MOISTURE CONTENT DETECTION OF SOYBEAN GRAINS BASED ON HYPERSPECTRAL IMAGING
/
基于高光谱成像的大豆籽粒含水率检测研究

Zhichang CHANG [1)], Man CHEN [1,2*)], Gong CHENG [1)], Chengqian JIN [1*)], Tengxiang YANG [1)]

[1)] Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing, Jiangsu / China;

[2)] National Digital Agriculture Equipment (South China Intelligence Agricultural Machine) Innovation Sub-center, Nanjing, Jiangsu / China;

Tel: +8602584346113; E-mail: chm_world@163.com; jinchengqian@126.com

DOI: https://doi.org/10.35633/inmateh-74-50

Keywords: *Soybean; Moisture content; Hyperspectral imaging; Detection*

ABSTRACT

Using hyperspectral imaging technology for rapid, non-destructive detection of soybean grain moisture content provides technical support for high-quality soybean harvesting. A total of 90 samples of soybean grains from different varieties were collected, with hyperspectral images acquired in the wavelength range of 900–1700 nm. The moisture content of each soybean grain sample was determined using the direct drying method as specified in GB 5009.3-2016. The samples were divided into a calibration set and a prediction set based on a 4:1 ratio using the sample partitioning method of Joint X-Y Distance. Eight preprocessing methods were applied to the raw spectral data, including baseline correction, moving average, Savitzky-Golay filtering, normalization, standard normal variate transformation, multiple scatter correction, first derivative, and deconvolution. Feature wavelengths were then extracted using the successive projections algorithm and the competitive adaptive reweighted sampling algorithm. Finally, a partial least squares regression model for predicting the moisture content of soybean grains was developed based on these feature wavelengths. The results show that the correlation coefficient and the root mean square error of the optimal model for the prediction set were 0.92 and 0.2371, respectively. The moisture spectrum inversion model can precisely and rapidly predict the moisture content of soybean grains non-destructively, thereby determining the timing of mechanical soybean harvesting and enhancing the quality of soybean harvesting, storage, and processing.

摘要

采用高光谱成像技术实现大豆籽粒含水率的快速无损检测，为大豆高质量收获提供技术支撑。采集了 90 个不同品种大豆籽粒样本在 900 ~1700 nm 的高光谱图像，采用 GB 5009.3-2016 中的直接干燥法测定每种大豆籽粒样本的水分含量。基于联合 X-Y 距离的样本划分法按照 4:1 的比例划分样品，建立校正集和预测集。采用基线校正、移动平均、Savitzky-Golay 滤波、归一化、标准正态变量变换、多元散射校正法、一阶导数、去卷积 8 种算法方法对原始光谱数据进行预处理，基于连续投影算法和竞争性自适应重加权算法提取特征波长，最后建立基于特征波长的偏最小二乘回归的大豆籽粒含水率预测模型。结果表明，最优模型预测集相关系数和均方根误差分别为 0.92 和 0.2371。水分光谱反演模型可以准确快速无损预测大豆籽粒含水率，从而制定大豆机收时间，提升大豆收获、存储、加工品质。

INTRODUCTION

China is one of the world's major producers and consumers of soybeans *(Li et al., 2023)*. The country's soybean planting area is approximately 9 to 10 million hectares, with an annual production of around 18 to 20 million tons *(Liu et al., 2022)*. Soybean cultivation in China is mainly concentrated in the Northeast region and the Huang-Huai-Hai Plain. In recent years, there has been a shift in China's soybean production model, moving from traditional, small-scale farming to a more large-scale, mechanized, and intelligent approach *(Li et al., 2023)*.

Moisture detection in soybeans is of great significance in mechanized production, as moisture content directly affects the efficiency and quality of various stages such as harvesting, transportation, storage, and processing. Moisture content is a key factor in determining the optimal harvest time for soybeans *(Mo et al., 2020)*.

Zhichang Chang, Stud. Eng.; Man Chen, Associate Prof. Ph.D. Eng.; Gong Cheng, Stud. Eng.; Chengqian Jin, Prof. Ph.D. Eng.; Tengxiang Yang, Research Associate. Ph.D. Eng.

When the moisture content is too high, harvesting machinery may face operational difficulties, which can cause machine blockages, reduced harvesting efficiency, and increased risks of grain breakage or damage. On the other hand, if the moisture content is too low, soybean pods are more likely to crack, leading to increased field losses. By monitoring the moisture content of soybeans in real time, farmers and machine operators can select the best harvest time, minimize losses, and improve harvesting efficiency.

