



ORIGINAL ARTICLE OPEN ACCESS

A Comparison of Spectral Preprocessing Methods and Their Effects on Nutritional Traits in Cowpea Germplasm

Siddhant Ranjan Padhi^{1,2}  | Racheal John² | Kuldeep Tripathi² | Dhammaprakash Pandhari Wankhede² | Tanay Joshi³ | Jai Chand Rana⁴ | Amritbir Riar³  | Rakesh Bhardwaj²

¹The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India | ²ICAR-National Bureau of Plant Genetic Resources, New Delhi, India | ³Department of International Cooperation, Research Institute of Organic Agriculture, FiBL, Frick, Switzerland | ⁴Alliance of Bioversity International and CIAT, India Office, New Delhi, India

Correspondence: Amritbir Riar (amritbir.riar@fiBL.org) | Rakesh Bhardwaj (rakesh.bhardwaj1@icar.gov.in)

Received: 19 October 2023 | **Revised:** 10 March 2024 | **Accepted:** 11 March 2024

Funding: This work was supported through CSIR–Senior Research Fellowship granted to SRP; Inhouse project PGR/GEV/BUR-DEL-01.06 funded by Indian Council of Agricultural Research; the Department of Biotechnology, Government of India, funded project on minor pulses (No. BT/Ag/Network/Pulse-I/2017–18, dated: 24 October 2018), and the international collaborative project on ‘Consumption of Resilient Orphan Crops and Products for Healthier Diets (CROPS4HD)’ co-funded by the Swiss Agency for Development and Cooperation, Global Programme Food Security (SDC GPFS), executed in India through FiBL (Research Institute of Organic Agriculture), Alliance of Bioversity and CIAT, with ICAR-NBPGR as a lead partner (LOA No. L22DEL211 dated 01 October 2022).

Keywords: legume | MPLS regression | multiplicative scatter correction | nutrition | standard normal variate with detrending

ABSTRACT

Cowpea (*Vigna unguiculata* L. (Walp)) is a multipurpose legume, which has good nutritional properties. Nutritional parameters assessed conventionally can be labour intensive, costly and time taking for germplasm screening. Near-infrared reflectance spectroscopy (NIRS) is a rapid and nondestructive method, which can facilitate high-throughput germplasm screening. In our study, estimation of amylose and sugars has been done using NIRS. Two preprocessing methods, that is, SNV-DT (standard normal variate with detrending) and MSC (multiplicative scatter correction), were performed for optimization of the original spectra. Subsequently, MPLS (modified partial least square) regression method was employed to construct the prediction models. In amylose, the best RSQ_{external} (coefficient of determination) (0.962) was found in SNV-DT with mathematical treatment 3,8,8,2. The same result was shown in sugar where the best RSQ_{external} (0.914) was found in SNV-DT with mathematical treatment 3,4,4,1. Overall, in the case of amylose and sugars, SNV-DT was found to be a good preprocessing treatment than MSC. Paired *t*-test values in all the treatments for both the preprocessing methods were >0.05 indicating their reliability. High RSQ_{external} values for both the traits imply the applicability of the prediction models. Thus, these models can facilitate high-throughput germplasm screening in different national and international crop improvement programmes focusing on quality traits.

1 | Introduction

Legumes play a crucial role in addressing malnutrition and enhancing global food security due to their nutritional value, sustainability and agronomic benefits. Legumes are a rich source of essential nutrients, including protein, dietary fibre, vitamins

(such as folate and vitamin B) and minerals (such as iron, zinc and calcium). Protein from legumes can be an excellent alternative for animal-based protein, making them valuable for people, especially in regions with limited access to animal products. Cowpea (*Vigna unguiculata* L. (Walp.)) is one such multipurpose legumes (use as both grain, vegetable and fodder) that belong to

Siddhant Ranjan Padhi and Racheal John have contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Legume Science* published by Wiley Periodicals LLC.

the Fabaceae family, which can facilitate to provide sustainable benefits. This crop is well known as a nutrient-rich indigenous vegetable in Africa and has the potential to enhance food and nutrition security in developing nations like India. Originally from West and Central Africa, cowpeas have spread to Latin America and South East Asia through cultivation and production (Ano and Ubochi 2008; Edeh and Igberi 2012). Cowpea is a legume that stands out for its elevated protein content (ranging from 19 to 27.3 g/100 g) (Padhi, Bartwal, et al. 2022) and low-fat content (1 to 1.2 g/100 g) (Jayathilake et al. 2018) along with good amount of carbohydrates, dietary fibre and micronutrients. Among carbohydrates, it has good quantities of amylose and sugars where amylose content affects the physicochemical and processing properties of pulse. Apart from these factors, it also affects the tendency to retrogradation in legume starch (Singh 2017). This property of legume starch leads to their resistance against digestive enzyme activity, which, in turn, causes a decrease in the glycaemic index (Singh, Dartois, and Kaur 2010). Hence, it is helpful for the diabetic patients who consume legume starch. Amylose content in cowpea generally ranges from 9 to 25 g/100 g (Adebooye and Singh 2008; Hamid et al. 2015; Hoover et al. 2010; Padhi, Bartwal, et al. 2022). Sugar plays a vital role in enhancing tolerance to abiotic stress, improving storability and providing energy for the human body. Additionally, it contributes to the desirable taste and texture experienced during cooking. The presence of soluble sugars is closely associated with carbohydrate metabolism, photosynthesis and overall seed development (Wilcox 2001). In previous studies, it ranged from 1.1 to 8.73 g/100 g (Nassourou et al. 2017; Omueti and Singh 1987; Padhi, Bartwal, et al. 2022; Weng et al. 2018). Apart from having these many multiple benefits of sugar, the presence of oligosaccharides limits the consumption. Due to the presence of α -galactosidic bonds, these oligosaccharides remain undigested in the human body, which causes flatulence and discomfort (Singh 2017). The current trend in the consumption of plant-based protein, along with the considerable yield, warrants a closer examination of the other important nutritional factors like sugar and amylose, which generally affects the physicochemical properties and palatability of cowpea. Understanding these aspects can have several advantages, including the ability to develop value-added products with accurate nutritional claims in the food industry and facilitating the selection of cowpea accessions with higher improved quality in cowpea breeding programmes. However, analysing the nutritional traits manually or conventionally is labour intensive, time consuming, costly and sometimes complicated. These factors can become a large drawback for screening huge germplasm in a quality breeding programme.

Near-infrared reflectance spectroscopy (NIRS) can overcome the above disadvantages for analysing the nutritional composition as it is a rapid, nondestructive, cost-effective, noninvasive, safe, simple method, finding increasing applications across various disciplines and matrices on a global scale and can screen huge germplasm (Hang et al. 2022; Johnson, Walsh, and Naiker 2020). This technique is based on various functional groups and bonds present in the sample absorbing electromagnetic radiation in the near-infrared (NIR) region, specifically within the wavelength range of 780–2500 nm (Osborne 2006). The presence of multiple overlapping NIR absorption bands in biological samples results from a combination of vibrations

and overtones of functional groups such as S–H, O–H, N–H and C–H (Tomar et al. 2021). These methods prove valuable in analysing molecular interactions among functional groups and obtaining chemical insights into the material (Shi et al. 2019). An issue commonly encountered in the practical implementation of the NIR technique is the interference caused by various factors during the signal acquisition process, which can disturb the spectra (both baseline shifts and nonlinearities). The various factors are spectral molecular vibration, mathematical treatments and statistical methods. Preprocessing the spectral data constitutes a critical step in developing a reliable prediction model, which results in an improved signal-to-noise ratio, increased signal variation and the elimination of irrelevant sources not related to the property of interest (Rinnan, Van Den Berg, and Engelsen 2009). Given the complexity of samples in food crops and the real measurement conditions, distinguishing interferences from spectra can be challenging. Spectral interferences can be perceived as a blend of numerous additive components, multiplicative elements, polynomial baseline shifts and spectral noise (Bi et al. 2016). Hence, empirical methods like derivatives, multiplicative scatter correction (MSC) and standard normal variate (SNV) are widely used for spectral preprocessing (Barnes, Dhanoa, and Lister 1989; Bellon-Maurel et al. 2010). A comparison of preprocessing treatments has been done for prediction of soil organic matter (Carvalho et al. 2022), soluble solid content in hard kiwi (Sarkar et al. 2020) and soil organic carbon (Dotto et al. 2018). Various types of derivatives effectively enhance the spectral resolution of a biological matrix like cowpea. This enhanced resolution not only benefits signals from regions with responses but also amplifies spectral areas without responses, which could lead to the models becoming over fitted. MSC and SNV possess the ability to eliminate both additive and multiplicative impacts within spectra. In SNV, each spectrum is centred and subsequently scaled by its corresponding standard deviation (Barnes, Dhanoa, and Lister 1989). The transformation via SNV helps reduce the multiplicative effects of scattering. On the other hand, MSC establishes a reference spectrum, often the mean spectrum of calibration data, and aims to rectify baseline and amplification effects concerning the reference spectrum for each individual spectrum (Geladi, MacDougall, and Martens 1985; Marten, Shenk, and Barton 1985). Postscatter correction and different regression methods such as MPLS (modified partial least square), PLS (partial least square) and PCR (principal component regression) are commonly used to correlate the biochemical components with spectral data, and the best model is selected on the basis of coefficients of determination, majorly, coefficient of determination (RSQ), residual prediction deviation (RPD) and bias values.

