

Research Article

Establishment and Verification of Near-infrared Spectral Prediction Model for Fat Content of *Xanthoceras Sorbifolia*

Ge Chao-hong¹ , Li Wei-ming^{1,*} , Zhao Hai-long²

¹Cotton Research Institute, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China

²Shijiazhuang Seed Management Station, Shijiazhuang, China

Abstract

In order to realize the nondestructive and rapid detection of fat content of *Xanthoceras sorbifolia* and meet the screening of breeding materials and industrial processing requirements of *X.sorbifolia*, 46 *X.sorbifolia* were selected as the standard sample set, the results showed that the fat content of 46 apricot kernel kernels was 49.38%~68.98% an average content of 61.62%. The fat content of the seed kernel was determined by the Soxhlet extraction method, and the spectral data of the sample was collected by the near-infrared spectroscopy (NIRS) technology, and the Unscrambler software was used to construct the NIRS prediction model of *X.sorbifolia* fat content by the partial least squares (PLS) method. The results showed that the regression curve R-Square (determination coefficient) of the model was 0.9856, and the RMSE (standard error) was 0.4149, which could be used for effective prediction. At the same time, 32 *X.sorbifolia* samples not participating in the modeling were selected as validation materials to further carry out external test on the prediction effect of the model. The results showed that the external test regression curve R-Square was 0.9014, RMSE was 0.8259, and the predicted value of fat content was in good agreement with the chemical value.

Keywords

Xanthoceras Sorbifolium, Fat Content, NIRS, Prediction Model

1. Introduction

Xanthoceras sorbifolium is a unique woody oil plant in China with both medicinal and edible value. The oil content of seed is as high as 60%. Besides, the fruit, root, and stem can be used as medicine, which is of great economic value [1, 2]. In addition to being processed into edible oil, can also be used to produce high-grade industrial crude oil. Currently, the determination of fat content in *X.sorbifolia* primarily relies on chemical analysis methods, which provide accurate results but the process is time-consuming, labor-intensive, and destructive to samples. Therefore, there is an urgent need

to develop a simple, rapid, non-destructive, and environmentally friendly method for determining fat content in *X.sorbifolia*. Therefore, it is imperative to develop a simple, non-destructive, and green determination method for the fat content of *X.sorbifolia* fruit. Near-infrared spectroscopy (NIRS) technology can quickly, efficiently, and non-destructively detect the content of various components in plants [3, 4]. In recent years, it has been widely applied in many fields such as agriculture, food, and medicine [5-13]. In agriculture, NIRS models have been established for ana-

*Corresponding author: 438366698@qq.com (Li Wei-ming)

Received: 30 September 2025; **Accepted:** 16 October 2025; **Published:** 22 November 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

lyzing relevant components in crops such as soybeans, peanuts, and corn, and the content of various components in different genotypes has been investigated [14-20]. However, as of now, there have been no reports on NIRS-based prediction of fat content in *X.sorbifolia*. In this study, spectra collected using the Perten Company's third-generation diode array near-infrared analyzer DA7250 were combined with corresponding chemical values. Using The Unscrambler X 10.4 (64-bit) software for analysis, a mathematical model for non-destructive and rapid detection of fat content in *X.sorbifolia* was established. The aim is to achieve rapid determination of fat content, to save time and costs, provide an efficient approach for screening breeding materials and industrial processing of *X.sorbifolia*.

2. Materials and Methods

2.1. Experimental Materials

The experimental materials were different types of *X.sorbifolia* from Hebei, Shandong, Henan, Liaoning Shanxi, and other places. Among them, 46 types of them were used for modeling, and 32 types were used for validation of the model. For each sample, 100 g of seeds were randomly selected, shelled and dried, and NIRS spectra were collected; then, they were ground, crushed, sieved through a round hole sieve with a diameter of 1.0 mm, and the material passing through the sieve was placed in an oven 80 °C to constant weight, and then packed in sealed bags for preservation, used for the determination of chemical values of fat content in *X.sorbifolia*.

