

RESEARCH

Open Access



Analyzing protein concentration from intact wheat caryopsis using hyperspectral reflectance

Xiaomei Zhang¹, Xiaoxiang Hou¹, Yiming Su¹, XiaoBin Yan¹, Xingxing Qiao¹, Wude Yang¹, Meichen Feng¹, Huihua Kong², Zhou Zhang³, Fahad Shafiq⁴, Wenjie Han¹, Guangxin Li^{1*}, Ping Chen^{2*} and Chao Wang^{1*}

Abstract

Background Winter wheat grain samples from 185 sites across southern Shanxi region were processed and analyzed using a non-destructive approach. For this purpose, spectral data and protein content of grain and grain powder were obtained. After combining six types of preprocessed spectra and four types of multivariate statistical models, a relationship between hyperspectral datasets and grain protein is presented.

Results It was found that the hyperspectral reflectance of winter wheat grain and powder was positively correlated with the protein contents, which provide the possibility for hyperspectral quantitative assessment. The spectral characteristic bands of protein content in winter wheat extracted based on the SPA algorithm were proved to be around 350–430 nm; 851–1154 nm; 1300–1476 nm; and 1990–2050 nm. In powder samples, SG-BPNN had the best monitoring effect, with the accuracy of $R_v^2 = 0.814$, $RMSE_v = 0.024$ g/g, and $RPD_v = 2.318$. While in case of grain samples, the SG-SVM model exhibited the best monitoring effect, with the accuracy of $R_v^2 = 0.789$, $RMSE_v = 0.026$ g/g, and $RPD_v = 2.177$.

Conclusions Based on the experimental findings, we propose that a combination of spectral pretreatment and multivariate statistical modeling is helpful for the non-destructive and rapid estimation of protein content in winter wheat.

Keywords Grain protein, Hyperspectral technology, Multivariate models, Spectral preprocessing, Wheat

*Correspondence:

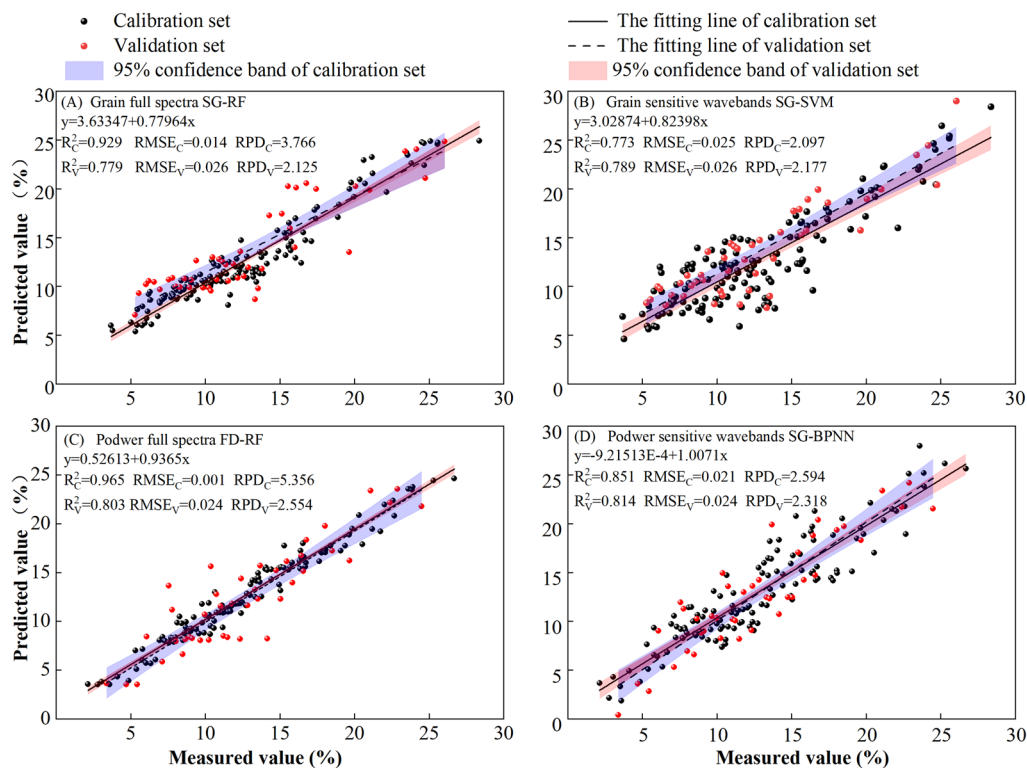
Guangxin Li
nkylgx@163.com
Ping Chen
chenping@nuc.edu.cn
Chao Wang
wcqxx2005@126.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Graphical Abstract



Introduction

Wheat is the second major food crop in the world and is a main source of energy, protein and trace minerals [1]. Its breeding, transportation, warehouse storage and quality rating are important aspects for grain quality assessments [2]. The commonly used method to measure wheat protein is the *N* method; however, conventional methods involving chemical analyses are time-consuming with a number of shortcomings [3]. Alternatively, there is a need to develop rapid quality detection techniques to meet the requirements of modern storage coupled with efficient and rapid grading [4, 5].

Hyperspectral technology could provide such information from the target sample accurately and in timely manner without the need of destructive methods. This technology has been successfully used for in crop growth monitoring; canopy nutrition diagnosis and estimation of grain starch and protein contents [6–8]. Although the spectral estimation of crop quality achieved by analyzing grain powder provide accurate estimation of grain quality [9], while the grain milling process is not resource efficient which consumes a lot of energy and time, which limits its wide application in production practice [10].

