Carbohydrate Polymers 368 (2025) 124257

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Rapid and nondestructive prediction of total starch and amylose contents in single sorghum kernel (SSK) based on near infrared (NIR) spectroscopy

Jianjun Zhou [a], Yonghui Li [a], Scott R. Bean [b], Paul R. Armstrong [b], Xiaorong Wu [b,*]

[a] Grain Science and Industry, Kansas State University, Manhattan, Kansas 66506, USA
[b] Grain Quality and Structure Research Unit, Center for Grain and Animal Health Research, USDA-ARS, Manhattan, KS 66502, USA

ARTICLE INFO

Keywords:
Sorghum single kernel
Starch
Amylose
NIR
PLS model
Spectral pretreatment

ABSTRACT

This study aimed to establish NIR spectroscopy models for fast predicting apparent amylose (AA) and total starch (TS) content in SSK. Reliable wet chemistry procedures for quantifying TS and AA in single sorghum kernel (SSK) were established, which achieved high accuracy with test errors below 1.0 %. The partial least squares (PLS) model with 2 latent variables (LVs) for AA prediction had coefficients of determination of 0.91 (R^2_{cal}) and 0.85 (R^2_{cv}), and root mean square errors (RMSE) of 1.90 % and 2.47 % for calibration (RMSEC) and cross-validation (RMSECV), respectively. It showed an R^2_{pred} of 0.83 and RMSE of 2.58 % for prediction (RMSEP) when validated with the independent validation set. The optimal SSK-TS NIR PLS calibration model was built from 187 calibration sorghum kernels with 10 LVs, which had a R^2_{cal} of 0.79, RMSEC of 2.76 % and RMSECV of 4.93 % and showed a R^2_{pred} of 0.72 and RMSEP of 3.19 % when applied to an independent validation set of 93 samples. Overall, this study successfully developed wet chemistry methods for measuring AA and TS contents in SSK and established NIR models for nondestructive prediction and sorting of sorghum kernels by their TS or AA content, serving as useful tools for sorghum breeding and application research.

1. Introduction

Sorghum (*Sorghum bicolor*), the fifth largest cereal crop in the world after wheat, maize, rice, and barley, is widely cultivated in various regions and valued for its resilience to drought and adaptability to diverse environments (Dykes et al., 2014). Sorghum is a staple food for many communities, particularly in Africa and Asia, while in the USA and Australia, it is primarily used as animal feed and for ethanol production (Rashwan et al., 2021). Recently, the global consumption of sorghum as a gluten-free food source has been on the rise, driven by their numerous health benefits, including cholesterol-lowering effects, anti-inflammatory properties, slow digestibility, and potential anti-cancer effects (Gasiński et al., 2023; Ofosu et al., 2020; Sullivan et al., 2018).

Starch is the main component of sorghum grain, which comprises two distinct categories of molecules: amylose (mainly linear, MW up to 10^6 Da with few branches) and amylopectin (highly branched with MW around 10^7–10^9 Da). Amylose content and structural features of amylopectin significantly affect physicochemical properties of starch as well as the quality and applications of cereal crops (Sheng & Wei, 2022). As a result, considerable efforts have been dedicated to managing and manipulating starch content and composition in sorghum (Chen et al., 2019; Yerka et al., 2016).

Research has demonstrated considerable compositional variations among cereal kernels of the same variety from the same field, driven by both genetic and environmental factors (Armstrong, 2014; Delwiche, 1995; Zheng et al., 2024). Such compositional variations among cereal kernels have not been studied due to the lack of technology to differentiate and sort kernels out by their composition. One of the main challenges is the ability to assess starch composition at the single-kernel level without destroying seeds, which is essential for early-stage breeding and potential application research. Traditional wet chemistry methods for assessing starch and amylose content are time-consuming and destructive to sorghum seeds (Chrastil, 1987; Megazyme, 2024). With the advancement of non-destructive analytical technologies, faster and more sustainable alternatives are replacing conventional methods (Delwiche, 1995; Delwiche & Massie, 1996; Liu et al., 2022).

Near infrared reflectance (NIR) spectroscopy is a quick and non-destructive technology, which has been applied for measuring major components of cereal grains and other agricultural materials for decades (Tomar et al., 2025). However, most NIR models were developed for predicting chemical compositions in bulk samples (Huang et al., 2021; Peiris et al., 2021; Zerihun et al., 2020). These samples are often ground

* Corresponding author.
E-mail address: shawn.wu@usda.gov (X. Wu).

https://doi.org/10.1016/j.carbpol.2025.124257
Received 3 June 2025; Received in revised form 1 August 2025; Accepted 18 August 2025
Available online 18 August 2025
0144-8617/Published by Elsevier Ltd.

before analysis, thus information about individual kernel characteristics is lost (Delwiche, 1995). Currently, several studies have only explored the use of NIR models to predict the chemical composition (mainly protein, oil, etc.) of single kernels in wheat, maize, and soybean, etc. (Bramble et al., 2006; Delwiche & Massie, 1996; Spielbauer et al., 2009), while no research has been conducted on sorghum starch at single kernel level. Therefore, developing non-destructive NIR-based detection models for differentiating chemical compositions in SSK could be beneficial to advancing sorghum breeding and sorghum application research. Moreover, the absence of reliable quantitative methods for assessing starch information in SSK poses a significant challenge in establishing stable and accurate NIR prediction models. This study hypothesizes that, with reliable procedures to measure starch and amylose contents in SSK, we would be able to effectively predict total starch or amylose levels in SSK using SK-NIR technology. This approach facilitates non-destructive detection of starch content in individual sorghum kernels and allows for accurate classification based on their starch composition.

This study has two main objectives: first, to establish wet chemistry methods for quantifying total starch (TS) and apparent amylose (AA) content in SSK, and second, to develop capable NIR models for predicting the TS and AA content in SSK. Specifically, NIR spectra are collected from SSK samples, followed by quantification of their AA and TS contents. Various spectral pretreatment techniques and PLS regression are used to develop calibration and prediction models to predict TS and AA contents in sorghum kernels, which will provide valuable tools for the screening of sorghum seeds for both sorghum breeding and application projects.

2. Materials and methods

2.1. Grain samples

Sorghum kernels were from 89 sorghum hybrids and inbred lines harvested from the 2018–2021 crop years in Texas, Nebraska, and Kansas.

2.2. Chemicals and reagents

Mazie starch, potato amylose reference, glucose standard, and K-TSTA-100 A kit were from Megazyme International Ireland Ltd. (Bray Co., Wicklow, Ireland). Dimethyl sulfoxide (DMSO, chromatographic grade) was purchased from Thermo Fisher Scientific Inc. (Waltham, MA, USA). NaOH, KI, urea, trichloroacetic acid (TCA), and amylopectin standard (#10120) were purchased from Sigma Chemical Co. (St. Louis, Mo., USA). Iodine was from Mallinckrodt Chemical. (St. Louis, Mo., USA).

2.3. NIR spectra acquisition

A custom-built NIR instrument was used to collect spectra of individual sorghum seeds in motion. A detailed overview of this instrument and the spectral data collection process can be found in a paper authored by Armstrong (2014). The system featured a 908–1689 nm InGaAs-based spectrometer (CD NIR-256-1.7 T1, Control Development, South Bend, IN). Two fiber-optic cables were positioned at each end of an inclined, 8 mm diameter glass tube to collect reflected light from seeds passing through. The tube was illuminated by 48 miniature halogen lamps positioned along its length and circumference. Three NIR spectra were collected for each kernel from 940 to 1640 nm at 1 nm intervals and the average of the three was stored in MATLAB (ver. R2024b; MathWorks Inc., Natick, MA, USA) for further analysis.

2.4. Determination of TS and AA in SSK

After NIR spectra were collected, the SSK was oven dried at 130 °C for 20 h (ASAE, 2017) and then crushed for three cycles (1 min per cycle) in a 4-mL polyethylene vial with one 9.5 mm stainless steel grinding ball on a Genogrinder (Model2010, Metuchen, NJ). The crushed sample was mixed thoroughly with 1.0 mL of 90 % DMSO solvent at room temperature for 30 min. To ensure complete dissolution of the starch, the mixture was heated to 100 °C for 30 min before sampled for TS and AA analysis. The heated SSK-90 % DMSO mixtures of a test sorghum sample and the standard maize starch were examined under a polarized light microscope to ensure complete dissolution. In this study, 1/8 of the weight of the SSK-90 % DMSO mixture solution was used for TS and the remaining 3/8 for AA content analysis. Each test was conducted in duplicate and the difference between two replicates kept within 1.0 % for both TS and AA content.

2.4.1. Determination of TS content in SSK

The TS content was determined using the Megazyme K- TSTA-100 A kit (Bray, Ireland) with minor modifications (Megazyme, 2024). The 1/8 mixture portion was mixed with 1.0 mL of 1.7 M sodium hydroxide for 15 min. Subsequently, 4.0 mL of sodium acetate buffer (pH 3.8) was added to adjust the pH to 5.0. The samples were then hydrolyzed with 0.1 mL thermostable α-amylase and 0.1 mL amyloglucosidase at 50 °C for 30 min. After centrifugation at 1300 rpm for 5 min, 0.1 mL of the hydrolysate was combined with 3.0 mL of GOPOD (glucose oxidase peroxidase) reagent and incubated at 50 °C for 20 min. The absorbance at 510 nm was measured against a reagent blank to calculate the percent starch content in the SSK sample.

