International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 4
ISSN: 2088-5334

# Estimation of Moisture and Protein Content in Pumpkin Seeds Using NIRS with Partial Least Square (PLS) Method

Ifmalinda [a,*], Andasuryani [a], Santosa [a], Iriwad Putri [a]

[a] *Department of Agricultural Engineering and Biosystem, Andalas University, Limau Manis, Padang, 25163, Indonesia*
*Corresponding author: [*]ifmalinda@ae.unand.ac.id*

***Abstract*—Most people do not use pumpkin seeds and only fruit for food production. Meanwhile, pumpkin seeds contain antioxidants, protein, carbohydrates, and other vitamins. Measurement of the chemical content of a material is usually carried out destructively (conventional methods) in the laboratory. This method is expensive, takes a long time to repair samples, and creates chemical waste. One method that is currently being developed to detect the chemical content of a material is the near-infrared spectroscopy (NIRS) method. Near-infrared spectroscopy (NIRS) is a non-destructive method used to quickly analyze and obtain information on a material's chemical content without using chemicals. The NIRS data calibration technique using partial least squares (PLS) is a step carried out to build a model that relates the response spectra of each sample at each wavelength to chemical concentrations known from laboratory analysis and protein content in pumpkin seeds with NIRS using PLS calibration. Spectrum data pre-treatment was carried out with GapDerivarive and Derivative Savitzky-Golay. The results showed that the best calibration model for moisture and protein content was obtained using DerivativeGap data processing with values of r = 0.92, R2 = 0.84, SEC = 0.11%, and RMSEC = 0.11%. RPD = 2.66 and using Latent Variable (LV) factor 2, while for protein content r = 0.95, R2 = 0.89, SEC = 0.71%, and RMSEC = 0.70%. RPD value = 3.33 and using Latent Variable (LV) factor 3.**

***Keywords*— Pumpkin seeds; NIRS; PLS; protein; moisture content.**

*Manuscript received 17 Aug. 2021; revised 8 Oct. 2021; accepted 13 Dec. 2021. Date of publication 31 Aug. 2022.*


## I. INTRODUCTION

Most people do not use pumpkin seeds and only use fruit for food production. Even pumpkin seeds contain various nutrients and can be used as drugs that can lower blood cholesterol [1]. In comparison, the nutritional content contained in pumpkin seeds includes protein, carbohydrates, and various other minerals. One of the main nutritional contents of pumpkin seeds is protein. Protein is a macromolecule formed from amino acids and consists of nitrogen, carbon, and oxygen atoms. In addition to the nutritional content of pumpkin seeds, the moisture content is also a factor that must be known from a food ingredient. Moisture content is one factor that can affect a product's quality and shelf life. Measurement of moisture and protein content is usually carried out in the laboratory by damaging the material and requires chemicals, skilled labor, expensive costs, and a long time. One method that is currently being developed is Near Infrared Spectroscopy (NIRS). In recent years, near-infrared (NIR) technology, which is combined with computer science, spectroscopy, and chemometrics, has developed rapidly. NIR has the advantages of rapid speed, less time spent, no sample destruction, and no need for sample pre-treatment, and it has been applied to the analysis of agricultural food products [2]. Using a certain wavelength, this method detects a material's chemical levels without damaging the product. NIR technique could be utilized in the development of authentication methods because able to detect small differences in the chemical composition present in some materials [3]. The advantages of this method are that it does not damage materials, has relatively easy sample preparation, does not require chemicals, is carried out accurately and quickly, as well as can predict the content of several heavy metals simultaneously [4].

According to Cen and He [5]. The NIRS method can estimate the content of a substance up to 0.1%. The NIRS data calibration technique uses partial least square regression (PLSR). According to Iskandar et al. [6], Infrared data will be used as data to predict the quality of the material by making a regression model between optical waves spectrum (X data) and quality parameters such as protein, fat, carbohydrates, and minerals (Y data). The regression of the model is to be built using principal component regression (PCR) and fractional

squares regression (PLSR) methods. PLS components are similar to principal components but are referred to as factors. PLS can be used for all data because it is a powerful analytical method, does not require many assumptions, and the sample size does not have to be large. These models were developed commonly by regressing spectra data (as X variable) and actually measuring quality attributes (as Y variable) through multivariate analysis. Partial least squares regression (PLSR) was applied to develop those models and validated using a full cross-validation approach [7].

Many studies related to the use of NIRS have been carried out, such as evaluating the quality of sapodilla fruit with the NIRS method [8]; non-destructive determination of taste-related compounds in tomatoes using NIR spectra [9]; determination of insect infestation on stored rice by near-infrared (NIR) spectroscopy [10]; contributions of Fourier-Transform Mid-Infrared (FT-MIR) spectroscopy to the study of fruit and vegetables [11]. This study aims to examine and apply Near Infrared Spectroscopy (NIRS) technology as a new method for predicting the protein content and moisture content of pumpkin seeds quickly and non-destructively.

## II. Materials and Method

The materials used in this study were pumpkin seeds of the bokor (cerme) variety with harvest criteria; namely, the skin of the pumpkin is dark yellow with a slightly greenish color [12]. The tools used in this study include one unit of NIR Spectrometer brand NIFLEX N-500 with a wavelength range of 1000 nm to 2500 nm or 10000 - 4000/cm, which is equipped with a laptop unit that is used to process spectral data, and LED lights and Spectroscopy Software. for data spectrum analysis.

### *A. Pumpkin Seed Preparation*

The pumpkin seed variety used in this study was the bokor variety. Pumpkin seeds are separated from the pumpkin, cleaned, and removed from the mucus, then washed and air-dried or drained. The next process is the pumpkin seeds are dried to a moisture content of <12% using an oven dryer with a gas heat source. Drying was carried out using a drying temperature of $60^0$ C. The number of samples prepared was 50 samples, then the NIRS spectrum was taken and then continued with the measurement of moisture content.

