

# Prediction of Acidity Level of Avomango (*Gadung Klonal 21*) Using Local Polynomial Estimator

M Ulya<sup>1</sup>, N Chamidah<sup>2\*</sup>

<sup>1</sup>PhD Student of Matematics and Natural Sciences Study Program, Faculty of Science and Technology, Airlangga University, Indonesia

<sup>1</sup>Study Program of Agroindustrial Technology, Faculty of Agriculture, University of Trunojoyo Madura, Indonesia

<sup>2</sup>Department of Mathematics, Faculty of Science and Technology, Airlangga University, Indonesia

\*Corresponding author: [nur-c@fst.unair.ac.id](mailto:nur-c@fst.unair.ac.id)

## ABSTRACT

The acidity level is one of the parameters that determine fruit maturity. Mature avomango has a low acidity level, indicated by its titratable acidity value. Non-destructive analysis using a near infrared (NIR) spectroscopy tool produces wavelength data that is used as a predictor variable with a titratable acidity value as a response variable. This study aims to predict the acidity level (titratable acidity) of avomangoes using nonparametric regression based on local linear estimators and compare them with the prediction using the parametric regression approach. This study uses 100 mango data as samples, which are divided into 80 into sample data for the training process and 20 out sample data for the prediction process. The results showed that the second degree local polynomial estimator gives the best results. From the 5 iterations, the 4th iteration data gives the mean square error value of prediction using local polynomial estimator and global polynomial regression method is 0.0651 and 0.0733, respectively. The mean absolute percentage error of prediction using local polynomial estimators and global polynomial regression is 39.6411 and 43.3676, respectively. It means that the nonparametric regression based on local polynomial estimator provides better predictive results compared to the parametric regression approach.

**Keywords:** *Titratable acidity, acidity level, avomango, local polynomial estimator.*

## 1. INTRODUCTION

Mango is known as the king of the fruit in East Asia [1]. One type of mango that most popular now is the avomango (*Gadung Klonal 21*) developed from the Pasuruan Regency. The avomango is in demand because it has thick fruit flesh, low fiber, sweet taste, and can be eaten like an avocado [2]. To reach broader marketing, avomango producers must pick avomango with optimal maturity so that the avomango received by consumers has good eating quality.

Avomango farmers generally determine the level of avomango maturity sensorily based on the texture (rough or smooth and hard or soft) of the surface of the tip of the avomango fruit. Sensory determination of avomango maturity has the disadvantage that avomango, which is predicted to be ripe only from a sensorial texture, does not necessarily have a good eating quality because many parameters affect the level of avomango maturity and eating quality. These parameters include pH, total acid, sugar content, total soluble solids (TSS), and firmness. This parameter is generally not tested by farmers because it is destructive, has a higher cost, and requires a relatively long test time. Sensory, ripe avomango taste sweet and have a low acidity level. The most straightforward laboratory test is the titratable acidity test to find out whether or not

avomango taste sour as a basis for determining the level of maturity of avomango.

Recently, there have been many studies conducted to obtain methods of predicting the maturity level of mangoes non-destructively. One of them is a laboratory test using the Near-Infra Red (NIR) spectroscopy method. Then, the results of this study were validated using destructive ways, such as pH, total acid test, sweetness level, titratable acidity, etc. Research on predicting the quality of mangoes using the NIR spectroscopy approach has been widely carried out. Jha *et al.* [3] with firmness and color parameters, Jha *et al.* [4] reviewed yellowness, Jha *et al.* [5] concerned the level of sweetness of mangoes, all the studies that have been carried out show the results that the NIR spectroscopy approach is good enough to predict the quality attributes of the mangoes.

Prediction models can be developed with parametric and nonparametric regression approaches. Most of the previous studies used the parametric regression approach, including the Simple Linear Regression (SLR) method [6, 7], Multiple Linear Regression (MLR) [3-5], Partial Least Square Regression (PLSR) [3-5, 8], and Principal Component Regression (PCR) [3, 4]. Several studies predict the non-linear (nonparametric) regression approach, including the use of Artificial Neural Network (ANN) by Schulze *et al.* [7] and the Kernel Partial Least Square Regression method [9]. Under these conditions, there have

not been many studies on the use of nonparametric regression to predict fruit quality attributes based on the results of the NIR spectroscopy test.