Second, hyperspectral estimation of milled grain powder is not ideal for repeated measurements for periodic quality assessments. By contrast, direct quantitative estimation of grain protein by using the hyperspectral technology might be valuable for providing real-time rapid and repeated assessments to ensure proper storage and maintenance of quality standards.

In addition, due to the influence of environmental factors and grain status during the spectrum acquisition process, the preprocessing method can overcome the influence of external factors, thereby improving the signal-to-noise ratio of spectral information and improving the modeling accuracy [11–13]. However, the principles of different modeling methods are quite different, which will also lead to different performances of the optimal model of wheat quality and content [14–16]. When the preprocessing method and the modeling method are different, there is no unified conclusion on the influence of the characteristic band on the modeling accuracy. Researcher depicted that the model based on characteristic bands can greatly simplify the complexity of model construction and further improved its prediction efficiency [17]. Therefore, it is necessary to study the

comprehensive effects of pretreatment methods combined with modeling methods, as well as full spectrum, characteristic wavelength, etc. on model accuracy to obtain the optimal combination method.

During this study, we analyzed winter wheat samples and collected grain and powder hyperspectral data to analyze grain protein contents. In addition, different pre-processing methods and multivariate modeling methods were used to elucidate response properties of winter wheat protein and hyperspectral data sets. Based on the information, we constructed hyperspectral quantitative monitoring model for the precise and real-time assessment of winter wheat grain protein.

Materials and methods

Sample collection

Samples were collected in Linfen City, Shanxi Province (36.0882° N, 111.5196° E), and Yuncheng City, Shanxi Province (35.0263° N, 111.0070° E) presented in Fig. 1.

Both areas are located in the middle reaches of the Yellow River Basin, which belongs to the typical landform area of the Loess Plateau, with an altitude of 500–1000 m, a temperate continental monsoon semi-arid climate, and an annual precipitation of about 500 mm, respectively [18]. Farmer fields across southern Shaanxi province were selected for experimentation and wheat grain

samples harvested at maturity were analyzed for grain protein content.

Spectral data acquisition

Wheat grain samples were evenly placed in a black plastic box (diameter 90 mm, height 15 mm), and the surface of the sample was kept flat. The ASD FieldSpec 3 portable spectrometer was then used to obtain the spectral data. The probe of the instrument was equipped with an analog solar light source, and spectral data was recorded in the range of 350–2500 nm. During the spectrum acquisition process, the influence of external light on the spectral information, the spectrometer probe was placed on the sample to reduce the effect of light. Each sample was then sampled using five-point method and a total of 9 spectral curves were obtained at each point corresponding to 9 spectral curves obtained for each sample. Finally, 45 spectral curves were averaged as the spectral curves of each sample.

Determination of protein contents

Determination of nitrogen content from grains was performed using Kjeldahl Method [19] and grain protein content (P) was calculated as follows:

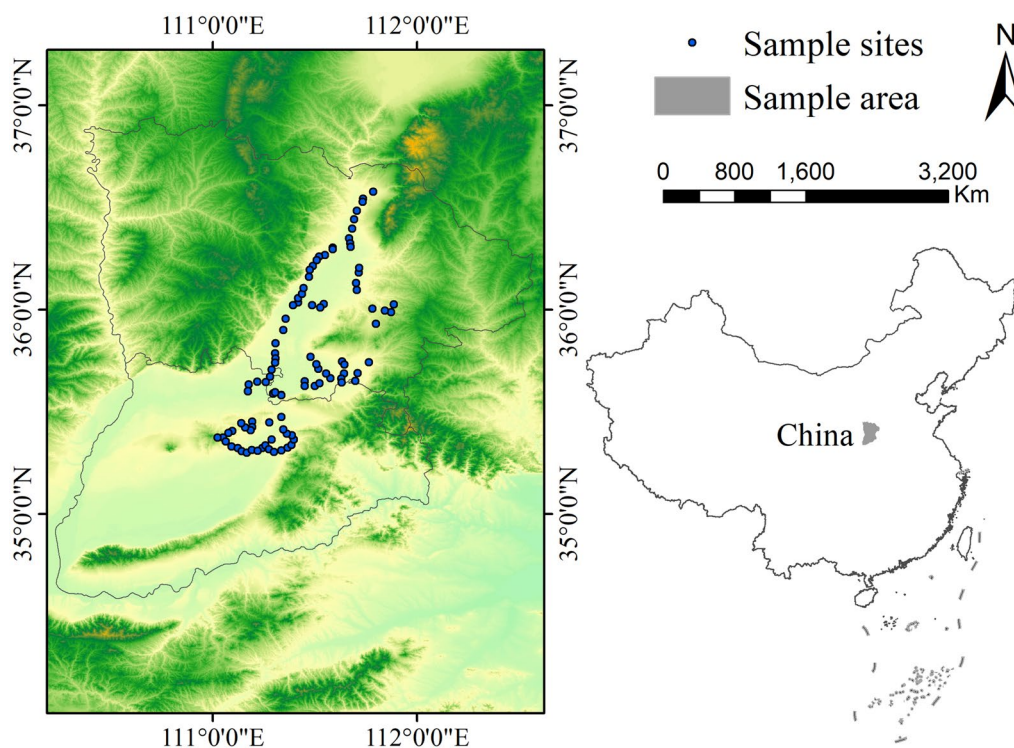


Fig. 1 Sampling sites and details of sampling point distribution across both the sites

$$P = \frac{5.7 \times C \times 100\%}{2M} \quad (1)$$

where M is the weight of the sample after crushing (g) and C is the nitrogen content in the digestate (mg/L).