Various NIRS-based prediction models have been developed and reported in multiple crops such as rice, pearl millet, amaranth, buckwheat, mung bean, cowpea and faba bean (Bartwal et al. 2023; John et al. 2022; Johnson, Walsh, and Naiker 2021; Padhi, John, et al. 2022; Shruti et al. 2023; Tomar et al. 2021). However, no literature could be traced concerning the development of NIRS calibration models for total amylose and sugar contents in cowpea. Therefore, the present study includes the quantification of amylose and sugar through NIRS prediction model development, simultaneously assessing two preprocessing methods (MSC and SNV) and their effects on each trait.

Based on the statistical results, the constructed models would be efficient in characterising cowpea for its high, low or medium amylose and sugar contents.

2 | Materials and Methods

2.1 | Sample Collection and Selection

A total of 475 cowpea accessions was taken from the medium-term storage of the National Gene Bank at ICAR-NBPGR, New Delhi, India, comprising both indigenous and exotic collections. The samples in this cowpea diversity panel were derived from various genotypes, exhibiting significant diversity in adaptation and other characteristics. These accessions were cultivated at the Issapur experimental farm in New Delhi, India, following standard agronomic practices using an augmented block design (Padhi, Bartwal, et al. 2022). Once matured, the seeds were collected and thoroughly sun dried to a grain moisture content of 8%–10% and stored at 4°C. All the 475 cowpea accessions were scanned using FOSS NIRS 6500, and the reflectance spectrum was recorded. A novel approach was used by John et al. (2022) where representative samples were selected based on the stratified purposive sampling. Hence, hierarchical clustering analysis was done by Ward's method using Euclidean squared distance of 5, and clusters were made using the reference spectrum. From those clusters and subclusters, 121 diverse cowpea accessions were selected for reference analysis and NIRS scanning. These samples underwent homogenisation, grinding and sieving using a 1-mm sieve in a FOSS Cyclotec machine. The resulting flour was utilised for scanning and wet chemistry analysis.

2.2 | Reference Analysis Generation

2.2.1 | Total Amylose Content

The total amylose content was estimated by iodometric method as described by Juliano et al. (1981) with several modifications (John et al. 2023). Fifty milligrams of the sample was weighed and placed in a centrifuge tube. Absolute ethanol (0.5 mL) was added, followed by the addition of 1-N NaOH (4.5 mL), and the mixture was vortexed and incubated at boiling temperature for 15 min. After rinsing with distilled water, the volume was adjusted to 25 mL, and 0.5 mL of the sample was transferred to amber tubes. To this, 100 µL of 1-N glacial acetic acid and 200 µL of iodine solution were added, and the volume was adjusted to 10 mL. The amber tubes were left in the dark for 20 min. The reactions generated a blue-coloured complex, and the absorbance was recorded at 620 nm.

2.2.2 | Total Sugar Content

Total soluble sugars were estimated using anthrone reagent method as described by Dubois et al. (1956). Sample (0.1 g) was taken in falcon tubes followed by the addition of 5 mL ethanol, vortexed and kept in water bath at 80°C for 30 min. The tubes were cooled and centrifuged at 12,000 rpm for 10 min.

The supernatant was collected and extracted two more times following the same method. The supernatant was collected in fresh tubes, and the final volume of the supernatant was made to 10 mL. Extract (0.1 mL) in triplicates was taken from the aliquot, evaporated at 100°C and reconstituted with 1 mL of double distilled water. A blank was prepared with the addition of 1-mL double distilled water. Standard solutions containing varying concentrations of D-glucose were also prepared. Ice-cold anthrone reagent (4 mL) was added to each set of test tubes (sample, blank and standards), followed by incubation in a water bath at 80°C for 8 min. The reaction generated a green-blue colour, spectrophotometrically quantified at 630 nm. The total soluble sugars were expressed as g/100 g.

2.3 | Spectral Acquisition

The FOSS NIRS 6500 spectrophotometer, operated with Win ISI III Project Manager Software 1.50, was calibrated using a 100% white reference tile. Approximately 5 g of homogenised sample was placed in a circular ring cup with a quartz window (3.8 cm in diameter and 1 mm in thickness) and scanned 32 times at wavelengths ranging from 400 to 2500 nm. The average spectrum was recorded as log (1/R) with increments of 2 nm, where R denotes the respective reflectance.

2.4 | Outlier Detection

The outlier detection method involved the use of neighbourhood Mahalanobis distance ($NH < 0.6$) and global H (GH), which represents the spectral distance from the mean spectrum of the population ($GH > 2.5$). NH calculates the proximity of each sample to all other samples in the population. Samples with scanning errors produce abrupt spectra, making them outliers for any trait. The removal of superfluous spectra from the calibration population was done by GH (Bartwal et al. 2023). Before calibration development, the sample set without having any outliers was carried forward to make calibration and validation set.

2.5 | Selection of the Calibration and Validation Set

The spectral data (without any outliers) and data from reference methods of 121 samples were imported in the WinISI III project manager 1.50 (Windows Infra Soft. International, USA), which was used to perform spectral data preprocessing, build calibration and validation models. The samples were arranged in an ascending order, and every second value was taken out to make the validation set. Hence, the calibration and validation set were set in a ratio of 2:1 after the arrangement ensuring uniform variability in the set (John et al. 2022). The samples were divided into two sets, that is, 80 in the training set and 41 in the validation set.

2.6 | Preprocessing of the Spectra

The preprocessing of spectral data plays a crucial role in eliminating or diminishing unwanted artefacts in the spectra

while maintaining the linear connections between multivariate signals (reflectance) and soil characteristics. These mathematical methods encompass several stages, including data cleansing, weighting, standardisation, scatter correction, elimination of nonlinear trends, smoothing and the calculation of derivatives. During the analytical stage of data processing, the spectra undergo a smoothing process to minimise noise or measurement inaccuracies. Following the removal of the noisy segment, the remaining spectra were then subjected to a smoothing procedure using the Savitzky–Golay transformation (Savitzky and Golay 1964) with a second-order polynomial applied over a sliding window of seven smoothing points. Several preprocessing techniques are there to improve the prediction of calibrated models like MSC, inverse MSC, extended MSC, detrending, SNV, normalisation and so forth. To evaluate which is better, MSC and SNV-DT were compared in our study as MSC followed by SNV is the most widely used preprocessing technique (Rinnan, Van Den Berg, and Engelsen 2009). The MSC preprocessing technique, first introduced by Marten, Shenk, and Barton (1985), alleviates nonlinearities present in spectral data due to the influence of both additive and multiplicative scattering effects caused by particles in the samples. This approach aligns each spectrum with a reference spectrum in a way that ensures consistent baseline and amplification effects, with their averages being identical in each spectrum (Isaksson and Næs 1988). SNV, a technique applied for mitigating scattering effects, achieves this by centring and scaling each spectrum, as outlined by Barnes, Dhanoa, and Lister (1989). In contrast, DT is a transformation designed to eliminate the mean value or linear trends within spectroscopic data.

2.7 | Calibration of the Model

WinISI project manager software v 1.5 was used to develop calibration equations using multivariate analysis by regressing spectral values with reference values. For both the preprocessing methods, MPLS regression was used to develop calibration equation with full spectra. Different mathematical treatments were used for SNV-DT and MSC ‘1,4,4,1’, ‘1,8,8,2’, ‘2,4,4,1’, ‘2,8,8,2’, ‘3,4,4,1’ and ‘3,8,8,2’. The first digit corresponds to the derivative where the first, second and third derivatives were used. The second digit corresponds to the gap (data calculated by the specific deviation) where the fourth and eighth derivatives have been used. The third and fourth data points correspond to smoothening 1 (S1) and 2 (S2). Various parameters, including coefficient of determination (RSQ_{internal}), standard error of cross-validation ($SEC(V)$), standard deviation (SD) and one minus variance ratio ($1 - VR$), were used to evaluate the developed calibration equations.

