stat.* 8 (2012), 342–347. <https://doi.org/10.3844/jmssp.2012.342.347>.
- [24] A. Delaigle, J. Fan, R.J. Carroll, A Design-Adaptive Local Polynomial Estimator for the Errors-in-Variables Problem, *J. Amer. Stat. Assoc.* 104 (2009), 348–359. <https://doi.org/10.1198/jasa.2009.0114>.
- [25] M. Francisco-Fernández, J.M. Vilar-Fernández, Local polynomial regression estimation with correlated errors, *Commun. Stat. - Theory Methods.* 30 (2001), 1271–1293. <https://doi.org/10.1081/sta-100104745>.

PREDICTION OF pH AND TOTAL SOLUBLE SOLIDS CONTENT OF MANGO

- [26] N. Chamidah, B. Lestari, Estimation of covariance matrix using multi-response local polynomial estimator for designing children growth charts: A theoretically discussion, *J. Phys.: Conf. Ser.* 1397 (2019), 012072.
<https://doi.org/10.1088/1742-6596/1397/1/012072>.
- [27] W. Cui, M. Wei, Strong consistency of kernel regression estimate, *Open J. Stats.* 3 (2013), 179–182.
<https://doi.org/10.4236/ojs.2013.33020>.
- [28] D. Aydin, Ö.I. Güneri, A. Fit, Choice of bandwidth for nonparametric regression models using kernel smoothing: a simulation study, *Int. J. Sci.: Basic Appl. Res.* 26 (2016), 47–61.
- [29] B. Lestari, Fatmawati, I.N. Budiantara, et al. Estimation of regression function in multi-response nonparametric regression model using smoothing spline and kernel estimators, *J. Phys.: Conf. Ser.* 1097 (2018), 012091.
<https://doi.org/10.1088/1742-6596/1097/1/012091>.
- [30] B. Lestari, Fatmawati, I.N. Budiantara, N. Chamidah, Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and kernel estimators approaches, *J. Phys.: Conf. Ser.* 1397 (2019), 012064. <https://doi.org/10.1088/1742-6596/1397/1/012064>.
- [31] J. Gao, P. Shi, M-type smoothing splines in nonparametric and semiparametric regression models, *Stat. Sinica*, 7 (1997), 1155–1169. <https://www.jstor.org/stable/24306178>.
- [32] N. Chamidah, B. Lestari, A. Massaid, et al. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach, *Commun. Math. Biol. Neurosci.* 2020 (2020), 72.
<https://doi.org/10.28919/cmbn/4963>.
- [33] Fatmawati, I.N. Budiantara, B. Lestari, Comparison of smoothing and truncated spline estimators in estimating blood pressures models, *Int. J. Innov. Creat. Change.* 5 (2019), 1177–1199.
- [34] Y. Wang, W. Guo, M.B. Brown, Spline smoothing for bivariate data with applications to association between hormones, *Stat. Sinica.* 10 (2000), 377–397. <https://www.jstor.org/stable/24306724>.
- [35] B. Lestari, Fatmawati, I.N. Budiantara, Spline estimator and its asymptotic properties in multiresponse nonparametric regression model, *Songklanakarin J. Sci. Technol.* 42 (2020), 533–548.
- [36] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Consistency and asymptotic normality of estimator for parameters in multiresponse multipredictor semiparametric regression model, *Symmetry.* 14 (2022), 336.
<https://doi.org/10.3390/sym14020336>.

- [37] B. Lestari, N. Chamidah, D. Aydin, et al. Reproducing kernel hilbert space approach to multiresponse smoothing spline regression function, *Symmetry*. 14 (2022), 2227. <https://doi.org/10.3390/sym14112227>.
- [38] N. Chamidah, M. Rifada, Estimation of median growth curves for children up two years old based on biresponse local linear estimator, *AIP Conf. Proc.* 1718 (2016), 110001. <https://doi.org/10.1063/1.4943348>.
- [39] N. Chamidah, E. Tjahjono, A.R. Fadilah, et al. Standard growth charts for weight of children in East Java using local linear estimator, *J. Phys.: Conf. Ser.* 1097 (2018), 012092. <https://doi.org/10.1088/1742-6596/1097/1/012092>.
- [40] M. Ulya, N. Chamidah, T. Saifudin, Local polynomial bi-responses multi-predictors nonparametric regression for predicting the maturity of mango (Gadung Klonal 21): a theoretical discussion and simulation, *Commun. Math. Biol. Neurosci.* 2022 (2022), 55. <https://doi.org/10.28919/cmbn/7499>.
- [41] J.J.M. Moreno, A.P. Pol, A.S. Abad, et al. El índice R-MAPE como medida resistente del ajuste en la previsión, *Psicothema*, 25 (2013), 500–506.
- [42] J. Luo, K. Ying, J. Bai, Savitzky–Golay smoothing and differentiation filter for even number data, *Signal Process.* 85 (2005), 1429–1434. <https://doi.org/10.1016/j.sigpro.2005.02.002>.
- [43] L. Xie, X. Ye, D. Liu, et al. Prediction of titratable acidity, malic acid, and citric acid in bayberry fruit by near-infrared spectroscopy, *Food Res. Int.* 44 (2011), 2198–2204. <https://doi.org/10.1016/j.foodres.2010.11.024>.