Spectral data processing

Removal of outliers from the collected spectral data using View Spec Pro Version software, and average processing, using the averaged spectral data as the final data was performed. Primarily, the original spectrum was pre-processed and signal-to-noise ratio was improved using Savitzky–Golay Smoothing (SG), Derivative [First Derivative (FD), Second Derivative (SD)], Standard Normal Variate (SNV), and Multiplicative Scattering Correction (MSC); and Continuous projection algorithm (SPA).

Savitzky–Golay smoothing (SG)

Savitzky–Golay is one of the most basic and commonly used spectral preprocessing methods. The principle of Savitzky–Golay is to divide spectral data into window distributions, set a window width, and average the data in the window by moving the window to eliminate noise. Smoothing can reduce the burr in the spectral data image and make the image smoother, and this method can keep the shape and width of spectral data unchanged [20].

Derivative

The derivative transformation method can enhance the spectral differences between samples and compensate for the baseline shift caused by light scattering, thereby improving the recognition of different sample spectra [21, 22]. Among them, the first derivative mainly reduces some linear or near-linear noise in the target spectrum, while the second derivative can eliminate baseline drift and interference. Both methods can improve the model monitoring accuracy of the sample [23, 24].

Standard normal variate (SNV)

When measuring the spectrum of a sample, if there are particles on the surface of the sample or the particles are unevenly distributed, scattering will occur. Standard Normal Variate (SNV) is often used to eliminate the scattering error and the interference of particle size [20].

Multiplicative scattering correction (MSC)

The basic principle of multiplicative scattering correction (MSC) is to obtain an ideal spectrum with high correlation by replacing the average spectral curve of the sample modeling set, so as to realize the scattering correction of spectral data [25]. This method can separate chemical light absorption from physical light scattering, and can eliminate

the scattering influence caused by uneven particle distribution and different particle sizes [26].

Continuous projection algorithm (SPA)

Continuous projection algorithm (SPA) is a forward variable selection method to minimize the collinearity of vector space. It uses vector projection analysis to eliminate redundant information in the spectrum to the greatest extent, and selects some characteristic bands from the whole spectral band, which can not only reduce the number of spectral bands involved in modeling, but also ensure the minimum collinearity between characteristic bands, thus improving the modeling efficiency. The calibration set and validation set used for extracting feature bands in SPA in this article are consistent with the test set and validation set used for building the model [11, 27].

Model method and evaluation index

We used back propagation neural network (BPNN), partial least squares (PLS), random forest (RF), and support vector machine (SVM) to construct multivariate statistical models [28], whereas R^2 , RMSE, and RPD were used to evaluate the effectiveness of each model using these formulas:

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSE}} \quad (4)$$

where n is the number of samples, \hat{y}_i is the predicted value of the sample, y_i is the measured value of the sample, and \bar{y} is the average measured value of the sample, and SD is the standard deviation of the sample measured value.

Data processing software

In this study, the Unscrambler 9.7 and Excel 2019 were used to preprocess the spectral data of the collected samples, Matlab R2010a was used to extract the characteristic bands and build the model, and Oringin 2021 was used to draw the map.

Results and analysis

Descriptive statistical analysis of winter wheat protein content

As per concentration gradient method, the acquired 185 sample data were divided into a calibration set (139) and

verification sets (46) with ratio of 3:1. The descriptive statistical analysis of winter wheat grain protein content in each data set is presented in Fig. 2. The minimum and maximum values of winter wheat protein content ranged 37 to 284 mg/g averaging 127 mg/g. The range difference between the correction set and verification set was 247 and 208 mg/g, respectively. In addition, the division of sample data was also reasonable and skewness values of each sample set were less than 1, indicating that the winter wheat grain protein content data in each data set exhibited characteristics of normal distribution.

Hyperspectral response of protein content in winter wheat

Average the four spectral data to obtain the spectral reflectance of grains and powders at four protein content levels is presented in Fig. 3. The overall change trends of the spectral curves of the two types of samples were similar with obvious peaks and troughs. However, the spectral reflectance of winter wheat powder was significantly higher than that of winter wheat grain. In addition, with the increase in grain protein content, the spectral reflectance of grain and powder samples was increased thereby indicating a positive correlation between protein content and spectral reflectance. Furthermore, changes in spectral reflectance with different protein levels were obvious and grain spectral reflectance was mainly in the ranges of 750–1200 nm and 1450–2400 nm, while the spectral reflectance of powder was mainly in the range of 500–750 nm, 1450–1800 nm, and 1900–2400 nm respectively.

Correlation between grain and powder spectrum of winter wheat

The correlation between the pretreatment spectra of grain and powder and the protein content of winter wheat is presented in Fig. 4. Among them, the correlation