### *B. Moisture Content Measurement*

Measurement of moisture content was carried out using an oven, starting with drying the aluminum cup for 30 minutes in the oven at a temperature of 100 - 105ºC, then the cup was put in a desiccator for 30 minutes and then weighed. A sample of 5 g was placed in a dry cup. The oven temperature used to heat the cup containing the sample was 105-110 ºC at 3-4 hours. When heating is complete, the cup is removed, put in a desiccator for 30 minutes, and then weighed. This step is repeated until a constant moisture content weight is obtained. The following equation can calculate the moisture content:

$$\text{Moisture Content (M)} = \frac{(B-C)}{(B-A)} \times 100\% \qquad (1)$$

Information:
MC = Moisture content (%)
A = Cup weight (g)
B = Weight of material before drying + cup (g)
C = Weight of material after drying + cup (g)

### *C. Protein Content Measurement*

A sample of 1 g was placed in a Kjeldahl flask. The sample was then added with 15 g of K2SO4, 1 mg of CuSO4 catalyst, 1 g of selen catalyst, boiling rock, and 25 mg of concentrated H2SO4. The mixture is then boiled until the sample changes color to clear green, then cooled and diluted using distilled water as needed; 75 ml of 30% NaOH solution is given before being distilled for 5-10 minutes until the solution reaches 150 ml, with 50 ml of 4% H3BO3 solution as a container. . The solution was titrated using 0.1 N HCl. The difference between the total value of the titrated sample and the blank was expressed as the total nitrogen value. The protein content contained in dragon fruit stem flour is obtained from the process of multiplying N% with a convection factor of 6.25 [13]. Protein content can be calculated by the equation:

$$N\% = \frac{(\text{HCL sample (ml)} - \text{HCL blanko (ml)}) \text{ x N HCl x } 1{,}4007}{\text{Sample weight (gram)}} \text{x} 100\% \qquad (2)$$

$$\text{Protein (\%)} = N\% \text{ x } 6{,}25 \qquad (3)$$

### *D. NIRS Measurement*

NIR measurements were carried out at the Food and Agricultural Product Processing Engineering Laboratory (TPPHP), Department of Agricultural Engineering, IPB. The number of samples used is 50 grams. The pumpkin seed samples were arranged into a Petri cup with a weight of 50 grams each (9.7 cm in diameter and 1.5 cm in height), then placed on a rotary plate on the FT-NIR spectrometer and scanned three times. The scanning results from the three replicates were transformed into absorbance spectra and then averaged. The reflectance of pumpkin seeds was measured at a wavelength of 1000-2500 nm.

### *E. Data Pre-Treatment*

Before analyzing the data to build a predictive model, the NIRS spectrum data was pre-processed on pumpkin seeds. This aims to minimize and eliminate interference or noise on the spectrum to make the resulting prediction model more accurate. The spectrum obtained from the results of the NIR reflectance measurement was then transformed into an absorbance spectrum. Then pre-treated, the data for further analysis using Partial Least Squares (PLS). In this study, the processing of the NIRS spectrum data was corrected by two kinds of pre-treatment. Pre-treatment is carried out before calibration and validation to eliminate disturbances during spectrum collection to obtain a more accurate and stable model because it can reduce noises and overcome radiation-related problems or the influence of wave interference. The data treatments used in this study were (1) GapDerivarive (second derivative) and (2) Savitzky-Golay Derivative (first derivative). Data processing with pre-treatment of the first and second derivatives was also carried out in [14] research to predict melon ripeness parameters using Near Infra-red Spectroscopy. According to Mathian et al. [15], derivatives function to separate correlated to produce lower reflectance values and the spectra display clearer peaks and valleys. The first and second derivatives serve to increase the spectral resolution and remove the background [5].

### F. NIRS Data Calibration Model Using the PLS Method

Protein content and moisture content of pumpkin seeds were predicted by building a model based on data from the acquisition of the NIRS spectrum (variable X) and moisture content data using the gravimetric method from the results of measurements in the laboratory (variable Y). Calibration was carried out to determine the correlation between destructively chemical content and NIRS reflectance. One of the calibration methods for analyzing NIR spectra data is the Partial Least Squares (PLS) method [16], One of the most promising and common regression methods in NIRS practices is partial least square regression (PLSR)

Prediction models performances were evaluated and judged for their accuracies and robustness using the following statistical indicators: the coefficient of determination (R2), correlation coefficient (r), root mean square error (RMSE), and the residual predictive deviation (RPD) defined as the ratio between the standard deviation and the RMSE [7]. Standard Error Calibration (SEC) indicates the accuracy of the calibration equation obtained from chemical and NIRS data. A low SEC value indicates a low predictive error of calibration. CV shows an error proportional to the chemical average of the material, and RPD shows the ratio between the standard deviation of chemical data and SEP [17].

Ifmalinda et al. [18] revealed that Standard Error Calibration (SEC) shows how well the calibration equation fits the data. The lower the SEC value indicates a lower predictive error of calibration. Standard Error Prediction indicates the inaccuracy of the validation model. A low value indicates a good model and should be low than the standard deviation of the actual data.

Munawar et al. [19] stated that the chemical content of materials could be predicted by building a model based on electro-optic properties (variable X) and measurement data in the laboratory (variable Y). The prediction model was built using the Partial Least Square regression (PLS) method. Several statistical parameters that explain the accuracy and precision of the NIRS method in predicting include the coefficient of determination (R2), which shows the correlation between the reference value and the predicted value. The maximum value is 1, which indicates a perfect prediction. The R2 value of 0.75 indicates the model can be used. The equation to calculate the value of R2 is used in the following equation

$$R^2 = \left( \frac{\sum (Xn - \bar{X}n)\,(Yn - \bar{Y}n)}{\sqrt{\sum (Xn - \bar{X}n)^2\,(Yn - \bar{Y}n)^2}} \right)^2 \quad (4)$$

Information:

$Xn$ = chemical data (laboratory test)
$\bar{X}n$ = chemical data (laboratory test) average
$Yn$ = chemical data (NIRS estimate)
$\bar{Y}n$ = chemical data (NIRS estimate) mean

## III. Result and Discussion

### A. Pumpkin Seed NIRS Reflectance Spectra

The initial Raw Spectrum takes the NIRS spectrum in the wavelength range of 1000 nm to 2500 nm without pre-treatment. In the spectra of pumpkin seeds, peaks and valleys of absorption of chemical compounds are seen, which serve to facilitate the interpretation of the data. The peaks that appear on the graph are caused by the vibrations that occur when chemical bonds interact with NIR rays [20]. The results of the original raw NIRS spectrum analysis without processing can be seen in Fig. 1.