2.8 | Model Validation and Accuracy

The validation set contained 41 samples, which were used to evaluate the predictive performance of the calibration equations. The accuracy of amylose and sugar prediction and the effectiveness of various preprocessing techniques were assessed based on the coefficient of determination (RSQ_{external}), SEP, RPD and bias values.

2.9 | Statistical Analysis

Win ISI III Project Manager software 1.50 was utilised to perform all calibrations and predictions by applying various mathematical treatments to both spectral and analysed data. Coefficient of determination ($RSQ_{\text{internal/external}}$) for reference versus predicted values of amylose and sugar other than WIN ISI was confirmed separately using MS Excel. To evaluate the prediction accuracy of the model, a paired *t*-test was conducted using Jamovi software at a 95% confidence interval. The result of the paired *t*-test was expressed in the form of a *p*-value.

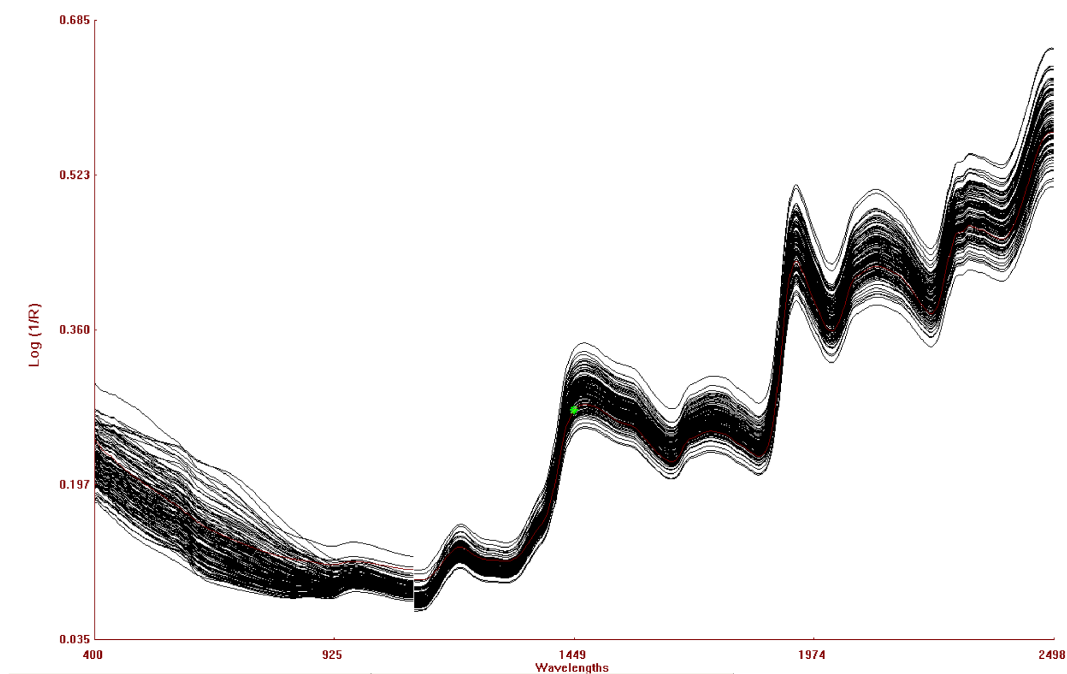
3 | Results and Discussion

3.1 | NIR Spectral Characteristics

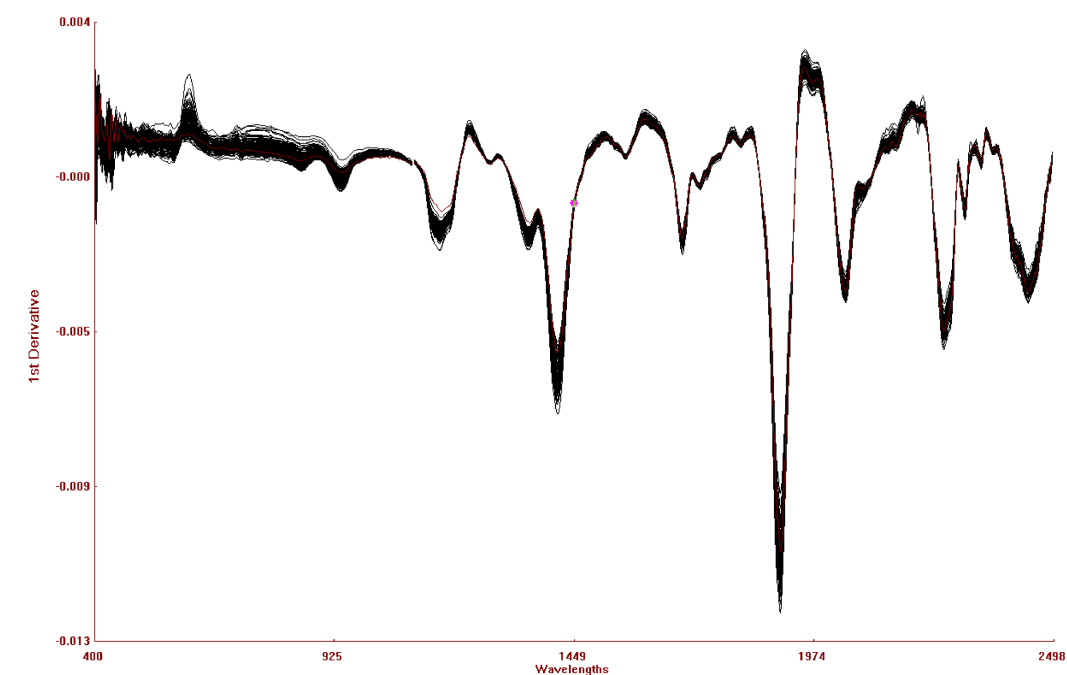
The raw spectra of 121 homogenised cowpea samples are given in Figure 1A over the spectral range of 1100–2400 nm. The absorption phenomena in the NIR region comprise overtone and combination bands arising from various molecular groups like C–H, O–H, N–H, C=O, S–H and so forth (Williams, Manley, and Antoniszyn 2019). The NIR spectrum can be categorised into three distinct regions. The initial region, labelled as Region I, spans from 800 to 1200 nm. This portion, also referred to as the ‘short-wave NIR region (SWNIR)’, ‘near-NIR region (NNIR)’ or ‘the Herschel region’, showcases bands originating from electronic transitions, overtones and combinations of modes. Region II encompasses the range from 1200 to 1800 nm and encompasses the first overtones of XH ($X = C, O, N$), stretching vibrations and diverse types of combination modes. Lastly, Region III (1800–2500 nm) is characterised by combination modes. Numerous applications make use of Regions II and III for their analysis (Ozaki 2012). NIR spectroscopy effectively demonstrated accurate outcomes concerning the distinct absorption bands of sugars, specifically at wavelengths of 1200, 1437, 2074 and 2320 nm (López, García-González, and Franco-Robles 2017). Distinctive absorption peaks were observed approximately at 1930 nm, indicating the combined effects of bending and stretching of O–H in amylose. Additionally, a peak in the vicinity of 1566 nm corresponds to the combined symmetric stretching of O–H in amylose (Tomar et al. 2021).

3.2 | Descriptive Statistics

The statistics and histogram generated by the reference values are given in Table 1 and Figure 2A,B, respectively. In order to establish robust calibrations, Williams, Manley, and Antoniszyn (2019) suggested that an even distribution of chemical measurements is preferable over a Gaussian distribution. This is because an equal representation of all values within the calibration set is deemed more favourable as done by Sánchez-Carnerero Callado et al. (2018). Hence, the effect by which the results of future analyses show to regress towards the mean will be avoided. The histograms of frequencies showed in general a continuous distribution of the reference values across the concentration intervals. Broad range of variability is shown in the statistics ensuring the complete representativeness. The values of amylose ranged from 10.02



A Raw spectra of 121 cowpea homogenised samples.



B First derivative spectra without pre-processing

FIGURE 1 | (A) Raw spectra of 121 cowpea homogenised samples. (B) First derivative spectra without preprocessing. (C) Second derivative spectra without preprocessing. (D) Raw spectra of 121 cowpea samples with MSC preprocessing treatment. (E) Raw spectra of 121 cowpea samples with SNV-DT preprocessing treatment.