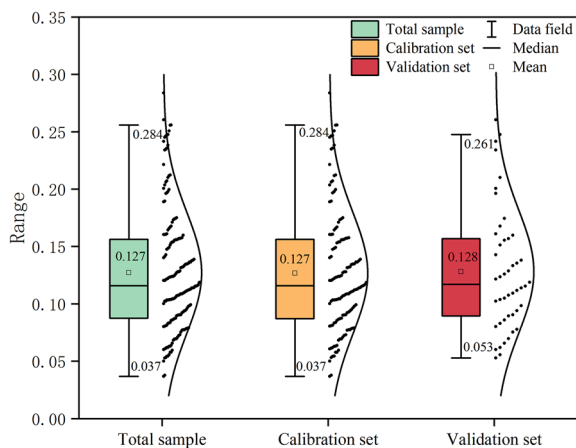


Fig. 2 Descriptive statistics of protein content in winter wheat

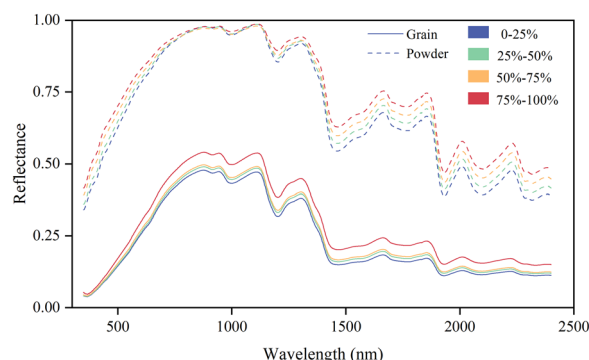


Fig. 3 Spectral reflectance of grain and powder under different protein content. The total sample was divided into four parts according to the quartile method, and each part of the data was 0–25%, 25–50%, 50–75%, and 75–100% of the total sample data, respectively

between R and protein in the grain state increased with the increase of wavelength and finally became stable. The correlation between R and the protein content of winter wheat grains in the powder state initially declined followed by an increase and became stable with the increase of wavelength. Overall, the trend of the correlation between the spectrum and the protein content of the two types of samples under SG and treatment were the same as that of the original spectrum and the protein content. In addition, the spectra of the two types of samples under SG pretreatment were positively correlated with the grain protein contents. The correlations at 1750 nm and 2000–2500 nm after SG pretreatment in the powder state were the highest. The fluctuation trend of the correlation between sample spectrum and protein content under FD pretreatment was like SD. The correlation between spectral data and protein fluctuated violently between

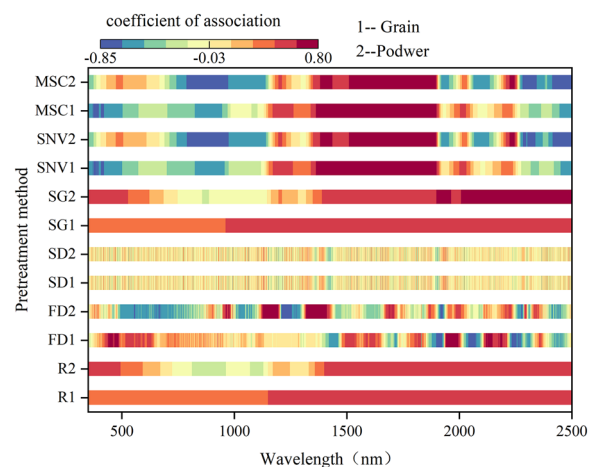


Fig. 4 Correlation analysis between the preprocessed spectral reflectance of grain, powder and protein content

positive and negative values after FD and SD transformations. Moreover, correlation of FD of grain samples increased gradually after 1400 nm, while that of powder samples decreased gradually after 1500 nm. Both spectra and protein contents of the two types of samples under SNV pretreatment were consistent with that of MSC pretreatment. In addition, powder spectra under MSC and SNV pretreatment exhibited negative correlations with protein content at 700–1200 nm and 1800–2500 nm, and positive correlations at 1300–1800 nm.

Feature band extracted of winter wheat grain proteins

We filtered characteristic bands and the best spectral bands using SPA (Table 1). Distributions of the spectral bands, characteristic bands and those obtained by these six preprocessing methods were plotted (Fig. 5). Spectral bands containing information on the protein content of the seed samples were concentrated in the ranges of 350–450 nm, 900–1160 nm, 1300–1500 nm, and 1901–2100 nm in the spectra of the seed samples, whereas spectral bands containing information on the protein content of the winter wheat seed samples were concentrated in the ranges of 330–430 nm, 550–600 nm, 1300–1400 nm and 1990–2050 nm for the spectra of powder samples. The analysis of the characteristic band distribution of the spectra of the seed samples and the spectra of the powder samples exhibited band range after 500 nm to a lesser extent than the spectra of the seed samples. Overall, there was an overlap between the two types of samples, and these overlapping bands were likely to be the common characteristic band distribution of the seed and powder samples.

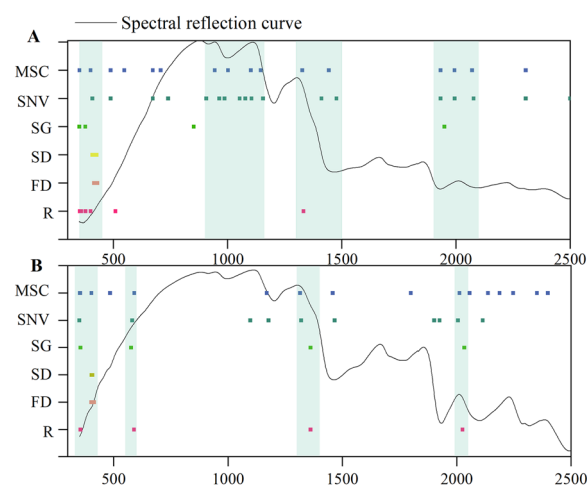


Fig. 5 Extraction and distributions for the important spectral wavelengths of grain protein based on SPA method. **A** is grain, **B** is powder. The shaded part represents the main distribution areas of characteristic bands of grain protein obtained by SPA under different soil spectral pretreatments. The solid line represents the spectral reflectance of all grain and powder samples

Hyperspectral estimation model for protein content of winter wheat seeds

The optimal model performance of different modeling methods in the quantitative estimation models of proteins in the full spectrum and characteristic bands is presented in Table 2. By comparing the values of R^2 and RMSE and RPD for the validation set of each model in the seed stage, where the optimal model in the full band was SG-RF ($R_v^2=0.779$, RMSE=0.026, RPD=2.125) and the optimal model in the characteristic band were SG-SVM