2.4.2. Determination of AA content in SSK

AA content in SSK samples was quantified colorimetrically based on a previous reported procedure (Chrastil, 1987; Peiris et al., 2021) with minor modifications to fit the small sample size of SSK. To accommodate AA contents in sorghum samples of different genotypes, varying volumes of a 90 % DMSO:0.6 M urea solution (0.2, 0.6–0.8, and 1.0 mL for waxy, heterowaxy, and regular) were used to dilute the SSK-90 % DMSO mixture to achieve appropriate absorbance values. The diluted solutions were then heated at 100 °C for 30 min. After cooling to room temperature, absorbance at 620 nm (A_{620}) was recorded after 100 μL of the above DMSO sample mixture was mixed with 5.0 mL TCA solution (0.5 %) and 100 μL of KI-I_2 solution (0.01 N) for 30 min. The amylose content in each sample kernel was calculated by comparing its A_{620} with that of a standard curve prepared with known amylose and amylopectin contents.

2.5. Spectral pretreatment and model optimization

In this study, NIR spectra were collected from a total of 430 sorghum kernels picked from samples of diverse genetic background. Results with a test error of less than 1.0 % between duplicate were used for AA (376) and TS (280) model development. Each dataset was then divided into calibration (70 %) and validation (30 %) sets prior to modeling. To correct the effects caused by light scattering, intensity differences, baseline shifts, and to enhance signal-to-noise ratio, dozens of individual and combined pretreatment methods were tested for model optimization using the DEVINER function of the PLS Toolbox 9.5 (Eigenvector Research Inc., Manson, WA) to preprocess the NIR spectra. Common pretreatment methods such as autoscale, baseline, scaling, extended multiplicative scatter correction (EMSC), external parameter orthogonalization (EPO) and extended mixture model (EMM) filter, generalized least squares weighting (glsw), mean center, multiplicative scatter correction (MSC, mean), normalize, smoothing (SavGol), SNV, etc. and different latent variables (LVs) were used, which are described in detail in supplementary materials (S1, S2). PLS models were evaluated by their RMSE of calibration, cross-validation, and prediction (RMSEC, RMSECV, and RMSEP) and determination coefficient for calibration, cross-validation, and validation (R^2_{cal}, R^2_{cv}, and R^2_{pred}). Those terms are defined by Eq. (1) and Eq. (2) below (Simeone et al., 2024):

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

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

Where $\hat{y}_i$ is the prediction value of the *i*-th sample, *yi* is the measured value of the sample, *n* is the number of samples, $\bar{y}$ is the mean of the measured values of the dependent variable, and Σ represents the sum of all the values. Eq. (1) can compute the R^2_{cal}, R^2_{cv}, and R^2_{pred}, and Eq. (2) compute the RMSEC, RMSECV, and RMSEP values.

2.6. Statistical analysis

Spectral data analyses were conducted using MATLAB (ver. R2024b; MathWorks Inc., Natick, MA, USA). Samples for the calibration set were selected and outliers identified from the score scatter plot for the calibration samples provided by the PCA, using Hotelling's ellipse at the 95 % confidence level and a plot of Hotelling T^2 (HT^2) vs. Q Residuals (*QRs*) (Cárdenas et al., 2015). Graphical presentations were done in Origin software (Origin2024, Northampton, MA, USA).

3. Results and discussion

3.1. Wet chemistry quantification of TS and AA in SSK

Moisture content (%) and dry weight (mg) of SSK were collected before wet chemistry analysis. Fig. 1 shows that the moisture content of SSK ranges from 4.0 % to 18.3 % and dry weight from 12.8 to 49.1 mg, with both following a normal distribution. Results in Fig. 1 show a mean of 10.63 % with a standard deviation of 2.38 % for moisture content and a mean of 26.63 mg with a standard deviation of 6.36 mg for dry weight.

For AA and TS measurement, the maize starch from Megazyme (85.0 % TS, 28.5 % AA) serves as references. A comparison of the laboratory test results with the reference values for TS and AA in commercial maize starch ($n = 10$) reveals an error of less than 1.0 % (data not shown), indicating that the methods for TS and AA quantification established in this study are reliable. The results show that the AA and TS contents of SSK are distributed between 0 and 22.0 % and 37.1–73.9 % (dw), respectively. The AA content in sorghum grains is related to wax gene regulation. The wax gene encodes granule-bound starch synthase (GBSS), an enzyme responsible for synthesizing amylose within starch granules (Wirnas et al., 2024). Generally, the endosperm of waxy sorghum contains three recessive waxy genes, heterowaxy sorghum contains at least one recessive gene, and normal sorghum does not contain recessive waxy genes (Sang et al., 2008; Yerka et al., 2016). This genetic variation could explain why SSK with amylose content between 5 and 10 % is rare. The correlation between TS and AA content in SSK is shown in Fig. 2. The results show that AA content in sorghum kernels does not correlate with TS content below 5 % ($R^2 = 0.001$) but shows a weak positive correlation ($R^2 = 0.392$) when TS content exceeds 5 %.

So far, methods for quantification of TS and AA in SSK have not been reported yet. The main reason is that SSK sample size is small and varies a lot, and it is difficult to crush and collect samples for subsequent analysis using general milling equipment. Accurately splitting the tiny amount of crushed SSK sample for simultaneously measuring both TS and AA presents additional challenge. In this study, the bead crushing process effectively breaks the sorghum grains into fine powder, which makes it possible for the starch to be fully dissolved in DMSO (Fang et al., 2006; Syahariza et al., 2010). DMSO acts as a hydrogen bond acceptor, effectively facilitating the dispersion of starch by disrupting both intermolecular and intramolecular hydrogen bonds among amylose and amylopectin molecules and between starch granules and water. It replaces the hydrogen bonds between starch and hydroxyl groups with new bonds between DMSO and starch, enhancing the solubility of the starch (Fang et al., 2006). Studies have also demonstrated that adding a small amount of water (10 % water in this study) or low-molecular-weight electrolytes (such as LiBr, Urea, or $NaNO_3$) in DMSO can enhance starch solubility, which is why a 90 % DMSO solution was used in this research (Chuang & Sydor, 1987; Lv et al., 2024).

Images in Fig. 3 showed the birefringence (Maltese cross) phenomenon of both SSK- (a, b, c) and maize starch samples (d, control) in 90 % DMSO solution. Fig. 3a, b, and c represent the 90 % DMSO solution of SSK crushed for 1, 2, and 3 min, respectively. It can be observed that as the crushing time increases from 1 to 3 min, the observed Maltese crosses decrease, suggesting that longer crushing times facilitate the dissolution of sorghum starch in the 90 % DMSO solution. After further heating at 100 °C for 30 min, the Maltese cross disappears completely (e, f, g, h). This observation confirms that the bead crushing procedure, combined with the use of 90 % DMSO as a solvent, effectively dissolves the starch granules in the SSK.

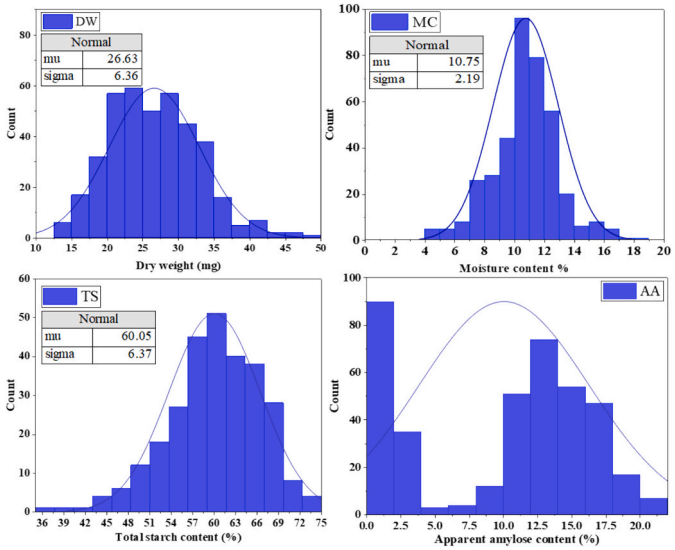

Fig. 1. Dry weight (DW, mg), moisture content (MC, %), total starch (TS, %) and apparent amylose (AA, %) content in sorghum single kernel (SSK) samples.

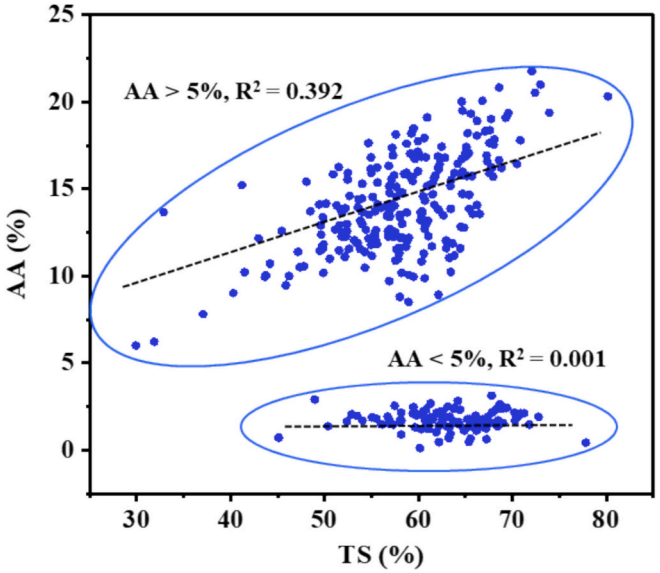

Fig. 2. Scatter plot between total starch (TS) and apparent amylose (AA) contents of sorghum grain.

