Development of calibration model for pH content of intact tomatoes using a low-cost Vis/NIR spectroscopy

R K Wati¹, M F R Pahlawan¹, and R E Masithoh¹*

¹Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Indonesia

*E-mail: evi@ugm.ac.id

Abstract. Determination of pH of intact tomatoes was investigated using a low-cost Vis/NIR spectroscopy in reflectance mode. The best calibration model measured pH in intact tomatoes using wavelength range of 527-799 nm with $R^2$ and SEC of 0.90 and 0.04, respectively. The prediction model obtained SEP, Bias, and RPD of 0.11, 0.007, and 1.17, respectively. The low-cost Vis/NIR type instrument is promising to be used for food and agricultural applications.

Keywords: pH; tomato; Vis/NIR; spectroscopy; PLSR

1. Introduction

One of quality parameters of fruit is pH which represents acidity level and is used as indicators of taste [1]. During ripening, fruits experience physiological changes which involve starch to sugars conversion, flavor formation, and pH increase [2]. In general, pH or acidity was determined using titration method, pH meter, or acid meter. Those methods required sample conversion to liquid thus involve intense-work and long-time analysis. Titration methods lead to subjective measurement and highly depends on the skills of the analysts. Moreover, for vast and huge samples, those conventional pH measurements take long-time and high cost for analysis.

At present, with the development of spectroscopy technology, methods for determining pH or acidity of fruits were carried out by infrared (IR) spectroscopy. IR spectrometers with a wide infra-red range (above 1000 nm) were used to determine acidity of bayberry [3][4], guava and passion fruit pulps [5], and purple passion fruits [6]. The new method which applies NIR hyperspectral imaging (HSI) was used to detect acidity in cherry fruits [1], mango [7], and limes [8]. The NIR spectroscopy or hyperspectral imaging methods provide rapid and accurate prediction results; however, require high costs for the purchase of instruments or operational costs for halogen lamps, LEDs, or lasers (in case for Raman HSI).

Other type of spectrometers working in the visible and short-wave near infra-red (Vis/ NIR) at 350 - 1000 nm is also available and can be used to determine quality parameters in fruits. Weak absorptions of Vis/NIR energy for water molecules make the wave range suitable for detecting small concentrations of objects since it will not be destructed by water spectrum which is the largest component in fruits [9]. Vis/NIR spectrometers have been used to detect quality parameters in several commodities, including detecting acidity, soluble solids, and firmness of satsuma mandarin [10], as well as measuring flavor and sweetness of orange and grapefruit [11] or total dissolved solids in kiwi, apricot, and nectar [12]. However, those spectrometers not only cover visible spectrum but also include longer wavelength usually up to 2000-2500 nm which eventually is not practical for applications in small industries. The
previous Vis/NIR spectrometers were bench-top models or hand-held instruments for field applications. They provided easy and accurate measurement, but the costs are still relatively expensive. Therefore, the use of low-cost spectrometer will be beneficial for farmers or industries with a low budget.

Vis/NIR spectrometer with a modular type that works in the range of 350-1000 nm is an alternative to determine quality parameters of fruits. The modular spectrometer has relatively lower cost compared to other types of spectrometers. In this study, the performance of the Vis/NIR spectrometer will be assessed to determine pH of tomatoes. From this research, it is expected that this low-cost Vis/NIR spectrometer can be used to speed up the process of detecting quality in food production chain.

2. Materials and methods

2.1. Fruit samples
A total of 99 tomatoes were purchased from local markets in different colors (green, orange, and red) which represent level of maturity to obtain sufficient variability of the samples. The fruits were stored at 25 °C prior to spectra measurements. 10-15 fruits per day were scanned using the Vis/NIR spectrometer and then followed by pH measurement. Of 99 samples, 50 samples were used for calibration and 49 samples were used for prediction.