Table 1 Extraction of important bands of winter wheat kernel protein-based SPA method

Spectrum types	Pretreatment methods	Number of variables	Characteristics of the band (nm)
Wavelengths for grain samples	R	6	377, 358, 508, 352, 399, 1332
	FD	8	421, 420, 429, 418, 414, 415, 409, 422
	SD	4	426, 425, 407, 424
	SG	4	1949, 851, 350, 376
	SNV	18	1411, 1052, 1476, 1154, 2305, 1104, 1932, 1994, 2500, 963, 986, 487, 407, 906, 672, 1077, 739, 2077
	MSC	16	706, 1327, 943, 1993, 1932, 1144, 672, 2070, 2305, 1001, 351, 399, 1443, 547, 487, 1101
Wavelengths for powder sample	R	4	2025, 589, 1361, 355
	FD	4	402, 407, 412, 415
	SD	4	403, 404, 405, 407
	SG	4	576, 2033, 355, 1361
	SNV	10	1925, 2005, 1901, 2113, 1097, 1320, 581, 1177, 350, 1466
	MSC	16	2398, 2350, 1799, 2136, 1169, 1458, 403, 590, 352, 2012, 2056, 1315, 484, 2246, 2187, 354

($R_v^2=0.789$, $RMSE=0.026$, $RPD=2.177$). whereas the optimal model in the powder state in the full band was RF-FD and the optimal model in the characteristic band was SG-BPNN, comparing the values of R^2 and RMSE and RPD for the validation sets of SG-RF, SG-SVM, FD-RF, SG-BPNN. The values of R^2 and RMSE and RPD of the sets, where SG-BPNN ($R_v^2=0.814$, $RMSE=0.024$, $RPD=2.318$) exhibited highest model accuracy. To facilitate a more visual understanding, a 1:1 fit of the measured and predicted values of winter wheat seed protein content for the four models, seed full band SG-RF, seed characteristic band SG-SVM, powder full band FD-RF, and powder characteristic band SG-BPNN has been presented (Fig. 6).

Discussions

In this study, the protein content of winter wheat grain had obvious absorption peaks at 851, 1443, 1458, 1476, and 2246 nm. The original spectrum of grain was positively correlated with protein content, and the correlation was the strongest at 851–1476 nm, while the original spectrum of powder was negatively correlated with protein content at 714–1154 nm and positively correlated with protein content at 1407–2500 nm. The characteristic spectra of wheat grain protein extracted by SPA were mainly distributed in 350–450 nm, 900–1160 nm, 1300–1500 nm, and 1901–2100 nm. The characteristic spectra of wheat grain powder protein contents were mainly distributed in 330–430 nm, 550–600 nm, 1300–1400 nm, and 1990–2050 nm. Considering the fact that the basic

unit of protein is an amino acid, which is mainly composed of C, H, O, and N; information provide by hyperspectral reflectance mainly comes from the frequency doubling absorption of C–H, O–H, and N–H groups, of which about 800 nm was related to the third harmonic generation of C–H and N–H [29]. In addition, values between 1200 and 1500 nm may be related to C–H triple frequency and O–H stretching vibration; around 2000 nm was combined with N–H stretching vibration frequency absorption [30, 31]. In summary, the spectral regions of 350–430 nm, 851–1154 nm, 1300–1476 nm, and 1990–2050 nm were closely related to winter wheat protein.

We found that the hyperspectral reflectance curves of the powder samples were built with higher model accuracy compared to the seed samples, and the correlation between the powder spectral data and protein content was higher. The reason for this was that the seed and powder samples have different particle sizes, and thus the spectral reflectance is very different, and the protein content is measured by the powder [10, 32]; therefore, the correlation and model accuracy was higher than that of the seeds. However, powder samples can damage seeds and seed coats when the model for powder samples was of similar accuracy to that for seed samples, it was more practical to choose seeds. In this study, the difference in R^2 between the model validation set of SG-BPNN in the powder state and SG-SVM in the seed state was 0.025 and considering the practical value, it was also feasible to choose the model in the seed state for monitoring the

Table 2 Estimated model performance of winter wheat grain protein under different spectral pretreatment

Sample types	Spectrum types	Models	Pretreatment methods	Calibration set			Validation set		
				R^2	RMSE	RPD	R^2	RMSE	RPD
Grain	Full spectra	BPNN	FD	0.796	0.027	2.215	0.739	0.034	1.956
		PLS	FD	0.832	0.022	2.437	0.764	0.026	2.058
		RF	SG	0.929	0.014	3.766	0.779	0.026	2.125
		SVM	FD	0.783	0.025	2.149	0.778	0.026	2.123
	Sensitive wavebands	BPNN	MSC	0.777	0.025	2.119	0.671	0.032	1.743
		PLS	SNV	0.754	0.027	2.016	0.714	0.029	1.871
		RF	SG	0.848	0.021	2.567	0.746	0.028	1.983
		SVM	SG	0.773	0.025	2.097	0.789	0.026	2.177
Powder	Full spectra	BPNN	SG	0.881	0.018	2.894	0.735	0.027	1.941
		PLS	SNV	0.901	1.643	3.172	0.727	0.027	1.915
		RF	FD	0.965	0.001	5.356	0.803	0.024	2.254
		SVM	FD	0.87	0.019	2.77	0.766	0.024	2.068
	Sensitive wavebands	BPNN	SG	0.851	0.021	2.594	0.814	0.024	2.318
		PLS	MSC	0.873	0.019	2.805	0.797	0.023	2.218
		RF	FD	0.935	0.013	3.923	0.777	0.025	2.118
		SVM	FD	0.864	0.019	2.71	0.789	0.023	2.176