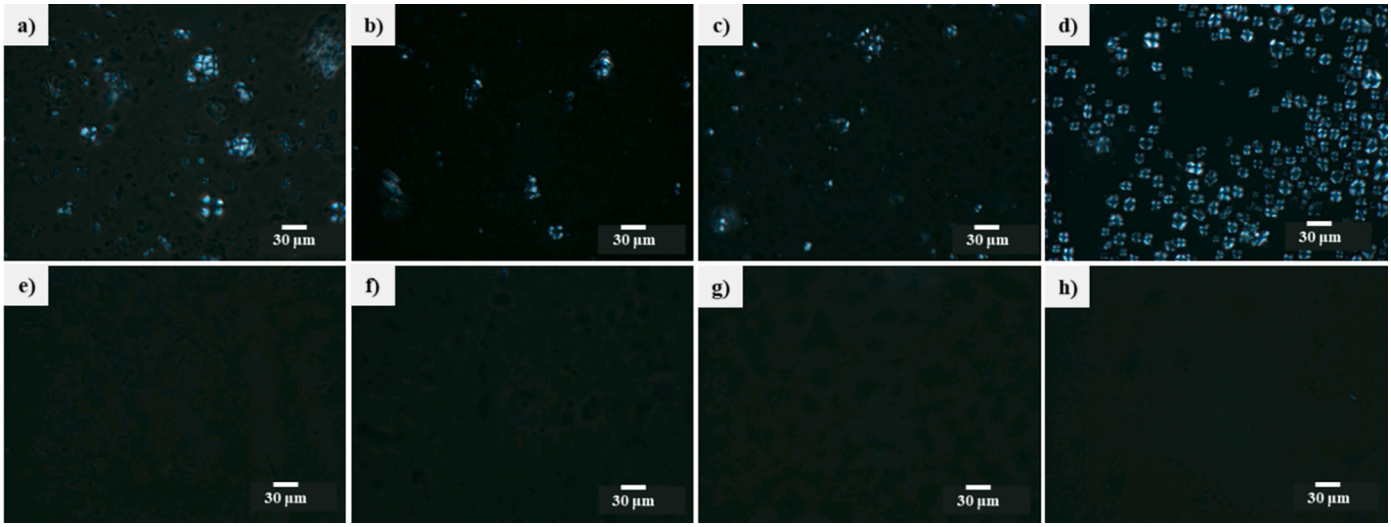

Fig. 3. Birefringence images for SSK and maize starch granules dissolved in 90 % DMSO solvents after different crush cycles (1,2,3) and heating process (100 °C, 30 min). Note: Image **a**), **b**), **c**) corresponds to SSK crushed 1, 2, and 3 cycle dissolved in 1.0 mL 90 % DMSO for 30 min respectively; Image **d**), maize starch dissolved in 1.0 mL 90 % DMSO for 30 min; Image **e**), **f**), **g**), and **h**) corresponds to the solution of **a**), **b**), **c**) and **d**) after heating at 100 °C for 30 min.

3.2. Performance evaluation of the NIR models

Scanning seeds in motion likely yield better results than static measurements, primarily because the scattering effects are minimized during the movement of the seeds, thereby reducing variability caused by differences in seed shape and size (Agelet & Hurburgh, 2014). After NIR raw data collection, pretreatment of spectral data is a crucial step before PLS algorithm modeling, which greatly reduce the baseline shifting and non-linearity caused by light scatter and improve the performance of multivariate regressions (Beć et al., 2025; Rinnan et al., 2009). Fig. 4 shows some of the common pretreatment techniques in NIR spectra, including various scatter correction methods and derivatization. To identify the most suitable NIR model for predicting AA and TS content in SSK, this study examines the effects of both individual and combined pretreatment techniques on the performance of the NIR calibration and prediction models.

A total of 6920 models with different pretreatment techniques and numbers of LVs (AA, see S1 file; TS, see S2 file) were evaluated using PLS_Tool box's DEVINER function based mainly on model's RMSEC, RMSECV, RMSEP values (Fig. 5**).** The SSK-AA-NIR models (Fig. 5a**)** with 1–10 PLS LVs from 376 sample spectra (calibration dataset 251, validation dataset 125) have RMSEC values between 0 % and 6 % (mostly within 1–4 %) and RMSEP values between 2 % and 8 % (mostly 2–4 %). The increase of LVs had little effect on RMSECV/RMSEC and PMSEP, which remained between 0 and 6 % and 2–6 %, respectively (Fig. 5b and c).

General criteria for model selection are low and comparable RMSEC and RMSEP values, as a small difference indicates good calibration and validation performance and a large discrepancy (RMSEP much larger than RMSEC) may suggest possible overfitting (Simeone et al., 2024). Close value between RMSEC and RMSECV, RMSEC and RMSEP indicate that the multivariate model estimates align well with the reference values, confirming the adequacy of spectral pretreatment and LV selection with little risk of overfitting (dos Santos et al., 2021). For better evaluation of the NIR model performance, the ratio between RMSEC and RMSECV, RMSEC and RMSEP were introduced. The results in Fig. 5d show that most models have a RMSECV/RMSEC between 0.8 and 1.2 %, and a RMSEC/RMSEP between 0.8 and 1.0 %. This suggests that the RMSECV and RMSEP value are very close to RMSEC, indicating no overfitting in most of the SSK-AA-NIR calibration models.

The SSK-TS-NIR models with 1–10 PLS LVs were built on 280 sample spectra with 187 for calibration and 93 for validation, which had RMSEC values between 0 % and 6.5 % (mostly 3–5 %), and RMSEP values between 4 % and 10 % (mostly 4–7 %) (Fig. 5e). The ratio of RESECV and RMSEC increases with the increase of LVs, but RMSEP has no significant increase (Fig. 5f and g). Results in Fig. 5h reveal that most models display an RMSECV/RMSEC ratio between 1.0 and 1.5, along with an RMSEC/RMSEP ratio between 0.8 and 1.0. The similarity of RMSECV and RMSEP values to RMSEC indicates that there is also no overfitting in most of the SSK-TS-NIR calibration models. Overall, most AA-NIR models have lower RMSEC and RMSEP values compared to TS-NIR model parameters, indicating that the AA NIR models in this study performed better than TS NIR models.

3.3. SSK-AA-NIR model determination

Six models were selected from the 6920 models (S1) to predict AA content in SSK where the RMSEC, RMSECV and RMSEP values were lower than 2.0, 2.5 and 3.0, respectively (Table 1). The results in Table 1 show that NIR spectrum after preprocessed by normalization (inf-norm, maximum = 1) or MSC (mean) + glsw (α = 0.02) resulted in models with good performance.

Among these pretreatment methods, normalization is a common method to address multiplicative scaling in NIR spectra, with techniques like 1-norm, 2-norm, and inf-norm. When using inf-norm (maximum = 1), each sample is scaled to the maximum value observed across all variables, normalizing the largest value to one and excluding smaller values from the scaling. In another study, inf-norm (maximum = 1) was used to preprocess NIR spectra and improve the model's predictive performance for tannins in *Acacia mearnsii* bark (Menezes et al., 2014). MSC is a transformation technique that compensates for both additive and multiplicative distortions in NIR spectra. It removes physical influences that do not provide meaningful chemical or physical information such as particle size and surface blaze, thus improves the accuracy of subsequent spectral analysis (Maleki et al., 2007). In addition, glsw utilizes the eigenvalues and eigenvectors of a covariance matrix to down-weight signals that are influenced by interferences or discrepancies between samples (Sun et al., 2023).

