EVALUATION OF INDONESIAN LOCAL SOYBEAN BASED ON CHEMICAL CHARACTERISTICS AND VISIBLE - NEAR INFRARED SPECTRA WITH CHEMOMETRICS

FARID R. ABADI, RUDIATI EVI MASITHOH*, LILIK SUTIARSO AND SRI RAHAYOE

[1]*Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*

Received 15 July 2023/ Revised 25 October 2023/ Accepted 30 October 2023

ABSTRACT

Soybean characterization is essential to ensure product quality during distribution according to internal values. In this context, non-destructive characterization method, such as spectroscopy, offer an effective and efficient approach to testing soybean quality in field applications. Among the instruments that are widely used for testing soybean quality, the semi-portable visible near-infrared (Vis-NIR) spectrometer operating at a specific range of 345 to 1033 nm has been proven effective. Therefore, this study aimed to investigate soybean seeds characterization using Vis-NIR spectroscopy with PCA and PLSR chemometric methods. The investigation was carried out using soybean seeds consisting of eight varieties locally produced on Java Island, Indonesia, including Dega1, Dena1, Deja2, Dering1, Devon1, Yellow Flap, Green, and Detam4, in the form of intact, crumble, flour, and paste. Several quality parameters such as protein, fat, crude fiber, carbohydrate, ash, water, chlorophyll, total carotene, vitamin C, and L^*, a^*, and b^* values were measured across intact, crumble, flour, and paste samples. The results of Principal Component Analysis (PCA) showed that sample form and genotypes affected soybean classification. Furthermore, Partial Least Squares Regression (PLSR) showed adequate model calibration for crude fiber, chlorophyll, total carotene, and vitamin C parameters. Based on this analysis, it could be concluded that Vis-NIR spectroscopy proved to be suitable for the classification and prediction of soybean characterization.

Keywords: *chemometrics, spectroscopy, soybean, Vis-NIR.*

INTRODUCTION

Soybean is a staple food in Indonesian society, progressively gaining popularity annually alongside rice and corn in daily consumption (Harsono *et al.* 2021). In Indonesia, the national soybean consumption rate from 2015 to 2021 has reached 11.5 kg/capita/year, with an average growth of 3.02 kg/capita/year. Due to this high demand, a consistent supply is required, as the main source of soybean is through import. A previous study has found that local contribution to meet national needs is below 50%, with Java Island being the most substantial contributor, accounting for more than 55% of the total soybean production in Indonesia (Kementan 2020).

*Corresponding author, email: evi@ugm.ac.id

The distribution of local soybean based on origin varies significantly according to geographic location, variety, and genotype, with the majority being distributed as seeds in the Indonesian market. Based on quality levels, the production of soybean strives to meet consumer demand according to value (Jia *et al.* 2020). Therefore, information about conditions and varieties, as well as the quality and characteristics of soybean, is required to determine the potential and quality measures at reasonable market prices. Quality testing is characterized by the use of destructive sample method during laboratory analysis, which is time-consuming and expensive. This phenomenon shows the need for a reliable, easy, fast, and cost-effective method to determine the characteristics of soybean.

A common non-destructive method in food characterization is spectroscopy, widely

recognized for the good measurement accuracy combined with chemometric analysis. This method has been used to explore the characterization of food ingredients in the spectrum of infrared (IR) (Manley 2014), ultraviolet (UV) (Farag *et al.* 2022), and visible (Vis) (Cortés *et al.* 2019; Wang *et al.* 2022). Specifically, the combination of visible and near-infrared (Vis-NIR) light spectrum is frequently used for characterization, along with spectrum absorption by functional groups (Alander *et al.* 2013). In Vis-NIR spectrometer, light absorption occurs due to the movement or excitation of subatomic particles relating to the specific element concentration of chemical compounds in the material (Mayerhöfer *et al.* 2019). Therefore, Vis-NIR spectrometer is used to measure the amount of light absorbed or reflected.

Vis-NIR spectrometer has been extensively investigated, emphasizing its suitability for field applications (Walsh *et al.* 2020) due to simplicity, speed, and cost-effectiveness for quality evaluation (Cortés *et al.* 2019) and classification (Monago-Maraña *et al.* 2021). Specifically, fiber optic-based Vis-NIR spectroscopy has been explored for soybean characterization (Abadi *et al.* 2022), showing good performance for classification (Pahlawan *et al.* 2022), water evaluation (Pahlawan *et al.* 2023), and authentication (Masithoh *et al.* 2023). The results showed the potential of this instrument to support the production process for soybean characterization in field applications. Therefore, this study aimed to evaluate the application of Vis-NIR spectroscopy for characterizing soybean in various forms such as intact, crumble, flour, and paste.

MATERIALS AND METHODS

Materials

The materials used included local soybean obtained from the Indonesian Legume and Tuber Crops Research Institute (ILETRI), as well as Yogyakarta Region for the first planting season 2021 (ILETRI, 2016). Soybean seeds were at mature pod stage of R8 labeled as (1) Dega1 = K1, (2) Dena1 = K2, (3) Deja2 = K3, (4) Dering1 = K4, (5) Devon1 = K5, (6) Yellow Flap = K6, (7) Green = K7, and (8) Detam4 = K8.

Spectra Measurement

Approximately, 15 g soybean seeds were placed in petri dish and the spectra were measured for intact, crumble, flour, and paste forms. Soybean flour was obtained by milling seeds for 2 minutes to reduce the size and sifting through a 60-mesh sieve. The samples that did not pass the 60 mesh sieve were used as crumble, while the paste samples were obtained by mixing 5 g soybean flour with 10 ml distilled water in petri dish. Therefore, four forms of soybean were used in this study, namely intact, crumble, flour, and paste.

Spectroscopic data were obtained using Vis-NIR spectrometer (Flame-T-VIS-NIR, Ocean Optics, USA) with 345-1033 nm wavelength range in reflectance mode. The light source was a tungsten-halogen lamp (HL-2000-HP-FHSA, Ocean Optics, USA) with a reflectance probe (QR400-7-VIS-NIR, Ocean Optics, USA).