The nonparametric regression approach does not assume a specific form (linear, quadratic, etc.) only assumed to be smooth so that it has high flexibility in determining the shape of its regression function [10]. Despite having high flexibility, this method must still be compared with the parametric approach to ensure that the nonparametric approach is suitable for Modelling related data. There are several estimators to estimate nonparametric regression curves, namely kernel estimator, local linear, local polynomial, spline, and Fourier series. Many previous studies on the use of various estimators in nonparametric regression Modelling, including the use of kernel estimators by Chamidah and Saifudin [11] and Lestari *et al.* [12]. Spline estimator is also widely studied by Budiantara *et al.* [13], Anwar and Chamidah [14], Islamiyati *et al.* [15], Adiwati *et al.* [16], Massaid *et al.* [17], Oktavetri *et al.* [18], Murbarani *et al.* [19], and Chamidah *et al.* [20]. One estimator that has a unique shape in nonparametric regression Modelling is the local polynomial estimator. If the degree of polynomial zero is equal to the kernel estimator, and if the degree of the polynomial is equivalent to the local linear estimator.

The use of nonparametric regression with local polynomial estimators has been applied in various fields by Chamidah and Rifada [21] about estimating child growth curves over two years with local bi-response local linear estimators, Chamidah *et al.* [22] about Modelling children's weight in East Java, and Puspitawati *et al.* [23] concerning solving classification cases with a local linear estimator. Previous research on determining the quality and maturity of fruit using a nonparametric approach so far has only been conducted by Nicolai *et al.* [9] by using kernel partial least square regression. However, there is no research to predict the quality and fruit maturity using nonparametric regression, mainly based on the local polynomial estimator. This study aims to predict the avomango acidity level using the nonparametric regression approach based on local polynomial estimator and compare the prediction performance with the parametric regression approach. The results of this study are expected to be useful for developing sensors that can predict the mango's maturity non-destructively.

## 2. METHODS

### 2.1. Research Materials

$$f(x_i) \approx \sum_{k=0}^p \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k = \sum_{k=0}^p \beta_k^*(x_0) (x - x_0)^k \quad (2)$$

where  $f^{(k)}(x_0)$  is the  $-k$  derivative of  $f(x_0)$  against  $x_0$ ,  $x \in (x_0 - h, x_0 + h)$ . If the equation (2) written in form matrix, we obtain:

The data used in this study were experimental data from 100 intact avomango samples tested using Near-Infra Red (NIR) spectroscopy with wavelengths between 900 - 1650 nm. There were 112 spectra data recorded and can be used as predictor variables. Because the spectra data have high dimensions and their number exceeds the number of observations, it is necessary to carry out a dimensional reduction process to facilitate data analysis. One method to reduce dimensions is the Principal Component Analysis (PCA) method. This method can produce a principal component that can represent the entire data. Only principal components that have eigenvalues more than one will be used. Then the observational data will be combined linearly with the principal component to get a new predictor variable, which will be processed at a later stage.

This study used a predictor variable ( $x$ ) absorbance spectra value from NIR spectroscopy and a response variable ( $y$ ) in the form of Total Titratable Acid (TTA) value, which describes the acidity level of avomango. The number of avomango samples was 100, divided into two parts, 80 as in-sample data and 20 as out-sample data. Avomango acidity level Modelling used two approaches, including nonparametric regression and parametric regression. The nonparametric regression approach used a local polynomial estimator, while the parametric regression approach used global polynomial regression. The two methods are compared in predicting performance using two goodness of fit criteria, namely the Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) values.

### 2.2. Regression Modelling

#### 2.2.1. Local Polynomial (Nonparametric Regression)

Nonparametric regression is a statistical method used to determine the functional relationship between variables  $x$  and  $y$  whose function is unknown. The nonparametric regression model of paired observation data  $(x_i, y_i)$  follows the following equation:

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (1)$$

where  $f$  is an unknown regression function,  $\varepsilon_i$  a random error assumed mean zero and variance  $\sigma^2$ . The regression function  $f$  in equation (1) estimated by using a local polynomial estimator, so it can be approximated by Taylor expansion as follows:

$$f(x_i) = \mathbf{x}_{x_0}^* \boldsymbol{\beta}^*(x_0) \tag{3}$$

where,

$$\mathbf{x}_{x_0}^* = [1 \quad (x - x_0) \quad \dots \quad (x - x_0)^p], x \in (x_0 - h, x_0 + h) \text{ and } \boldsymbol{\beta}^*(x_0) = [\beta_0^*(x_0) \quad \beta_1^*(x_0) \quad \dots \quad \beta_p^*(x_0)]^T.$$

Based on equation (3), we can write an equation (1) to be:

$$\mathbf{y}_i = \mathbf{x}_{x_0}^* \boldsymbol{\beta}^*(x_0) + \boldsymbol{\varepsilon}_i \tag{4}$$