The NIR model preprocessed by normalization (inf-norm, maximum = 1) + glsw (α = 0.02) shows the best AA prediction performance when using PLS with 2 LVs. The scatter plots (predicted AA content against measured AA) of NIR data are shown in Fig. 6. HT^2 and *QRs* are calculated at the 95 % confidence level. As can be seen from Fig. 6a, none of the NIR spectra exceeds the 95 % confidence level of HT^2 and *QRs*. In

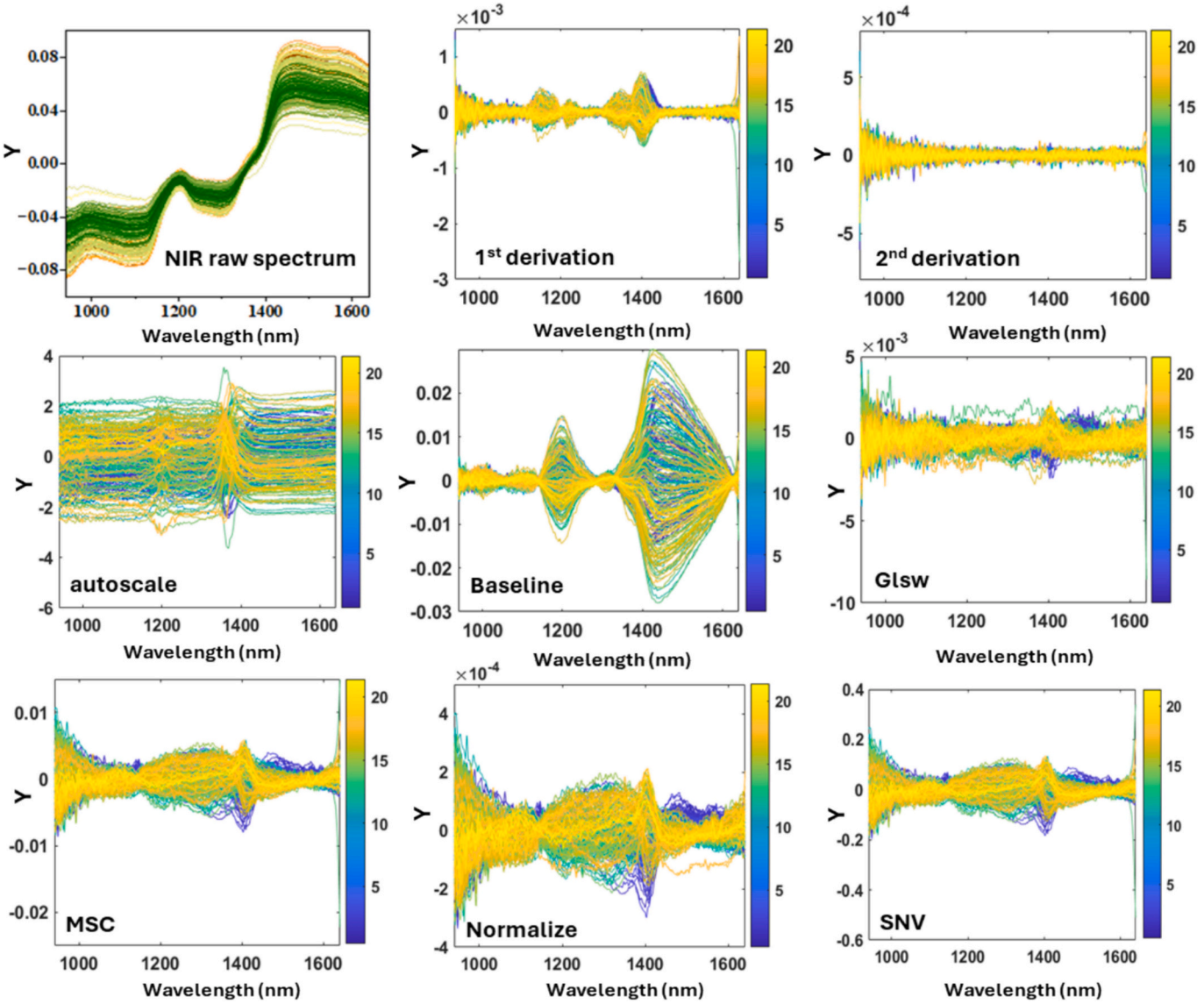

Fig. 4. NIR raw spectral curve (940–1640 nm) and preprocessing using stated techniques for modeling.

addition, leverage versus Studentized residuals charts are built in Fig. 6b and there are no points in critical region over ±3 limit Studentized residual and 3LV/m limit Leverage. Samples presenting high-Studentized residuals and Leverage values above a critical value are considered outliers (Pedro & Ferreira, 2007), the results indicates that all spectra are acceptable for subsequent analysis. The results in Fig. 6c show the AA calibration model from 251 sorghum kernel spectra with 2 PLS LVs had a R^2_{cal} of 0.91, R^2_{cv} of 0.85, RMSEC of 1.90 % and RMSECV of 2.47 %, which achieved a R^2_{pred} of 0.83 and RMSEP of 2.58 % when used to predict the AA content in an independent set of 125 sorghum kernels. Fig. 6c and Fig. 6d show that the calibration and validation sets are evenly distributed in the feature space, ensuring the model captures sufficient information across the entire data range (0–5 %, 10–22 %), thus ensuring prediction reliability. These results demonstrate that a reliable NIR model for predicting AA content at the single-kernel level of sorghum has been successfully established, which fills the gap in the application of NIR for predicting starch content at the single-kernel level in sorghum, a method that has previously only been reported for wheat, rice, corn, and soybean. Wu and Shi (2004) established a prediction model for single kernel rice weight, brown rice weight and polished rice amylose content using NIR, with prediction standard errors of 2.82, 1.09 and 1.30, and R^2 of 0.85, 0.71 and 0.67, respectively, which has potential advantages for rice breeding. Agelet et al. (2012) reported the use of NIR to differentiate normal or vigorously germinating corn kernels and soybean seeds from abnormal or dead seeds. In addition, NIR-based prediction of sorghum amylose content has been reported, but it has been limited to bulk or flour samples rather than at the single-kernel level. For example, the previous study developed a model for predicting AA content in sorghum bulk meal samples using NIR spectra, achieving an R^2cal of 0.84, an RMSECV of 2.96 %, an R^2pred of 0.76, and an RMSEP of 2.60 % based on 102 calibration and 51 independent samples (Peiris et al., 2021). Zerihun et al. (2020) developed a model for predicting AA content in sorghum grains using NIR spectra of sorghum flour samples, achieving a R^2cal of 0.76, RMSEC of 0.27 %, and a R^2pred of 0.69. Huang et al. (2021) employed hyperspectral imaging (HSI) to establish a model for predicting AA contents in sorghum, with the best model achieving a residual predictive deviation (RPD) value of 5.59 and RMSEP of 0.47 %. These models were based on the average spectrum and AA content of bulk sorghum samples. Our study successfully developed applicable models for predicting AA content in SSK.

The results in Fig. 6c also show that the predicted AA values for some SSK samples are negative, while the measured AA values in these sorghum grains are very low. This indicates that NIR can relatively accurately predict grains with high AA content, but the predictions become

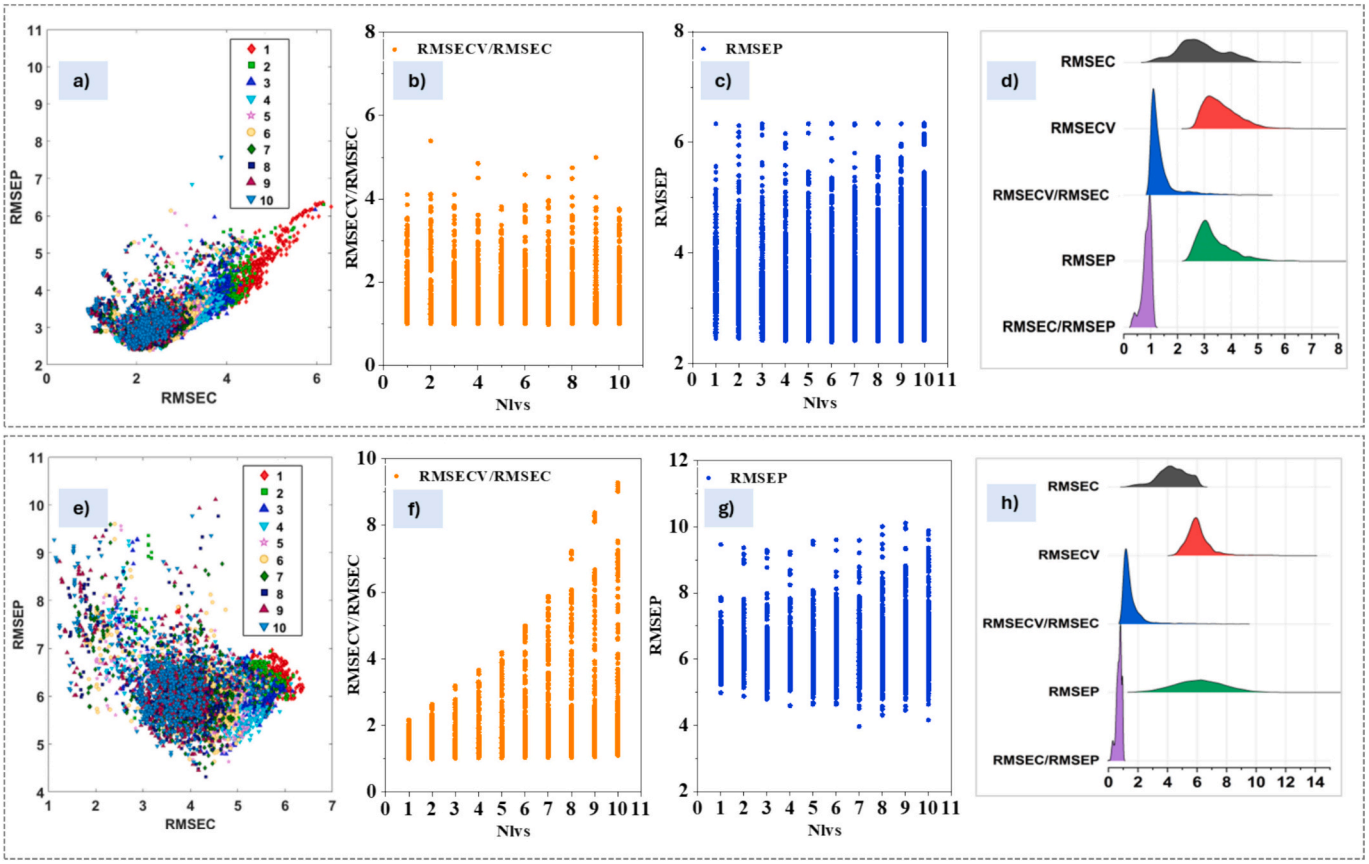

Fig. 5. AA (**a, b, c, d**) and TS (**e, f, g, h**) NIR models (totaling 6920 models) performance evaluation parameters visualization. Note: AA (apparent amylose), TS (total starch), RMSEC, RMSECV, and RMSEP are root mean square error of prediction, cross-validation, and prediction, respectively, Nlvs, number of latent variables.