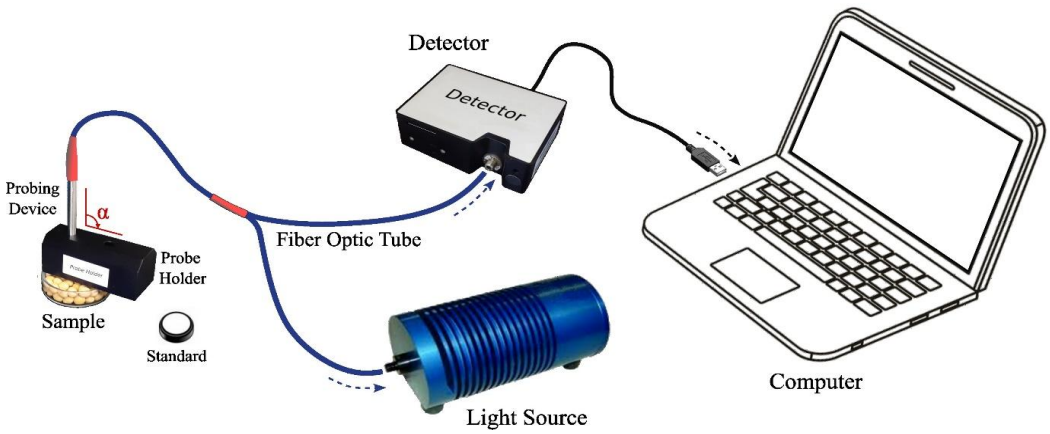


Figure 1 A portable Vis-NIR spectrometer arrangement

Bulk, flour, and paste soybean samples were poured into petri dish (Ø=5 cm), followed by spectra measurementdescribed by Abadi *et al.* (2022). The fiber optic sensor was placed in the probe holder, which was arranged perpendicularly ($\alpha=90^{o}$) to the sample at a distance of 2 mm, as shown in Figure 1. The probe sensor was connected to a light source and a detector capturing reflectance data, which was collected by OceanView 1.6.7 software (Ocean Insight, USA). Before reflectance spectra measurement, instrument calibration procedures were performed by measuring the white reflectance spectra standard, followed by black reference spectra. This calibration was performed on every 10 samples, which were individually measured in triplicate, as described by Masithoh *et al.* (2023). Subsequently, color and chemical measurements were carried out after spectra measurement.

Color and Chemical Parameters Measurement

Color parameters denoted by L^*, a^*, and b^* values were measured using a handheld color meter (TES 135A, Taiwan). L^* value represented lightness plotted on the vertical axis on a color space diagram with values ranging from 0 (black) to 100 (white). Moreover, a^* and b^* were color coordinates with a^* (positive) and a^* (negative) representing red and green, while b^* (positive) and b^* (negative) represented yellow and blue, respectively (Berns 2019). After color measurement, the destructive analyses were carried out to determine water content, crude protein, crude fiber, ash content, carbohydrates, fat, chlorophyll, total carotene, and vitamin C using the official AOAC methods (AOAC 2007). One-way analysis of variance (ANOVA) was performed for all chemical parameters and soybean samples. Subsequently, color and chemical parameters were analyzed using a single-factor analysis of variance with SPSS (SPSS Inc., Chicago, IL, USA), and mean comparisons were evaluated with Duncan's test.

Chemometric Analysis

Chemometric analysis, including Principal Component Analysis (PCA) and Partial Least Square Regression (PLSR), was conducted after measuring quality parameters such as color and chemical aspects. The data comprised four sample types and eight soybean varieties with five replications for each sample, totaling 160 data generated by spectra acquisition. The spectra and quality parameters data were used as x-variables in PCA analysis. For PLSR analysis, the reflectance spectra of all samples were used as x-variables, while color and chemical parameters served as y-variables. Compared to supervised PLSR, PCA is unsupervised chemometrics that does not require y-variables as predictors. During PLSR analysis, the 160 data obtained were divided into calibration and prediction data of 112 and 48, respectively. The performance of PLSR calibration model was determined from the determination of calibration (R^2c) coefficient and root mean square error of calibration (RMSEC). The calibration model was validated using a full cross-validation (CV) method. Subsequently, the best calibration PLSR model was applied to predict color and quality parameters using prediction data presented in the form of R^2p and RMSEP. PLSR model was improved using various spectra preprocessing methods such as smoothing, normalization, and derivation. As shown by the workflow of chemometric analysis in Figure 2, the final output was the best PLSR model, which was validated using the 30% prediction data. The performance of the validation model was determined from R^2 of prediction and RMSEP.

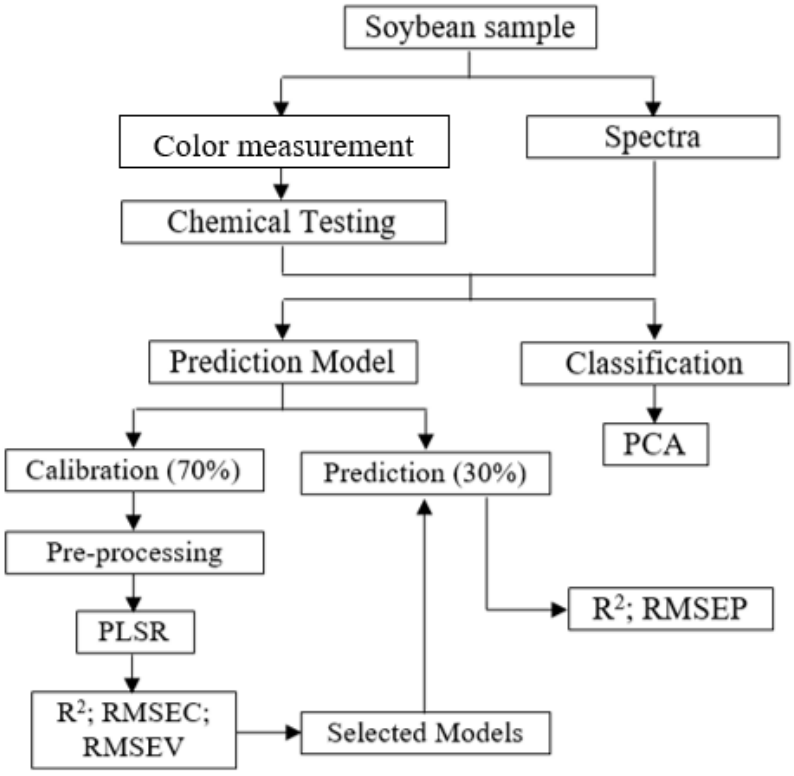


Figure 2 Workflow of chemometric analysis in soybean characterization

RESULTS AND DISCUSSION

Color and Chemical Parameters

Soybean samples were harvested ripe or at a late-stage color, with K1-K6 being yellow, while other genotypes were green (K7) and black (K8). Although intact soybean seeds in Figure 3 had different peel colors when observed using human eyes, a similar color was identified in crumble, flour, and paste forms.