Traditional methods for detecting soybean moisture content include the drying method and the Karl Fischer titration method. These methods are recognized as standard due to their high detection accuracy and reliable results. However, they are time-consuming, involve complex procedures, and require specialized equipment and laboratory conditions, making them unsuitable for on-site, real-time detection. In recent years, with the rapid advancement of sensing technology, online detection techniques for soybean moisture content have developed quickly *(Zhang et al., 2023)*. These techniques include methods such as capacitance *(Zhang et al., 2021)*, resistance *(Zhao et al., 2018)*, near-infrared spectroscopy *(Chen et al., 2023)*, microwave measurement *(Song et al., 2023)*, infrared thermal radiation *(Shi et al., 2024)*, and machine vision and imaging technology *(Wu et al., 2022)*. These methods offer advantages like fast detection speed, high accuracy, and the ability to perform non-destructive, real-time measurements, making them widely used in modern agriculture. A rapid detection method for tomato water stress based on terahertz spectroscopy was proposed, integrating three-dimensional terahertz features and using support vector machines for the rapid detection of tomato water stress (*Zhang et al., 2021*). A spectral classification model was developed to distinguish the moisture levels of different alfalfa seeds (*Yang et al., 2024*). A method based on near-infrared spectroscopy and random forest variable selection was applied to identify soybean seed cracks, achieving an accuracy rate of 84% (*Wang et al., 2023*). A microwave radar measurement system was employed for fast, non-destructive, and high-precision detection of grain moisture content, with standard errors of 0.564% and 0.983% for predicting moisture content in soybeans and wheat, respectively (*Li et al., 2022*). A portable near-infrared and visible light spectrometer, combined with multivariate analysis, was used for non-destructive detection of rice moisture content, and the root mean square error of the optimal model was 0.388 (*Zhang et al., 2023*). Additionally, a method based on terahertz imaging technology was applied to measure the moisture content of soybean leaves, with a root mean square error of 0.0465 (*Bu et al., 2018*). Most existing technologies focused on a single variety, and the model's generalization was weak, making it unable to accurately detect parameters of multiple varieties. Given the wide variety of soybean cultivars in China, constructing a robust moisture detection model suitable for multiple cultivars is the focus of this study.

Therefore, this study focuses on 90 different soybean cultivars and utilizes hyperspectral imaging technology combined with chemometric methods to establish and identify the optimal prediction model for detecting soybean moisture content in the range of 900–1700 nm. This approach provides a new method for monitoring moisture content during soybean harvesting, storage, and processing.

MATERIALS AND METHODS

Sample Collection for the Experiment

In order to improve the generalization of the model, the samples used in the experiment were collected from major soybean-producing regions in China, including the Northeast Plain, Huang-Huai Plain, and Yangtze River Basin. These samples included 90 different varieties such as Heihe 43, Suinong 28, Zhonghuang 301, Wanhuang 506, Zhonghuang 37, Zhongdou 43, Zhongdou 41, Nandou 5, and Xiangchun Bean 26. During the soybean harvest season, the samples were mechanically harvested, cleaned to remove impurities, and then sealed in bags. They were labeled and brought back to the laboratory for further analysis.

Determination of Moisture Content in Experimental Samples

The moisture content of each soybean sample was measured using the direct drying method as outlined in GB 5009.3-2016 "National Food Safety Standard for the Determination of Moisture in Foods.", as shown in Fig.1.

The calculation formula for moisture content is as follows:

$$X = \frac{m_1 - m_2}{m_1 - m_3} \times 100 \quad (1)$$

where: X is moisture content in the sample, expressed as g/100 g;

m_1 is mass of the weighing bottle (with sea sand and glass rod) and the sample, in grams (g);

m_2 is mass of the weighing bottle (with sea sand and glass rod) and the sample after drying, in grams (g);

m_3 is mass of the weighing bottle (with sea sand and glass rod), in grams (g);

100 is conversion factor to express moisture content as a percentage.

When the moisture content is ⩾1 g/100 g, the result is reported with three significant figures. When the moisture content is <1 g/100 g, the result is reported with two significant figures *(Zhang et al., 2023)*. Each soybean variety was measured three times, and the average value was used as the moisture content for that variety.

a) Sample drying ***b) Sample weighing***

Fig. 1 – Measurement of moisture content in soybean samples

Spectral Measurement of Experimental Samples

The hyperspectral data acquisition system for soybean samples consists of three main components: a near-infrared hyperspectral imaging system, a halogen lamp light source, and a motion control system, as illustrated in Figure 2. This system uses the Pika NIR-320 Hyperspectral Imaging Camera from Resonon Inc., USA. It features a spectral range of 900–1700 nm, a spectral resolution of 8.8 nm, a spectral bandwidth of 4.8 nm, and 168 spectral channels. The system employs a halogen lamp to provide the necessary illumination for hyperspectral imaging across the full wavelength range. It includes a finned heat dissipation structure, complemented by a high-performance cooling fan, and is equipped with a voltage and current stabilizer to enhance the stability and reliability of the light source. This system integrates the light source and sensor into a single unit, allowing for high-precision synchronization between the sensor and the linear light source. It also facilitates easy adjustment of the camera's relative position to the target sample, accommodating the needs of capturing diverse and complex samples. This setup ensures comprehensive and accurate hyperspectral imaging by providing stable illumination, precise spectral measurements, and flexible positioning for various sample types.