To estimate  $\boldsymbol{\beta}^*(x_0)$  based on the local polynomial estimator, from  $n$  sample paired data  $\{x_t, y_t\}_{t=1}^n$ , the equation (6) is defined as follows:

$$\left. \begin{aligned} y_1 &= \beta_0^*(x_0) + \beta_1^*(x_0)(x_1 - x_0) + \beta_2^*(x_0)(x_1 - x_0)^2 + \dots + \beta_p^*(x_0)(x_1 - x_0)^p + \varepsilon_1 \\ y_2 &= \beta_0^*(x_0) + \beta_1^*(x_0)(x_2 - x_0) + \beta_2^*(x_0)(x_2 - x_0)^2 + \dots + \beta_p^*(x_0)(x_2 - x_0)^p + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0^*(x_0) + \beta_1^*(x_0)(x_n - x_0) + \beta_2^*(x_0)(x_n - x_0)^2 + \dots + \beta_p^*(x_0)(x_n - x_0)^p + \varepsilon_n \end{aligned} \right\} \tag{5}$$

Expression (5) in matrix form

$$\mathbf{y}^* = \mathbf{X}_{x_0}^* \boldsymbol{\beta}^*(x_0) + \boldsymbol{\varepsilon}^* \tag{6}$$

where

$$\mathbf{X}_{x_0}^* = \begin{bmatrix} 1 & (x_1 - x_0) & \dots & (x_1 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & \dots & (x_n - x_0)^p \end{bmatrix}, \mathbf{y}^* = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\varepsilon}^* = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{7}$$

Obtaining  $\boldsymbol{\beta}^*(x_0)$  an estimator is done by minimizing the Weighted Least Square (WLS) criteria as follows:

$$Q(x_0) = (\mathbf{y}^* - \mathbf{X}_{x_0}^* \boldsymbol{\beta}^*(x_0))^T \mathbf{K}_{\mathbf{h}^*}(x_0) (\mathbf{y}^* - \mathbf{X}_{x_0}^* \boldsymbol{\beta}^*(x_0)) \tag{8}$$

where

$$\mathbf{K}_{\mathbf{h}^*}(x_0) = \begin{bmatrix} K_{h^*}(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_{h^*}(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_{h^*}(x_n - x_0) \end{bmatrix} \tag{9}$$

$\mathbf{K}_{\mathbf{h}^*}(x_0)$  is matrix contain a weighted function that  $K_{h^*}(\cdot)$  is Kernel function with bandwidth  $h^*$  is defined as follows:

$$K_{h^*}(x) = \frac{1}{h} K\left(\frac{x}{h}\right); -\infty < x < \infty \text{ and } h^* > 0$$

This study used the Gaussian Kernel as follows:

$$\mathbf{K}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{10}$$

Bandwidth ( $h$ ) is a smoothing parameter that controls the smoothness of the curve. The selection of optimum bandwidth is based on the bandwidth that has the minimum Cross-Validation (CV) value [31].

$$h_{opt} = \arg \min \frac{1}{n} \sum_{i=1}^n \left( \widehat{f}_{n,i}^{(h)}(x_i) - y_i \right)^2 \tag{11}$$

*2.2.2. Polynomial Regression (Parametric Regression)*

Parametric regression is a regression approach to determine the pattern of relationship between the response variable and predictor variable where the shape of the curve is known. The equation of the global model using polynomial regression as follows [25].

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \varepsilon_i \quad , i = 1, 2, \dots, n \quad (12)$$

In the matrix, we obtain the regression model:

$$y = X\beta + \varepsilon \quad , \varepsilon \sim N(0, \sigma^2 I)$$

To estimate a regression coefficient  $\beta$ , we use OLS (Ordinary Least Square) that is minimize a sum of square error. We have estimator

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (13)$$

### 2.2.3. The Goodness of Fit

The goodness of fit used in this research was the MSE [28] and MAPE [26] values that can be calculated by the following formula:

$$MSE = \frac{SSE}{n} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (14)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|\hat{y}_i - y_i|}{y_i} \right) \times 100\% \quad (15)$$

### 2.2.4. Research Work Steps

The stages of this research are as follows:

1. Input data pairs (112 predictor variables  $z$  and one response variable  $y$ )
2. Reduce dimensions using the PCA method to obtain new predictor variables ( $x$ )
3. Make a scatter plot of predictor vs. response variable
4. Determine optimal bandwidth values ( $h$ ) at first and second polynomial degrees using Eq. (11)
5. Fit model in in-sample data using optimal bandwidth values ( $h$ ) at each degree of local polynomial
6. Calculate the MSE and MAPE values of in-sample data using Eq. (14) and (15)
7. Predict the TTA value on the out-sample data based on the optimal bandwidth ( $h$ ) value
8. Calculate the values of MSE and MAPE of out-sample data using Eq. (14) and (15)
9. Average the values of MSE and MAPE of overall data
10. Do point 5-9 using a polynomial regression approach
11. Compare the MSE and MAPE values between nonparametric approaches based on local polynomial

estimators with parametric regression (polynomial regression).