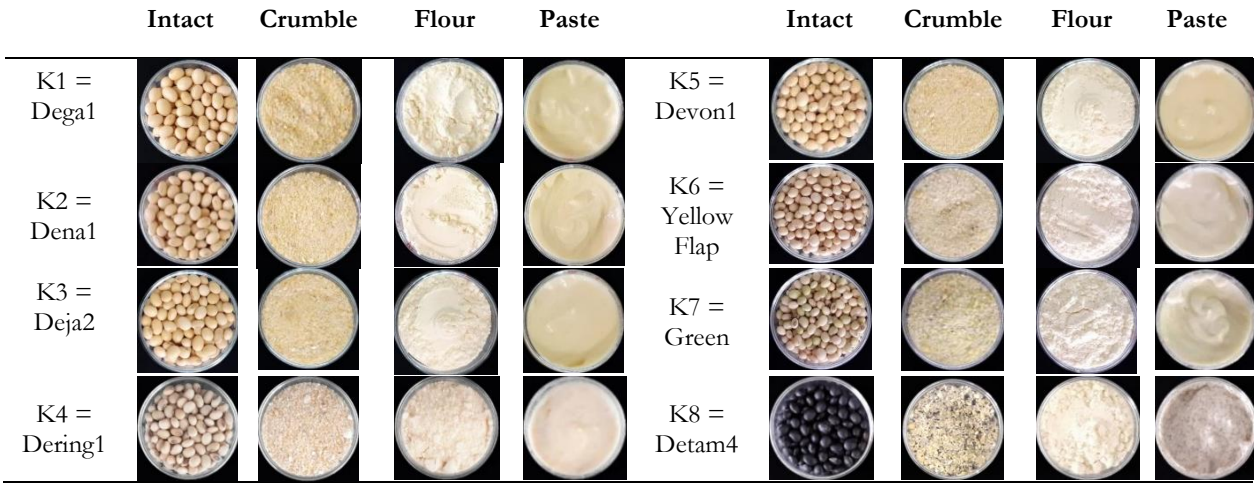


Figure 3 Figures of different soybean genotypes in bulk, crumble, flour, and paste forms

Table 1 shows the results of one-way ANOVA for different samples. The statistical analysis based on parameters L^*, a^*, and b^* showed that p-value was below the significance level $\alpha = 0.05$. A significant difference was observed in the average L^*, a^*, and b^* values between groups K1 to K8 in bulk, crumble, flour, and paste samples. However, all groups in paste form had low lightness due to the addition of excess water, which decreased surface reflection.

Table 2 shows significant differences between groups regarding protein, ash, water, chlorophyll, and vitamin C. Black color of the K8 soybean variety had the highest protein, chlorophyll, and total carotene content. However, there were no significant differences between groups regarding fat, crude fiber, carbohydrate, chlorophyll, and total carotene.

Principal Component Analysis (PCA)

PCA Based on Soybean Characteristics

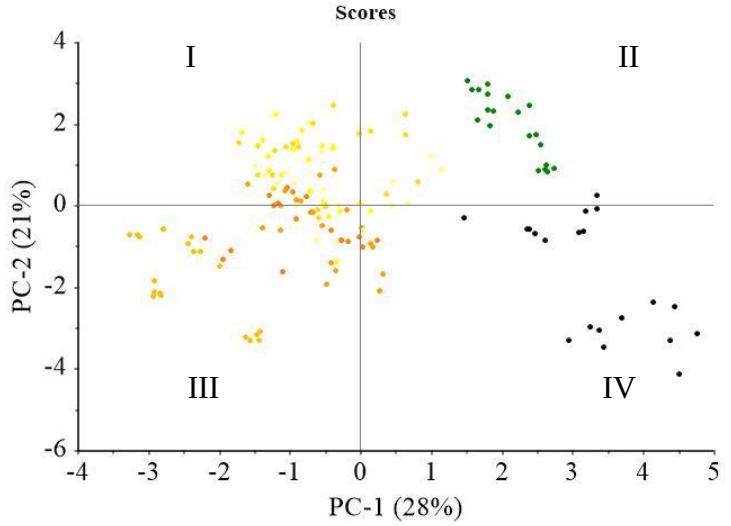
Figure 4 shows PCA plot based on soybean characteristics. The data obtained were treated in the preprocessing stage, including a moving average smoothing with 150 segments. Subsequently, the transformation stage, including normalization, derivative, baseline correction, and Standard Normal Variate (SNV), showed that range normalization yielded the most appropriate performance. PCA scores showed distinct clustering among more resilient colored soybean, specifically the yellowish, greenish, and black varieties, as explained by PC-1 (28%) and PC-2 (21%). In Figure 2a, yellowish soybean (K1-K6) predominantly occupied Quadrants I and IV, with PC-1 values primarily negative and PC-1 positive <1.5. Greenish soybean (K7) was positioned in Quadrant I, marked by positive PC-1 and PC-2 values, while black soybean (K8) was located in Quadrant IV. Figure 2b shows the significance of various parameters in differentiating samples in each quadrant. In Quadrant I, parameters such as crude fiber, chlorophyll, ash, protein, and total carotene played a crucial role. A significant influence was observed in Quadrant II due to L^*, a^*, and b^* color values along with fat content. Quadrant IV showed carbohydrate parameters, while water and vitamin C content were observed in Quadrant III. The clustering of greenish soybeans (K7) appeared to be influenced by parameters such as crude fiber, chlorophyll, ash, protein, and total carotene, while water and vitamin C content had a notable impact on the black soybeans (K8). This showed that various characteristics, based on the analyzed parameters, significantly influenced the grouping of samples.

Table 1 L^*, a^*, and b^* values of intact, crumble, flour, and paste of different soybean genotypes

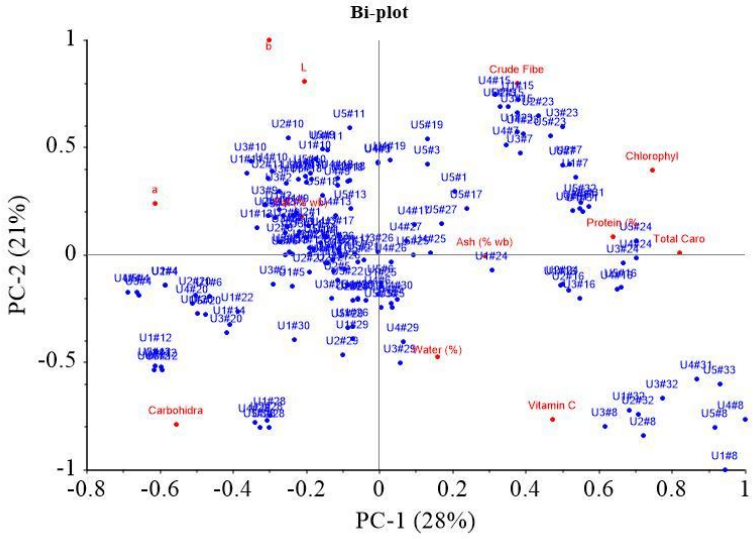
Color	Sample	K1	K2	K3	K4	K5	K6	K7	K8
L^*	Intact	70.56	49.86	60.06	66.88	54.33	58.95	60.3	23.24
	Crumble	72.98	76.75	77.76	82.24	83.14	62.16	83.51	64.97
	Flour	70.83	72.62	94.19	75.28	92.37	56.78	92.15	87.46
	Paste	52.52	49.67	54.64	31.33	48.61	64.31	59.63	41.76
a^*	Intact	5.16	6.51	8.34	10.83	8.17	9.14	0.99	-2.05
	Crumble	7.81	6.94	9.93	1.96	6.96	9.92	2.21	0.32
	Flour	3.12	0.97	1.91	2.21	2.07	8.6	-1.99	-3.87
	Paste	2.79	2.74	2.5	0.98	3.15	2.96	-1.4	-3.78
b^*	Intact	34.08	31.76	34.38	37.08	31.35	30.98	35.67	0.15
	Crumble	40.34	38.82	38.44	29.34	41.93	29.62	44.81	31.36
	Flour	27.69	32.29	35.2	30.65	31.9	27.81	36.4	33.06
	Paste	32.49	26.2	31.54	22.14	28.54	27.97	33.04	16.14