Table 1
Key chemometric terms and features of selected SSK-NIR models for AA prediction.

		Calibration set		Cross-validation set			Validation set		
PLS LVs	Pretreatment techniques	RMSEC %	R^2	RMSECV %	R^2	RMSECV/RMSEC	RMSEP %	R^2	RMSEC/RMSEP
3	N + G + M	1.77	0.92	2.38	0.86	1.34	2.74	0.82	0.65
5	N + G + M	1.69	0.93	2.43	0.85	1.44	2.99	0.79	0.57
3	M + G + M	1.93	0.91	2.48	0.85	1.45	2.670	0.82	0.56
2	M + G + M	2.02	0.90	2.45	0.85	1.21	2.66	0.82	0.76
4	N + G + M	1.71	0.93	2.45	0.85	1.43	2.92	0.80	0.59
2	N + G + M	1.90	0.91	2.47	0.85	1.30	2.580	0.83	0.74

Note: N + G + M: Normalization (inf-norm, maximum = 1) + glsw (α = 0.02) + Mean center; M + G + M: MSC (mean) + glsw (α = 0.02) + Mean center. R^2 (coefficient of determination); RMSEC, RMSECV, and RMSEP are root mean square error of calibration, validation, and prediction, respectively.

less accurate when the measured AA values are very low. It also should be noted that SSK samples with AA content ranging from 5 to 10 % are relatively rare. Adding sorghum seeds with 5–10 % AA content to the model's calibration dataset may improve the model's prediction performance.

3.4. Contribution of the spectral variables to SSK AA-NIR model

Evaluating the contribution of spectral variables to PLS models is crucial to ensure that the key wavelengths are related to the spectroscopic signals of the target molecule, which confirms the validity of the NIR spectroscopy model (Peiris et al., 2021). The NIR spectrum preprocessed with normalize (inf-norm, maximum = 1) + glsw (α = 0.02) is shown in Fig. 7a.

The variable influence on projection (VIP) score is applied to highlight the contribution of spectral variables to the PLS models, as shown in Fig. 7b. The circle marks in the spectrum represent the characteristic peaks that have a great contribution to the NIR model for SSK-AA content prediction. The peaks near 960 and 1000 nm are the characteristic wavelengths of amylopectin, and the peaks near 980, 1400, and 1600 nm are the characteristic wavelengths of amylose (de Alencar Figueiredo et al., 2006; Huang et al., 2021). The characteristic peak differences between amylose and amylopectin arise from their distinct molecular structures. Amylose is composed of a linear chain of glucose units connected by α (1–4) glycosidic bonds, while amylopectin consists of a similar linear chain with approximately 5 % α (1–6) branching bonds, resulting in a branched structure (Tester et al., 2004). The α (1–6) bond in amylopectin involves a sixth carbon atom of glucose that forms a linkage with a CH_2 group at one end, which connects to an oxygen atom, and to the fifth carbon of the adjacent glucose unit at the branching point. This CH_2 group may exhibit different vibrational frequencies compared to the CH_2 groups in the linear glucose units (Peiris et al., 2021). Overall, the AA model can predict the AA content by utilizing the interaction of key NIR wavelengths with AA molecules in SSK, which can

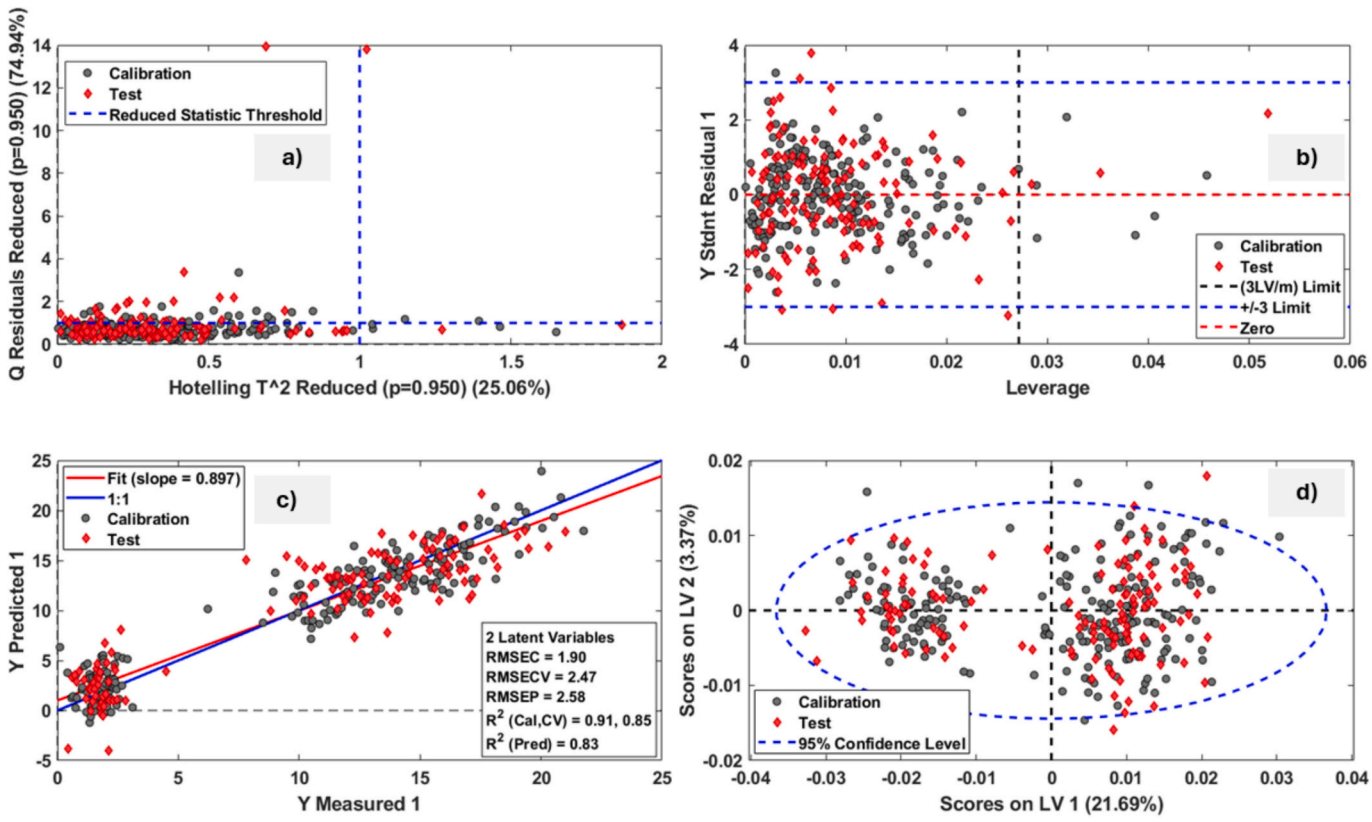

Fig. 6. Scatter plots of the predicted AA content against the measured AA content in SSK, with NIR calibration data (grey circles) and validation data (red circles). Chemometric techniques used: Normalization (inf-norm, maximum = 1), glsw (α = 0.02), mean center. Note: AA (apparent amylose), R^2 (coefficient of determination); RMSEC, RMSECV, and RMSEP are root mean square error of calibration, validation, and prediction, respectively.

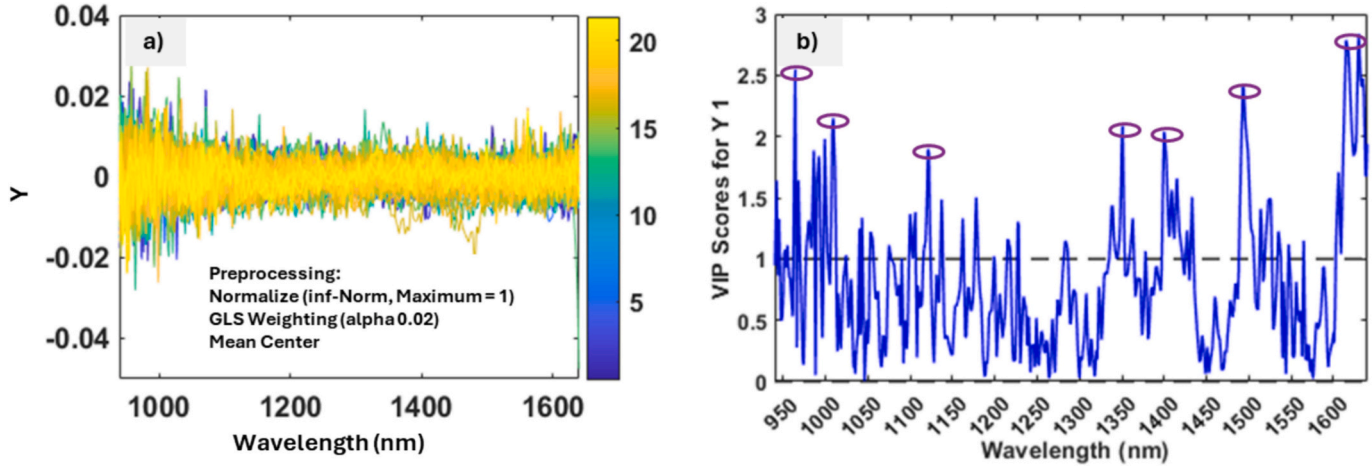

Fig. 7. (**a**) NIR spectra of AA-NIR after pretreatment with normalization (inf-norm, maximum = 1), glsw (α = 0.02), mean center; (**b**) Vip scores for AA prediction model. Note: AA (apparent amylose).

further improve the efficiency of data collection and processing.