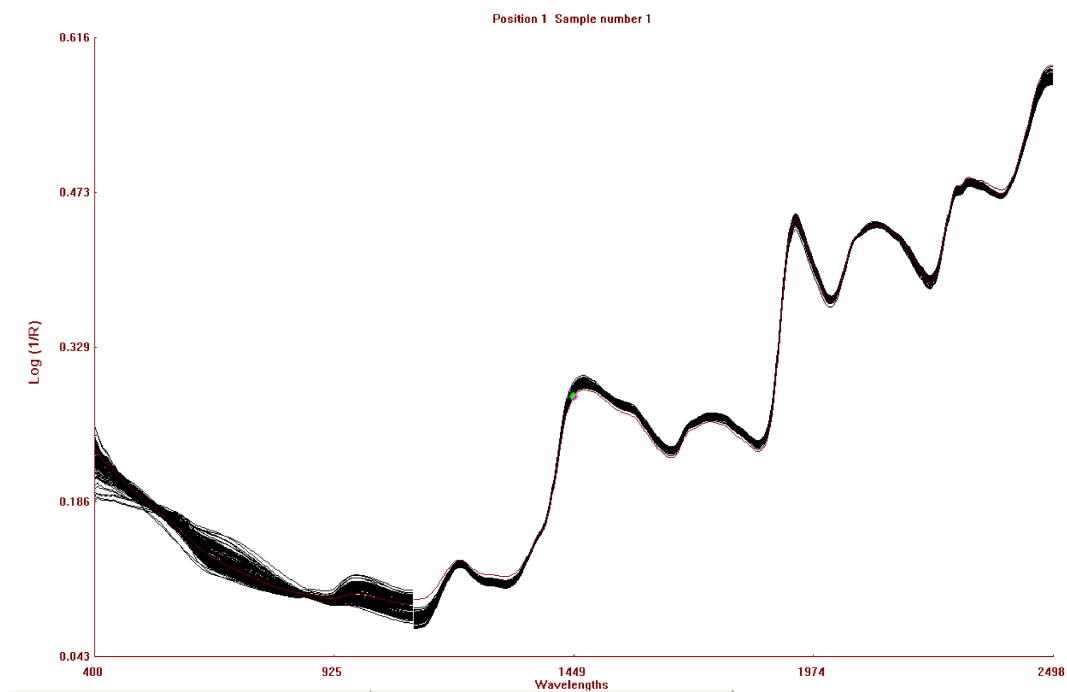
to 21.7 g/100 g having a mean value of 15.7 g/100 g. The higher amylose values in our study correspond to the results obtained by Hamid et al. (2015) for red and black cowpea cultivars. Sugar content in our study ranged from 2.54 to 8.13 g/100 g with a mean value of 5.38 g/100 g. The results were in agreement with the work done in 113 cowpea samples by Weng et al. (2018) (3.2–8.6 g/100 g).

3.3 | Application of the Two Preprocessing Treatments on the Data

Figure 1B,C depicts the first and second derivatives derived from the unprocessed absorbance spectra of the calibration samples given in Figure 1A. These two figures evidently illustrate a general enhancement in addressing baseline



C Second derivative spectra without pre-processing

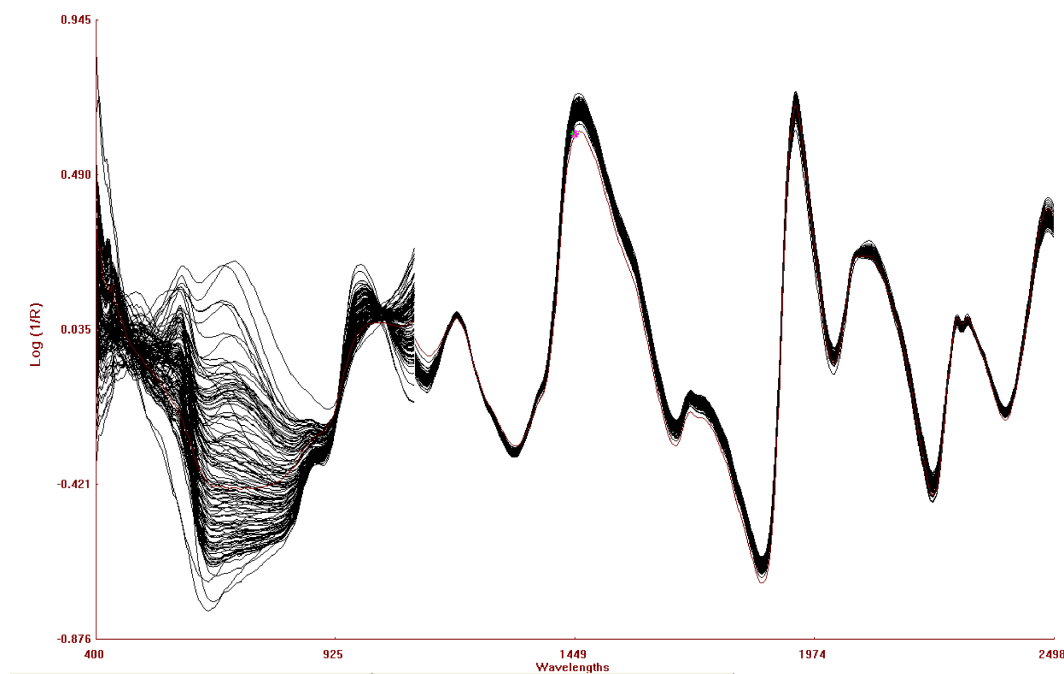


D Raw spectra of 121 cowpea samples with MSC pre-processing treatment

FIGURE 1 | (Continued)

interference and signal overlap. To mitigate the impact of scattering and systematic noise, various preprocessing methods were employed in the experimental data. Figure 1D presents spectra that have undergone multiple scatter correction (MSC), resulting in a significant reduction of light scattering interferences compared to the original experimental spectra. In Figure 1E, SNV spectra is displayed, where, similar to derivative spectra, most baseline offsets have been eliminated.

Notably, the advantage here is that there is no requirement for preselecting a wavelength interval, as is necessary when applying Savitsky–Golay smoothing in derivative calculations. Furthermore, there are no additional peaks introduced, as occurred in the case of the initial and secondary derivative spectra. Consequently, the direct interpretation of spectra is more straightforward for SNV spectra (Figure 1E) in comparison to the first and second derivative spectra (Figure 1B,C). When



E Raw spectra of 121 cowpea samples with SNV-DT pre-processing treatment

FIGURE 1 | (Continued)

TABLE 1 | Descriptive statistics of reference values for developing NIRS prediction models.

Traits	N	Minimum	Maximum	Mean	Std dev.
Amylose	121	10.0	21.7	15.7	1.57
Sugars	121	2.54	8.13	5.39	1.25

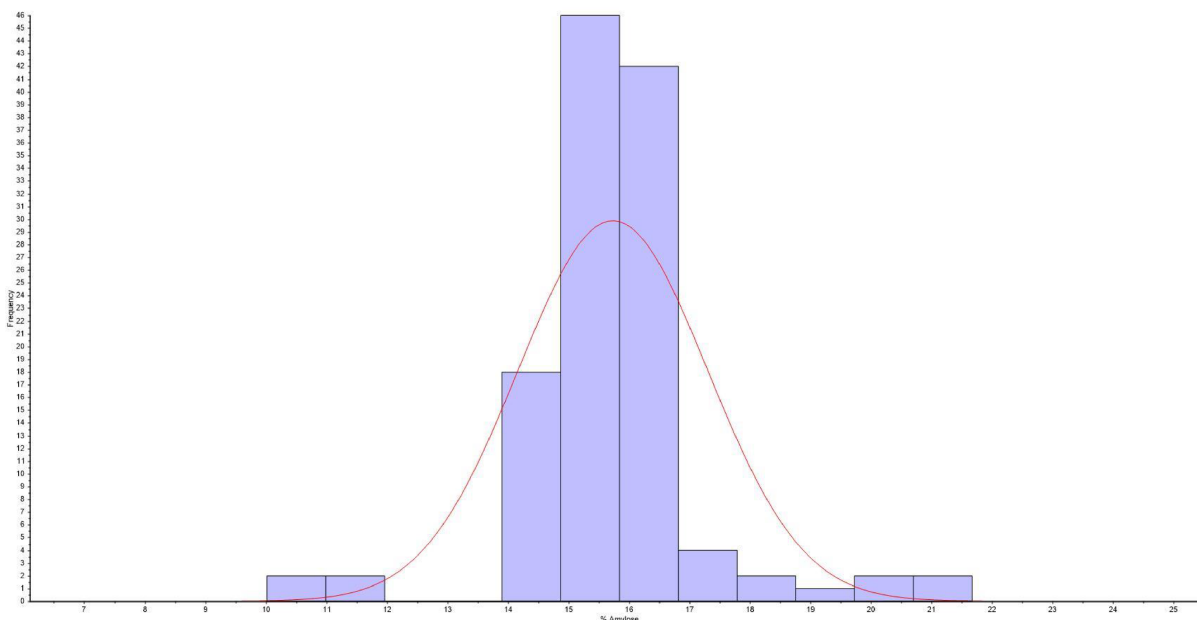
Note: All values are expressed in g/100 g.

we visually compare MSC and SNV spectra (Figure 1D,E), we observe a decrease in the slope of the SNV spectra, which is associated with scatter-related multiplicative interferences and variations in particle size, as compared to the slopes of the raw and MSC spectra. Additionally, SNV spectra does not depend on the spectral mean values.