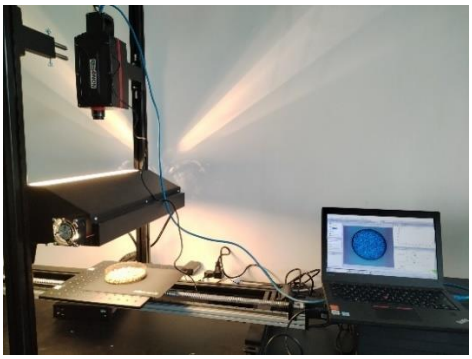


Fig. 2 – Hyperspectral data acquisition system

To minimize the impact of dark current and uneven light source intensity on hyperspectral images, black and white calibration of the hyperspectral images is required *(Jin et al., 2022; Guo et al., 2023)*. Place a white reference board with a reflectance of 0.99 in the sample acquisition area *(Chen et al., 2019)*. Capture an image of this white board to obtain the white calibration image *(Chen et al., 2019)*, denoted as I_{w}. Cover the hyperspectral camera lens and capture an image to obtain the black calibration image, denoted as I_{d}.

The calibration formula for correcting hyperspectral images can be expressed as follows:

$$I_t = \frac{I - I_d}{I_w - I_d} \quad (2)$$

where, I_{t} is calibrated hyperspectral image, I is raw hyperspectral image of the sample.

During the sample scanning process, the exposure time was set to 7.3 ms, and the movement speed of the displacement platform was 3.512 cm/s. The hyperspectral data were analyzed using the SpectrononPro software that comes with the Pika NIR-320 hyperspectral imaging system.

For each hyperspectral image, a region of interest (ROI) with a side length of 200 pixels was selected. The average spectrum of this ROI was extracted and used as the spectral information for the soybean sample. This approach ensures that the spectral data accurately represent the sample within the defined area of interest.

Spectral Data Preprocessing

During the spectral data acquisition with the Pika NIR-320 hyperspectral imaging system, due to differences in particle size and absorbance of the sample, variations occur during the diffuse reflection process, resulting in decreased spectral repeatability. Derivative is the most commonly used method to eliminate additive effects. Standard normal variate transformation (SNV) and multiple scatter correction (MSC) are commonly used methods to eliminate multiplication effects, which can to some extent reduce the impact of solid particle size or scattering effects on spectral data. Orthogonal signal correction (OSC) and net analyte signal (NAS) are non-independent methods of response variables, mainly used to remove spectral variations that are unrelated to the response variable, with the aim of simplifying the analysis model and improving its predictive ability. Therefore, baseline correction (BC), moving average, Savitzky-Golay filtering, normalization, SNV, MSC, first derivative, and deconvolution methods were used to preprocess the raw spectral data and remove irrelevant information.

Moisture Content Estimation Based on Spectral Information

Using the sample set partitioning based on joint X-Y distance (SPXY) method, the soybean samples were divided into a calibration set and a prediction set in a 4:1 ratio.

The raw spectral data contain a large amount of redundant and useless spectral information as well as various noises, which can reduce computation speed and affect the accuracy and precision of the model predictions. Therefore, the continuous projection algorithm (SPA) and competitive adaptive reweighted sampling (CARS) were used for dimensionality reduction and to select feature wavelengths.

Partial least squares regression (PLSR) was employed to establish the mathematical relationship between spectral information and soybean moisture content. PLSR constructs a linear regression model that indirectly describes the relationship between independent and dependent variables, effectively addressing the issue of high linear correlation among spectral variables.

The performance of the soybean moisture content PLSR model was evaluated using the coefficient of determination (R_c^2 , R_{cv}^2 and R_p^2) and the root mean square error (R_{msec} , R_{msecv} and R_{msep}) for the calibration set, cross-validation set, and prediction set. Generally, a well-performing model should have a high coefficient of determination and a low root mean square error.

Data analysis

The results were processed by Microsoft Office Excel (version 2021, Microsoft Corp., USA) and Matlab (version 2021a, MathWorks Corp., USA). Single factor analysis of variance (ANOVA) was carried out with IBM SPSS Statistics (version 24, IBM Corp., USA).

RESULTS

Spectral Feature Analysis of Soybean Samples with Different Moisture Contents

Figure 3 shows the spectral reflectance curves of soybean samples. The distinct reflectance dip near 1210 nm is attributed to the second overtone vibration band of C-H bonds in organic compounds.