The main instruments or reagents include near-infrared analyzer DA7250 (manufactured by Perten Instruments, USA), Soxhler extractor (produced by Shanghai Hongji Instrument Equipment Co., LTD.), and petroleum ether (boiling point 60~90 °C).

2.2. Experimental Methodology

(1) Determination of chemical value of fat content in sample set: The fat content of *X.sorbifolia* kernels was determined by Sox extraction method [21] according to GB 5009.6-2016.

(2) Collection of NIRS Spectral Profiles for Sample Sets: Forty-six varieties of *X.sorbifolia* kernels were selected. Using the DA7250 near-infrared analyzer in advanced mode, spectral data collection was performed with 5 nm resolution and a spectral range of 950-1650 nm. To ensure experimental accuracy, each sample was measured three times with repeated measurements for a total of three replicates. Spectral data from each variety's kernels were processed into DX format files.

Import the dx format file into The Unscrambler X 10.4 (64-bit) software. After successful import, add the chemical values corresponding to each sample. Define the database range using spectral wavelength as the X-variable and

chemical values as the Y-variable. View the original spectrum, remove abnormal spectra and their corresponding outliers, and generate the sample spectral profile.

(3) NIRS Model Establishment: Utilizing The Unscrambler X 10.4 (64-bit) software, the spectra after deleting the abnormal values were pre by 1st Derivative, SNV, and Detrending to eliminate the effects of scattered light and path length variation, and the chemical value and the acquired NIR data were fitted respectively; the chemometrics method of partial least squares (PLS) was used to establish a mathematical model, and the model was optimized by repeatedly adopting the internal-validation to eliminate the influence of the abnormal value and the number of factors; the best model was screened through comparing the R-Square (coefficient of determination), RM (standard error), and the influence of the number of factors of the model. The smaller the RMSE, the larger the R-Square, and the better the prediction of the model.

(4) External Validation of the NIRS Model: 32 non-modelled samples were randomly selected as validation materials. The screened NIRS model was applied to predict fat content, with chemical values measured using Soxhlet extraction. The model's predictive performance was evaluated by comparing the correlation and accuracy between NIRS predicted values and chemical measurements for each sample.

2.3. Data Analysis

Data processing and analysis were performed using Excel 2017 and The Unscrambler X 10.4 (64-bit) software.

3. Results and Analysis

3.1. Chemical Values of Sample Set Kernel

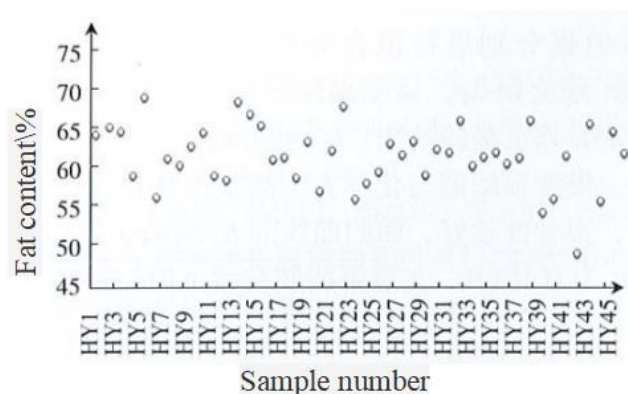


Figure 1. Chemical value of fat content of the samples.

The Soxhlet extraction results (Figure 1) indicate that the fat content of 46 varieties of *X.sorbifolia* kernels ranges from 49.38% to 68.98%, with an average of 61.62%. The fat con-

tent distribution of the samples essentially covers the normal range of production and breeding materials, demonstrating good continuity and representativeness that meets the requirements for NIRS calibration. The selected *X.sorbifolia* kernels in this study exhibit a wide variation in fat content, making them suitable for establishing NIRS models.