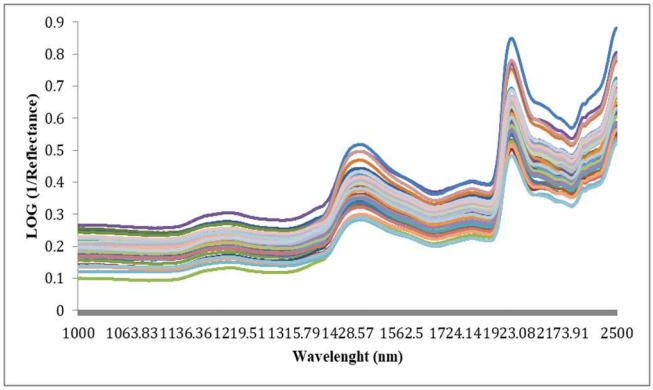


Fig. 1 Spectrum of Raw NIRS Original Pumpkin Seeds

Fig. 1 shows that the original spectra resulting from NIRS measurements on pumpkin seeds still have distances between the spectral data. There is noise and the possibility of cavities in the pile of pumpkin seeds in the Petri cup and interference in the form of unwanted signals [17]. According to Ma et al. [21], Once outlier data were removed, spectra data were corrected and enhanced using multiplicative scatter correction (MSC) to eliminate noises due to light scattering, sensor curvature and inside temperature.. Based on the results of the original spectra, it is necessary to carry out Pre-treatment with the Savitzky-Golay Derivative method (first derivative) and Gap Derivative Pretreatment (second derivative). According to Xu et al [22] to compensate the disadvantages of NIR, spectra were first transformed by the Savitsky-Golay (SG) smoothing to eliminate noise, 1st or 2nd derivative is used to remove background and increase spectral resolution.

According to Samadi et al. [23], to obtain an accurate and robust prediction model, it is possible to process the NIR spectral data with mathematical transformations of the spectrum using various types of pre-treatments, reducing and eliminating the effects unrelated to the property of interest. The spectra in this study were processed using the Derivative Savitky-Golay method and the results can be seen in Fig. 2.

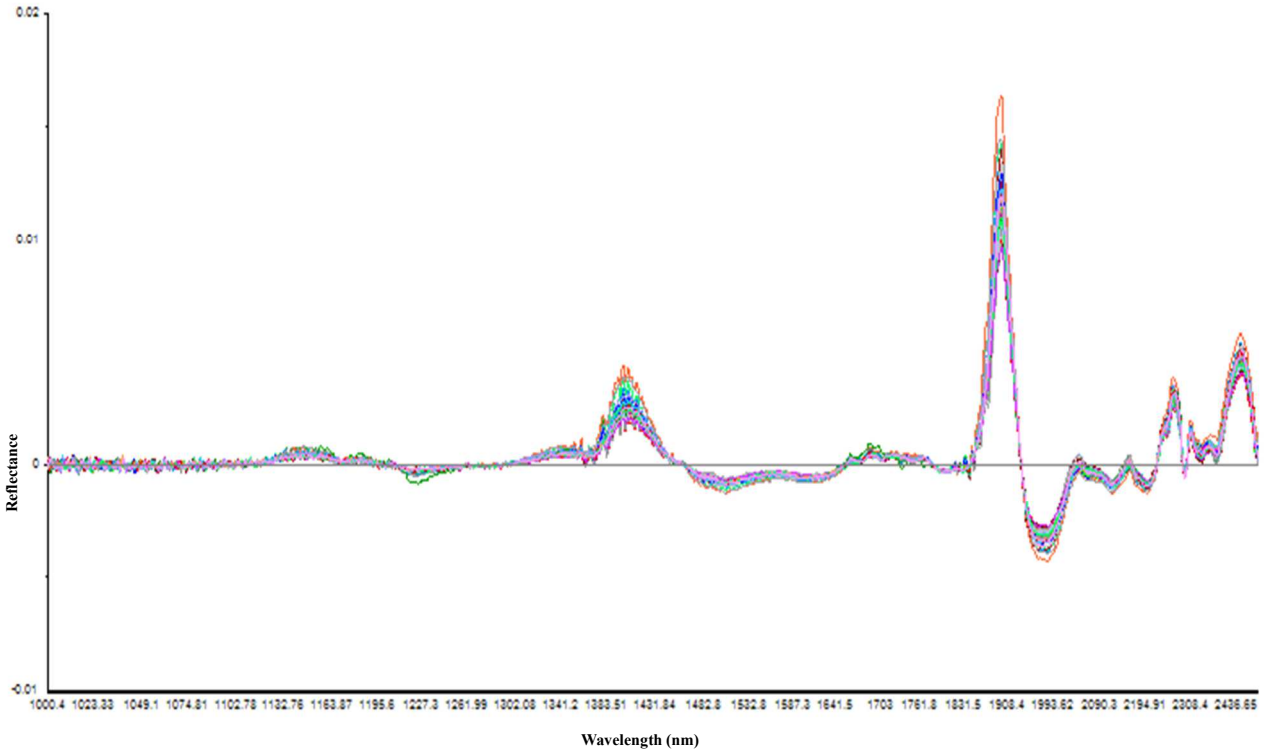


Fig. 2 Spectrum of Pumpkin Seeds with Savitzky-Golay Derivative Pretreatment

In Fig. 2, the results of the spectra look closer than the original spectra. The original spectra were transformed using the Derivative Savitzky-Golay method (first derivative) in order to reduce high-frequency noise in a signal due to its smoothing properties and reduce low-frequency signal (e.g., due to offsets and slopes) using differentiation [24]. According to Biancolillo et al. [14], spectra needed to be pre-treated before building the classification model to remove the unwanted variability possibly present in the data. Specifically, the effect of using first and second derivatives. The Savitzky-Golay Derivative Pretreatment can also eliminate multiplication interference in the distribution of data and particle size and can reduce the effect of scattering at NIR wavelengths so that the spectral results can be thinner. The same thing is stated by Budiastra et al. [17] that the first guidance method can separate overlapping spectra (overlapped).