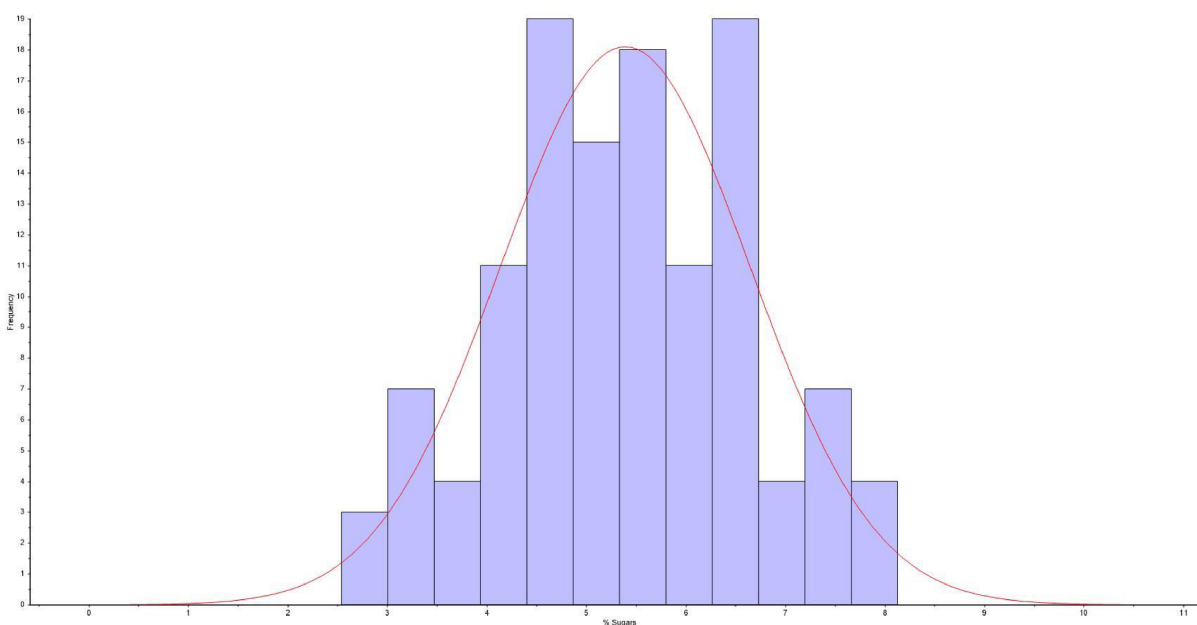
3.4 | Comparison of Calibration Data of the Two Preprocessing Treatments

Preprocessing of NIR spectral data has become an integral part of NIRS prediction modelling. The objective of the preprocessing is to remove physical phenomena in the spectra in order to improve the subsequent multivariate regression, classification model or exploratory analysis (Rinnan, Van Den Berg, and Engelsen 2009). Several physical factors may involve the size and organisation of the particle causing a variation in the pathlength (Barnes, Dhanoa, and Lister 1989). The application of different preprocessing techniques on spectral data before modelling serves multiple purposes. It not only helps reduce the impact of noise and external interference but also enhances the spectral features associated with the properties of interest. As a result, this improvement in spectral information leads to increased prediction accuracy for calibrated models (Nocita et al. 2015). First derivative can remove

a constant baseline (offset), and a second derivative can additionally eliminate slope. To mitigate noise, it is often necessary to apply smoothing before performing the derivation process. While dealing with complicate interferences or if incorrect smoothing parameters are applied, the outcome of the derivative process could become ineffective (Bi et al. 2016). A total of 24 equations for predicting amylose and sugars was generated by combining six spectral derivative mathematical treatments, that is, '1,4,4,1', '1,8,8,2', '2,4,4,1', '2,8,8,2', '3,4,4,1' and '3,8,8,2' along with two scatter correction algorithms (SNV-DT and MSC). Individual MPLS calibrations were conducted for each parameter, that is, amylose and sugar, utilising the calibration set consisting of 80 samples. Several statistics (SEC(V), RSQ_{internal}) were used to describe the performance of the calibration equation as given in Tables 2 and 3, respectively. For amylose, the highest RSQ_{internal} (0.793) was obtained in the case of MSC in the mathematical treatment of 2,4,4,1 followed by RSQ_{internal} (0.792), which was found in the case of SNV-DT in the mathematical treatment of 1,4,4,1. In both these treatments, the principal components (PCs) were 3 (Table 2). Unlike amylose, for sugars, the highest RSQ_{internal} (0.943) was found in the case of SNV-DT in the mathematical treatment of 3,4,4,1 while in MSC, the highest RSQ_{internal} (0.864) was found in the mathematical treatment of 2,4,4,1 having PCs 8 in both the cases (Table 3). Overall, in calibration set, SNV-DT provided a better result than MSC while



A Histogram of amylose



B Histogram of sugars

FIGURE 2 | (A) Histogram of amylose. (B) Histogram of sugars.

in amylose, MSC has provided better results than SNV-DT. The number of PCs varied from 2 (amylose: MSC; 1,4,4,1 and 2,4,4,1) to 8 (sugars: SNV-DT and MSC; 2,4,4,1 and 3,4,4,1).

3.5 | Comparison of Validation Data of the Two Preprocessing Treatments

The validation was done on 41 samples for the given traits in two preprocessing treatments. We have tried not to remove

any outliers in the validation process ensuring robustness of the model. Different statistical parameters for amylose and sugars were used to choose best fit models like RSQ_{external} , RPD and SEP(C) as given in Tables 4 and 5, respectively. In the case of amylose, the highest RSQ_{external} (0.962) was found in third derivative treatment in SNV-DT (Ano and Ubochi 2008; Bagchi, Sharma, and Chattopadhyay 2016; Callado et al. 2018). Whereas in the case of MSC pretreatment, the highest RSQ_{external} (0.959) was found in the third derivative (Ano and Ubochi 2008; Bagchi, Sharma, and Chattopadhyay 2016;

TABLE 2 | Calibration statistics for amylose content.

Mathematical treatments	SNV-DT			MSC		
	RSQ	PCs	SEC(V)	RSQ	PCs	SEC(V)
1441	0.792	3	0.527	0.769	2	0.539
1882	0.779	3	0.528	0.765	2	0.542
2441	0.774	3	0.531	0.793	3	0.528
2882	0.776	3	0.530	0.777	3	0.530
3441	0.710	3	0.674	0.710	3	0.675
3882	0.784	3	0.541	0.783	3	0.542

Abbreviations: MSC, multiplicative scatter correction; PCs, principal components; RSQ, coefficient of determination; SEC(V), standard error of calibration; SNV-DT, standard normal variate with detrend.

TABLE 3 | Calibration statistics for sugar content.

Mathematical treatments	SNV-DT			MSC		
	RSQ	PCs	SEC(V)	RSQ	PCs	SEC(V)
1441	0.418	3	1.126	0.555	5	1.043
1882	0.402	3	1.139	0.507	5	1.054
2441	0.868	8	1.035	0.864	8	1.025
2882	0.577	5	1.157	0.583	5	1.148
3441	0.943	8	0.961	0.712	8	0.951
3882	0.676	5	1.049	0.686	5	1.059

Abbreviations: MSC, multiplicative scatter correction; PCs, principal components; RSQ, coefficient of determination; SEC(V), standard error of calibration; SNV-DT, standard normal variate with detrend.

TABLE 4 | Validation statistics for amylose content.