3.2. NIRS Mapping of Sample Set Kernel

After acquiring spectral data from 46 varieties of *X.sorbifolia* kernels, we generated Near-Infrared Spectral

(NIRS) curves using The Unscrambler X 10.4 (64-bit) software. The results (Figure 2) demonstrate that all 46 samples exhibited nearly identical spectral trends, with distinct absorption peaks observed across the 950-1650 nm wavelength range. While multiple absorption peaks were present in individual samples, significant variations were observed at the same peak positions between different specimens. This indicates that the NIR absorption spectrum of Chinese wingnut seeds can effectively support both qualitative and quantitative analysis of their constituent components.

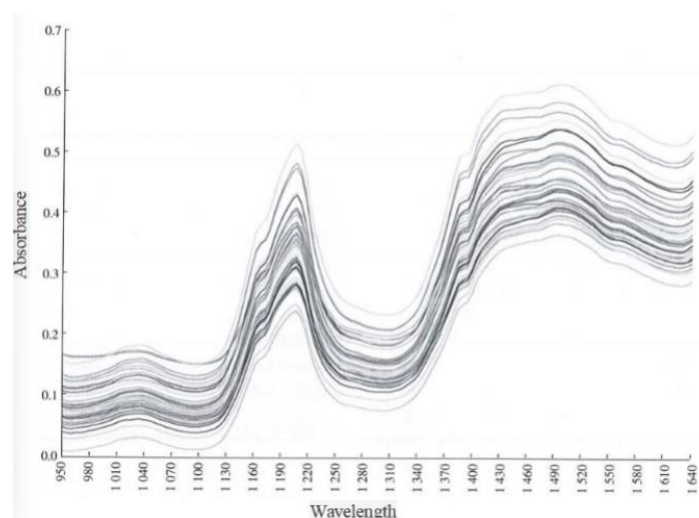


Figure 2. NIR spectra of *X.sorbifolia* samples.

3.3. Construction of NIRS Model

Spectral data were collected using a near-infrared analyzer, and Unscrambler software was employed to perform fitting processes on both the chemical values of seed fat content and the acquired NIRS data. A partial least squares (PLS) model was established through iterative outlier removal, with the best model selected based on R-square and root mean square error (RMSE). Results (Figure 3) demonstrate that the model

predictions exhibit a strong linear relationship with chemical method measurements, showing excellent fitting performance. The regression curve achieved an R-square value was 0.9856 and RMSE was 0.4149. The predicted value of the model of the fat content *X.sorbifolia* kernels by NIRS is in good agreement with the measured value of the chemical method, model is credible, and the prediction of the fat content of *X.sorbifolia* kernel has high reference value.

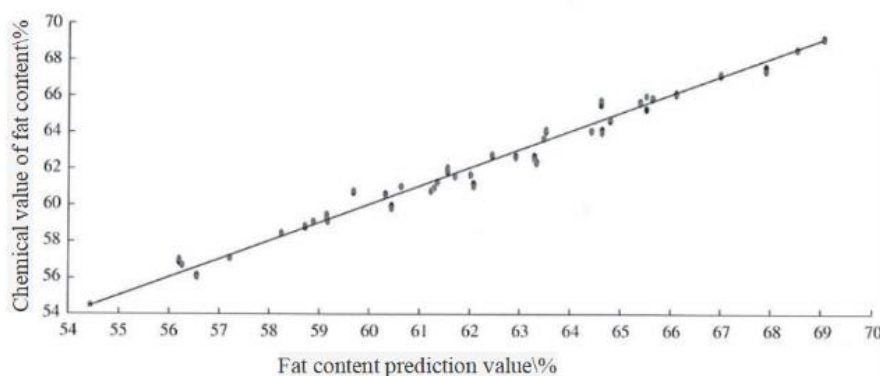


Figure 3. Regression curve of the predicted model.

3.4. External Validation of the NIRS Model

Thirty-two *X.sorbifolia* kernels not included in the modeling process were randomly selected for external validation of the established Near-Infrared Spectroscopy (NIRS) model. The results (Figure 4) demonstrate a strong linear correlation between NIRS predicted values and chemical analysis measurements, with sample data clustering closely around the

central axis. The model exhibits an R-squared value of 0.9401 and root mean square error (RMSE) of 0.8259, indicating high reliability of the predictions. The prediction value of fat content of *X.sorbifolia* kernel obtained by NIRS model was more accurate, and this model can be used for the screening of materials and industrial production of *X.sorbifolia*.