2.2. Vis/NIR spectroscopy set up and measurements
The spectroscopy equipment consisted of a Flame-T-VIS-NIR spectrometer (Ocean Optics, USA), with tungsten halogen lamp (HL-2000-HP-FHSA Ocean Optics, USA) as a light source, reflection probe (QR400-7-VIS-NIR Ocean Optics, USA), and a sample holder. The wavelength range of the spectrometer was 350–1050 nm. The fiber optic probe was clamped by a retort stand and the sample was placed 1 cm above the probe on a self-developed sample holder as arranged in Figure 1. A white reference was used to obtain the relative reflectance before spectra acquisition. Spectra were collected using Oceanview software (version 1.6.7, Ocean Optics, USA). In total, 99 spectra were used for multivariate analysis.

![Figure 1. Sample dan spectrometer arrangement showing 1 cm distance from the fiber optic probe to the sample.](image)

2.3. pH reference analysis
After spectra measurements, each sample was blended into juice including the peels. Acidity level in term of pH was determined from the liquid of the tomato juice using a digital pH Meter (KL-009(I), China). Measurement of pH was done in triplicate and averaged.

2.4. Chemometrics and data analysis
2.4.1. Spectra pre-processing. The reflectance spectra of tomatoes were transferred into xls. file (Microsoft® Excel®) which were then imported to the Unscrambler software v10.5.1 (CAMO Software AS, Norway) for spectra pre-processing and analysis. Several pre-processing methods were applied, namely normalization (such as area, unit vector, mean, maximum, and range normalization), baseline
correction, Savitzky Golay derivative, multiplicative scatter correction (MSC), and standard normal variate correction (SNV).

2.4.2. Multivariate analysis. Original and pre-processing spectra were used for multivariate analysis, namely principal component analysis (PCA) and partial least squares regression (PLSR). PCA was used for data exploration to find deeper insights of a large data obtained from spectroscopy instrument, while PLSR models developed using full cross-validation method were used to quantify chemical content. Performance of calibration models were assessed in terms of coefficient determination of calibration ($R^2_c$), standard error or calibration and prediction (SEC and SEP), Bias, and ratio of performance to deviation (RPD).

3. Results and Discussion

3.1. pH content and spectra exploration

Of 99 tomato samples contained pH in average of 4.5, minimum of 4.2, maximum of 4.9, and standard of deviation of 0.2. Most tomatoes had pH of 4.4-4.6 with a peak of 4.5. The pH distributions in this study is similar to that reported by [13] at range of 3.8-4.4. However, this study used narrower range of pH compared to study reported by [14] at range of 3.8–4.8.

Figure 2 shows average spectra of all tomatoes which were classified into three classes based on color, i.e. green, orange, and red. Only spectra between 500-900 nm were presented since spectra below 500 nm and above 900 nm were heavily noisy. In general, there were increase in reflectance values from 500 nm to 550 nm for green tomatoes, from 500 nm to 600 nm for orange tomatoes, and 500 nm to 620 nm for red tomatoes. Those increase were then followed by decrease to low values around 675 nm. Relatively low and stable absorbance were noticeable at 720 – 900 nm similar to those reported by [14]. Those weak absorptions were water molecules which were captured as low values by the Vis/NIR system. The system had a low spectral resolution (0.22 nm) thus could not detect small absorptions [15].

![Figure 2. Average spectra of green, orange, and red tomatoes captured by Vis/NIR spectrometer](image)

Figure 3 was spectra of MSC pre-processing result of original spectra. It can be inferred from Figure 4, at 675 nm green tomatoes had the lowest reflectance values meaning they had the highest absorbance values followed by orange and red tomatoes. Those absorbance values were due to the absorption of chlorophyll showed by green tomatoes. Green tomatoes with the highest chlorophyll had higher absorbance compared to orange and red tomatoes which had lower chlorophyll. Similar findings reported by [15] reporting high absorbance of apple, peach, pear, and kiwi at 675 nm due to chlorophyll.
The lowest reflectance values between 500 and 580 nm were found at red tomatoes followed by orange and green tomatoes meaning red tomatoes had the highest absorbance values at 500-580 nm while green tomatoes had the least. The absorbance at 500-580 nm was due to lycopene pigment owned by red and orange tomatoes [16].