The second pre-treatment carried out in this study was data processing with Gap Derivatives, as shown in Fig. 3.

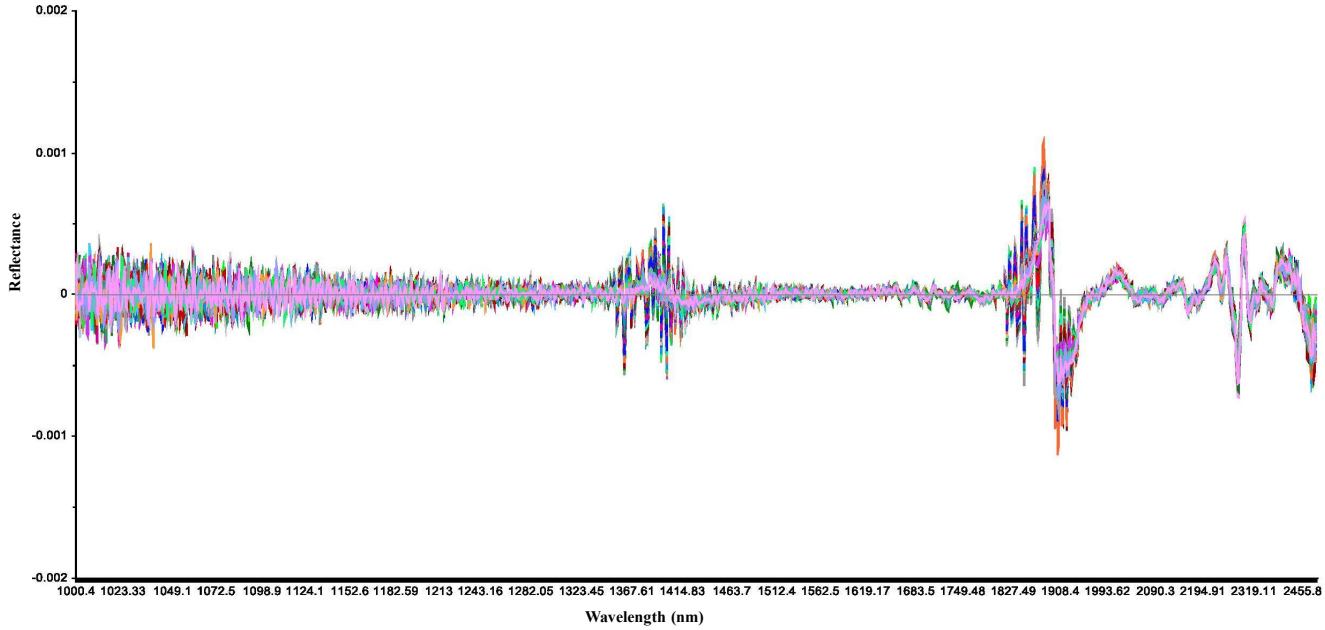


Fig. 3 Pumpkin Seed Spectrum with Gap Derivative Pre-treatment

In Fig. 3, the spectra results show that the peaks and valleys of absorption of the chemical content produced are clearer and more abundant when compared to the Savitzky-Golay Derivative Pretreatment. According to Munawar et al. [25], the derivative data processing function works to separate overlapping spectra. All spectral features seen in the image of each pre-treatment still overlap with each other. The absorption band influences the spectral features of NIRS, and a combination occurs due to changes in molecular energy during absorption, so it is necessary to develop a calibration model using the Partial Least Square (PLS) method.

### *B. Pumpkin Seed Moisture and Protein Content*

The chemical content of pumpkin seeds in the form of moisture and protein content is shown in Table 1. The minimum and maximum values obtained in the chemical data are used as calibration data sets in building the calibration model.

TABLE I
THE STATISTICAL DESCRIPTION OF MOISTURE CONTENT AND PROTEIN CONTENT

| Chemical Content | Mean | Min | Max | Standard Deviation |
|---|---|---|---|---|
| Moisture Content | 12.30 | 11.05 | 13.01 | 0.32 |
| Protein | 29.17 | 24.15 | 34.15 | 2.33 |

### *C. Loading Pumpkin Seed Spectra Plot*

The loading plot of the resulting spectra shows the absorption wave of the desired chemical content to identify the presence of moisture and protein content. The NIRS spectrum of pumpkin seeds has peaks and valleys at certain wavelengths. The moisture content consisting of O-H bonds absorbs at a wavelength of 1147-1901 nm. The peak moisture content formed in pumpkin seeds was between 1147 - 1150 nm, 1400 - 1410 nm, and 1895 - 1901 nm. Meanwhile, the peak protein content (CHON) in pumpkin seeds was at the wavelengths of 1944-1950 nm, and 2024-2047 nm.

The calibration results of the NIRS absorbance spectra on pumpkin seeds show that infrared with a wavelength of 1000-2500 nm can be used to identify the chemical content of the material. The loading plots obtained to identify the moisture and protein content can be seen in Fig. 4 and Fig. 5.

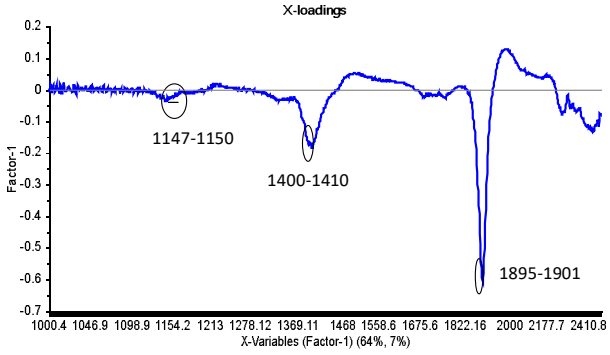


Fig. 4 Loading Plot of Derivative Savitzky-Golay Moisture Content

The absorption peak at wavelength 1147-1150 O-H bonds looks not sharp. The presence of a high-intensity absorption peak is seen at the wavelength of 1895-1901. The presence of a high-intensity absorption peak is seen at the wavelength of 1895-1901. Strong NIR absorption bands occurring at 1400-1440 and 1900-1950 nm are often applied for quantitative analysis of moisture content in foods. According to Zhao et al. [26], materials with high moisture content can be seen at a wavelength of 1450 nm. The results of the research by Chen et al. [27] are moisture absorption in black rice at wavelengths of 1932 nm and 1443 nm.