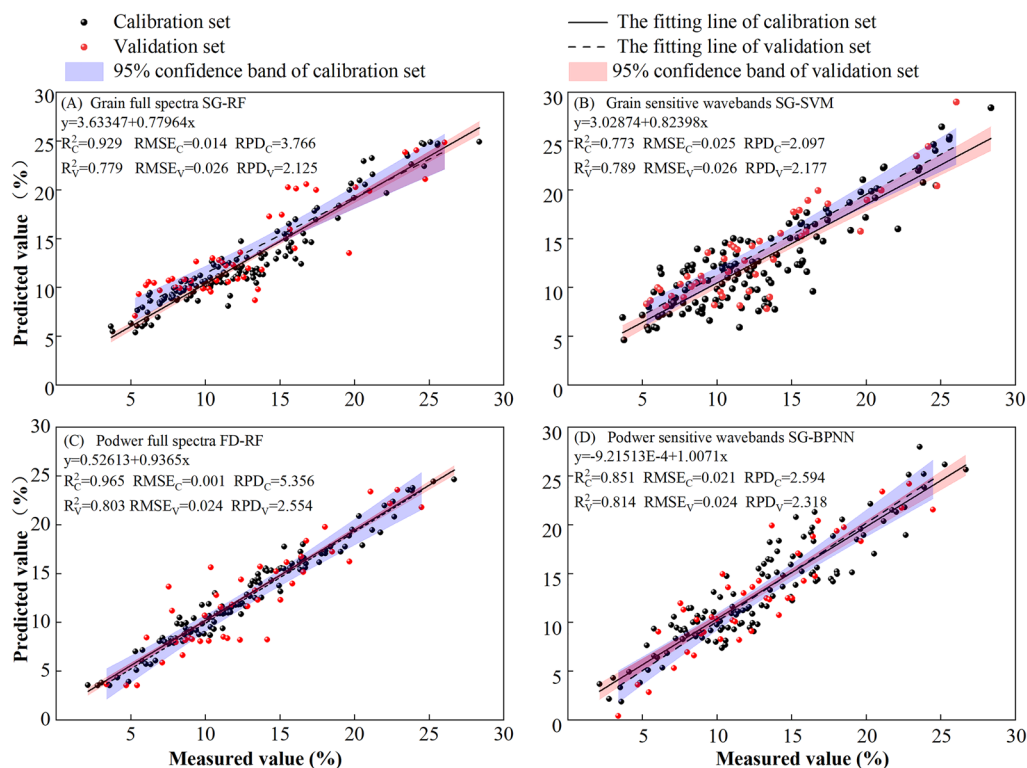


Fig. 6 The measured and predicted values of the optimal estimation model for grain protein content in winter wheat were fitted 1:1

protein content of wheat. Therefore, the potential mechanism between the accuracy of the spectral prediction of the sample treatment on quality and the explanation of such subtle differences remains to be further investigated.

In addition, when performing the correlation analysis between pretreatment methods and protein content, it was found that the highest correlation was FD [33]. It is reported that FD can remove linear and near-linear components in the original spectrum, highlighting the increase and deceleration rate of spectral reflectance. It can also capture the inflection point and extreme point of the original spectral curve, and accurately locate the peak valley characteristics of protein absorption in the spectral curve [34]. In addition, FD can also separate the absorption characteristics and change trends of the protein spectrum in the infrared region, achieving better prediction results than the original spectrum [35, 36]. However, the optimal model in this paper is SG-BPNN based on characteristic bands in the powder state, and the pre-processing method used is SG instead of FD. In addition, the influence of seed coat on spectral information is greatly reduced in the powder state [37–39]. In addition, BPNN has a strong nonlinear fitting ability, which can effectively analyze and use rich data sets to simulate the complex relationship of the internal mechanism of variables, greatly improving the accuracy of the model [40,

41]. The featured band training model will not train invalid information, which improves the recognition accuracy [42–45]. Therefore, the combination of sample state, band selection, and model algorithm may have a great impact on the accuracy of the model.

Conclusion

We established a quantitative estimation model for the protein content of winter wheat grains based on spectral preprocessing methods such as SG, FD, SD, SNV, and MSC, combined with BPNN, PLS, RF, and SVM. The hyperspectral and protein content of both types of samples of winter wheat seeds and seed powder were positively correlated, in which the spectral regions 350–430 nm, 851–1154 nm, 1300–1476 nm, and 1990–2050 nm bands were closely related to winter wheat protein. Among the various hyperspectral estimation models constructed for the protein content of winter wheat, SG-BPNN based on the spectral feature bands of powders was the best ($R_c^2=0.851$, $RMSE_c=0.021$, $RPD_c=2.594$; $R_v^2=0.814$, $RMSE_v=0.024$, $RPD_v=2.318$). However, the SG-SVM in the seed feature band was of more practical value ($R_c^2=0.073$, $RMSE_c=0.025$, $RPD_c=2.097$; $R_v^2=0.789$, $RMSE_v=0.026$, $RPD_v=2.177$). We found that the protein content of winter wheat seeds could be effectively

monitored using hyperspectral techniques in combination with different spectral preprocessing methods affected the accuracy of the quantitative estimation models. In short, this study provides a reference for future use of hyperspectral techniques for accurate and rapid estimation of wheat grain protein and provides technical support for future studies.