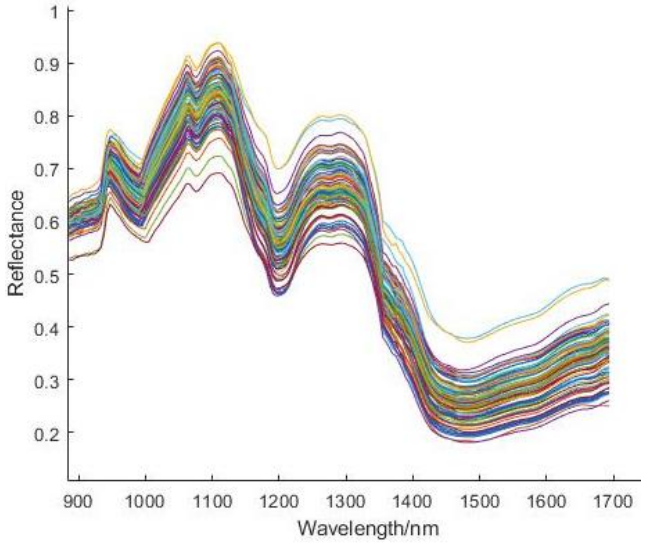


Fig. 3 – Original spectral curves of near infrared of 90 soybean samples

The reflectance dip at 1450 nm is related to the overtone vibration of O-H bonds, both of which are characteristic spectral bands for moisture content. The raw spectrum contains background information and noise, and the spectral data before 950 nm and after 1650 nm do not provide useful sample information. Therefore, spectral bands from 950 to 1650 nm, totaling 145 bands, are retained for modeling.

Sample Set Partitioning

The SPXY method divided the 90 samples into calibration and prediction sets in a 4:1 ratio, resulting in 72 samples for calibration and 18 samples for prediction. The moisture content range in the soybean samples is shown in Table 1. The calibration set covers the full range of moisture content present in the prediction set, indicating that the sample set partitioning is scientific and reasonable.

Table 1

Statistical data of soybean water content in sample set divided by SPXY method

Sample Set	*Number of Samples*	Moisture Content			
		Maximum Value	Minimum Value	Average Value	Standard Deviation
Calibration Set	72	12.114	6.124	8.451	1.619
Prediction Set	18	10.952	6.344	8.237	1.652

Spectral Data Preprocessing

In this experiment, quantitative analysis of soybean moisture content was performed by establishing PLSR models based on both the raw spectra and spectra preprocessed using eight different methods. The results are shown in Table 2.

From Table 2, it can be observed that the PLSR model based on the raw near-infrared spectra has the R_p^2 of 0.7995 and the R_{msep} of 0.4205, indicating good stability of the model. After preprocessing with baseline correction, standard normal variate transformation, multiple scatter correction, and first derivative, the models exhibited the R_p^2 values higher than those of the raw spectrum-based model and the R_{msep} values lower than 0.4205. The performance of the models improved with these four preprocessing methods, and the stability of the cross-validation set was better.

The three preprocessing methods with the best performance (BC, SNV and MSC) were selected for further model processing.

Table 2

PLSR model based on different pretreatment methods

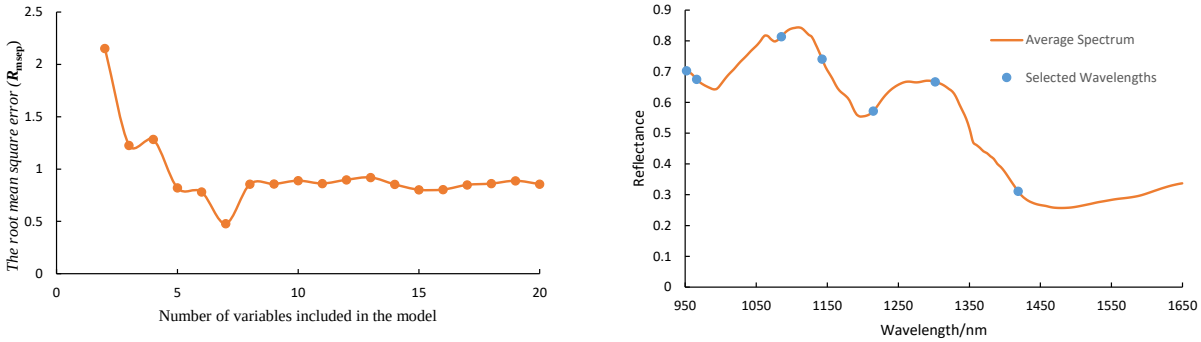
Preprocessing Methods	*PCs*	Calibration Set		Cross-Validation Set		Prediction Set	
		R_c^2	R_{msec}	R_{cv}^2	R_{msecv}	R_p^2	R_{msep}
Raw spectral	9	0.8181	0.5072	0.6916	0.8598	0.7995	0.4205
Baseline Correction	14	0.8780	0.3267	0.5906	1.0961	**0.8535**	**0.3612**
Moving Average	9	0.8080	0.5353	0.6922	0.8581	0.7876	0.4454
Savitzky-Golay filtering	9	0.8095	0.5311	0.6889	0.8675	0.7799	0.4617
Normalization	10	0.8215	0.4640	0.7093	0.7557	0.8515	0.4230
Standard Normal Variate	9	0.8347	0.4329	0.6734	0.8555	**0.8929**	**0.2838**
Multiplicative Scatter Correction	9	0.8312	0.4393	0.6754	0.8447	**0.9052**	**0.2646**
Derivative Spectroscopy	6	0.8369	0.4525	0.7186	0.7809	0.8100	0.4051
Deconvolution	9	0.8157	0.5139	0.6919	0.8590	0.7935	0.4331