![Figure 3](image.png)

**Figure 3.** MSC pre-processed spectra of green, orange, and red tomatoes captured by Vis/NIR spectrometer

3.2. **PCA analysis**

PCA was used to visualize large variables such as spectroscopy data (17Guillén-Casla et al., 2011). Table 1 was the variance percentage and cumulative percentage for the first 3 principal components (PC) which accounted for 100% of the total variance. By using PC-2 and PC-3, tomatoes were clearly classified into green, orange, and red groups, as shown in Figure 4(a). The PC-2 separated tomatoes into green, orange, and red. Green and orange tomatoes had negative values while red tomatoes had positive value of PC-2.

| Principle components | Variance percentage | Cumulative percentage |
|----------------------|---------------------|-----------------------|
| PC-1                 | 89                  | 89                    |
| PC-2                 | 10                  | 99                    |
| PC-3                 | 1                   | 100                   |

The peaks at 675 nm (Figure 4(b)) showed positive loadings of PC-2 for red tomatoes and negative loadings of PC-3 for orange and green tomatoes. Positive loading at 675 nm indicated high reflectance but low absorbance while negative loading indicated low reflectance but high absorbance. The absorbance values at 675 nm were absorption of chlorophyll [15], therefore tomatoes reflected red colors. This phenomenon was reflected in Figure 4(a) by which using PC-2 tomatoes were grouped in green and orange located in the left side of PC-2 and red located in the right side of PC-2.
3.3. PLSR calibration for determination of pH

The original spectra yielded the best results for pH determination at wavelength range of 527-799 nm. The calibration results were shown in Table 2. The coefficient determination of calibration ($R^2_C$) of 0.90 suggested that the model was applicable for screening and approximate calibration [18]. The $R^2_C$ obtained here were better than those reported by [13][14] which were 0.76, even significantly higher compared to $R^2_C$ of 0.32 reported by [19]. The SEP value for pH (0.07) were similar with result of [19] but higher compared to that reported by [19]. RPD (1.17) was lower compared to those reported by [19] for tomatoes and [20] for grapes. The low prediction results might be due to narrow range of pH used in this study which were 4.2-4.9.

**Figure 4.** PCA of the Vis/NIR spectra at 500-800 nm of tomatoes: (a) score plot of PC-2 (10%) and PC-3 (1%) and (b) loadings of PC-2 and PC-3 with the bands that contributed to differentiation of tomatoes.
Table 2. Calibration and prediction results of partial least square regression (PLSR) using original spectra for determination of pH tomatoes measured with Vis/NIR spectroscopy at wavelength range of 527-799 nm.

| Component | Calibration | Prediction |
|-----------|-------------|------------|
| pH        | Samples     | 𝑅^2_C      | SEC | Samples | SEP | Bias | RPD |
|           | 50          | 0.90       | 0.04 | 49      | 0.11 | 0.007 | 1.17 |

𝑅^2_C : coefficient determination of calibration; SEC: standard error of calibration; SEP: standard error of prediction; Bias: the average of difference between actual and instrument predicted values; RPD: ratio of performance to deviation.

Figure 5 illustrated regression coefficient of PLSR using original spectra for the determination of pH tomatoes. The spectra were noisy, but several peaks could be noticed at 560, 670, 730, and 760 nm which might had important roles for the calibration model. Wavebands at 560 was assigned to lycopene [16] while wavebands at 670 and 760 were assigned to chlorophyll-a and water [19]. However, it is difficult to identify which absorption bands specifically contributed to pH prediction.