3.5. SSK-TS-NIR model determination

Four models are obtained from all pretreatment methods (S2) when the RMSEC, RMSECV, and RMSEP values were all set lower than 5.0 (Table 2). The results in Table 2 demonstrate that various pretreatment techniques, including autoscale, mean center, epo/emm filter + mean center, and smoothing (order: 0, window: 21 pt., tails: polyinterp) + mean center, enhanced the development of the SSK-TS-NIR model. Compared to other pretreatment techniques, the NIR model preprocessed with the epo/emm filter + mean center exhibits lower RMSEC and RMSEP values when using PLS with 10 LVs. The epo filter is used to eliminate or at least minimize the effects of extraneous variables or variables (such as moisture) not relevant to the system under study on the NIR spectra, while the emm filter is able to identify unwanted covariance structures and remove these sources of variance from the data before calibration or prediction, thereby strengthening the TS-NIR prediction model performance (Kunze et al., 2021; Liu et al., 2015).

The scatter plots of predicted TS content from NIR versus measured

Table 2
Key chemometric terms and features of selected SSK-NIR models for TS prediction.

		Calibration set		Cross-validation set			Validation set		
PLSLVs	Pretreatment techniques	RMSEC%	R^2	RMSECV%	R^2	RMSECV/RMSEC	RMSEP %	R^2	RMSEC/RMSEP
7	Autoscale	4.05	0.56	4.55	0.46	1.12	3.97	0.60	1.02
10	Mean center	3.59	0.69	4.50	0.51	1.25	4.16	0.55	0.86
8	S + Mean center	4.43	0.52	4.90	0.42	1.16	4.30	0.53	1.03
10	Epo/emm filter (1 pcs), Mean center	2.77	0.79	4.93	0.40	1.78	3.19	0.72	0.86

Note: S + Mean center: Smoothing (order: 0, window: 21 pt., tails: polyinterp) + Mean center. R^2 (coefficient of determination); RMSEC, RMSECV, and RMSEP are root mean square error of calibration, validation, and prediction, respectively.

TS content are shown in Fig. 8. As shown in Fig. 8a, none of the NIR spectra exceed the 95 % confidence level for HT^2 and *QRs*, and Fig. 8b shows no points in the critical region beyond the ±3 limit for Studentized residuals or the 3LV/m limit for Leverage, indicating that all spectra are acceptable for building the SSK-TS-NIR models. The results in Fig. 8c show the TS calibration model with 10 PLS LVs based on 187 SSK spectra have a R^2_{cal} of 0.79, R^2_{cv} of 0.40, RMSEC of 2.77 % and RMSECV of 4.93 %, which obtains a R^2_{pred} of 0.72 and RMSEP of 3.19 % when applied to predict the TS content on an independent set of 93 sorghum kernels. Although 10 LVs were selected, the RMSEC/RMSEP ratio of 0.87 indicates that the model is not overfitted. The results in Fig. 8c and Fig. 8d indicate that the calibration and validation sets are uniformly spread across the feature space, which ensure the model to effectively capture the spectral information in the full data range (TS content from 40 to 75 %) and lead to a more reliable prediction model.

Several studies have also reported using NIR models to predict starch content in bulk sorghum samples, but none was on starch content of single sorghum kernel. For instance, Zerihun et al. (2020) developed an NIR model to predict starch content in sorghum flour with R^2_{cal} of 0.98, RMSE of 0.44, and R^2_{pred} of 0.91. Our previous study used sorghum grain samples (20 g of grains each sample) to establish sorghum TS calibration model (11 PLS LVs) with a R^2_{cal} of 0.87 and RMSECV of 1.57 % and R^2_{pred} of 0.76 and RMSEP of 2.13 % on the validation set (Peiris et al., 2021). These models only work on bulk sorghum kernels or ground sorghum meals, not for SSK. Furthermore, Fig. 8 reveals a limited distribution of grains with TS below 50 % and above 70 %, suggesting that incorporating additional data in these ranges could further enhance the model's predictive accuracy.

3.6. Contribution of the spectral variables to SSK TS-NIR model

The NIR spectrum preprocessed with epo/emm filter + mean center are shown in Fig. 9a. The VIP score is applied to highlight the contribution of spectral variables to the SSK TS-NIR PLS models (Fig. 9b). In Fig. 9b, key absorption peaks are identified at approximately 960, 990, 1030, 1130, 1370, 1490, and 1620 nm.

These peaks are likely associated with starch molecular vibrations, particularly C—H and O—H bonds. A comparison with known spectral features suggests that the peak near 990 nm corresponds to the second overtone of the O—H stretch in starch (Huang et al., 2021), while peaks

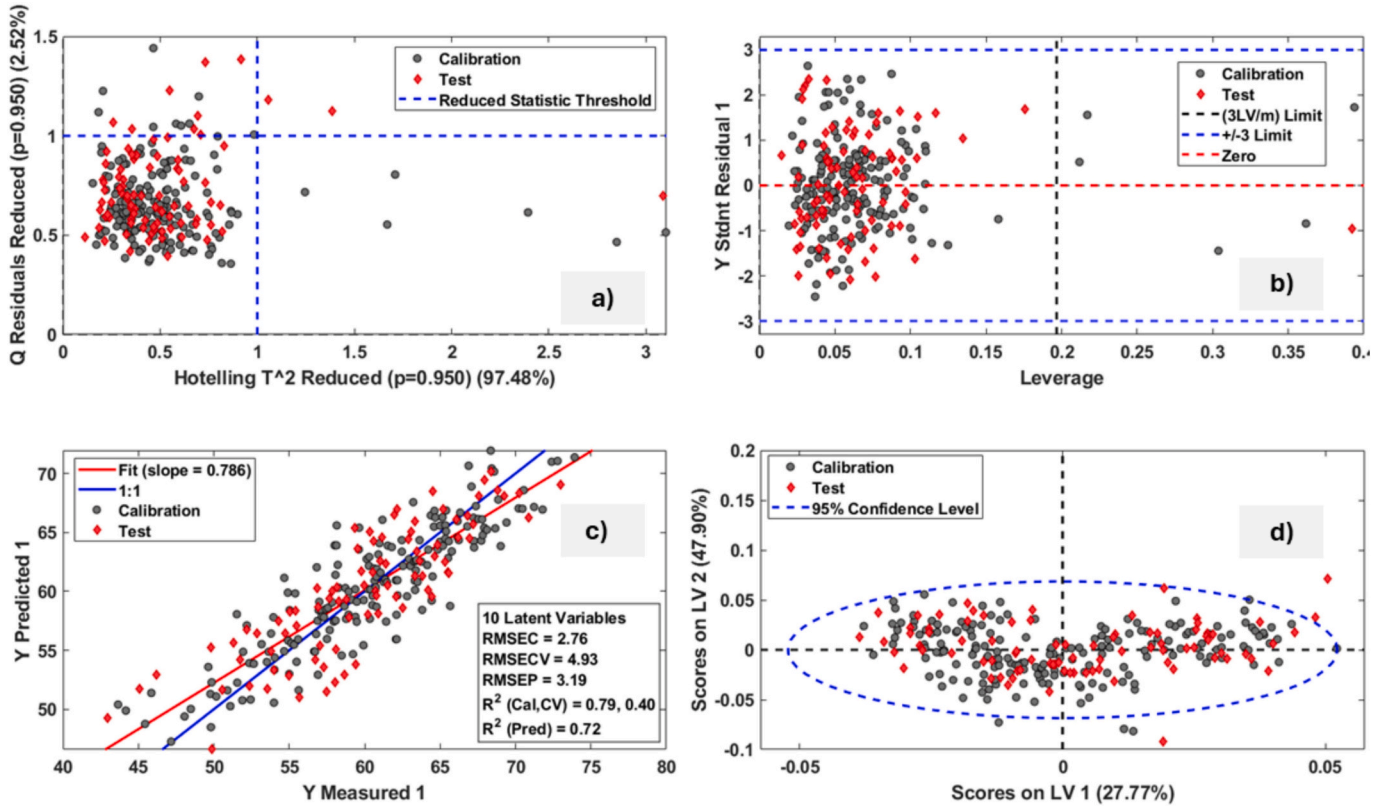

Fig. 8. Scatter plots of the predicted TS content against the measured TS content in SSK, with NIR calibration data (grey circles) and validation data (red circles). Note: TS (total starch), R^2 (coefficient of determination); RMSEC, RMSECV, and RMSEP are root mean square error of calibration, validation, and prediction, respectively.