Table 2 Biochemical parameters of different soybean genotypes

Code	Protein (% wb)	Fat (% wb)	Crude Fiber (% wb)	Carbohydrate (% wb)	Ash (% wb)	Water (% wb)	Chlorophyll (mg/100g)	Total Carotene (mg/100g)	Vitamin C (mg/100gr)
K1	35.17	17.04	8.87	34.37	5.13	8.65	3.41	2.47	113.53
K2	33.42	18.38	9.35	33.51	5.31	9.99	3.56	1.98	117.72
K3	35.05	16.74	9.46	33.93	5.16	8.76	3.36	1.95	111.52
K4	32.39	17.64	6.6	36.31	5.09	8.65	0.45	1.48	185.64
K5	35.06	15.43	8.19	35.37	5.76	9.02	3.47	2.16	143.07
K6	34.21	15.15	8.36	34.41	6.01	8.52	3.84	1.56	141.82
K7	34.53	16.71	9.5	32.54	5.59	10.21	15.1	2.76	138.34
K8	36.29	16.41	8.62	34.03	5.56	8.42	9.32	3.54	178.17



(a)



(b)

Figure 4 PCA plot of (a) score and (b) bi-plot of score and loading based on soybean characteristics

PCA Based on Soybean Spectra

Figure 5 shows the average reflectance data for all preparation methods of each variety. The most effective preprocessing method included the application of a smoothing filter with 150 segments, followed by range normalization. This spectral preprocessing method yielded more consistent and dependable quantitative values for other analyses. The graph shows a typical pattern, where a valley is evident in the spectral range of 390 to 430 nm for all treatment types, followed by a peak in reflectance between 500 and 560 nm, decreasing towards the end of the spectral band. Specifically, paste samples showed relatively higher reflectance levels compared to others.

Furthermore, the addition of water had a significant effect on the reflex response, leading to an increase in the reflectance values.

As shown in Figures 3a and 3b, the difference in reflectance was significant between K1 and K2. The average reflectance values for intact, crumble, and flour treatments differed significantly from the paste treatment, showing lower reflectance. For K2, K3, and K4, the flour samples tended to show higher reflectance, while K5 was greater in the crumble samples. Intact soybean consistently showed lower reflectance levels across all preparation methods. An anomaly was observed for K7, where a small valley in the reflectance spectrum appeared in the 650 to 690 nm range.

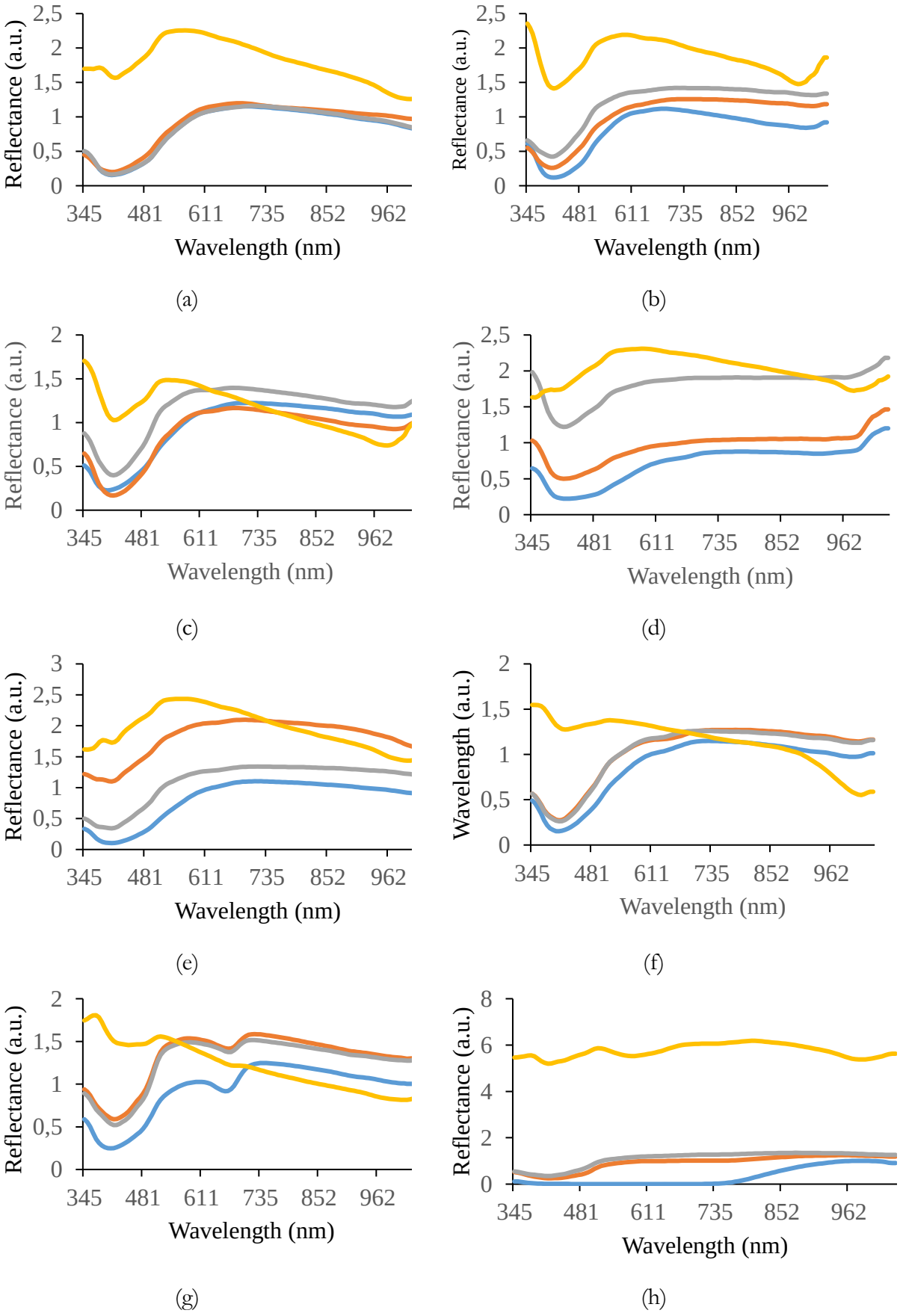


Figure 5 The reflectance spectra in various sample types of all soybean varieties (a) K1, (b) K2, (c) K3, (d) K4, (e) K5, (f) K6, (g) K7, and (h) K8.