Figure 5. Regression coefficient of PLSR using original spectra for the determination of pH

4. Conclusion
Acidity level in term of pH tomatoes were successfully measured using a Vis/NIR spectrometer in reflectance mode. Calibration model resulted in 𝑅^2_C and SEC of 0.90 and 0.04, while prediction model resulted in SEP, Bias, and RPD of 0.11, 0.007, and 1.17, respectively. The Vis/NIR spectrometer has proven to be a promising tool for measurements of pH of intact tomatoes.

References
[1] Li, Xiaoli, Yuzhen Wei, Jie Xu, Xuping Feng, Feiyue Wu, Ruiqing Zhou, Juanjuan Jin, Kaiwen Xu, Xinjie Yu, and Yong He. 2018. “SSC and PH for Sweet Assessment and Maturity Classification of Harvested Cherry Fruit Based on NIR Hyperspectral Imaging Technology.” Postharvest Biology and Technology 143 (May): 112–18. https://doi.org/10.1016/j.postharvbio.2018.05.003.

[2] Maftoonazad, Neda, and Hosahalli Ramaswamy. 2019. “Design and Testing of an Electrospun Nanofiber Mat as a PH Biosensor and Monitor the PH Associated Quality in Fresh Date Fruit (Rutab).” Polymer Testing 75 (December 2018): 76–84.
[3] Li, Xiaoli, and Yong He. 2006. “Non-Destructive Measurement of Acidity of Chinese Bayberry Using Vis / NIRS Techniques.” Eur. Food Res. Technol., 731–36. https://doi.org/10.1007/s00217-006-0260-x.

[4] Xie, Lijuan, Xingqian Ye, Donghong Liu, and Yibin Ying. 2011. “Prediction of Titratable Acidity, Malic Acid, and Citric Acid in Bayberry Fruit by near-Infrared Spectroscopy.” Food Research International 44 (7): 2198–2204. https://doi.org/10.1016/j.foodres.2010.11.024.

[5] Alamar, Priscila D., Elem T.S. Caramés, Ronei J. Poppi, and Juliana A.L. Pallone. 2016. “Quality Evaluation of Frozen Guava and Yellow Passion Fruit Pulps by NIRS Spectroscopy and Chemometrics.” Food Research International 85: 209–14. https://doi.org/10.1016/j.foodres.2016.04.027.

[6] Maniwara, Phonkrit, Kazuhiro Nakano, Shintaroh Ohashi, Danai Boonyakiat, Pimjai Seehanam, Parichat Theanjumpol, and Pichaya Poonlarp. 2019. “Evaluation of NIRS as Non-Destructive Test to Evaluate Quality Traits of Purple Passion Fruit.” Scientia Horticulturae 257 (July): 108712. https://doi.org/10.1016/j.scienta.2019.108712.

[7] Rungpichayapichet, Parika, Marcus Nagle, Pasinee Yuwanbun, Pramote Khwiijitjaru, Busarakorn Mahayothee, and Joachim Müller. 2017. “Prediction Mapping of Physicochemical Properties in Mango by Hyperspectral Imaging.” Biosystems Engineering 159 (2011): 109–20. https://doi.org/10.1016/jbiosystemseng.2017.04.006.

[8] Teerachaichayut, Sontisuk, and Huong Thanh Ho. 2017. “Non-Destructive Prediction of Total Soluble Solids, Titratable Acidity and Maturity Index of Limes by near Infrared Hyperspectral Imaging.” Postharvest Biology and Technology 133 (March): 20–25. https://doi.org/10.1016/j.postharvbio.2017.07.005.

[9] Manley, Marena, Elizabeth Joubert, Lindie Myburgh, and Martin Kidd. 2007. “Prediction of Soluble Solids Content and Post-Storage Internal Quality of Bulida Apricots Using Near Infrared Spectroscopy” 188: 179–88. https://doi.org/10.1255/jnirs.725.