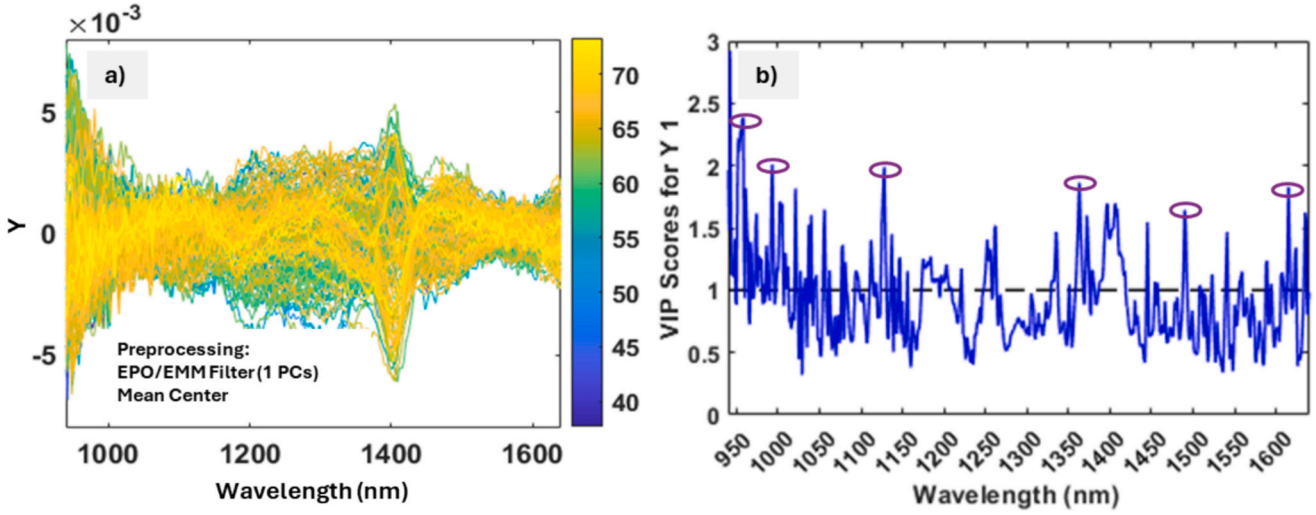

Fig. 9. (**a**) NIR spectra of TS-NIR after pretreatment with EPO/EMM Filter (1 PCs), Mean center; (**b**) Vip scores for TS prediction model. Note: TS (total starch).

at 1130 nm and 1370 nm relate to the combination of C—H stretching and C—H deformation, and the second overtone of C—H stretching (Camilotti et al., 2014). Similarly, 1030 nm and 1490 nm are likely attributed to C—H stretching and deformation overtones. The 1620 nm peak is associated with the first overtone of C—H stretching (Bantadjan et al., 2020; Williams, 2001). Starch molecules feature many C—H or O—H bonds that contribute greatly to NIR absorption (Lanjewar et al., 2024). Therefore, it is possible to develop an NIR model to predict TS content of SSK samples by utilizing these key NIR wavelengths.

4. Conclusion

This study successfully establishes protocols to accurately measure both AA and TS content of SSK, which are critical in developing reliable NIR models for predicting TS and AA in sorghum seeds. This method can be further used to measure other chemical components in grain seeds. By using PLS algorithms and various pretreatment methods, we established NIR models for predicting TS and AA at a single kernel level for the first time. Different pretreatment methods greatly affect the model prediction parameters, the normalize (inf-norm, maximum = 1) + glsw (α = 0.02) and epo/emm filter + mean centering are found to be most suitable for AA and TS prediction models, respectively. The AA-NIR prediction model showed high accuracy (R^2 of 0.8–0.9) and can be used to predict AA content in sorghum breeding programs and to sort waxy, heterowaxy, and normal sorghum kernels. The TS-NIR prediction model performs less accuracy (R^2 of 0.7–0.8) than the AA-NIR model, while its performance could be enhanced by including SSK samples with less than 50 % and more than 70 % TS content. Overall, the developed NIR models in this study could be useful tools for estimating and sorting sorghum kernels by their AA and TS contents and describe their characteristic distribution profiles, which have great potential in sorghum breeding programs and application research.

CRediT authorship contribution statement

Jianjun Zhou: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Yonghui Li:** Writing – review & editing, Supervision, Resources. **Scott R. Bean:** Writing – review & editing, Resources, Project administration. **Paul R. Armstrong:** Writing – review & editing, Methodology. **Xiaorong Wu:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The contents are the sole responsibility of the authors. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer. This project was funded by USDA-ARS project #3020-43440-001-00D.

Data availability

Data will be made available on request.

References

Agelet, L. E., Ellis, D. D., Duvick, S., Goggi, A. S., Hurburgh, C. R., & Gardner, C. A. (2012). Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. *Journal of Cereal Science, 55*(2), 160–165. https://doi.org/10.1016/j.jcs.2011.11.002

Agelet, L. E., & Hurburgh, C. R. (2014). Limitations and current applications of near infrared spectroscopy for single seed analysis. *Talanta, 121*, 288–299. https://doi.org/10.1016/j.talanta.2013.12.038

Armstrong, P. R. (2014). Development and evaluation of a near-infrared instrument for single-seed compositional measurement of wheat kernels. *Cereal Chemistry, 91*(1), 23–28.

ASAE. (2017). *Moisture measurement-unground grains and seeds: ASAE S352.2 APR 1988 (R2017)*. USA: Michigan.

Bantadjan, Y., Rittiron, R., Malithong, K., & Narongwongwattana, S. (2020). Establishment of an accurate starch content analysis system for fresh cassava roots using short-wavelength near infrared spectroscopy. *ACS Omega, 5*(25), 15468–15475. https://doi.org/10.1021/acsomega.0c01598

Beć, K. B., Grabska, J., & Huck, C. W. (2025). Interpretability in near-infrared (NIR) spectroscopy: Current pathways to the long-standing challenge. *TrAC, Trends in Analytical Chemistry, 189*. https://doi.org/10.1016/j.trac.2025.118254. Article 118254.

Bramble, T., Dowell, F. E., & Herrman, T. J. (2006). Single-kernel near-infrared protein prediction and the role of kernel weight in hard red winter wheat. *Applied Engineering in Agriculture, 22*(6), 945–949.

Camilotti, J. G., Somer, A., Costa, G. F., Ribeiro, M. A., Bonardi, C., Cruz, G. K., … Novatski, A. (2014). The phase-resolved photoacoustic method to indicate chemical assignments of paracetamol. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 121*, 719–723. https://doi.org/10.1016/j.saa.2013.11.099

Cárdenas, V., Cordobés, M., Blanco, M., & Alcalà, M. (2015). Strategy for design NIR calibration sets based on process spectrum and model space: An innovative approach for process analytical technology. *Journal of Pharmaceutical and Biomedical Analysis, 114*, 28–33. https://doi.org/10.1016/j.jpba.2015.05.002

Chen, B., Wang, C., Wang, P., Zhu, Z., Xu, N., Shi, G., Yu, M., Wang, N., Li, J., Hou, J., Li, S., Zhou, Y., Gao, S., Lu, X., & Huang, R. (2019). Genome-wide association study for starch content and constitution in sorghum (*Sorghum bicolor* (L.) Moench). *Journal of Integrative Agriculture, 18*(11), 2446–2456. https://doi.org/10.1016/s2095-3119(19)62631-6

Chrastil, J. (1987). Improved colorimetric determination of amylose in starches or flours. *Carbohydrate Research, 159*, 154–158.

Chuang, J. Y., & Sydor, R. J. (1987). High performance size exclusion chromatography of starch with dimethyl sulfoxide as the mobile phase. *Journal of Applied Polymer Science, 34*, 1739–1748. https://doi.org/10.1002/app.1987.070340431

de Alencar Figueiredo, L. F., Davrieux, F., Fliedel, G., Rami, J. F., Chantereau, J., Deu, M., … Mestres, C. (2006). Development of NIRS equations for food grain quality traitsthrough exploitation of a core collection of cultivated sorghum. *Journal of Agricultural and Food Chemistry, 54*, 8501–8509. https://doi.org/10.1021/jf061054g

Delwiche, S. R. (1995). Single wheat kernel analysis by near-infrared transmittance: Protein content. *Cereal Chemistry, 72*(1), 11–16.

Delwiche, S. R., & Massie, D. R. (1996). Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chemistry, 73*(3), 399–405.

dos Santos, V. J., Baqueta, M. R., Neia, V. J. C., de Souza, P. M., Março, P. H., Valderrama, P., & Visentainer, J. V. (2021). MicroNIR spectroscopy and multivariate calibration in the proximal composition determination of human milk. *LWT - Food Science and Technology, 147*. https://doi.org/10.1016/j.lwt.2021.111645. Article 111645.