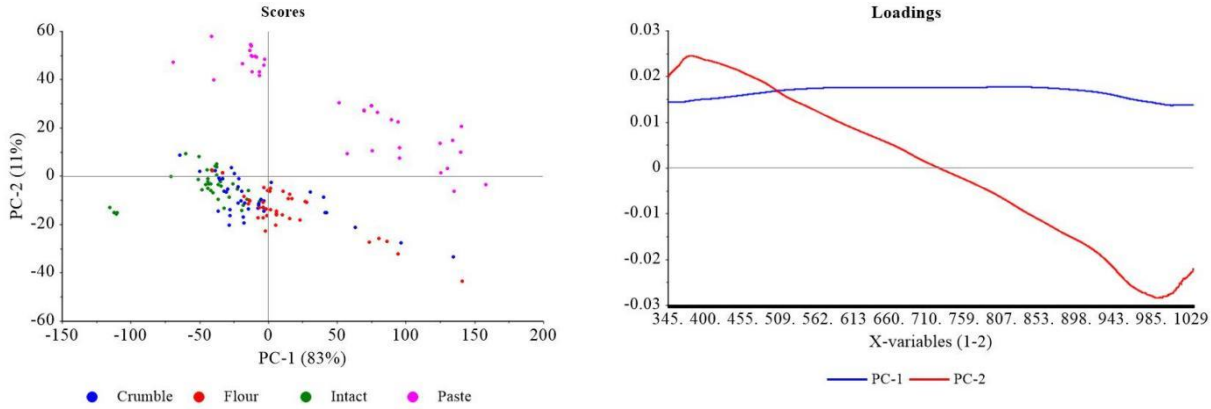
PCA analysis of the comprehensive reflectance dataset showed distinct patterns based on different preparation methods, as indicated in Figure 4a. Generally, two main data groups were found depending on the treatment applied to the samples. These included Group I, consisting of intact, crumble, and flour, and Group II, comprising paste samples. Specifically, Group II accounted for 83% of the variance in PC-1 and 11% in PC-2. Most samples in Group I tended to have negative PC-2 values, primarily in Quadrants III and IV. A significant distinction appeared when water was added to soybean paste in Group II, resulting in positive PC-2 values in Quadrants I and II. This phenomenon showed a spectrum response significantly different from other treatments.

Various treatments showed distinct positions along the axes, with intact samples primarily located in the negative range of PC-1, spanning from <-25 to >-150. Meanwhile, crumble samples extended in the positive direction, ranging from >-50 to approximately 0. An outlier, representing the intact black soybean (K8) sample, was found at the left end of Quadrant III. The flour samples ranged from >-25 to <50 in terms of PC-1, with significant overlap.

In Figure 6(b), PCA loading plot showed a prominent peak at approximately 0.02 to 0.03 on PC-1 at 350 nm wavelength, which gradually decreased to 998 nm. Although no significant anomalies were observed, the peak showed wavelength with the most substantial effect on PCA grouping.

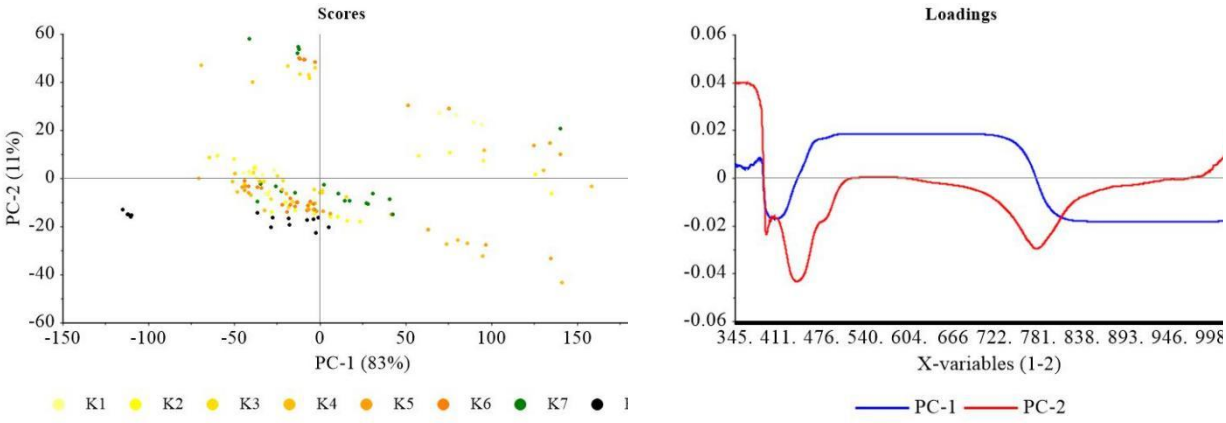
As shown in Figure 6a, the application of Multiplicative Scatter Correction (MNC) was used for optimal visualization to show the loading performance of PC-1 and PC-2 concerning genotypes. The distinction among the various groups became evident, primarily characterizing the yellow soybeans (K1-6), with significant contributions from greenish (K7) and black soybean (K8). Except for paste form, in the case of intact, crumble, and flour forms, soybean samples were dispersed across PC-2 negative region in Quadrants III and IV. Greenish soybean (K7) showed a more extensive spread, while black (K8) was positioned with the lowest PC-2 negative values.

Figure 6b shows the loading value performance of PCA results. Anomalies in PC-1 were evident in peaks and valleys spanning from 380 nm to 790 nm, while PC-2 ranged from 430 nm to 780 nm. In PC-1, approximately 82% of variances were described, and the line descended to -0.02 at its lowest point around 370 nm, followed by an increase to 0.02 at 460 nm. This value remained relatively constant until approximately 750 nm, which gradually decreased to -0.02 and continued steadily to the end of the range. The pattern observed in PC-2 followed a similar trajectory but with distinct lowest and peak points. The line initially descended from 380 nm to almost -0.05 at the lowest point and ascended to peak value of 0.02, which was maintained until 700 nm. Subsequently, a decline to -0.02 was observed, followed by an increase to approximately -0.01 at about 820 nm. These anomalies affected sample grouping, with pronounced deviations in the spectral regions of 380-470 nm and 700-800 nm due to variations in pigment levels, specifically carotene and chlorophyll, in the samples.



(a) (b)

Figure 6. PCA result of the classification of soybean, based on sample types: (a) score plot (b) loading plot using range normalization spectra preprocessing method.