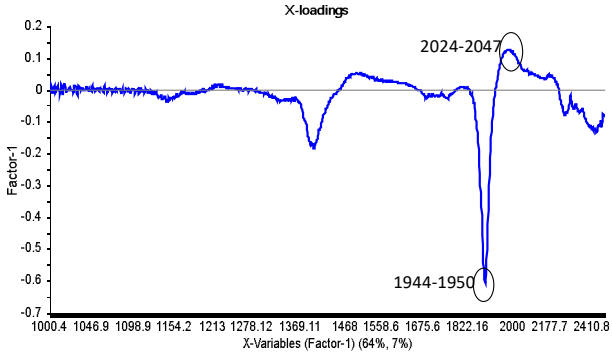


Fig. 5 Loading Plot with Derivative Savitzky-Golay Protein

Pumpkin seeds' peak protein content (CHON) was at 1944-1950 nm and 2024-2047 nm wavelengths. According to Athfiyah et al. [28], the spectra show more of the peak of absorbance, meaning that more chemical contents of the sample were detected. The peaks and valleys of the spectrum are also influenced by the physical properties of the material to be analyzed. Another study by Chen et al. [29] on rice flour obtained wavelength values observed of 1466, 1934, 2104, 2288, and 2322 nm.

### *D. Prediction of Model Calibration Using the PLS (Partial Least Square Regression) Method*

PLS is a promising procedure used to conduct consecutive wavelength selection for building an optimal calibration model; this method has proven effective for waveband selection in analyzing many objects [30]. Partial Least Squares (PLS) will provide spectral information relevant to the required chemical content. As one of the models, the Partial Least Squares (PLS) regression model looks for the correlation between the x variable and the y variable.

The model evaluation indicators used in this paper are coefficient determination of calibration (R2c), root mean square error of calibration (RMSEC), coefficient determination of prediction ($R^2p$) and root mean square error of prediction (RMSEP). The smaller the RMSEP, the larger the R2p, and the smaller the MRE, the better the quantitative model performance [31].

The plot of calibration data for the pumpkin seed moisture content test using PLS method data processing with non-pretreatment can be seen in Fig. 6. The prediction results of the calibrated moisture content of pumpkin seeds with non-pretreatment still show a low prediction model, this is indicated by the values: r = 0.60, R2 = 0.37, SEC = 0.24%, and RMSEC = 0.24%, RPD = 1.33 and using Latent Variable (LV) factor 7. This is classified in the category of predictions that are still rough. The SEC value is a value that indicates the accuracy of the calibration equation.

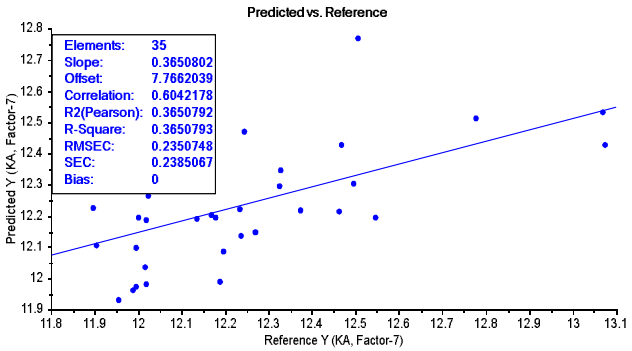


Fig. 6 Plot of Calibration Data for Moisture Content Test of Non-Pretreatment Pumpkin Seeds

According to Zhao et al. [26], The best model should have higher Rc2, Rcv2, and Rp2 values and lower RMSEC, RMSECV, and RMSEP. The resulting model is still not good because the R2 value is still very low. According to Deng et al. [2], The coefficient of determination (R2), correlation coefficient ®, root mean square error (RMSE), and the residual predictive deviation (RPD) are defined as the ratio between the standard deviation and the RMSE. Good and excellent prediction model performance should have high R2, r coefficient (equal to or above 0.8), high RPD index (above 3), and low RMSE.

The plot of calibration data for the pumpkin seed moisture content test using PLS method data processing with Savizky-Golay pre-treatment can be seen in Fig. 7.

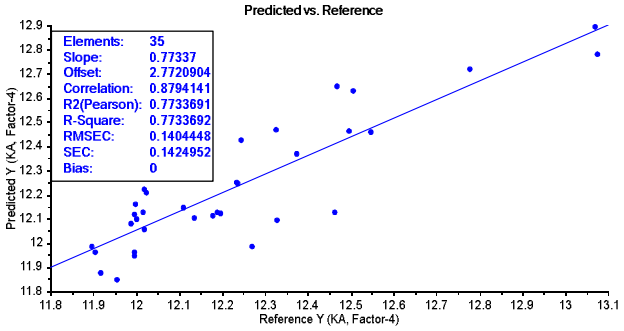


Fig. 7 Plot of Calibration Data Moisture Content of Pumpkin Seed Pretreatment Derivative Savitzky-Golay

The prediction results of the calibrated moisture content of pumpkin seeds with Derivative Savitzky-Golay pre-treatment improved the prediction model to be better than without pre-treatment. This is indicated by the values: r = 0.88, R2 = 0.77, SEC = 0.14, and RMSEC = 0.14%. RPD = 2.29 and using the Latent Variable (LV) factor 4. The results of the prediction of PLS calibration predictions are good. The value of R indicates the model's ability to explain the diversity of the values of the dependent variable. The higher the value of R, the more accurate the model [17]. The estimation results related to the number of PLS factors used are smaller than those without pre-treatment, namely with LV 4. This is in line with Pasquini [32] that weak non-linearities may be overcome by increasing the number of latent variables included in the PLS regression model. SEC is a value that indicates the accuracy of the calibration equation. The smaller the SEC value indicates a low prediction error of calibration, a small value indicates a good model [25].

The plot of calibration data for the pumpkin seed moisture content test using PLS method data processing with Derivative Gap pre-treatment can be seen in Fig. 8.