Dykes, L., Hoffmann, L., Jr., Portillo-Rodriguez, O., Rooney, W. L., & Rooney, L. W. (2014). Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. *Journal of Cereal Science, 60*(1), 138–142. https://doi.org/10.1016/j.jcs.2014.02.002

Fang, Z., Wallace, Y., Qian, W., & Charles, F. S. (2006). Rice starch, amylopectin, and amylose: Molecular weight and solubility in dimethyl sulfoxide-based solvents. *Journal of Agricultural and Food Chemistry, 54*(6), 2320–2326. https://doi.org/10.1021/jf051918i

Gasiński, A., Kawa-Rygielska, J., Spychaj, R., Opiela, E., & Sowiński, J. (2023). Production of gluten-free beer brewing from sorghum malts mashed without external enzyme preparations. *Journal of Cereal Science, 112*. https://doi.org/10.1016/j.jcs.2023.103693. Article 103693.

Huang, H., Hu, X., Tian, J., Jiang, X., Sun, T., Luo, H., & Huang, D. (2021). Rapid and nondestructive prediction of amylose and amylopectin contents in sorghum based on hyperspectral imaging. *Food Chemistry, 359*. https://doi.org/10.1016/j.foodchem.2021.129954. Article 129954.

Kunze, D. C. G. C., Pastore, T. C. M., Rocha, H. S., Lopes, P. V. D. A., Vieira, R. D., Coradin, V. T. R., & Braga, J. W. B. (2021). Correction of the moisture variation in wood NIR spectra for species identification using EPO and soft PLS2-DA. *Microchemical Journal, 171*, Article 106839. https://doi.org/10.1016/j.microc.2021.106839

Lanjewar, M. G., Asolkar, S., Parab, J. S., & Morajkar, P. P. (2024). Detecting starch-adulterated turmeric using Vis-NIR spectroscopy and multispectral imaging with machine learning. *Journal of Food Composition and Analysis, 136*. https://doi.org/10.1016/j.jfca.2024.106700. Article 106700.

Liu, G., Wei, J., Li, X., Tian, M., Wang, Z., Shen, C., Sun, W., Li, X., Lv, E., Tian, S., Wang, J., Xu, S., & Zhao, B. (2022). Near-infrared-responded high sensitivity nanoprobe for steady and visualized detection of albumin in hepatic organoids and mouse liver. *Advanced Science, 9*(26), Article e2202505. https://doi.org/10.1002/advs.202202505

Liu, Y., Pan, X., Wang, C., Li, Y., & Shi, R. (2015). Predicting soil salinity with Vis-NIR spectra after removing the effects of soil moisture using external parameter orthogonalization. *PLoS One, 10*(10), Article e0140688. https://doi.org/10.1371/journal.pone.0140688

Lv, X., Hong, Y., Gu, Z., Cheng, L., Li, Z., Li, C., & Ban, X. (2024). Effect of solution on starch structure: New separation approach of amylopectin fraction from gelatinized native corn starch. *Carbohydrate Polymers, 329*. https://doi.org/10.1016/j.carbpol.2023.121770. Article 121770.

Maleki, M. R., Mouazen, A. M., Ramon, H., & De Baerdemaeker, J. (2007). Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosystems Engineering, 96*(3), 427–433. https://doi.org/10.1016/j.biosystemseng.2006.11.014

Megazyme. (2024). Total starch (α-amylase /amyloglucosidase) assay procedure. K-TSTA-100A, Procedure (b). https://www.megazyme.com/documents/Assay_Protocol/K-TSTA-100A_DATA.pdf Accessed April 23, 2025.

Menezes, C. M., da Costa, A. B., Renner, R. R., Bastos, L. F., Ferrão, M. F., & Dressler, V. L. (2014). Direct determination of tannins in *Acacia mearnsii* bark using near-infrared spectroscopy. *Analytical Methods, 6*(20), 8299–8305. https://doi.org/10.1039/c4ay01558d

Ofosu, F. K., Elahi, F., Daliri, E. B. M., Yeon, S. J., Ham, H. J., Kim, J. H., … Oh, D. H. (2020). Flavonoids in decorticated sorghum grains exert antioxidant, antidiabetic and antiobesity activities. *Molecules, 25*(12), Article 2854. https://doi.org/10.3390/molecules25122854

Pedro, A. M. K., & Ferreira, M. M. C. (2007). Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy. *Analytica Chimica Acta, 595*(1–2), 221–227. https://doi.org/10.1016/j.aca.2007.03.036

Peiris, K. H. S., Wu, X., Bean, S. R., Perez-Fajardo, M., Hayes, C., Yerka, M. K., … Bean, B. (2021). Near infrared spectroscopic evaluation of starch properties of diverse sorghum populations. *Processes, 9*(11), Article 1942. https://doi.org/10.3390/pr9111942

Rashwan, A. K., Yones, H. A., Karim, N., Taha, E. M., & Chen, W. (2021). Potential processing technologies for developing sorghum-based food products: An update and comprehensive review. *Trends in Food Science & Technology, 110*, 168–182. https://doi.org/10.1016/j.tifs.2021.01.087

Rinnan, Å., Berg, F. V. D., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry, 28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Sang, Y. J., Bean, S., Seib, P. A., Pedersen, J., & Shi, Y. C. (2008). Structure and functional properties of sorghum starches differing in amylose content. *Journal of Agricultural and Food Chemistry, 56*(15), 6680–6685.

Sheng, W., & Wei, C. (2022). Screening methods for cereal grains with different starch components: A mini review. *Journal of Cereal Science, 108*. https://doi.org/10.1016/j.jcs.2022.103557. Article 103557.

Simeone, M. L. F., Pimentel, M. A. G., Queiroz, V. A. V., Santos, F., Brito, A., Aquino, L. F. M., … Trindade, R. D. S. (2024). Portable near-infrared (NIR) spectroscopy and multivariate calibration for reliable quality control of maize and sorghum grain chemical composition. *Journal of Food Composition and Analysis, 134*. https://doi.org/10.1016/j.jfca.2024.106502. Article 106502.

Spielbauer, G., Armstrong, P., Baier, J. W., Allen, W. B., Richardson, K., Shen, B., & Settles, A. M. (2009). High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chemistry, 86*(5), 556–564. https://doi.org/10.1094/cchem-86-5-0556

Sullivan, A. C., Pangloli, P., & Dia, V. P. (2018). Kafirin from *Sorghum bicolor* inhibition of inflammation in THP-1 human macrophages is associated with reduction of intracellular reactive oxygen species. *Food and Chemical Toxicology, 111*, 503–510. https://doi.org/10.1016/j.fct.2017.12.002

Sun, X., Wang, Z., Aydin, H., Liu, J., Chen, Z., & Feng, S. (2023). First step for hand-held NIRS instrument field use: Table grape quality assessment consideration of temperature and sunlight chemometrics correction. *Postharvest Biology and Technology, 201*. https://doi.org/10.1016/j.postharvbio.2023.112374. Article 112374.

Syahariza, Z. A., Li, E., & Hasjim, J. (2010). Extraction and dissolution of starch from rice and sorghum grains for accurate structural analysis. *Carbohydrate Polymers, 82*(1), 14–20. https://doi.org/10.1016/j.carbpol.2010.04.014

Tester, R. F., Karkalas, J., & Qi, X. (2004). Starch-composition, fine structure and architecture. *Journal of Cereal Science, 39*(2), 151–165. https://doi.org/10.1016/j.jcs.2003.12.001

Tomar, M., Bhardwaj, R., Singh, P., Kaur, S., Singh, S. P., Dahuja, A., … Sachdev, A. (2025). From grain to Ggain: Bridging conventional methods with chemometric innovations in cereal quality analysis through near-infrared spectroscopy (NIRS). *Food Control, 178*, Article 111482. https://doi.org/10.1016/j.foodcont.2025.111482

Williams, P. C. (2001). *Implementation of near-infrared technology* (pp. 145–169). St. Paul: American Association of Cereal Chemists.

Wirnas, D., Trikoesoemaningtyas, Rini, E. P., Marwiyah, S., Sopandie, D., & Nur, A. (2024). Genetic study of amylose content and yield-related traits in sorghum germplasm. *SABRAO Journal of Breeding and Genetics, 56*(3), 951–962. https://doi.org/10.54910/sabrao2024.56.3.5

Wu, J. G., & Shi, C. H. (2004). Prediction of grain weight, brown rice weight and amylose content in single rice grains using near-infrared reflectance spectroscopy. *Field Crops Research, 87*(1), 13–21. https://doi.org/10.1016/j.fcr.2003.09.005

Yerka, M. K., Toy, J. J., Funnell-Harris, D. L., Sattler, S. E., & Pedersen, J. F. (2016). Evaluation of interallelic waxy, heterowaxy, and wild-type grain sorghum hybrids. *Crop Science, 56*(1), 113–121. https://doi.org/10.2135/cropsci2015.03.0151

Zerihun, M., Fox, G., Nega, A., Seyoum, A., Minuye, M., Jordan, D., Taddese, T., & Assefa, A. (2020). Near-infrared reflectance spectroscopy (NIRS) for tannin, starch and amylase determination in sorghum breeding programs. *International Journal of Food and Nutritional Science, 7*(1), 45–50. https://doi.org/10.15436/2377-0619.20.2716

Zheng, R., Jia, Y., Ullagaddi, C., Allen, C., Rausch, K., Singh, V., … Kamruzzaman, M. (2024). Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy. *Food Chemistry, 456*. https://doi.org/10.1016/j.foodchem.2024.140062. Article 140062.