(a) (b)

Figure 7 PCA result of the classification of soybean, based on sample genotypes: (a) score plot (b) loading plot using multiplicative scatter correction spectra preprocessing method

Black soybean is characterized by black outer skin, which contains anthocyanins, as documented by Yang *et al.* (2022). These anthocyanins significantly impact the interaction between black skin layer and light spectra, either reflecting a specific spectrum of light or being absorbed by particular molecules. Specifically, when examining the distribution of genotypes through PCA scores, some dissimilarities are frequently observed between paste and other forms.

The implementation of physical treatment led to significant patterns in PCA plots. The condition of intact soybean, which was not subjected to any physical treatment, primarily reflected the composition of outer skin, resulting in a marked divergence from other soybean types, particularly the yellow ones. In contrast, physical treatments such as grinding exposed the inner layers of the soybeans, leading to a more uniform composition compared to other varieties.

The results of PCA analysis showed that adding water served as a discriminative factor, segregating the samples into two clearly defined groups. Specifically, other preparation methods did not show different clusters, indicating that the specific crushing or sample size reduction had no significant influence on sample response to Vis-NIR spectrum.

Partial Least Square Regression (PLSR)

Table 3 shows the coefficient of determination (R^2) and root mean square error (RMSE) of calibration (C) and prediction (P) of various quality parameters for intact, crumble, flour, and paste of soybean obtained from Partial Least Squares (PLS) analysis. Several parameters, including protein, fat, ash, carbohydrate, and water contents as well as *a** color, yielded inadequate models. In all cases, R^2 values were consistently below 0.7, showing relatively low performance levels. This showed that the parameters could not be effectively detected in the specific Vis-NIR range of observation.

Despite being a protein source, soybean showed a low R^2 prediction across all groups, with the highest of 0.64 obtained in the paste form. Protein identification in spectroscopy is often detected by amino acids referring to the nitrogen content above 300 nm, specifically at 346 nm in the visible range. However, traces of histidine were found in the Dika nut, contributing to the failure to detect amino acid responses at the energy levels observed at 380-400 nm (Okoronkwo *et al.* 2017). R^2 value for fat was also low at 0.50, but peaks related to fat were found at 537-770 nm without contributing to PLSR model. In contrast, Lapčíková *et al.* (2018) detected wavelength range of 300 to 550 nm for unsaturated fatty acid, with ash showing the highest R^2 value of 0.37. Preece *et al.* (2009) reported significant wavelength for predicting ash content at 350, 360, 390, 410, 430, 480, and 630 nm. Based on the results, similar peaks were found at 460-590 nm, with low concentrations of ash contributing to PLSR model. Although Chandaka *et al.* (2012) reported that the significant 750 nm could determine the amount of carbohydrate, wavelength at 630-750 nm found in this study did not affect the PLSR mode.

Wavelength contributing to PLSR model for determining water was found at 882 nm and 760–970 nm, in line with the second overtones of O-H bands, which indicated water absorption (Devianti *et al.* 2023). The calibration and prediction PLSR model for water was also low due to reduced intensity at wavelength above 900 nm which was typical Vis-NIR spectroscopy. Based on Figure 1, the model demonstrated low ability in detecting *a** value, as indicated by green or yellow with a specific black color shown by soybean samples.

Crude fiber parameters, including chlorophyll, total carotene, and vitamin C showed effective model performance, achieving relatively high R^2 values across all groups, ranging from 0.80 to 0.98. In this study, wavelength contributing to crude fiber appeared at 510-630 nm, while OIV (2021) reported the carboxymethyl cellulose absorbance at 540 nm.

Chlorophyll in soybean seeds is mainly chlorophyll-a (C55H70MgN4O5) and chlorophyll-b (C55H70MgN4O6). In the visible range, the absorption peaks of chlorophyll-a occur at 430 nm and 662 nm, while chlorophyll-b occur at 453 nm and 642 nm (Lopes *et al.* 2017; Shi *et al.* 2022; Yang *et al.* 2022; Zhu *et al.* 2018). Longoni *et al.* (2020) and Gebregziabher *et al.* (2022) reported chlorophyll and carotenes substances in the visible range, where R^2 values in calibration and prediction played a significant role in specific sample classification. At 680 nm, soybean samples showed chlorophyll content that distinguished green color (Pahlawan *et al.* (2022). In the visible range, the absorption peaks of chlorophyll-a occur at 430 nm and 662 nm, while chlorophyll-b was found at 453 nm and 642 nm (Lopes *et al.* 2017; Shi *et al.* 2022; Yang *et al.* 2022; Zhu *et al.* 2018). Peak in absorbance of around 670 nm found in the olive oil spectrum confirmed the presence of chlorophyll pigments (Lapčíková *et al.* 2018). At 500 nm, soybean samples indicated carotenoid content that distinguished yellow or darker (Pahlawan *et al.* 2022).

R^2 values of vitamin C ranged from 0.87-0.98 for calibration and prediction, with the highest observed for intact samples, particularly K8, as shown in Table 2. However, there is uncertainty regarding the impact of soybean seeds coat as the primary factor contributing to the elevated vitamin C levels. According to (Riscahyani *et al.* 2019), the analysis of vitamin C can be obtained at the wavelength of 494 nm. The validation of the well-performing model ranged from 0.44 for L^* parameter in the crumble sample to 0.96 for vitamin C in paste form. The only parameter with less affirmative validation was L^* parameter, registering the lowest R^2 value. However, the prediction accuracy at 0.71 indicated a more reliable model outcome, suggesting that physical treatments, such as milling, induced variations in the color characteristics of soybeans. The distinct darkening effect observed with water addition compared to the flour sample could be attributed to the interaction or bonding of water molecules with soybean particles.

PLSR model for determining total carotene yielded high R^2 values for calibration and prediction at wavelength 390, 536, 870, and 990 nm. Moreover, the most relevant spectral variables for carotenoids were detected at the wavelength of 449, 448, and 450 nm (Afonso *et al.* 2017). The results also showed that PLSR model for predicting L^* and b^* values had high accuracy. The contrast between green (K7) and black (K8) soybeans was particularly pronounced, as indicated in the L^* and b^* parameters in the intact sample group, with R^2 values of 0.80 and 0.92. This significant difference in the characteristics of whole soybeans enhanced the model performance, allowing for the effective differentiation of color (green and black) from other yellowish samples. Subsequently, the results were adjusted by considering R^2 values for validation and prediction. L^* indicated lightness value in the range 0 (pure black) to 100 (pure white), representing light intensity that increased the absorption at the 400-500 nm spectrum band (Afonso *et al.* (2017). The a^* value was identified as reddish colored sample that presents three peaks of great absorption in the 400-500 nm region of the spectrum. Meanwhile, yellow colored sample represented in b^* profile showed significant peaks at 400-500 nm (Afonso *et al.* 2017).