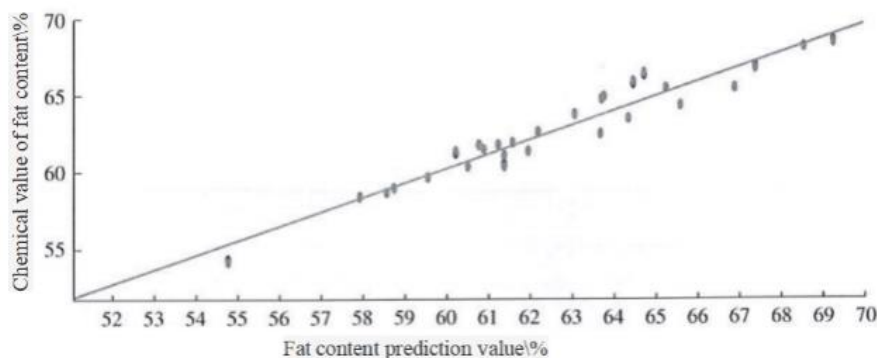


Figure 4. Regression curve of externally validation for NIRS model.

4. Conclusion and Discussion

The *X.sorbifolia*, a unique woody oil crop in China, yields seeds containing over 60% oil with more than 90% unsaturated fatty acids. These neuroactive compounds, which can repair brain nerve cells and promote nerve fiber regeneration, play a crucial role in preventing and treating neurodegenerative diseases and nervous system disorders. In recent years, Near-Infrared Spectroscopy (NIRS) technology has been widely applied in agriculture, with predictive models established for various crops. For instance, Geng Liguang et al. [14] developed a non-destructive NIRS model to assess soybean seed viability, Ji Hongchang et al. [18] created an oil content prediction model for peanut kernels using NIRS, and Fang Yan [20] established a corn kernel crude starch content prediction model through NIRS analysis. However, current methods for determining Chinese wingnut fat content typically involve Soxhlet extraction or acid hydrolysis—processes requiring sample pulverization and chemical reagents that are time-consuming, damage sample integrity, and cause environmental pollution. This study employed a DA7250 NIRS spectrometer combined with Unscramble software to develop a predictive model for seed fat content. The model enables accurate and rapid measurement of fat content in whole seeds, semi-seeds, and powdered forms. Compared to traditional methods, this approach preserves sample integrity, avoids material waste, and allows efficient fat content determination without chemical reagent contamination.

Due to the influence of multiple factors such as measurement environment, technical conditions, background composition, and target components of samples on near-infrared spectroscopy collection [22–24], the spectral information often exhibits significant overlap and abnormal complexity. This results in complicated parameter configurations for predictive models, making prediction accuracy the most critical challenge limiting.

their development. Therefore, scientific sampling and error reduction are crucial for improving the predictive accuracy of spectral analysis models. To achieve ideal prediction outcomes, this study randomly selected excluded, while pre-processing eliminated the effects of scattered light and path variations. Through repeated fitting of chemical values and spectral data, we established and progressively optimized the model, ultimately achieving an optimized *X.sorbifolia* kernel fat content determination model with R-squared value of 0.9856 and root mean square error (RMSE) of 0.4149. In conclusion, the developed near-infrared spectroscopy prediction model for *X.sorbifolia* kernel fat content demonstrates stable and accurate performance. The detection method is economical, rapid, efficient, and non-polluting, providing a swift and effective pathway for *X.sorbifolia* breeding material screening and industrial processing.