Mathematical treatments	SNV-DT				MSC			
	RSQ	RPD	SEP(C)	<i>p</i>	RSQ	RPD	SEP(C)	<i>p</i>
1441	0.931	3.71	0.291	0.135	0.922	3.57	0.303	0.059
1882	0.931	3.74	0.289	0.167	0.918	3.47	0.311	0.063
2441	0.962	5.10	0.172	0.025	0.922	3.57	0.304	0.059
2882	0.948	4.14	0.261	0.732	0.947	4.11	0.263	0.767
3441	0.913	3.14	0.279	0.269	0.941	3.72	0.236	0.512
3882	0.960	4.98	0.176	0.136	0.959	4.93	0.178	0.200

Abbreviations: MSC, multiplicative scatter correction; RPD, residual prediction deviation; RSQ, coefficient of determination; SEP(C), standard error of prediction; SNV-DT, standard normal variate with detrend.

TABLE 5 | Validation statistics for sugar content.

Mathematical treatments	SNV-DT				MSC			
	RSQ	RPD	SEP(C)	<i>p</i>	RSQ	RPD	SEP(C)	<i>p</i>
1441	0.537	1.47	0.800	1.00	0.668	1.73	0.657	0.819
1882	0.513	1.42	0.766	0.502	0.594	1.56	0.667	0.455
2441	0.848	2.52	0.468	0.809	0.830	2.38	0.491	0.892
2882	0.758	2.02	0.575	0.663	0.748	1.98	0.587	0.644
3441	0.914	3.39	0.359	0.445	0.718	1.87	0.625	0.123
3882	0.820	2.29	0.527	0.258	0.824	2.35	0.518	0.235

Abbreviations: MSC, multiplicative scatter correction; RPD, residual prediction deviation; RSQ, coefficient of determination; SEP(C), standard error of prediction; SNV-DT, standard normal variate with detrend.

Callado et al. 2018) (Table 4). Previous studies have been carried out for amylose prediction in rice with best model in the second derivative using SNV-DT, that is, $RSQ_{\text{external}} = 0.540$ (Bagchi, Sharma, and Chattopadhyay 2016), while Xie et al. (2014) found the best equation in MSC than SNV-DT for both brown ($RSQ = 0.920$) and milled flour ($RSQ = 0.920$) in the second-order derivative. These discrepancies may be found due to spectral characteristics, the spectral measurement methods, preprocessing techniques and calibration methods. However, as compared to these studies, we have got the high RSQ_{external} in both SNV-DT and MSC. In the case of sugars, the best RSQ_{external} (0.914) was found in the third derivative treatment (Adebooye and Singh 2008; Bagchi, Sharma, and Chattopadhyay 2016; Barnes, Dhanoa, and Lister 1989) in SNV-DT, while for MSC, the best RSQ_{external} (0.830) was found in the second derivative (Adebooye and Singh 2008; Ano and Ubochi 2008; Barnes, Dhanoa, and Lister 1989) (Table 5). Overall, in the case of amylose and sugars, SNV-DT was found to be a good preprocessing treatment but in higher derivatives. The limitation of MSC lies in the fact that it attempts to match a spectrum to an idealised reference spectrum. This can pose challenges when dealing with spectra from diverse fractions that need to conform to the same reference spectrum (De Groot et al. 2001). Similar studies of comparison of preprocessing treatment have been done by Miloš, Bensa, and Japundžić-Palenkić (2022) in studying soil organic carbon, cation exchange capacity and clay. This study has found DT more robust than MSC in all the parameters as it removes the mean value or linear trends in spectroscopic data of densely packed solids more efficiently than MSC (Barnes, Dhanoa, and Lister 1989). Kho et al. (2020) did a similar kind of study with

two regression algorithms, that is, PLS and SVM, and found SNV-DT preprocessing better than MSC along with least LVs, that is, 2. One more reason with respect to reference and baseline correction is that, in MSC, both corrections are applied simultaneously not consecutively. Hence, MSC will generally give a smaller baseline correction than SNV-DT (Rinnan, Van Den Berg, and Engelsen 2009).

Detrending is known to reduce the curve linearity of powdered and packed samples along with correction of baseline shifts. Simultaneously, SNV-coupled detrend could also compensate additional baselines shifts. Williams, Manley, and Antoniszyn (2019) have given a suitable criterion to know the accuracy of model prediction. According to Williams, Manley, and Antoniszyn (2019), models are deemed excellent if their RSQ_{external} value surpasses 0.91, and they are considered good for prediction if the RSQ_{external} falls between 0.82 and 0.90. RSQ_{external} values ranging from 0.66 to 0.81 suggest approximate quantitative predictions, while values between 0.50 and 0.65 indicate that over 50% of the variance in Y is explained by variance in X, enabling differentiation between high and low concentrations. So, in case of amylose, RSQ_{external} surpasses 0.91 for both SNV-DT and MSC indicating the excellent prediction capacity of the models. However, in case of sugars, RSQ_{external} surpasses 0.91 for SNV-DT only indicating the excellent prediction capacity in this pretreatment, but in the case of MSC, RSQ_{external} falls between 0.82 and 0.90 indicating a good prediction model. SNV-DT was found to be superior for prediction modelling in the case of sugars. The regression plots of best prediction models for both the traits and preprocessing treatments are given in Figure 3A–D.

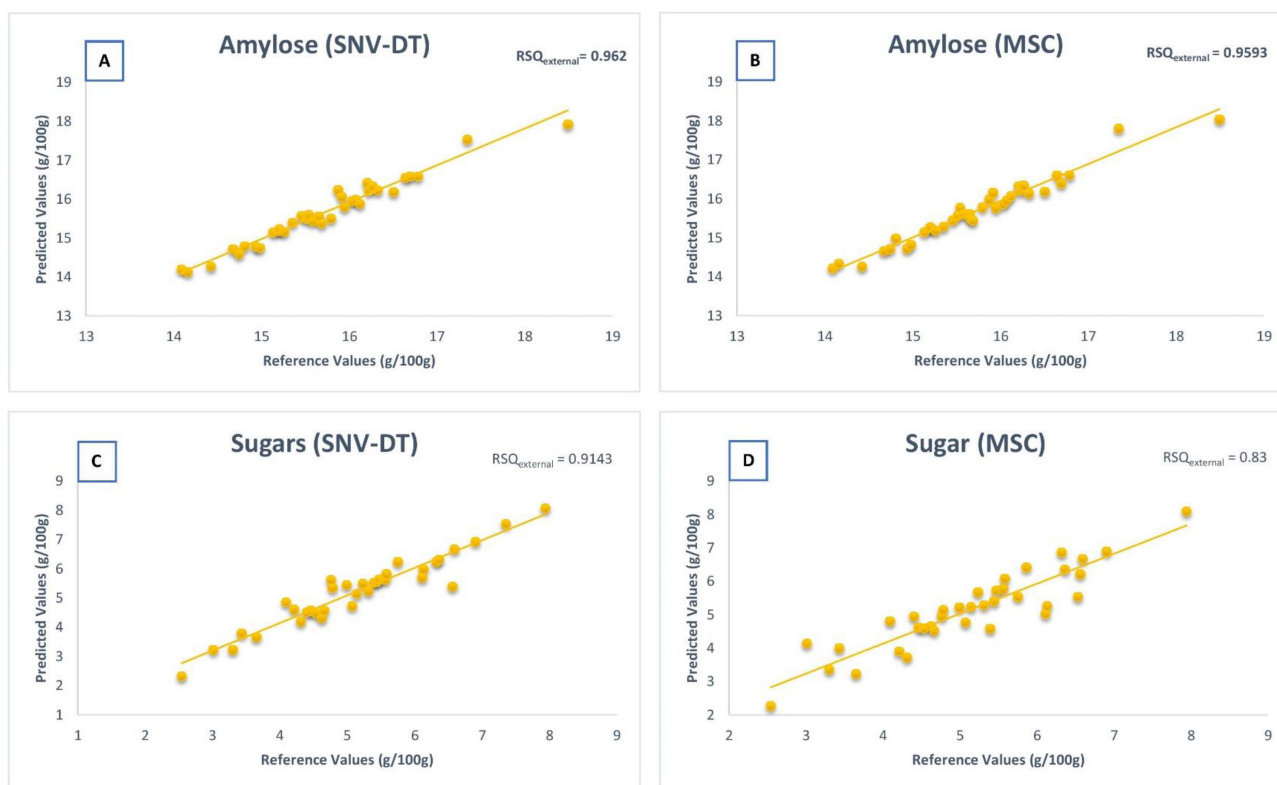


FIGURE 3 | (A–D) Regression plot for amylose and sugar.