Data analysis performed using Open Source Software - R (OSS-R).

## 3. RESULTS AND DISCUSSION

### 3.1. Description of Avomango Acidity Level

Fruit acidity is closely related to fruit maturity [29]. The lower the acidity, the more ripe the fruit is. Sensory tests conducted by farmers from mango texture on 100 samples showed that from 100 samples, there were unripe, ripe, and overripe avomango of 35%, 55%, and 10%, respectively. The sensory test was then validated by NIR spectroscopy and the TTA value test. The following are the actual value of TTA as a response variable in this research.

**Table 1.** Descriptive Statistics of the actual TTA value

	Value
The number of observations	100
Mean	1.067 %
Maximum	2.335 %
Minimum	0.338 %
Standard deviation	0.414 %

Table 1 reveals that the maximum value of the actual TTA was 2.335%, and the minimum one was 0.338%. So, the range of the actual TTA value was 1.997%. The range of actual TTA values is wide enough according to the maturity level of the mangoes used. The results of this study are following Jha *et al.* [29], which stated the range of the TTA values of mango was between 0.065-7.26% (overripe until unripe mango). The higher the TTA value of mango, the more acidic the mango is.

### 3.2. Summary of Dimension Reduction using PCA

In the initial stage, the dimension reduction of the initial data pairs of the 112 variables  $z$  and 100 observations were carried out using the PCA method. This data pair is formed into a matrix with the size  $M_{100 \times 112}$ . In the dimensional reduction process, only one principal component that has an eigenvalue  $>1$  is obtained to construct the  $N_{112 \times 1}$  matrix. 98.83% of the variation in the overall data can be explained by one principal component. Thus, it performed a linear combination between the initial data and the principal components to obtain new data with a matrix size  $D_{112 \times 1}$ . That is, there is only one predictor variable ( $x$ ) that will be used in regression modelling to predict one response variable ( $y$ ).

**Table 2.** Descriptive Statistics of the Absorbance Spectra Value (reduced by PCA method)

	The $x$ value
The number of observations	100
Mean	-12.909
Maximum	-9.729
Minimum	-16.355
Standard deviation	1.666

Table 2 shows that the range of absorbance spectra data values is 6.626. The wide range illustrates that the avomango used as samples vary in their maturity. In the test using NIR spectroscopy, the spectrum of white light absorbed by the fruit can represent the fruit's content, such as TTA, total soluble solids, firmness, pH value [30].

### 3.3. Regression Modelling

#### 3.3.1. Local polynomial (nonparametric approach)

This study uses a local polynomial estimator to model the avomango acidity level (response variable) based on the obtained absorbance spectra value (predictor variable). The data used for the fitting model are 80 samples of in-sample data. The first step is to find the optimal bandwidth value ( $h$ ) at first and second order of local polynomial based on

the minimum value of Cross-Validation (CV). The following are tables and figures that show the results of calculating the optimal bandwidth for the first order and the second-order local polynomial (See **Table 3**, **Table 4**, and **Fig.1**).

The bandwidth parameter ( $h$ ) has an essential role in determining the optimal model [24]. If  $h$  is large, the estimated function obtained will be smoother. Conversely, if  $h$  is small, then the estimated function obtained will be rougher. Therefore it is necessary to determine the optimum  $h$  so that the estimation of the model is also optimum. Table 3 shows the optimum bandwidth value for the first order local polynomial ( $h_1$ ) is 1.1, with the minimum CV value of 0.174281. Table 4 shows the optimum bandwidth value for the second-order local polynomial ( $h_2$ ) is 6.3, with the minimum CV value of 0.171504. **Fig. 1** a dan b) clearly show the optimum bandwidth of the first and second-order local polynomial. Thus, fitting the model can be obtained based on the optimum bandwidth and the order of the local polynomial. So, the regression coefficient values ( $\beta_i$ ) can be estimated.