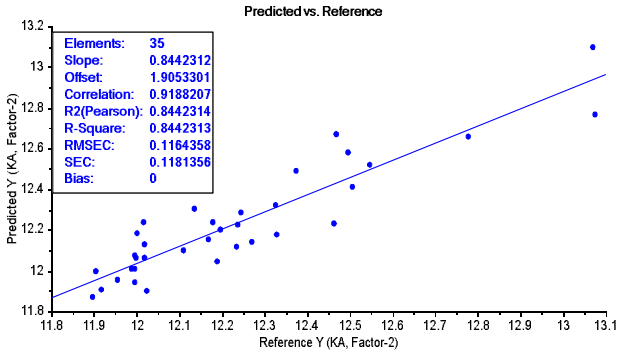


Fig. 8 Plot of Calibration Data for Pumpkin Seed Moisture Content Pretreatment Derivative Gap

The prediction results of the calibrated moisture content of pumpkin seeds with DerivativeGap pre-treatment were able to improve the prediction model to be better than without pre-treatment. This is indicated by the values: r = 0.92, R2 = 0.84, SEC = 0.11%, and RMSEC = 0.11%. RPD = 2.66 and using the Latent Variable (LV) factor 2. According to Munawar et al. [25], the prediction model with the least number of latent variables is better and more efficient than the prediction model with more LV or PC by Ibáñez et al. [9]. Minimum RMSECV values and the number of latent variables were used as the selection criteria for the number of latent variables to be included in the model. New latent variables were included if they provided a reduction of RMSECV higher than 2%. An RPD value greater than 1.5 indicates the model is said to be good and can predict correctly. The prediction results of PLS calibration predictions are very good. According to research by Budiastra et al. [17], the greater the RPD, the better the accuracy of the prediction model and the model that will be chosen as the best model.

The plot of calibration data for pumpkin seed protein content test using PLS method data processing with non-pretreatment can be seen in Fig. 9.

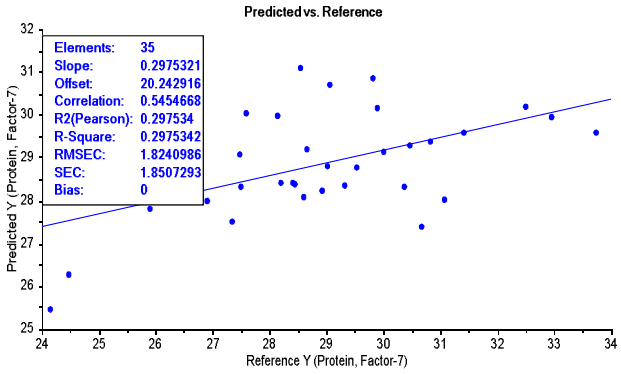


Fig. 9 Plot of Calibration Data Test for Pumpkin Seed Protein Content Non Pre-treatment

The prediction results of the calibrated moisture content of pumpkin seeds with non-pretreatment still show a very low prediction model, this is indicated by the values: r = 0.55, R2

= 0.29, SEC = 1.85%, and RMSEC = 1.82%. RPD value = 1.27 and using Latent Variable (LV) factor 7. This is classified in the category of predictions that are still rough. A high PLS factor value can reduce the ability to predict. The selection of the number of factors in PLS is one of the important stages in making the NIR calibration model. The PLS factor is needed to overcome the overfitting model due to the SEP value being higher than the SEC value and the underfitting model due to the SEC being higher than the SEP value [33]. The performance of the prediction model was evaluated using an internal cross-validation method, which incorporates root mean square error of calibration (RMSEC), standard error of cross-validation (SECV), and correlation coefficient of cross-validation (Rcv). Smaller RMSEC and SECV and higher values of Rcv indicate better prediction model performance [34].

The plot of calibration data for pumpkin seed protein content test using PLS method data processing with Derivative Savizky-Golay pre-treatment can be seen in Fig. 10.

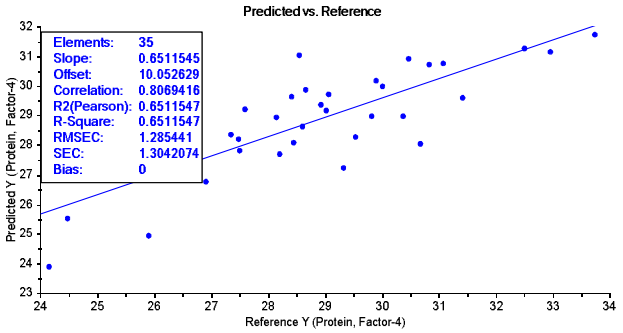


Fig. 10 Plot of Pumpkin Seed Protein Test Calibration Data with Savitzky-Golay Derivative Pretreatment

The prediction results of the calibrated moisture content of pumpkin seeds with the Derivative Savitzky-Golay pre-treatment improved the prediction model, the values showed this: r = 0.81, R2 = 0.65, SEC = 1.30%, and RMSEC = 1.28%. RPD value = 1.81 and using Latent Variable (LV) factor 4. This is classified in the category of predictions that are still rough. The SEC value is a value that indicates the accuracy of the calibration equation. Good and excellent prediction model performance should have high R2, r coefficient (equal to or above 0.8), high RPD index and low RMSE [7]. The plot of calibration data for pumpkin seed protein content test using PLS method data processing with DerivativeGap pre-treatment can be seen in Fig. 11.