Selection of Feature Wavelengths

Among the 145 retained spectral bands, there is still a significant amount of redundant information. To improve modeling speed and model robustness, feature wavelengths need to be extracted from the spectral data. This study applied the continuous projection algorithm and competitive adaptive reweighted sampling algorithm to extract feature wavelengths from spectral data preprocessed by baseline correction, standard normal variate transformation, and multiple scatter correction.

Figure 4 illustrates the process of extracting feature wavelengths from the spectral data preprocessed with multiple scatter correction using the continuous projection algorithm. The selection process is determined by the size of the root mean square error, with smaller values indicating better model performance. The R_{msep} varies with the number of relevant wavelengths.

As shown in Figure 3(a), when the minimum R_{msep} is 0.4786, the number of extracted feature wavelengths is 7. The corresponding wavelengths in the original data are shown in Figure 3(b), with the 7 feature wavelengths being 966.15, 1301.8, 1142.68, 1085.21, 1418.51, 1214.81, and 951.93 nm, accounting for 4.83% of the total wavelengths.



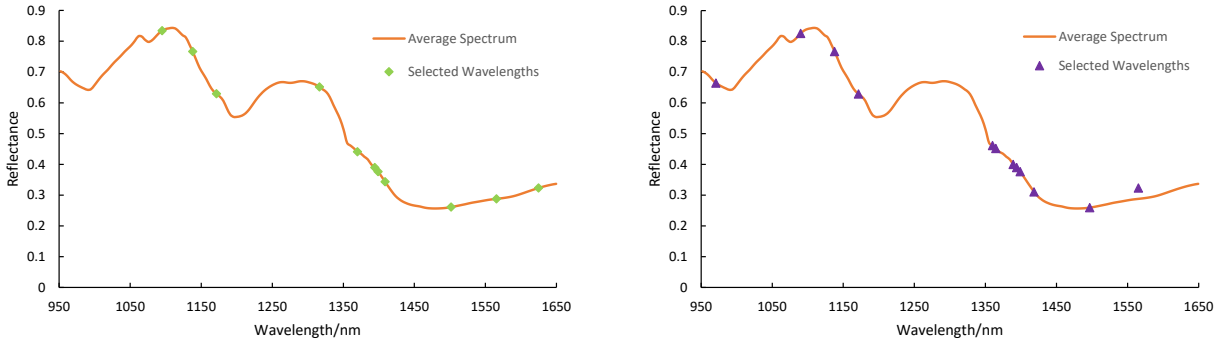
a) RMSECV of variable: *b) Sequence number of selected characteristic wavelength*

Fig.4 Process of SPA algorithm in screening characteristic wavelength of soybean water content

Similarly, for the spectral data preprocessed by standard normal variate transformation, 2 feature wavelengths were extracted when the minimum R_{msep} is 0.5601. These 2 feature wavelengths are 1496.77 and 951.93 nm, accounting for 1.38% of the total wavelengths.

For the spectral data preprocessed by baseline correction, 5 feature wavelengths were extracted when the minimum R_{msep} is 0.9474. These 5 feature wavelengths are 1330.9, 1457.59, 1219.63, 1142.68, and 1418.51 nm, accounting for 3.45% of the total wavelengths.

Figure 5 shows the results of extracting feature wavelengths from spectral data preprocessed using the competitive adaptive reweighted sampling algorithm after baseline correction, standard normal variate transformation, and multiple scatter correction.



a) SNV spectral data *b) BC spectral data*

Fig. 5 - Result of CARS algorithm in screening characteristic wavelength of soybean water content

For the spectral data preprocessed by multiple scatter correction, 19 feature wavelengths were extracted. These wavelengths are 1085.21, 1089.99, 1133.09, 1137.89, 1176.3, 1306.64, 1311.49, 1316.34, 1369.78, 1374.64, 1389.25, 1394.12, 1399.00, 1413.63, 1496.77, 1501.68, 1570.49, 1580.34, and 1624.75 nm, accounting for 13.10% of the total wavelengths.