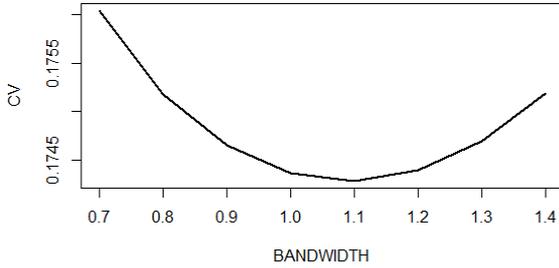
**Table 3.** The bandwidth and minimum CV values at the first order of the local polynomial

No.	Bandwidth	CV	MSE	MAPE
1	0.7	0.176041	0.152184	40.15841
2	0.8	0.175179	0.15342	40.55809
3	0.9	0.174651	0.154547	40.96832
4	1	0.174364	0.155595	41.34927
5	<b>1.1</b>	<b>0.174281</b>	<b>0.156606</b>	<b>41.69411</b>
6	1.2	0.174392	0.157627	42.00871
7	1.3	0.174696	0.158695	42.30408
8	1.4	0.175181	0.159826	42.55927

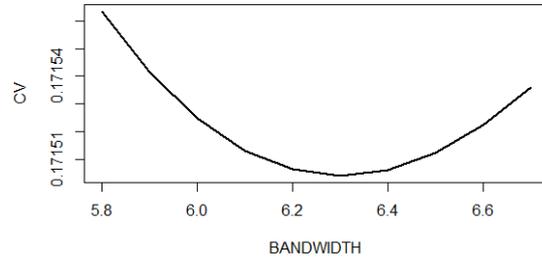
**Table 4.** The bandwidth and minimum CV values at the first order of the local polynomial

No.	Bandwidth	CV	MSE	MAPE
1	5.8	0.171564	0.153452	39.78648
2	5.9	0.171541	0.153524	39.81503
3	6	0.171525	0.153596	39.84304
4	6.1	0.171513	0.15367	39.87051
5	6.2	0.171506	0.153745	39.89742
6	<b>6.3</b>	<b>0.171504</b>	<b>0.153821</b>	<b>39.92376</b>
7	6.4	0.171506	0.153898	39.94954
8	6.5	0.171512	0.153976	39.97475

9	6.6	0.171522	0.154054	39.99938
10	6.7	0.171536	0.154133	40.02344



a). Bandwidth optimum at 1<sup>st</sup> order local polynomial



b). Bandwidth optimum at 2<sup>nd</sup> order local polynomial

**Figure 1.** The Determination of bandwidth optimum based on the minimum CV value

The process of modelling nonparametric regression based on local polynomial estimator is carried out using OSS-R software. The following is a summary of the prediction

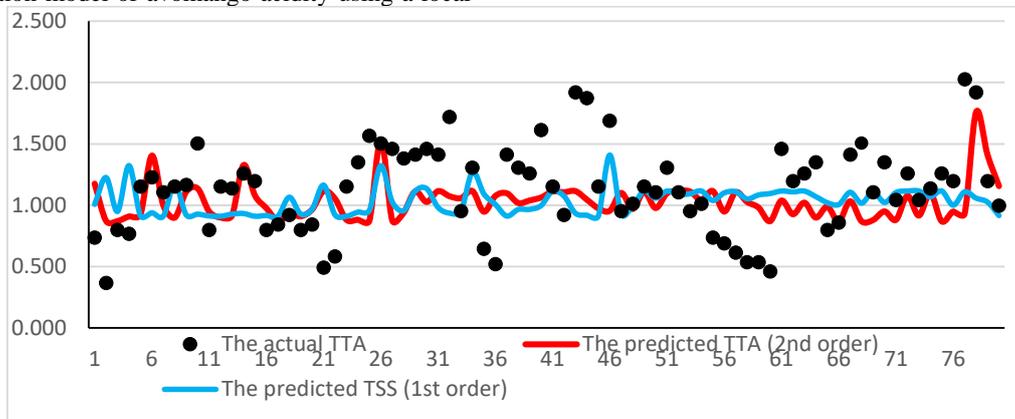
performance using nonparametric regression based on local polynomial in the first and second order.

**Table 5.** Optimum bandwidth (*h*), MSE and MAPE values in in-sample data

Degree	<i>h</i>	In sample Data		Out sample Data		Overall Data	
		MSE	MAPE	MSE	MAPE	MSE	MAPE
1	1.100	0.157	41.694	0.130	20.490	0.151	37.454
2	6.300	0.154	39.924	0.145	22.060	<b>0.152</b>	<b>36.351</b>

**Table 5** reveals that local quadratic (2<sup>nd</sup> order local polynomial) gives a better MSE and MAPE value than local linear (1<sup>st</sup> order local polynomial). The lower MSE and MAPE values indicate the goodness of a better fit. It means that a prediction model of avomango acidity using a local

quadratic estimator is better than a local linear estimator. Figure 2 shows the model estimation of the TTA values of Avomango (*y*) based on absorbance spectra values (*x*) using a local polynomial estimator.