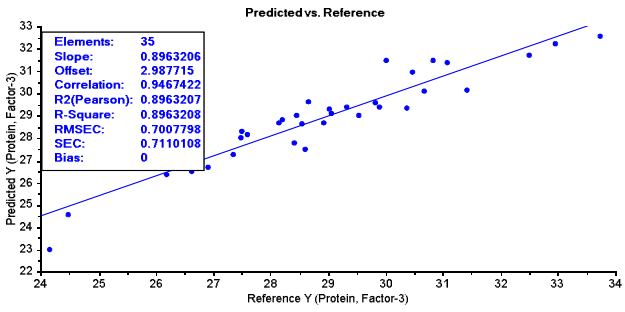


Fig. 11 Plot of Pumpkin Seed Protein Test Calibration Data with Derivative Gap Pre-treatments

The prediction results of the calibration value of pumpkin seed protein content with DerivativeGap pre-treatment were able to improve the prediction model to be better than without pre-treatment. According to Mathian et al [15], the use of DerivativeGap pre-treatment or this second derivative has The second derivative procedure was employed to enhance the resolution of sharply defined features (which are hereafter named components) where these overlap with broad bands.. This is indicated by the values: r = 0.95, R2 = 0.89, SEC = 0.71%, and RMSEC = 0.70%. RPD value = 3.33 and using Latent Variable (LV) factor 3. RPD value higher than 1.5 indicates the model is said to be good and can predict correctly. According to Genisheva et al. [35] RPD value greater than 3 is considered fair and recommended for screening purposes, and an RPD value greater than 5 is considered good for quality control. All the RPD values calculated in the present report are higher than 3 confirming the good performance of the developed models.

The most precise prediction results of PLS estimation for pumpkin seed protein content are data processing using Savitzky-Golay pre-treatment and DerivativeGap pre-treatment able to reduce the number of factors used in data processing [36]. The consideration of all latent variables identifies this performance with coefficient of determination R2 of the linear regression.

## IV. Conclusion

The NIRS method can be used to predict the moisture content and protein content of pumpkin seeds properly. The best calibration model for moisture and protein content was obtained using DerivativeGap data processing with values of r = 0.92, R2 = 0.84, SEC = 0.11%, and RMSEC = 0.11%. RPD = 2.66 and using Latent Variable (LV) factor 2, while for protein content r = 0.95, R2 = 0.89, SEC = 0.71%, and RMSEC = 0.70%. RPD value = 3.33 and using Latent Variable (LV) factor 3.

## Acknowledgments

The authors thank Andalas University for providing the Grant of this research (PDU-KRP2GB- Unand) through Andalas University BOPTN DIPA fund, with contract number 1297/UN16.R/XII/KPT/2021, Maret 29, 2021.

## References

[1] E. Puspaningdyah and D. Herawati, “Kombinasi Bilakupu (Biji Labu Kuning Dan Kunyit Putih) dalam Menurunkan Hiperkolesterolemia,” *J. Sains Heal.*, vol. 4, no. 1, 2020.

[2] Y. Deng, Y. Wang, G. Zhong, and X. Yu, "Simultaneous quantitative analysis of protein, carbohydrate and fat in nutritionally complete formulas of medical foods by near-infrared spectroscopy," *Infrared Phys. Technol.*, vol. 93, pp. 124–129, 2018, doi: 10.1016/j.infrared.2018.07.027.

[3] Rachmawati, E. Rohaeti, and M. Rafi, "Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour," *J. Phys. Conf. Ser.*, vol. 835, no. 1, 2017, doi: 10.1088/1742-6596/835/1/012011.

[4] C. Gardana, A. Scialpi, C. Fachechi, and P. Simonetti, "Near-infrared spectroscopy and chemometrics for the routine detection of bilberry extract adulteration and quantitative determination of the anthocyanins," *J. Spectrosc.*, vol. 2018, 2018, doi: 10.1155/2018/4751247.

[5] H. Cen and Y. He, "Theory and application of near infrared reflectance spectroscopy in determination of food quality," *Trends Food Sci.*

*Technol.*, vol. 18, no. 2, pp. 72–83, 2007, doi: 10.1016/j.tifs.2006.09.003.

[6] C. D. Iskandar, Zainuddin, and A. A. Munawar, "Rapid assessment of frozen beef quality using near infrared technology," *Int. J. Sci. Technol. Res.*, vol. 9, no. 5, pp. 156–160, 2020.

[7] A. A. Munawar, Kusumiyati, and D. Wahyuni, "Near infrared spectroscopic data for rapid and simultaneous prediction of quality attributes in intact mango fruits," *Data Br.*, vol. 27, p. 104789, 2019, doi: 10.1016/j.dib.2019.104789.

[8] Kusumiyati *et al.*, "Non-Destructive Method for Predicting Sapodilla Fruit Quality Using Near Infrared Spectroscopy," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 334, no. 1, 2019, doi: 10.1088/1755-1315/334/1/012045.

[9] G. Ibáñez, J. Cebolla-Cornejo, R. Martí, S. Roselló, and M. Valcárcel, "Non-destructive determination of taste-related compounds in tomato using NIR spectra," *J. Food Eng.*, vol. 263, no. April, pp. 237–242, 2019, doi: 10.1016/j.jfoodeng.2019.07.004.

[10] A. Biancolillo, P. Firmani, R. Bucci, A. Magrì, and F. Marini, "Determination of insect infestation on stored rice by near infrared (NIR) spectroscopy," *Microchem. J.*, vol. 145, no. August 2018, pp. 252–258, 2019, doi: 10.1016/j.microc.2018.10.049.

[11] S. Bureau, D. Cozzolino, and C. J. Clark, "Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review," *Postharvest Biol. Technol.*, vol. 148, no. September 2018, pp. 1–14, 2019, doi: 10.1016/j.postharvbio.2018.10.003.

[12] N. P. Sari and W. D. R. Putri, "Pengaruh Lama Penyimpanan dan Metode Pemasakan Terhadap Karakteristik Fisikokimia Labu Kuning (Cucurbita moschata)," *J. Pangan dan Agroindustri*, vol. 6, no. 1, pp. 17–27, 2018, doi: 10.21776/ub.jpa.2018.006.01.3.

[13] J.S.H., "Association of official agricultural chemists," *J. Franklin Inst.*, vol. 219, no. 2, p. 236, 1935, doi: 10.1016/s0016-0032(35)91522-8.

[14] A. Biancolillo *et al.*, "Authentication of an Italian PDO hazelnut ('Nocciola Romana') by NIR spectroscopy," *Environ. Sci. Pollut. Res.*, vol. 25, no. 29, pp. 28780–28786, 2018, doi: 10.1007/s11356-018-1755-2.

[15] M. Mathian *et al.*, "Identifying the phyllosilicate minerals of hypogene ore deposits in lateritic saprolites using the near-IR spectroscopy second derivative methodology," *J. Geochemical Explor.*, vol. 186, pp. 298–314, 2018, doi: 10.1016/j.gexplo.2017.11.019.