Abbreviations

NIRS	Near-Infrared Spectroscopy
RMSE	Root Mean Square Error
PLS	Partial Least Squares

R-Square Determination Coefficient

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Chen Ou, Dong Fengliang, Ma Haiyuan, Zheng Shuqing, Ma Tjun, Jia Changxi. Extraction and Analysis of Physicochemical Properties and Fatty Acid Composition of Xinghua Fruit Oil. *Journal of Beijing College*, 2013, 28(1): 78-80. <https://doi.org/BNxb.0.2013-01-023>
- [2] Sun Linlin, Zhao Dengchao, Hanuanming, Luan Senian, Hou Liqun. Genetic Analysis of Economic Fruit Traits of Xinghua Fruit Tree Plants. *Shand Agricultural Science*, 2012, 44(1): 25-28. <https://doi.org/10.14083/j.issn.1001-4942.2012.01.021>
- [3] Hao Xuesong, Chen Pu, D Jiawei, Wang Haipeng, Liu Dan, Li Jingyan, Xu Yupeng, Chu Xiaoli. Application Progress and Prospect of Miniature NearInfrared Spectrometer. *Journal of Analytical Testing*, 2022, 41(9): 1301-1313. <https://doi.org/10.19969/j.fxcxb.22031804>
- [4] Zheng Niannian, Luan Xiaoli, Liu Fei. Near-Infrared Spectroscopy Modeling Technology Based on Adaptive Elastic Net. *Spectroscopy and Spectral Analysis*, 2019, 39(1): 319-324. <https://doi.org/GUAN.0.2019-01-055>
- [5] Shang Jing, Zhang Yan, Meng Qinglong. Identification of Apple Varieties by Spectral Technology Combined with Chemometrics. *Northern Horture*, 2019(16): 66-71. <https://doi.org/BFYy.0.2019-16-011>
- [6] Wang Siling, Cai Chen, Ma Huiqing, Longilin. Non-Destructive Detection of Apple Bitter Pit Based on Hyperspectral Imaging. *Northern Horticulture*, 2015(08): 124-130. <https://doi.org/BFYy.0.2015-08-033>
- [7] Zhang Shengsheng. Research on Quality Control of Traditional Chinese Medicine Shui Niu Jiao Based on Near-Infrared Spectroscopy. Guiyang: Guizhou Normal University, 2022. <https://doi.org/10.27048/d.cnki.ggzsu.2022.000487>
- [8] Chen Ziyun, Huang Xiaoxia, Yao Wanqing, Peng Mengxia. Rapid detection of the content of citral in Litsea cubeba essential oil by near-infrared spectroscopy combined with partial least squares method. *Anhui Chemical Industry*, 2022, 48(1): 121-124. <https://doi.org/AHHG.0.2022-01-030>
- [9] Wang Shengpeng, Zheng Pengcheng, Gong Zhiming, Liu Panpan, Gao Shiwu, Teng Jing, Gui Anhui, Wang Xueping, Ye Fei, Zheng Lin. Rapid non-destructive evaluation of the taste quality of tea soup in green brick tea by near-infrared spectroscopy. *Journal of Central China Agricultural University*, 2020, 39(3): 113-119. <https://doi.org/10.13300/j.cnki.hnlkxb.2020.03.015>
- [10] Yan Kewei, Wang Fu, Mei Guorong, Lu Junyu, Zhang Lan, Fu Guilan, Chen Lin, Liu Youping, Chen Hongping. Establishment of a rapid qualitative discrimination model for Guangxiangpi based on near-infrared spectroscopy technology. *Chinese Journal of Chinese Materia Medica*, 2015, 46(20): 3096-3099. <https://doi.org/ZCYO.0.2015-20-027>
- [11] Wang Yun, Xu Kexin, Chang Min. Establishment of calibration model for the detection of fat and protein content in milk by near-infrared spectroscopy technology. *Optical Instruments*, 2006(3): 3-7. <https://doi.org/GXYQ.0.2006-03-000>