Only RSQ_{external} would not be sufficient for testing the prediction accuracy, and RPD is used for checking the robustness of the models. Based on their RPD values, it is suggested that values ranging from 2.5 to 2.9 are classified as fair and suitable for preliminary assessment, while values between 3 and 3.4 are deemed suitable for quality control. Those exceeding 3.5 are considered fitting for process control, and those surpassing 4.1 are regarded as excellent (Chadalavada et al. 2022). Conversely, calibrations with RPD values below 2 are labelled as inadequate and not advisable. In this study, for mathematical treatments '2,4,4,1', '2,8,8,2' and '3,8,8,2' in amylose for SNV-DT preprocessing, RPD values were found to be more than 4 indicating excellent prediction models, while, for MSC preprocessing, mathematical treatments '2,8,8,2' and '3,8,8,2' RPD values were found to be more than 4 indicating excellent prediction models while for the rest of treatments, the RPD values exceeds 3.5, which will be considered fitting for process control.

However, in the case of sugars in SNV-DT preprocessing for mathematical treatment '3,4,4,1', the RPD value was in between 3 and 3.4, which is deemed suitable for quality control. For mathematical treatments '2,4,4,1', '2,8,8,2' and '3,8,8,2', the RPD values come in between 2 to 3.0, which is considered as fair and suitable for screening purpose. For the rest of mathematical treatments, the RPD was less than 2, which was not advisable. While, in case of MSC, the RPD values were not good as SNV-DT in sugars, for mathematical treatments '2,4,4,1' and '3,8,8,2', the RPD values were in between 2 and 3, which is considered as fair and suitable for screening purpose. And for the rest of the treatments, RPD values were less than 2, which is not considered advisable. Clearly in the case of sugars, SNV-DT got excellent results than MSC for prediction.

Paired *t*-test has been done to determine if the average of a dependent variable matches the analytical and predicted values of the tested biochemical parameters, at a 95% confidence level. In our investigation, the resulting *p*-value exceeded 0.05, affirming the precision and dependability of the models. The *p* values of amylose and sugars have been given in Tables 4 and 5, respectively. A *p*-value greater than 0.05 indicates the rejection of the null hypothesis, suggesting that the difference between the means of predicted and reference values is not significantly different, while *p*-values below 0.05 indicate otherwise.

4 | Conclusion

In the present study, the ability of NIRS to quantify amylose and total sugars in cowpea was assessed. Through the course of present study, it can be said that NIRS chemometrics coupled with SNV-DT were proved to be efficient in extracting the spectral features of cowpea spectra. The spectroscopic models of amylose and sugars in cowpea performed well in external validation as evident by their respective coefficients of determinations. Spectral preprocessing proved to be the major contributor in the development of robust prediction models for both the traits. In the future, it will be required to select specific wavelength regions, multivariate regression techniques and coupled scatter correction techniques, which would increase the prediction ability of NIRS models.

Author Contributions

Siddhant Ranjan Padhi: Data curation; Writing – original draft. **Racheal John:** Data curation; Writing – original draft. **Kuldeep Tripathi:** Resources; Writing – review and editing. **Dhammaprakash Pandhari Wankhede:** Writing – review and editing. **Tanay Joshi:** Resources; Visualization. **Jai Chand Rana:** Funding acquisition; Supervision. **Amritbir Riar:** Funding acquisition; Project administration; Writing – review and editing. **Rakesh Bhardwaj:** Conceptualization; Methodology; Supervision.

Acknowledgements

The authors acknowledge the support received from HOD-DGE, Professor P.G.R., the Graduate School ICAR-IARI and Director, ICAR-National Bureau of Plant Genetic Resources, in carrying out these studies.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author.

References

- Adebooye, O. C., and V. Singh. 2008. "Physico-Chemical Properties of the Flours and Starches of Two Cowpea Varieties (*Vigna unguiculata* (L.) Walp)." *Innovative Food Science & Emerging Technologies* 9, no. 1: 92–100. <https://doi.org/10.1016/j.ifset.2007.06.003>
- Ano, A. O., and C. I. Ubochi. 2008. "Nutrient Composition of Climbing and Prostrate Vegetable Cowpea Accessions." *African Journal of Biotechnology* 7, no. 20: 3795–3798.
- Bagchi, T. B., S. Sharma, and K. Chattopadhyay. 2016. "Development of NIRS Models to Predict Protein and Amylose Content of Brown Rice and Proximate Compositions of Rice Bran." *Food Chemistry* 191: 21–27. <https://doi.org/10.1016/j.foodchem.2015.05.038>
- Barnes, R. J., M. S. Dhanoa, and S. J. Lister. 1989. "Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra." *Applied Spectroscopy* 43, no. 5: 772–777. <https://doi.org/10.1366/0003702894202201>
- Bartwal, A., R. John, S. R. Padhi, et al. 2023. "NIR Spectra Processing for Developing Efficient Protein Prediction Model in Mungbean." *Journal of Food Composition and Analysis* 116: 105087. <https://doi.org/10.1016/j.jfca.2022.105087>
- Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J. M. Roger, and A. McBratney. 2010. "Critical Review of Chemometric Indicators Commonly Used for Assessing the Quality of the Prediction of Soil Attributes by NIR Spectroscopy." *TrAC Trends in Analytical Chemistry* 29, no. 9: 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>
- Bi, Y., K. Yuan, W. Xiao, et al. 2016. "A Local Pre-Processing Method for Near-Infrared Spectra, Combined With Spectral Segmentation and Standard Normal Variate Transformation." *Analytica Chimica Acta* 909: 30–40. <https://doi.org/10.1016/j.aca.2016.01.010>
- Callado, C. S. C., N. Núñez-Sánchez, S. Casano, and C. Ferreira-Vera. 2018. "The Potential of Near Infrared Spectroscopy to Estimate the Content of Cannabinoids in *Cannabis sativa* L.: A Comparative Study." *Talanta* 190: 147–157. <https://doi.org/10.1016/j.talanta.2018.07.085>
- Carvalho, J. K., J. M. Moura-Bueno, R. Ramon, et al. 2022. "Combining Different Pre-Processing and Multivariate Methods for Prediction of Soil Organic Matter by Near Infrared Spectroscopy (NIRS) in Southern