**Figure 2.** The Model Estimation of TTA values of Avomango (*y*) based on absorbance spectra values (*x*) using a local polynomial estimator

After modelling using in-sample data, the prediction of avomango acidity is then performed using the model that has been obtained. The prediction process used 20 out-sample data. The prediction performance was measured using the MSE and MAPE values (See **Table 5**). In local

polynomial regression, each observation has a regression equation:

$$\hat{y} = \beta_0 + \beta_1(x - x_0) + \beta_2(x - x_0)^2 \text{ where } x \in (x_0 - h, x_0 + h)$$

(16)

The following is an example of a regression equation model for the 10<sup>th</sup> data, with optimum bandwidth = 6.3 and  $x_0 = -11.8601$  produces the following estimation equation:

$$\hat{y}_5 = 0.8925 + 0.1312(x - (-11.8601)) + 0.079(x - (-11.8601))^2 \quad x \in (-18.1601; -5.5601)$$

(17)

If the value of  $x$  (absorbance spectra value) = -11.8601, then Eq. (17) became:

$$\hat{y}_5 = 0.8925 + 0.1312(-11.8601 - (-11.8601)) + 0.079(x - 18.601 - (-11.8601))^2$$

$$\hat{y}_5 = 0.8925$$

So, with  $x = -11.8601$ , the predicted value of TTA was 0.8925 %. The actual TTA value in the 10<sup>th</sup> data is 0.845%, so that there is an error percentage of 5.65%.

Overall, the values of MSE and MAPE using local quadratic regression are 0.151 and 36.351%, respectively. The smaller the MSE value, the better the predictive performance of the model. The MAPE value in this model is 36.351%, which indicates that the prediction performance is included in the category of reasonable

prediction. Moreno *et al.* [27] stated that the MAPE value between 20-50 % was included. This too high MAPE value may be because the relationship between the variables  $x$  and  $y$  is not too high. The actual TTA pattern is too diffuse or spreads out (See Fig. 2), so the error tends to be high. For further research, the research could add the number of samples to make a more visible relationship between  $x$  and  $y$ .

Based on the MSE and MAPE values in Table 5, it is known that the prediction of avomango acidity using local quadratic regression (local polynomial of 2 degrees) gives a lower MSE and MAPE value than the local linear regression. Therefore, the next step is a process of comparing the goodness of fit of local quadratic regression with parametric regression, global quadratic regression. Although nonparametric regression is said to be more flexible in modelling, it is not necessarily the best. So, it still needs to be compared with parametric regression so that it is known whether the local polynomial regression approach is suitable for modelling the prediction of acidity level of avomango.

modelling the acidity level (TTA value) using global quadratic regression was also carried out in this study. The graph of the global quadratic regression model between the value of TTA ( $y$ ) with the spectrum value ( $x$ ) can be seen in Fig. 3. While the MSE and MAPE values of the model using the global quadratic regression approach can be seen in Table 6.

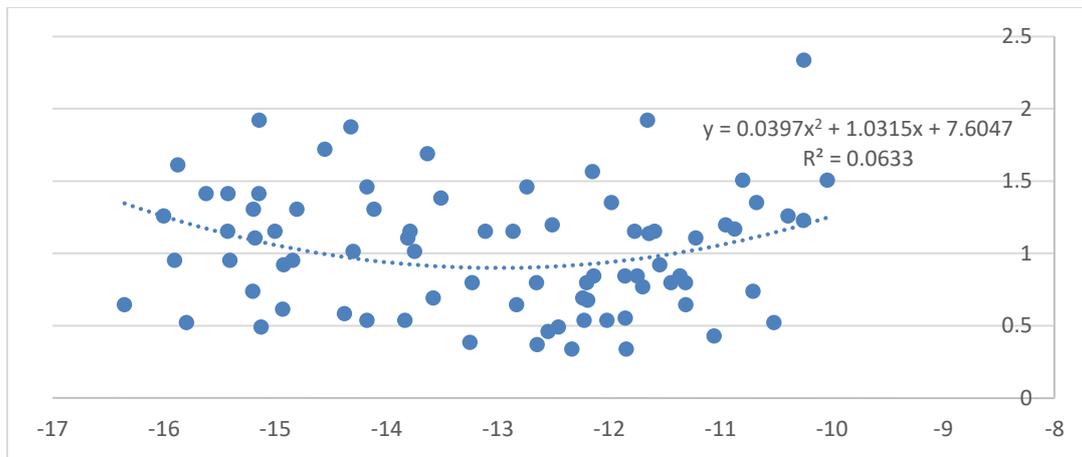


Figure. 3. Global quadratic regression model of TSS value based on absorbance spectra values