[16] A. A. Munawar, Kusumiyati, Hafidh, R. Hayati, and D. Wahyuni, "The application of near infrared technology as a rapid and non-destructive method to determine vitamin C content of intact mango fruit," *INMATEH - Agric. Eng.*, vol. 58, no. 2, pp. 1–12, 2019, doi: 10.35633/INMATEH-58-31.

[17] I. W. Budiastra, S. Sutrisno, S. Widyotomo, and P. C. Ayu, "Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 147, no. 1, 2018, doi: 10.1088/1755-1315/147/1/012004.

[18] Ifmalinda, M. Makky, and F. Y. S, "Uji Non Destruktif Kandungan Kafein Biji Kopi Arabika Solok Rajo dengan Menggunakan Near Infrared Spectroscopy," 2017.

[19] A. A. Munawar, D. von Hörsten, J. K. Wegener, E. Pawelzik, and D. Mörlein, "Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics," *Eng. Agric. Environ. Food*, vol. 9, no. 3, pp. 208–215, 2016, doi: 10.1016/j.eaef.2015.12.004.

[20] Sudarjat, Kusumiyati, Hasanuddin, and A. A. Munawar, "Rapid and non-destructive detection of insect infestations on intact mango by means of near infrared spectroscopy," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 365, no. 1, 2019, doi: 10.1088/1755-1315/365/1/012037.

[21] Y. Ma *et al.*, "Characteristics of groundwater pollution in a vegetable cultivation area of typical facility agriculture in a developed city," *Ecol. Indic.*, vol. 105, no. November 2018, pp. 709–716, 2019, doi: 10.1016/j.ecolind.2018.10.056.

[22] R. Xu *et al.*, "Use of near-infrared spectroscopy for the rapid evaluation of soybean [Glycine max (L.) Merri.] water soluble protein content," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 224, p. 117400, 2020, doi: 10.1016/j.saa.2019.117400.

[23] Samadi, S. Wajizah, and A. A. Munawar, "Rapid and simultaneous determination of feed nutritive values by means of near infrared spectroscopy," *Trop. Anim. Sci. J.*, vol. 41, no. 2, pp. 121–127, 2018, doi: 10.5398/tasj.2018.41.2.121.

[24] J. B. Jianwen Luo, Kui Ying, "Savitzky–Golay smoothing and differentiation filter.pdf," no. January, 2020.

[25] A. A. Munawar, Y. Yunus, Devianti, and P. Satriyo, "Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties," *Data Br.*, vol. 30, p. 105469, 2020, doi: 10.1016/j.dib.2020.105469.

[26] X. Zhao, W. Wang, X. Ni, X. Chu, Y. F. Li, and C. Sun, "Evaluation of near-infrared hyperspectral imaging for detection of peanut and walnut powders in whole wheat flour," *Appl. Sci.*, vol. 8, no. 7, 2018, doi: 10.3390/app8071076.

[27] H. Chen, C. Tan, and Z. Lin, "Authenticity Detection of Black Rice by Near-Infrared Spectroscopy and Support Vector Data Description," *Int. J. Anal. Chem.*, vol. 2018, 2018, doi: 10.1155/2018/8032831.

[28] S. N. Athfiyah, I. W. Budiastra, Sutrisno, and Y. A. Purwanto, "Prediction of Piperine content in white pepper by NIR spectroscopy using Partial Least Square (PLS) method," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 542, no. 1, 2020, doi: 10.1088/1755-1315/542/1/012020.

[29] J. Y. Chen, Y. Miao, S. Sato, and H. Zhang, "Near infrared spectroscopy for determination of the protein composition of rice flour," *Food Sci. Technol. Res.*, vol. 14, no. 2, pp. 132–138, 2008, doi: 10.3136/fstr.14.132.

[30] P. S. Sampaio, A. Soares, A. Castanho, A. S. Almeida, J. Oliveira, and C. Brites, "Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms," *Food Chem.*, vol. 242, no. February 2017, pp. 196–204, 2018, doi: 10.1016/j.foodchem.2017.09.058.

[31] C. Chen, H. Li, X. Lv, J. Tang, C. Chen, and X. Zheng, "Application of near infrared spectroscopy combined with SVR algorithm in rapid detection of cAMP content in red jujube," *Optik (Stuttg).*, vol. 194, no. April, p. 163063, 2019, doi: 10.1016/j.ijleo.2019.163063.

[32] C. Pasquini, "Near infrared spectroscopy: A mature analytical technique with new perspectives – A review," *Anal. Chim. Acta*, vol. 1026, pp. 8–36, 2018, doi: 10.1016/j.aca.2018.04.004.

[33] A. Andasuryani, Y. A. Purwanto, I. W. Budiastra, and K. Syamsu, "Determination of Catechin Content in Gambir Powder from Dried Gambir Leaves Quickly using FT NIR PLS Model," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 4, no. 5, p. 303, 2014, doi: 10.18517/ijaseit.4.5.423.

[34] Z. Zhu, S. Chen, X. Wu, C. Xing, and J. Yuan, "Determination of soybean routine quality parameters using near-infrared spectroscopy," *Food Sci. Nutr.*, vol. 6, no. 4, pp. 1109–1118, 2018, doi: 10.1002/fsn3.652.

[35] Z. Genisheva, C. Quintelas, D. P. Mesquita, E. C. Ferreira, J. M. Oliveira, and A. L. Amaral, "New PLS analysis approach to wine volatile compounds characterization by near infrared spectroscopy (NIR)," *Food Chem.*, vol. 246, no. November 2017, pp. 172–178, 2018, doi: 10.1016/j.foodchem.2017.11.015.

[36] M. Bezara and S. Taibi, "An approach of PLS method applied to model the rice self-sufficiency of peasant households in Atsinanana Madagascar," *19th Int. Sci. Conf. "Economic Sci. Rural Dev. 2018". Integr. Sustain. Reg. Dev. Mark. Sustain. Consum.*, vol. 48, no. 48, pp. 321–327, 2018, doi: 10.22616/esrd.2018.101.