- Brazil." *Geoderma Regional* 29: e00530. <https://doi.org/10.1016/j.geodrs.2022.e00530>
- Chadalavada, K., K. Anbazhagan, A. Ndour, et al. 2022. "NIR Instruments and Prediction Methods for Rapid Access to Grain Protein Content in Multiple Cereals." *Sensors* 22, no. 10: 3710. <https://doi.org/10.3390/s22103710>
- De Groot, P. J., G. J. Postma, W. J. Melssen, and L. M. C. Buydens. 2001. "Influence of Wavelength Selection and Data Preprocessing on Near-Infrared-Based Classification of Demolition Waste." *Applied Spectroscopy* 55, no. 2: 173–181.
- Dotto, A. C., R. S. D. Dalmolin, A. ten Caten, and S. Grunwald. 2018. "A Systematic Study on the Application of Scatter-Corrective and Spectral-Derivative Pre-Processing for Multivariate Prediction of Soil Organic Carbon by Vis-NIR Spectra." *Geoderma* 314: 262–274. <https://doi.org/10.1016/j.geoderma.2017.11.006>
- DuBois, M., K. A. Gilles, J. K. Hamilton, P. T. Rebers, and F. Smith. 1956. "Colorimetric Method for Determination of Sugars and Related Substances." *Analytical Chemistry* 28, no. 3: 350–356. <https://doi.org/10.1021/ac60111a017>
- Edeh, H. O., and C. O. Igberu. 2012. "Assessment of Vegetable Cowpea Production Among Small Holder Farmers in Ebonyi State, Nigeria." *ARNP Journal of Agricultural and Biological Science* 7, no. 3: 215–222.
- Geladi, P., D. MacDougall, and H. Martens. 1985. "Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat." *Applied Spectroscopy* 39, no. 3: 491–500. <https://doi.org/10.1366/0003702854248656>
- Hamid, S., S. Muzzafar, I. A. Wani, and F. A. Masoodi. 2015. "Physicochemical and Functional Properties of Two Cowpea Cultivars Grown in Temperate Indian Climate." *Cogent Food & Agriculture* 1, no. 1: 1099418. <https://doi.org/10.1080/23311932.2015.1099418>
- Hang, J., D. Shi, J. Neufeld, K. E. Bett, and J. D. House. 2022. "Prediction of Protein and Amino Acid Contents in Whole and Ground Lentils Using Near-Infrared Reflectance Spectroscopy." *LWT* 165: 113669. <https://doi.org/10.1016/j.lwt.2022.113669>
- Hoover, R., T. Hughes, H. J. Chung, and Q. Liu. 2010. "Composition, Molecular Structure, Properties, and Modification of Pulse Starches: A Review." *Food Research International* 43, no. 2: 399–413. <https://doi.org/10.1016/j.foodres.2009.09.001>
- Isaksson, T., and T. Næs. 1988. "The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy." *Applied Spectroscopy* 42, no. 7: 1273–1284.
- Jayathilake, C., R. Visvanathan, A. Deen, et al. 2018. "Cowpea: An Overview on its Nutritional Facts and Health Benefits." *Journal of the Science of Food and Agriculture* 98, no. 13: 4793–4806. <https://doi.org/10.1002/jsfa.9074>
- John, R., R. Bhardwaj, C. Jeyaseelan, et al. 2022. "Germplasm Variability-Assisted Near Infrared Reflectance Spectroscopy Chemometrics to Develop Multi-Trait Robust Prediction Models in Rice." *Frontiers in Nutrition* 9: 946255. <https://doi.org/10.3389/fnut.2022.946255>
- John, R., H. Bollinedi, C. Jeyaseelan, et al. 2023. "Mining Nutri-Dense Accessions From Rice Landraces of Assam, India." *Heliyon* 9, no. 7: e17524. <https://doi.org/10.1016/j.heliyon.2023.e17524>
- Johnson, J. B., K. Walsh, and M. Naiker. 2020. "Application of Infrared Spectroscopy for the Prediction of Nutritional Content and Quality Assessment of Faba Bean (*Vicia faba* L.)." *Legume Science* 2, no. 3: e40. <https://doi.org/10.1002/leg3.40>
- Johnson, J. B., K. B. Walsh, and M. Naiker. 2021. "Assessment of Bioactive Compounds in Faba Bean Using Infrared Spectroscopy." *Legume Science* 5: e203. <https://doi.org/10.1002/leg3.203>
- Juliano, B. O., C. M. Perez, A. B. Blakeney, et al. 1981. "International Cooperative Testing on the Amylose Content of Milled Rice." *Starch/Stärke* 33, no. 5: 157–162. <https://doi.org/10.1002/star.19810330504>
- Kho, E. A., J. N. Fernandes, A. C. Kotze, et al. 2020. "Visible-Near Infrared Spectroscopy for Detection of Blood in Sheep Faeces." *Journal of Near Infrared Spectroscopy* 28, no. 5–6: 255–266. <https://doi.org/10.1177/0967033520927519>
- López, M. G., A. S. García-González, and E. Franco-Robles. 2017. "Carbohydrate Analysis by NIRS-Chemometrics." *Developments in Near-Infrared Spectroscopy* 10: 67208.
- Marten, G. C., J. S. Shenk, and F. E. Barton. 1985. "Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality (No. 643)." US Department of Agriculture, Agricultural Research Service.
- Miloš, B., A. Bensa, and B. Japundžić-Palenkić. 2022. "Evaluation of Vis-NIR Preprocessing Combined With PLS Regression for Estimation Soil Organic Carbon, Cation Exchange Capacity and Clay From Eastern Croatia." *Geoderma Regional* 30: e00558. <https://doi.org/10.1016/j.geodrs.2022.e00558>
- Nassourou, M. A., T. J. B. Noubissié, Y. N. Njintang, and J. M. Bell. 2017. "Diallel Analyses of Soluble Sugar Content in Cowpea (*Vigna unguiculata* L. Walp.)." *The Crop Journal* 5, no. 6: 553–559. <https://doi.org/10.1016/j.cj.2017.05.005>
- Nocita, M., A. Stevens, B. van Wesemael, et al. 2015. "Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring." *Advances in Agronomy* 132: 139–159. <https://doi.org/10.1016/bs.agron.2015.02.002>
- Omueti, O., and B. B. Singh. 1987. "Nutritional Attributes of Improved Varieties of Cowpea (*Vigna unguiculata* (L.) Walp.)." *Human Nutrition. Food Sciences and Nutrition* 41, no. 2: 103–112. <https://doi.org/10.1080/09528954.1987.11904105>
- Osborne, B. G. 2006. "Near-Infrared Spectroscopy in Food Analysis." In *Encyclopedia of Analytical Chemistry*, edited by R. A. Meyers and R. J. McGorin. <https://doi.org/10.1002/9780470027318.a1018>
- Ozaki, Y. 2012. "Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry." *Analytical Sciences* 28, no. 6: 545–563. <https://doi.org/10.2116/analsci.28.545>
- Padhi, S. R., A. Bartwal, R. John, et al. 2022. "Evaluation and Multivariate Analysis of Cowpea [*Vigna unguiculata* (L.) Walp] Germplasm for Selected Nutrients—Mining for Nutri-Dense Accessions." *Frontiers in Sustainable Food Systems* 6: 888041. <https://doi.org/10.3389/fsufs.2022.888041>
- Padhi, S. R., R. John, A. Bartwal, et al. 2022. "Development and Optimization of NIRS Prediction Models for Simultaneous Multi-Trait Assessment in Diverse Cowpea Germplasm." *Frontiers in Nutrition* 9: 1001551. <https://doi.org/10.3389/fnut.2022.1001551>
- Rinnan, Å., F. Van Den Berg, and S. B. Engelsen. 2009. "Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra." *TrAC Trends in Analytical Chemistry* 28, no. 10: 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Sarkar, S., J. K. Basak, B. E. Moon, and H. T. Kim. 2020. "A Comparative Study of PLSR and SVM-R With Various Pre-Processing Techniques for the Quantitative Determination of Soluble Solids Content of Hardy Kiwi Fruit by a Portable Vis/NIR Spectrometer." *Food* 9, no. 8: 1078. <https://doi.org/10.3390/foods9081078>
- Savitsky, A., and M. J. Golay. 1964. "Smoothing and Differentiation of Data by Simplified Polynomial Smoothing." *Analytical Chemistry* 36: 1627–1639.
- Shi, H., Y. Lei, L. L. Prates, and P. Yu. 2019. "Evaluation of Near-Infrared (NIR) and Fourier Transform Mid-Infrared (ATR-FT/MIR) Spectroscopy Techniques Combined With Chemometrics for the Determination of Crude Protein and Intestinal Protein Digestibility of Wheat." *Food Chemistry* 272: 507–513. <https://doi.org/10.1016/j.foodchem.2018.08.075>

- Shruti, S., A. Shukla, S. S. Rahman, et al. 2023. "Developing an NIRS Prediction Model for Oil, Protein, Amino Acids and Fatty Acids in Amaranth and Buckwheat." *Agriculture* 13, no. 2: 469. <https://doi.org/10.3390/agriculture13020469>
- Singh, J., A. Dartois, and L. Kaur. 2010. "Starch Digestibility in Food Matrix: A Review." *Trends in Food Science & Technology* 21, no. 4: 168–180. <https://doi.org/10.1016/j.tifs.2009.12.001>
- Singh, N. 2017. "Pulses: An Overview." *Journal of Food Science and Technology* 54: 853–857. <https://doi.org/10.1007/s13197-017-2537-4>
- Tomar, M., R. Bhardwaj, M. Kumar, et al. 2021. "Development of NIR Spectroscopy-Based Prediction Models for Nutritional Profiling of Pearl Millet (*Pennisetum glaucum* (L.) R. Br: A Chemometrics Approach." *LWT* 149: 111813. <https://doi.org/10.1016/j.lwt.2021.111813>
- Weng, Y., W. S. Ravelombola, W. Yang, et al. 2018. "Screening of Seed Soluble Sugar Content in Cowpea (*Vigna unguiculata* (L.) Walp)." *American Journal of Plant Sciences* 9, no. 7: 1455–1466. <https://doi.org/10.4236/ajps.2018.97106>
- Wilcox, J. R. 2001. "Sixty Years of Improvement in Publicly Developed Elite Soybean Lines." *Crop Science* 41, no. 6: 1711–1716. <https://doi.org/10.2135/cropsci2001.1711>
- Williams, P., M. Manley, and J. Antoniszyn. 2019. *Near Infrared Technology: Getting the Best out of Light*. Stellenbosch, South Africa: African Sun Media.
- Xie, L. H., S. Q. Tang, N. Chen, et al. 2014. "Optimisation of Near-Infrared Reflectance Model in Measuring Protein and Amylose Content of Rice Flour." *Food Chemistry* 142: 92–100. <https://doi.org/10.1016/j.foodchem.2013.07.030>