Table 3 also shows the validation results of the calibration and prediction models, specifically in terms of RMSE, including RMSEC and RMSEP. Most proximate parameters, such as protein, fat, ash, and water, did not correlate significantly. Regarding RMSE values, crude fiber parameters showed higher values in intact preparation mode,

while paste form had the lowest RMSE. This phenomenon suggested that adding water to paste form significantly affected samples grouping.

Based on chlorophyll, the lowest RMSE was observed in the flour form, potentially indicating more uniform sample conditions between the outer and inner layers of soybean containing chlorophyll. A similar trend was observed for total carotene, with the lowest RMSE values in the flour and paste forms. This pattern was also found in vitamin C parameter, where the paste form showed the lowest RMSE value.

Regarding color parameters, physical treatment did not yield significant differences in RMSE. The comprehensive Mean Absolute Percentage Error (MAPE) assessment shown in Figure 8 further explored this aspect. The calibration model parameters suggested that the model accuracy was good for proximate parameters, while water showed less accurate measurements. In contrast, chlorophyll, total carotene, and L^* were assessed as showing good accuracy, vitamin C and b^* as reasonably accurate, and a^* as not accurate.

Table 3. R^2 and RMSE of calibration and prediction of various quality parameters for intact, crumble, flour, and paste soybean

Parameter	Unit	Calibration								Prediction							
		Intact		Crumble		Flour		Paste		Intact		Crumble		Flour		Paste	
		R^2C	RMSEC	R^2	RMSEC	R^2	RMSEC	R^2	RMSEC	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
Protein	(% wb)	0.5	1.13	0.59	0.76	0.61	0.87	0.64	0.85	0.56	0.91	0.55	0.97	0.63	0.93	0.68	0.85
Fat	(% wb)	0.53	0.83	0.29	0.84	0.13	1.09	0.69	0.59	0.5	0.89	0.28	0.81	0.11	0.98	0.47	0.7
Crude Fiber	(% wb)	0.8	0.46	0.82	0.44	0.83	0.42	0.95	0.23	0.72	0.57	0.78	0.49	0.85	0.41	0.92	0.29
Carbohydrate	(% wb)	0.66	0.69	0.68	0.7	0.58	0.8	0.8	0.54	0.33	0.98	0.72	0.72	0.52	0.94	0.78	0.55
Ash	(% wb)	0.09	0.36	0.22	0.38	0.23	0.38	0.15	0.38	0.02	0.37	0.37	0.37	0.16	0.41	0.1	0.27
Water	(%)	0.66	0.39	0.5	0.43	0.78	0.31	0.61	0.75	0.56	0.43	0.4	0.41	0.91	0.21	0.76	0.66
Chlorophyll	(mg/100g)	0.93	1.11	0.85	1.63	0.95	0.96	0.88	1.46	0.91	1.28	0.77	2.06	0.97	0.76	0.87	1.6
Total Carotene	(mg/100g)	0.85	0.25	0.83	0.28	0.86	0.33	0.92	0.18	0.77	0.32	0.83	0.28	0.91	0.2	0.92	0.18
Vitamin C	(mg/100g)	0.95	13.67	0.92	18.9	0.82	29.12	0.98	9.36	0.94	15.65	0.88	23.54	0.87	24.16	0.95	14.5
L^*		0.8	6.43	0.66	4.56	0.64	7.76	0.77	5	0.86	4.83	0.71	4.28	0.59	7.41	0.61	6
a^*		0.57	3.52	0.71	2.06	0.66	2.29	0.74	1.37	0.4	4.47	0.73	2.12	0.9	1.08	0.6	1.36
b^*		0.92	3.33	0.47	4.6	0.43	2.96	0.73	3.04	0.92	3.47	0.38	3.71	0.42	3.53	0.7	2.96

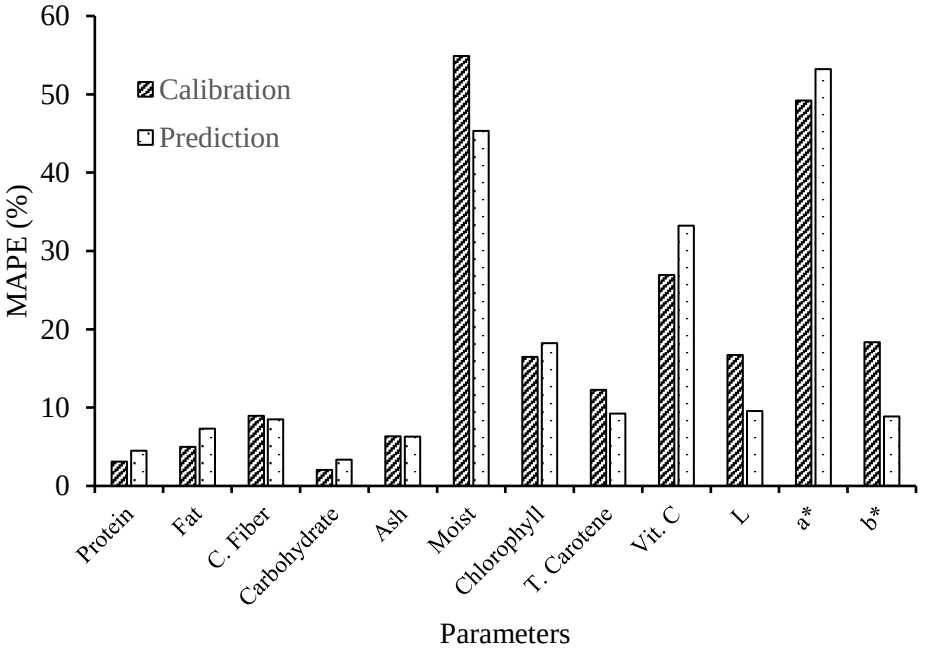


Figure 8 MAPE of calibration and prediction for all parameters

CONCLUSION

In conclusion, this study investigated the relationship between soybean color and chemical parameters, examining the impact of various sample preparation methods. The results showed that processing reduced color differences compared to intact soybeans. Statistical analysis showed significant color variations among genotypes and preparation methods due to varying composition and pigmentation. Furthermore, PCA differentiated soybean samples based on specific components and showed the role of water addition. PLSR models predicted some parameters effectively, but protein and fat showed weaker correlations. Reflectance spectra showed specific patterns, with intact soybean having lower reflectance. The results provided valuable insights into the relationship between soybean color, sample preparation method, and spectral characteristics, including the potential for predicting biochemical parameters using spectroscopic data. However, model accuracy varied by parameter and preparation method, emphasizing the need for further investigation to improve predictions, specifically for protein and fat.

REFERENCES

Abadi FR, Masithoh RE, Sutiarso L and Rahayoe S. 2022. A study of characterization procedure for non-destructive testing of soybean seed based on spectroscopy. IOP Conference Series: Earth and Environmental Science 1059: 012015. doi: 10.1088/1755-1315/1059/1/012015.