For the spectral data preprocessed by standard normal variate transformation, 11 feature wavelengths were extracted. These wavelengths are 1094.78, 1137.89, 1171.5, 1316.34, 1369.78, 1394.12, 1399.00, 1408.75, 1501.68, 1565.56, and 1624.75 nm, accounting for 7.59% of the total wavelengths.

For the spectral data preprocessed by baseline correction, 12 feature wavelengths were extracted. These wavelengths are 970.9, 1089.99, 1137.89, 1171.5, 1360.05, 1364.91, 1389.25, 1394.12, 1399.00, 1418.51, 1496.77, and 1565.56 nm, accounting for 8.28% of the total wavelengths.

Model Optimization

PLSR models for predicting soybean moisture content were established using the original spectral data from the 950–1650 nm wavelength range, as well as spectral data preprocessed by baseline correction, standard normal variate transformation, and multiple scatter correction.

The root mean square error of prediction for the prediction set was used as the metric to evaluate model performance, with lower R_{msep} values indicating better prediction accuracy. As shown in Table 3, the MSC-PLSR model has the lowest R_{msep} value of 0.2646, indicating that the PLSR model based on the 950–1650 nm wavelength spectra with MSC preprocessing has good prediction performance and stability.

To improve modeling speed and model robustness, SPA and CARS algorithms were used to select feature wavelengths for building soybean moisture content prediction models. These include SPA-BC-PLSR, SPA-MSC-PLSR, SPA-SNV-PLSR, CARS-BC-PLSR, CARS-MSC-PLSR, and CARS-SNV-PLSR. As shown in Table 3, compared to the SPA and full-band models, the PLSR models using feature wavelengths selected by the CARS algorithm had lower R_{msep} values. Among the models based on feature wavelengths, CARS-MSC-PLSR and CARS-SNV-PLSR had the lowest R_{msep} values of 0.2401 and 0.2371, respectively. This suggests that the models using feature wavelengths selected by the CARS algorithm performed better, likely due to CARS's effectiveness in reducing spectral collinearity.

Table 3

PLSR model based on different pretreatment methods

NO	Models	*Number of Wavelengths*	Calibration Set		Cross-Validation Set		Prediction Set	
			R_c^2	R_{msec}	R_{cv}^2	R_{msecv}	R_p^2	R_{msep}
1	PLSR	145	0.8181	0.5072	0.6916	0.8598	0.7995	0.4205
2	BC- PLSR	145	0.8780	0.3267	0.5906	1.0961	0.8535	0.3612
3	MSC- PLSR	145	0.8312	0.4393	0.6754	0.8447	0.9052	0.2646
4	SNV- PLSR	145	0.8347	0.4329	0.6734	0.8555	0.8929	0.2838
5	SPA-BC- PLSR	5	0.6673	0.9014	0.6222	1.0236	0.837	0.3683
6	SPA-MSC- PLSR	7	0.6797	0.8543	0.6231	1.0053	0.8191	0.4497
7	SPA-SNV- PLSR	2	0.5804	1.1109	0.5454	1.2034	0.7098	0.5601
8	CARS -BC- PLSR	12	0.811	0.503	0.7289	0.7214	0.8771	**0.3053**
9	CARS -MSC- PLSR	19	0.8547	0.3826	0.786	0.5638	**0.9105**	**0.2401**
10	CARS -SNV- PLSR	11	0.845	0.3984	0.7877	0.5459	0.92	**0.2371**

DISCUSSION

Moisture content is a critical factor influencing the quality of soybean grains. However, determining the moisture content of soybean seeds is both time-consuming and expensive. Therefore, a rapid and non-destructive method for determining the moisture content in soybean seeds is essential. Guo Zhen applied visible near-infrared hyperspectral imaging technology (400–1000 nm) in conjunction with a wavelength selection algorithm to determine the moisture content of soybean seeds. They found that the combination of the interval variable iterative space shrinkage approach and the successive projections algorithm, based on the original spectra, was the most suitable model for determining the moisture content of soybean seeds *(Guo et al., 2023)*. Jin Chengqian employed near-infrared hyperspectral imaging technology for rapid and non-destructive detection of soybean moisture content. They found that the SPA algorithm provided better predictive performance for feature wavelength modeling, and selected the Normalize-SPA-PCR model, which demonstrated good stability and predictability, allowing for accurate prediction of soybean moisture content *(Jin et al., 2022)*. These findings suggest that the moisture inversion model constructed using spectral imaging technology can enable rapid detection of soybean moisture. This method provides a new approach for soybean moisture detection.