Table 6. The prediction performance using parametric regression (global quadratic regression)

Data	The Goodness of Fit	
	MSE	MAPE
In-sample	0.168	42.372
Out-sample	0.154	21.855

Overall	0.165	38.269
---------	-------	--------

Modelling using global quadratic regression produces a regression equation as follows:

$$\hat{y} = 7.6047 + 1.0315x + 0.0397x^2$$

(18)

If  $x$  value (absorbance spectra value) in the 10<sup>th</sup> data was -11.8601, then equation (18) became:

$$\hat{y} = 7.6047 + 1.0315(-11.8601) + 0.0397(-11.8601)^2$$

$$\hat{y} = 7.6047 - 12.2337 + 5.5843$$

$$\hat{y} = 0.9553$$

So with  $x = -11.8601$ , the estimated value of TTA was 0.9553 %. The actual TTA value in the 10<sup>th</sup> data is 0.845% so that there is an error percentage of 13.078%. This error percentage value is higher than the error percentage using local quadratic regression.

Based on the MSE and MAPE values in **Table 5** and **Table 6**, it is known that the prediction of avomango acidity using a nonparametric regression approach based on a local quadratic estimator gives a lower MSE and MAPE value than the global quadratic approach. The results of this study are in line with Fan and Gijbels [28]. While global polynomial regression has been widely used, it suffers from drawbacks. One is that individual observations can have a considerable influence on remote parts of the curve so that the nonparametric regression approach with local polynomial estimator was better in modelling data that does not have a specific pattern compared to the parametric regression approach.

#### 4. CONCLUSION

Based on the analysis, it can be concluded that the nonparametric regression approach using a local quadratic estimator can predict the avomango acidity better than the global polynomial regression approach because it has a lower MSE value of 0.152 and the MAPE value of 36.351. It means that the nonparametric regression based on local polynomial estimator provides better predictive results compared to the parametric regression approach

#### ACKNOWLEDGMENT

The first author would like to thank the Directorate General of Indonesian Higher Education for supporting this research through a BPPDN scholarship and also thanks to the owner of the avomango's garden, Mr. Sugik, as an expert who has supported this research and provided knowledge about sensory Determination of avomango maturity.

#### REFERENCES

- [1] Pott, I., Marx, M., Neidhart, S., Mühlbauer, W., Carle, R. (2003). Quantitative Determination of  $\beta$ -carotene Stereoisomers in Fresh, Dried, and Solar-dried Mangoes (*Mangifera indica* L.). *Journal of Agricultural and Food Chemistry*. 51(16): 4527-4531.
- [2] Karsinah. Rebin. Tasliah. 2017. Varietas Unggul Mangga Gadung 21: Daging Buah Tebal, Berserat Rendah, Rasa Manis, dan Dapat Dimakan Seperti Alpukat. *Iptek Hortikultura*. 39-44.
- [3] Jha, S. N., Kingsly, A. R. P., Chopra, S. (2006). Non-destructive Determination of Firmness and Yellowness of Mango During Growth and Storage using Visual Spectroscopy. *Biosystems Engineering*. 94(3): 397-402.
- [4] Jha, S. N., Chopra, S., Kingsly, A. R. P. (2007). Modelling of Color Values for Nondestructive Evaluation of Maturity of Mango. *Journal of Food Engineering*. 78: 22-26.
- [5] Jha, S. N., Jaiswal, P., Nasaiah, K., Gupta, M., Bhardwaj, R., Singh, A. K. (2012). Non-destructive Prediction of Sweetness of Intact Mango Using Near Infra Red Spectroscopy. *Scientia Horticulturae*. 138: 171-175.
- [6] Watanawan, C., Wasusri, T., Srilaong, V., Wongs-Aree, C. And Kanlayanarat, S. 2014. Near Infrared Spectroscopic Evaluation of Fruit Maturity and quality of export Thai Mango (*Mangifera indica* L. Var. Namdokmai). *International Food Research Journal*. 21(3): 1109-1114.
- [7] Schulze, K., Nagle, M., Spreer, W., Mahayothee, Müller, J. (2015). Development and Assessment of Different Modelling Approaches for Size-mass Estimation of Mango Fruits (*Mangifera indica* L., cv. 'Nam Dokmai'). *Computers and Electronics in Agriculture*. 134: 269-276.
- [8] Rungpichayapichet, P., Mahayothee, B., Nagle, M., Khuwijitjaru, P., Müller, J. (2016). Robust NIRS Models for Non-destructive Prediction of Postharvest Fruit Ripeness and Quality in Mango. *Postharvest Biology and Technology*. 111: 31-40.
- [9] Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., Lammertyn, J. (2007). Nondestructive Measurement of Fruit and Vegetable Quality by Means of NIR spectroscopy: A Review. *Postharvest Biology and Technology*. 46(2): 99-118.
- [10] Eubank, R. (1999). *Spline Smoothing and Nonparametric Regression* Second Edition. Marcel Dekker. New York.
- [11] Chamidah, N., Saifudin, T. (2013). Estimation of Children Growth Curve Based on Kernel Smoothing in Multi-response Nonparametric regression. *Applied Mathematical Sciences*. 7(37): 1839-1847.
- [12] Lestari, B., Fatmawati, B. IN, Chamidah, N. (2018). Estimation of Regression Function in Multi-response Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators. In *Journal of Physics: Conference Series*. 1097(1).