[10] Gómez, Antihus Hernández, Yong He, and Ania García Pereira. 2006. “Non-Destructive Measurement of Acidity, Soluble Solids and Firmness of Satsuma Mandarin Using Vis / NIRS-Spectroscopy Techniques.” Journal of Food Engineering 77: 313–19. https://doi.org/10.1016/j.jfoodeng.2005.06.036.

[11] Neama, K., U L. Opara, S Z. Tesfay, O A. Fawole, and L S. Magwaza. 2017. “Application of Vis/NIR Spectroscopy for Predicting Sweetness and Flavour Parameters of ‘Valencia’ Orange (Citrus Sinensis) and ‘Star Ruby’ Grapefruits (Citrus x Paradisi Macfie).” Journal of Food Engineering, no. 193: 86–94. https://doi.org/10.3390/app7010010.

[12] Guo, W., W. Li, B. Yang, Z.Z. Zhu, D. Liu, and X. Zhu. 2019. “A Novel Noninvasive and Cost-Effective Handheld Detector on Soluble Solids Content of Fruits.” Journal of Food Engineering 257: 1–9.

[13] Rahman, Anisur, Lalit Mohan Kandpal, Santosh Lohumi, Moon S Kim, and Hoonsoo Lee. 2017. “Nondestructive Estimation of Moisture Content, PH and Soluble Solid Contents in Intact Tomatoes Using Hyperspectral Imaging.” Applied Sciences, no. January. https://doi.org/10.3390/app7010109.

[14] Huang, Yaping, Renfu Lu, and Kunjie Chen. 2018. “Assessment of Tomato Soluble Solids Content and PH by Spatially-Resolved and Conventional Vis/NIR Spectroscopy.” Journal of Food Engineering 236 (May): 19–28. https://doi.org/10.1016/j.jfoodeng.2018.05.008.

[15] Qin, Jianwei, and Renfu Lu. 2008. “Measurement of the Optical Properties of Fruits and Vegetables Using Spatially Resolved Hyperspectral Diffuse Reflectance Imaging Technique.” Postharvest Biology and Technology 49: 355–65. https://doi.org/10.1016/j.postharvbio.2008.03.010.

[16] Choudhary, R., T. J. Bowser, P. Weckler, N. O. Maness, and W. McGlynn. 2009. “Rapid Estimation of Lycopene Concentration in Watermelon and Tomato Puree by Fiber Optic Visible Reflectance Spectroscopy.” Postharvest Biology and Technology 52 (1): 103–9.
https://doi.org/10.1016/j.postharvbio.2008.10.002.

[17] Guillén-Casla, V., N. Rosales-Conrado, M.E. León-González, L.V. Pérez-Arribas, and L.M. Polo-Diez. 2011. “Principal Component Analysis (PCA) and Multiple Linear Regression (MLR) Statistical Tools to Evaluate the Effect of E-Beam Irradiation on Ready-to-Eat Food.” Journal of Food Composition and Analysis 24: 456–64.

[18] Williams, P.C. 2001. “Implementation of Near-Infrared Technology.” In Near-Infrared Technology in the Agricultural and Food Industries, edited by P.C. Williams and K. Noris, 2nd ed. Minnesota: The American Association of Cereal Chemist, Inc.

[19] Clément, A., M. Dorais, and M. Vernon. 2008. “Nondestructive Measurement of Fresh Tomato Lycopene Content and Other Physicochemical Characteristics Using Visible- NIR Spectroscopy.” J. Agric. Food Chem. 56: 9813–18. https://doi.org/10.1021/jf801299r.

[20] Cozzolino, D, M B Esler, R G Dambergs, W U Cynkar, D R Boehm, I L Francis, and M Gishen. 2004. “Prediction of Colour and PH in Grapes Using a Diode Array Spectrophotometer (400 – 1100 Nm).” J. Near Infrared Spectrosc. 12, 111: 105–11.