- [12] Geng Xiang, Zhou Liping, Ma Xinxin, Jiang Longfa, Yang Weigen. Research on rapid determination method of oil content and moisture in tea cake based on near-infrared spectroscopy. *Anhui Agricultural Science Bulletin*, 2017, 23(16): 130-132 171. <https://doi.org/10.16377/j.cnki.issn1007-7731.2017.16.049>
- [13] Na Rong, Ren Jindong, Hu Bo, Zhao Qiang, Lei Xi. Research on the establishment of rapid analysis model of alfalfa nutritional components based on different pretreatment methods. *Journal of Animal Ecology*, 2021, 42(12): 37-43. <https://doi.org/JCST.0.2021-12-006>
- [14] Geng Lig, Song Chunfeng, Wang Lina, An Xuesong, Sun Juan. Research on the method of nondestructive determination of soybean seed vitality by near-infrared spectroscopy. *Journal of Plant Genetic Resources*, 2013, 14(6): 1208-1212. <https://doi.org/10.13430/j.cnki.jpgr.2013.06.026>
- [15] Yang Shuhan. Design and Research of Cotton Moisture Content Detection System Based on Near-Infrared. Xi'an: Xi'an University of Technology, 2021. <https://doi.org/10.27398/d.cnki.gxalu.2021.001650>
- [16] Wang Zhiwei, Wang Xiuzhen, Ma Lang Liu Ting, Tang Yueyi, Wu Qi, Sun Quanxi, Wang Chuangang. Construction of Near-Infrared Analysis Model for Edible Sens Quality of Peanut Kernel. *Peanut Science*, 2022, 51(3): 77-82. <https://doi.org/10.14001/j.issn.1002-4093.2022.03.010>
- [17] Mzimhiri Rehema Idriss. Research on the Method of Determining the Content of Oleic Acid and Linoleic Acid in Peanut Seeds and Peanut Oil Based on Hyperspectral Imaging Technology and Near-Infrared Spectroscopy. Beijing: Chinese Academy of Agricultural Sciences, 016.
- [18] Ji Hongchang, Qiu Xiaocheng, Liu Wenhao, Hu Changli, Kong Ming, Hu Xiaui, Huang Jianbin, Yang Xue, Tang Yan, Zhang Xiajun, Wang Jingshan, Qiao Lipian. Construction and Application of Near-Infrared for Oil Content in Peanut Kernels. *Journal of Chinese Oil Crop*, 2022, 44(5): 189-1097. <https://doi.org/10.19802/j.issn.1007-9084.2021205>
- [19] Zhou Qingmei, Wang Gang, Wang Rong, Chen Shuang, Guo Liy. Feasibility Analysis of Near-Infrared Technology Applied to the Determination of Moisture and Acid Value of Rice. *China and Foreign Liquor*, 2021(21): 14-17. <https://doi.org/PJKJ.0.2021-21-003>
- [20] Fang Yan. Research on the Non-Destructive Determination of Cornel Crude Starch Content by Near-Infrared Spectroscopy. *Crop Journal*, 2011, 141(2): 5-27. <https://doi.org/10.16035/j.issn.1001-7283.2011.02.010>
- [21] GB 5009.6—2016, National Food Safety Standard Determination of Fat in Food [S]. <https://max.book118.com/html/2020/1107/7152053151003014.shtml>

- [22] Li Junxia, Zhang Hongliang, Yan Yanlu, Min Shungeng, Li Zichao. Creation and Application of Near-Infrared Predictive Model for Rice Protein in Breeding. *Chinese Agricultural Science*, 2006, 39(4): 836-81. <https://doi.org/ZNYK.0.2006-04-029>
- [23] Lv Longfei, Zhang Miaomiao, Bao Xiang, et al. Construction of a Model for Predicting the Content Routine Nutritional Components in Whole Corn Silage Based on Near-Infrared Reflectance Spectroscopy. *Animal Nutrition*, 1-1 [2025-10-16]. <https://doi.org/10.12418/CJAN2025.59>
- [24] Hu Yichao, Dai Suming, Sun Jiansheng, et al. Research on the Early Aging Process of Tobacco Leaves on Near-Infrared Spectroscopy Technology. *Biomass Chemical Engineering*, 2025, 59(05): 87-4. <https://doi.org/LCHG.0.2025-05-010>