- 012091).
- [13] Budiantara, I. N., Ratna, M., Zain, I., Wibowo, W. (2012). Modelling the Percentage of Poor People in Indonesia Using Spline Nonparametric Regression Approach. *International Journal of Basic & Applied Sciences*. 12: 119-124.
- [14] Anwar, A. E., Chamidah, N. (2019). Glaucoma Identification on Fundus Retinal Images Using Statistical Modelling Approach. In *IOP Conference Series: Materials Science and Engineering*. 546(5): 052010).
- [15] Islamiyati, A., Chamidah, N. (2018). Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression. *JPhCS*. 979(1): 012093.
- [16] Adiwati, T., Chamidah, N. (2019). Modelling of Hypertension Risk Factors Using Penalized Spline to Prevent Hypertension in Indonesia. In *IOP Conference Series: Materials Science and Engineering*. 546(5): 052003. IOP Publishing.
- [17] Massaid, A., Hanif, M., Febrianti, D., Chamidah, N. (2019). Modelling of Poverty Percentage of Non-Food Per Capita Expenditures in Indonesia Using Least Square Spline Estimator. In *IOP Conference Series: Materials Science and Engineering*. 546(5): 052044. IOP Publishing.
- [18] Oktavitri, N. I., Kuncoro, E. P., Purnobasuki, H., Chamidah, N. (2019). Prediction of Suspended and Attached Process Behavior in Anaerobic Batch Reactor Using Nonparametric Regression Model Approach Based on Spline Estimator *Eco. Env. & Cons.* 25: 103-107.
- [19] Murbarani, N., Swastika, Y., Dwi, A., Aris, B., Chamidah, N. (2019). Modelling of the Percentage of Aids Sufferers in East Java Province with Nonparametric Regression Approach Based on Spline Truncated Estimator. *Indonesian Journal of Statistics and Its Applications*. 3(2): 139-147.
- [20] Chamidah, N., Zaman, B., Muniroh, L., Lestari, B. (2019). Estimation of Median Growth Charts for Height of Children in East Java Province of Indonesia Using Penalized Spline Estimator, *International Conference Proceedings GCEAS*. 5: 68-78.
- [21] Chamidah, N., Rifada, M. (2016, March). Estimation of Median Growth Curves for Children Up Two Years Old Based on Biresponse Local Linear Estimator. In *AIP Conference Proceedings*. 1718(1): 110001. AIP Publishing LLC.
- [22] Chamidah, N., Tjahjono, E., Fadilah, A. R., Lestari, B. (2018). Standard Growth Charts for Weight of Children In East Java Using Local Linear Estimator. In *Journal of Physics: Conference Series*. 1097(1): 012092.
- [23] Puspitawati, A., Chamidah, N. (2019). Choroidal Neovascularisation Classification on Fundus Retinal Images Using Local Linear Estimator. In *IOP Conference Series: Materials Science and Engineering*. 546(5): 052056. IOP Publishing.
- [24] Reusch, W. (2019). *Infrared Spectroscopy*. Internet article. <https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/InfraRed/infrared.htm> (accessed August 26th 2019).
- [25] Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R. (2018). *Data Mining dengan R Konsep Serta Implementasi*. Jakarta: InMedia.
- [26] Eubank, R. L., Huang, C., Maldonado, Y. M., Wang, N., Wang, S., Buchanan, R. J. (2004). Smoothing Spline Estimation in Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 66(3): 653-667.
- [27] Suparti., Santoso, R., Prahutama, A., Devi, A. R., (2018). *Regresi Nonparametrik*. Wade Group. Ponorogo.
- [28] Moreno, J. J. M, Pol, A. P., Abad, A. S. and Blasco, B. C. 2013. Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*. 25( 4): 500 – 506.
- [29] Fan, J., Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall. London.