Afonso T, Moreco R, Uarrota VG, Navarro BB, Nunes EC, Maraschin M and Rocha M. 2017. UV-Vis and CIELAB Based Chemometric Characterization of *Manihot esculenta* Carotenoid Contents. J Integr Bioinform 14(4): 20170056.

Alander JT, Bochko V, Martinkauppi B, Saranwong S, and Mantere, T. 2013. A Review of Optical Non-destructive Visual and Near-Infrared Methods for Food Quality and Safety. International Journal of Spectroscopy 2013: 1-36. doi: 10.1155/2013/341402.

Association of Official Analytical Chemists (AOAC). 2007. Official Methods of Analysis of AOAC International (18th ed.). Washington: Association of Official Analytical Chemists.

Berns R.S. 2019. Numerical Color Specification: Colorimetry, in: Billmeyer and Saltzman's Principles of Color Technology. John Wiley & Sons Ltd Hoboken NJ: 51-84.

Chandaka M, Kumar SA, Reddy PS, Rajamohitha K, Reddy NS and Sudeshna R. 2012. Quantitative evaluation of carbohydrate levels in different natural food stuffs by UV-visible spectrophotometer. *Pharmanest* 3(3): 239-242.

Cortés V, Blasco J, Aleixos N, Cubero S, and Talens P. 2019. Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. Trends in Food Science and Technology 85 (January): 138-148. doi: 10.1016/j.tifs.2019.01.015.

Farag M A, Sheashea M, Zhao C, and Maamoun AA. 2022. UV Fingerprinting Approaches for Quality Control Analyses of Food and Functional Food Coupled to Chemometrics: A Comprehensive Analysis of Novel Trends and Applications. Foods, 11(18). doi: 10.3390/foods11182867.

Gebregziabher BS, Sheng-rui Z, Azam M, Jie Q, Boateng KGA, Yue F, Yi-tian L, Jing L, Bin L and Jun-ming S. 2022. Natural variation and geographical distribution of seed carotenoids and chlorophylls in 1167 Chinese soybean accessions. Journal of Integrative Agriculture. doi: 10.1016/ j.jia.2022.10.011.

Harsono A, Harnowo D, Ginting E, and Elisabeth DAA. 2021. Soybean in Indonesia: current status, challenges and opportunities to achieve self-sufficiency. In Legumes: Vol. December. doi: 10.5772/intechopen.101264.

ILETRI. 2016. Deskripsi Varietas Unggul aneka kacang dan umbi [Description of Legume and Tuber Crops Superior Varieties]. Badan Penelitian dan Pengembangan Pertanian, Kementerian Pertanian.

Jia F, Peng S, Green J, Koh L, and Chen X. 2020. Soybean supply chain management and sustainability: a systematic literature review. Journal of Cleaner Production 255: 120254. doi: 10.1016/j.jclepro.2020.120254.

Kementan (Kementerian Pertanian). 2020. Outlook kedelai: komoditas pertanian tanaman pangan kedelai [SOybean Outlook: Soybean The Agricultural Commodity of Food Crops] (A. A. Susanti & A. Supriyatna, eds.). Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian.

Longoni M, Freschi A and Cicala N. 2019. Non-invasive identification of synthetic organic pigments in contemporary art paints by visible–excited spectrofluorimetry and visible reflectance spectroscopy, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy (2019) DOI:10.1016/ j.saa.2019.117907.

Lopes EJ, Zepka LQ and Queiro MI. 2017. Chlorophyll. InTechOpen.

Manley M. 2014. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. Chemical Society Reviews 43(24): 8200-8214. doi: 10.1039/ c4cs00062e

Masithoh R E, Pahlawan MFR, Saputri DAS, and Abadi FR. 2023. Visible-Near-Infrared Spectroscopy and Chemometrics for Authentication Detection of Organic Soybean Flour. Pertanika Journal of Science and Technology 31(2): 671-688. doi: 10.47836/pjst.31.2.03

Mayerhöfer TG, Pipa AV, and Popp J. 2019. Beer's Law-Why Integrated Absorbance Depends Linearly on Concentration. ChemPhysChem 20(21): 2748-2753. doi: 10.1002/cphc.201900787

Monago-Maraña O, Eskildsen CE, Galeano-Díaz T, Muñoz de la Peña A, and Wold JP. 2021. Untargeted classification for paprika powder authentication using visible – Near infrared spectroscopy (VIS-NIRS). Food Control 121(June 2020). doi: 10.1016/ j.foodcont.2020.107564.

Okoronkwo NE, Mba KC and Nnorom IC. 2017. Estimation of Protein Content and Amino Acid Compositions in Selected Plant Samples Using UV-Vis Spectrophotometeric Method. American Journal of Food Science and Health Vol. 3, No. 3 (2017) p: 41-46.

Pahlawan MFR, Murti BMA, and Masithoh RE. 2022. The potency of Vis/NIR spectroscopy for classification of soybean based of colour. IOP Conf. Series: Earth and Environmental Science, 1018. Yogyakarta: IOP Publishing Ltd.

Pahlawan MFR, Saputri DAS, and Masithoh RE. 2023. Non-Destructive Evaluation of Moisture Content in Single Soybean Seed Using Vis-NIR Spectroscopy. Proceedings of the International Conference on Sustainable Environment, Agriculture and Tourism *(ICOSEAT 2022) 26*: 396-400.

Riscahyani NM, Ekawati ER, and Ngibad K. 2019. Identification of Ascorbic Acid Content In *Carica papaya* L. Using Iodimetry and UV-VIS Spectrophotometry. Indonesian Journal of Medical Laboratory Science and Technology 1(2): 58-64.

Shi D, Hang J, Neufeld J, Zhao S, and House JD. 2022. Estimation of crude protein and amino acid contents in whole, ground and defatted ground soybeans by different types of near-infrared (NIR) reflectance spectroscopy. Journal of Food Composition and Analysis 111 (2022): 104601.

Walsh KB, Blasco J, Zude-Sasse M, and Sun X. 2020. Visible-NIR' point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. Postharvest Biology and Technology 168(April 2019): 111246. doi: 10.1016/j.postharvbio.2020. 111246

Wang M, Xu Y, Yang Y, Mu B, Nikitina MA, and Xiao X. 2022. Vis/NIR optical biosensors applications for fruit monitoring. Biosensors and Bioelectronics*:* X 11(March): 100197. doi: 10.1016/ j.biosx.2022.100197.

Yang L, Wang S, Zhang H, Du C, Li S, and Yang J. 2022. Effects of black soybean powder particle size on the characteristics of mixed powder and wheat flour dough. LWT-Food Science and Technology 167 (2022) 113834.

Zhu Z, Chen S, Wu X, Xing C and Yuan J. 2018. Determination of soybean routine quality parameters using near-infrared spectroscopy. Food Sci Nutr. (2018): 1-10.