Furthermore, the CARS algorithm utilizes adaptive reweighted sampling technology to quickly identify variables that are most sensitive to moisture changes, thereby obtaining the optimal variable combination and reducing the consumption of computational resources. SNV enhances the stability of spectral data by performing a standard normal variate transformation, making the data more closely align with a normal distribution. Therefore, the combination of SNV and CARS algorithms for processing soybean spectral data can yield reliable prediction results.

CONCLUSIONS

(1) Eight preprocessing methods were used to develop a PLSR model for the spectral data in the 900–1700 nm range. The model with the highest R_c^2 value and lowest R_{msep} value was achieved after applying MSC, indicating that MSC performed best in full-band modeling and prediction.

(2) Two methods, SPA and CARS, were used to extract characteristic wavelengths, effectively reducing the spectral dimensionality. The characteristic wavelengths extracted by SPA accounted for 4.83%, 1.38%, and 3.45% of the spectral data, while those extracted by CARS accounted for 13.10%, 7.59%, and 8.28%.

(3) By combining preprocessing methods, feature wavelength extraction methods, and modeling methods, the effectiveness of 10 models was compared, and the CARS-SNV-PLSR model was selected as the optimal one. The CARS-SNV-PLSR model showed a high R_c^2 value of 0.92 and a low R_{msep} value of only 0.2371, indicating good stability and prediction performance. This model enables rapid, accurate, and non-destructive detection of soybean moisture content.

ACKNOWLEDGEMENT

We greatly appreciate the careful and precise reviews by the anonymous reviewers and editors. This research was financially supported by the National Natural Science Foundation of China, grant number 32272004, 32171911; the National Key Research and Development Program of China, grant number 2021YFD2000503; the Natural Science Foundation of Jiangsu, grant number BK20221188.

REFERENCES

[1] Bu, Z.Y., Li, Z.F., Song, F.H., Li, B. & Li, J. (2018). Determination of moisture content in soybean leaves based on terahertz imaging (基于太赫兹成像技术的大豆叶片水分含量测定). *Acta Agriculturae Zhejiangensis*, Vol.30, pp.1420-1426. DOI: 10.6041/j.issn.1004-524.2018.08.21

[2] Chen, J.W., Zhou, D.Q., Cui, C.C., Ren, Z.J. & Zuo, W.J. (2023). Prediction Model of Farinograph Characteristics of Wheat Flour Based on Near Infrared Spectroscopy (近红外光谱的小麦粉粉质特性预测模型研究). *Spectroscopy and Spectral Analysis*, Vol.43, pp.3089-3097. DOI: 10.3964/j.issn.1000-0593(2023)10-3089-09

[3] Chen, M., Ni, Y.L., Jin, C.Q., Xu, J.S., Yuan, W.S. (2019). High spectral inversion of wheat impurities rate for grain combine harvester (谷物联合收割机收获小麦含杂率高光谱反演研究). *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, Vol.35, pp.22-29. DOI: 10.11975/j.issn.1002-6819.2019.14.003

[4] Chen, M., Xu, J.S., Jin, C.Q., Zhang, G.Y, Ni, Y.L. (2019). Inversion model of soybean impurity rate based on hyperspectral (基于高光谱的大豆含杂率反演模型). *Journal of China Agricultural University*, Vol.24, pp.160-167. DOI: 10.11841/j.issn.1007-4333.2019.09.17

[5] Guo, Z., Zhang, J., Ma, C.Y., Yin, X., Guo, Y.M., Sun, X. & Jin, C.Q. (2023). Application of visible-near-infrared hyperspectral imaging technology coupled with wavelength selection algorithm for rapid determination of moisture content of soybean seeds. *Journal of Food Composition and Analysis*, Vol.116, pp.105048. DOI: 10.1016/j.jfca.2022.105048

[6] Jin, C.Q., Guo, Z., Zhang, J., Ma, C.Y., Tang, X.H., Zhao, N. & Yin, X. (2022). Non-destructive detection and visualization of soybean moisture content using hyperspectral technique (大豆水分含量的高光谱无损检测及可视化研究). *Spectroscopy and Spectral Analysis*, Vol.42, pp.3052-3057. DOI: 10.6041/j.issn.1000-0593(2022)10-3052-06

[7] Li, C.X., Zhao, C.Y., Ren, Y., He, X., Yu, X.T. & Song, Q. (2022). Microwave traveling-standing wave method for density-independent detection of grain moisture content. *Measurement*, Vol.198, pp.111373. DOI: 10.1016/j.measurement.2022.111373

[8] Li, Y., Wang, M.M., Lü, B., & Yu, H.S. (2023). Comparative analysis of soybean production efficiency in the main production areas in the context of rural revitalization (乡村振兴背景下主产区大豆生产效率比较分析). *Soybean Science*, Vol.43, pp.120−128.