Acknowledgements

We thank the Scientific Experimental Observatory for Crop Cultivation and Arable Land Conservation in the Loess Plateau Region of North China, Ministry of Agriculture, College of Agriculture, Shanxi Agricultural University for help with experimental materials.

Author contributions

XBY and CW: writing—review and editing. XMZ: writing—original draft preparation. XMZ, GXL ZZ, YMS and XXH: thesis Translation. WJH and XMZ: data curation, FS and WHK: formal analysis. GXL, CW, PC and XXQ: supervision, writing—review and editing. WDY, PC and MCF: project administration. All the authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (31371572, 31871571); Basic research program of Shanxi Province (202203021211275), and the Opening Foundation of Shanxi Key Laboratory of Signal Capturing & Processing (2022-006); Supported by the earmarked fund for Modern Agro-industry Technology Research System (2023CJYJTX02-23); Scientific and Technological Innovation Fund of Shanxi Agricultural University, (2018YJ172020BQ32).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹College of Agriculture, Shanxi Agricultural University, Taigu, Jinzhong, Shanxi, China. ²Shanxi Key Laboratory of Signal Capturing & Processing, North University of China, Taiyuan, Shanxi, China. ³Biological Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA. ⁴Department of Botany, Government College University Lahore, Lahore, Pakistan.

Received: 3 May 2023 Accepted: 10 August 2023

Published online: 28 August 2023

References

- Han J, Zhang Z, Cao J, Luo Y, Zhang L, Li Z, et al. Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sens-Basel*. 2020;12(2):236.
- McGovern CM, Snyders F, Muller N, Botes W, Fox G, Manley M. A review of triticale uses and the effect of growth environment on grain quality. *J Sci Food Agr*. 2011;91(7):1155–65.
- Bao Y, Mi C, Wu N, Liu F, He Y. Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics. *Appl Sci*. 2019;9(19):4119.
- Cai J-H. Near-infrared spectrum detection of wheat gluten protein content based on a combined filtering method. *J AOAC Int*. 2017;100(5):1565–8.
- Kim J-E, Baik B-K, Park Chul S, Son J-H, Choi C-H, Mo Y, et al. Relationship between physicochemical characteristics of Korean wheat flour and quality attributes of steamed bread. *J Integr Agr*. 2019;18(11):2652–63.
- Fan L, Fang W, Guiping L, Xin L, Jiayi Y. Prediction of oleic acid content of rapeseed using hyperspectral technique. *Appl Sci*. 2021;11(12):5726.
- Lantao L, Luyi C, Yanru J, Ding Q, Shuyu F, Xinyue F, et al. Quantification and dynamic monitoring of nitrogen utilization efficiency in summer maize with hyperspectral technique considering a non-uniform vertical distribution at whole growth stage. *Field Crop Res*. 2022;281:108490.
- Xie X, Li YX, Li R, Zhang Y, Huo Y, Bao Y, et al. Hyperspectral characteristics and growth monitoring of rice (*Oryza sativa*) under asymmetric warming. *Int J Remote Sens*. 2013;34(23):8449–62.
- Caporaso N, Whitworth MB, Fisk ID. Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Appl Spectrosc Rev*. 2018;53(8):667–87.
- Hu N, Li W, Du C, Zhang Z, Gao Y, Sun Z, et al. Predicting micronutrients of wheat using hyperspectral imaging. *Food Chem*. 2021;343:128473.
- Lee LC, Liang C-Y, Jemain AA. Effects of data pre-processing methods on classification of ATR-FTIR spectra of pen inks using partial least squares-discriminant analysis (PLS-DA). *Chemometr Intell Lab Syst*. 2018;182:90–100.
- Xu L, Xie D, Fan F. Effects of pretreatment methods and bands selection on soil nutrient hyperspectral evaluation. *Procedia Environ Sci*. 2011;10:2420–5.
- Xitong X, Shengbo C, Zhengyuan X, Yan Y, Sen Z, Rui D. Exploring appropriate preprocessing techniques for hyperspectral soil organic matter content estimation in black soil area. *Remote Sens-Basel*. 2020;12(22):3765.
- Zheli W, Xi T, Shuxiang F, Chi Z, Jiangbo L. Maturity determination of single maize seed by using near-infrared hyperspectral imaging coupled with comparative analysis of multiple classification models. *Infrared Phys Technol*. 2021;112:103596.
- Baohua Y, Jifeng M, Xia Y, Weixing C, Yan Z. Estimation of leaf nitrogen content in wheat based on fusion of spectral features and deep features from near infrared hyperspectral imagery. *Sens Basel*. 2021;21(2):613.
- Baichuan J, Chu Z, Liangquan J, Qizhe T, Lu G, Guangwu Z, et al. Identification of rice seed varieties based on near-infrared hyperspectral imaging technology combined with deep learning. *ACS Omega*. 2022;7(6):4735–49.
- Ziran Y, Yin Y, Lifei W, Xin Y, Can H. Study on the optimization of hyperspectral characteristic bands combined with monitoring and visualization of pepper leaf SPAD value. *Sens Basel*. 2021;22(1):183.
- Ding Z, Yao S. Theory and valuation of cross-regional ecological compensation for cultivated land: A case study of Shanxi province, China. *Ecol Indic*. 2022;136:108609.
- Salo-väänänen PP, Koivistoinen EP. Determination of protein in foods: comparison of net protein and crude protein (Nx 6.25) values. *Food Chem*. 1996;57(1):27–31.
- Shuang L, Ye YH, Jia PZ, Chen CM, Tong Y, Juan KL, et al. Study on extracting characteristic wavelength of soybean physiological information based on hyperspectral technique. *Spectrosc Spect Anal*. 2020;40(11):3542–8.
- Vašát R, Kodešová R, Klement A, Borůvka L. Simple but efficient signal pre-processing in soil organic carbon spectroscopic estimation. *Geoderma*. 2017;298:46–53.
- Sun X, Subedi P, Walker R, Walsh KB. NIRS prediction of dry matter content of single olive fruit with consideration of variable sorting for normalisation pre-treatment. *Postharvest Biol Technol*. 2020;163:111140.
- Dehua G, Lang Q, Lulu A, Hong S, MinZan L, Ruomei Z, et al. Diagnosis of maize chlorophyll content based on hybrid preprocessing and wavelengths optimization. *Comput Electron Agric*. 2022;197:106934.
- Bin-Bin G, Ya-Lan F, Chao M, Jun Z, Xiao S, Meng-Yuan W, et al. Suitability of different multivariate analysis methods for monitoring leaf N accumulation in winter wheat using in situ hyperspectral data. *Comput Electron Agric*. 2022;198:107115.
- Yin S, Chunjiang Z, Bin L, Guanglin L, Yanxin Y, Binshuang P. Determination of wheat moisture using terahertz spectroscopy combined with the tabu search algorithm. *Anal Methods Adv Methods Appl*. 2021;13(36):4120–30.