[9] Liu, L.L., Li, J.F., Shu, Y., Chen, X.Y., & Tang, G.X. (2022). Current situation of soybean production and consumption in China and strategies to improve self-sufficiency rate (我国大豆生产消费现状及提升自给率策略). *Chinese Journal of Oil Crop Sciences*, Vol.44, pp.242-248. DOI: 10.1016/j.issn.1007-9084.2022015

[10] Li, Y., Xie, Q.Z., Liu, B.Q., He, S.Q., Wu, X.Z., Yang, Q., Liu, Z., Shi, X.L., Zhang, M.C., Yang, C.Y., Yan, L., Zhang, R.F., & Tao, P.J. (2023). Life cycle analysis of soybean production in typical district of the North China Plain (华北平原典型区大豆生产全生命周期分析). *Chinese Journal of Eco-Agriculture*, Vol.31, pp.1416−1427.

[11] Mo, F., Wang, G.X., & Hu, M.Z. (2020). Analysis of soybean production status in northeast China based on cost (基于成本视角的东北地区大豆生产现状分析). *Soybean Science*, Vol.39, pp.947−953. DOI: 10.11861/i.issn.10009841.2020.06.0947

[12] Shi, W.Q., Li, Y.H., Wei, Z., Yu, J.K., Zhao, C. & Qiu, J.K. (2024). Monitoring and zoning soybean maturity using UAV remote sensing. *Industrial Crops & Products*, Vol.222, pp.119470. DOI: 10.1016/j.indcrop.2024.119470

[13] Song, Y.H., Gao, S.S., Chu, X.X., Zhou, Y.M., Xu, Y.Q., Sun, T., Zhou, G.X. & Liu, X.Q. (2023). Non-destructive detection of moisture and fatty acid content in rice using hyperspectral imaging and chemometrics. Journal of Food Composition and Analysis, Vol.121, pp.105397. DOI: 10.1016/j.jfca.2023.105397

[14] Wang, L.S., Huang, Z.L. & Wang, R.J. (2023). Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Physics and Technology*, Vol.115, pp.103731. DOI: 10.1016/j.infrared.2021.103731

[15] Wu, J.Z., Zhang, L., Li, J.B., Liu, C.L., Sun, X.R. & Yu, L. (2022). Detection model of moisture content of single maize seed based on hyperspectral image and ensemble learning (基于高光谱与集成学习的单粒玉米种子水分检测模型). *Transactions of the Chinese Society for Agricultural Machinery*, Vol.53, pp.302-308. DOI: 10.6041/i.issn.1000-298.2022.05.031

[16] Yang, S.F., Jia, Z.C., Yi, K., Zhang, S.H., Zeng, H.G., Qiao, Y., Mao, P.S. & Manli Li (2024). Rapid prediction and visualization of safe moisture content in alfalfa seeds based on multispectral imaging technology. *Industrial Crops & Products*, Vol.222, pp.119448. DOI: 10.1016/j.indcrop.2024.119448

[17] Zhang, J., Guo, J., Wang, S.H., Yue, M.H., Zhang, S.S., Peng, H.H., Yin, X., Du, J., Zhao, N. & Ma, C.Y. (2023). Comparison of methods for water content in rice by portable near-infrared and visible light spectrometers (便携式近红外和可见光光谱仪检测水稻水分含量方法比较研究). *Spectroscopy and Spectral Analysis*, Vol.43, pp.2059-2066. DOI: 10.6041/j.issn.1000-0593(2023)07-2059-08

[18] Zhang, T., Guan, H.O., Ma, X.D. & Shen, P.P. (2023). Drought recognition based on feature extraction of multispectral images for the soybean canopy. *Ecological Informatics*, Vol.77, pp.102248. DOI: 10.1016/j.ecoinf.2023.102248

[19] Zhang, X.D., Duan, Z.H., Mao, H.P., Gao, H.Y., Shi, Q., Wang, Y.F., Shen, B.G.& Zhang, Q. (2021). Tomato water stress state detection model by using terahertz spectroscopy technology (利用太赫兹光谱技术构建番茄水分胁迫状态检测模型). *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, Vol.43, pp.121-128. DOI: 10.11975/j.issn.1002-6819.2021.15.015

[20] Zhao, X.T., Zhang, S.J., Li, B. &Li, Y.K. (2018). Study on Moisture Content of Soybean Canopy Leaves under Drought Stress Using Terahertz Technology (太赫兹光谱技术用于干旱胁迫下大豆冠层含水量检测研究). *Spectroscopy and Spectral Analysis*, Vol.43, pp.3089-3097. DOI: 10.3964/j.issn.1000-0593(2018)08-2350-05