26. Ting-ting Z, Ying-ying X, Li-ming Y, Jian-hua W, Qun S. Wavelength Variable Selection Methods for Non-Destructive Detection of the Viability of Single Wheat Kernel Based on Hyperspectral Imaging. *Spectroscopy And Spectral Analysis*. 2018;39(05):1556–62.
27. Lei T, Lin X-H, Sun D-W. Rapid classification of commercial Cheddar cheeses from different brands using PLS-DA, LDA and SPA-LDA models built by hyperspectral data. *J Food Meas Charact*. 2019;13(4):3119–29.
28. Abdi H. Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdiscip Rev Comput Stat*. 2010;2(1):97–106.
29. Pu W, Ping W, HanWei W, JiXin C. Hyperspectral vibrational photoacoustic imaging of lipids and collagen. *Proc Spie*. 2012;8223:45. <https://doi.org/10.1117/12.908714>.
30. Xiuye L, Mingguang Z, Xiting S, Chunfeng S, Hongfu Y, Longhai G, et al. Online real time determination of free formaldehyde content during polymerization process of phenolic resin by NIR spectra and a free modeling method. *Polym Test*. 2020;93:106584.
31. Wang X, Bao Y, Liu G, Li G, Lin L. Study on the best analysis spectral section of NIR to detect alcohol concentration based on SiPLS. *Procedia Eng*. 2012;29:2285–90.
32. Apan A, Kelly R, Phinn S, Strong W, Lester D, Butler D, et al. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. *Int J Geoinform*. 2006;2(1):93–108.
33. Shi H, Yu P. Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. *Food Control*. 2017;82:57–65.
34. Chen S, Hu T, Luo L, He Q, Zhang S, Li M, et al. Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods. *Infrared Phys Technol*. 2020;111:103542.
35. Jin X, Xiao Z-Y, Xiao D-X, Dong A, Nie Q-X, Wang Y-N, et al. Quantitative inversion model of protein and fat content in milk based on hyperspectral techniques. *Int Dairy J*. 2022;134:105467.
36. Zhu S, Chao M, Zhang J, Xu X, Song P, Zhang J, et al. Identification of soybean seed varieties based on hyperspectral imaging technology. *Sens Basel*. 2019;19(23):5225.
37. Liu Y, Gao L, Xiao C, Qu Y, Zheng K, Marinoni A. Hyperspectral image classification based on a shuffled group convolutional neural network with transfer learning. *Remote Sens-Basel*. 2020;12(11):1780.
38. Hou L, Li X, Li F. Hyperspectral-based inversion of heavy metal content in the soil of coal mining areas. *J Environ Qual*. 2019;48(1):57–63.
39. Li J-M, Ye X-J, Wang Q-N, Zhang C, He Y. Development of prediction models for determining N content in citrus leaves based on hyperspectral imaging technology. *Spectrosc Spect Anal*. 2014;34(1):212–6.
40. Lei Y, Hu X, Tian J, Zhang J, Yan S, Xue Q, et al. Rapid resolution of types and proportions of broken grains using hyperspectral imaging and optimisation algorithm. *J Cereal Sci*. 2022;108:103565.
41. Jie D, Wu S, Wang P, Li Y, Ye D, Wei X. Research on *Citrus grandis* granulation determination based on hyperspectral imaging through deep learning. *Food Anal Methods*. 2021;14(2):280–9.
42. Wan G, Liu G, He J, Luo R, Cheng L, Ma C. Feature wavelength selection and model development for rapid determination of myoglobin content in nitrite-cured mutton using hyperspectral imaging. *J Food Eng*. 2020;287:110090.
43. Sun J, Yang W, Zhang M, Feng M, Xiao L, Ding G. Estimation of water content in corn leaves using hyperspectral data based on fractional order Savitzky-Golay derivation coupled with wavelength selection. *Comput Electron Agric*. 2021;182:105989.
44. Verdú S, Vásquez F, Grau R, Ivorra E, Sánchez AJ, Barat JM. Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: the specific cases of flour and bread. *Food Control*. 2016;62:373–80.
45. Rivard B, Feng J, Gallie A, Sanchez-Azofeifa A. Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sens Environ*. 2008;112(6):2850–62.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
