Prediction and Analysis of Bamboo heating value Near Infrared Spectroscopy Based on Competitive Adaptive Weighted Sampling Algorithm

Peiwei Jin¹, Yan Yao³*, Qian Qiu¹, Huijun Liu¹, Xiaoyu Liang¹, Jinhui Cai¹
¹college of metrology & measurement engineering, China Jiliang university, Zhejiang, 310018, China
³Corresponding author’s e-mail: yaoyan@cjlu.edu.cn

Abstract. China is one of the largest bamboo producers in the world. Bamboo has the characteristics of large biomass and fast growth, which is considered to be an ideal biomass energy source. In order to satisfy the needs of modern biomass conversion processes, there is an urgent need to develop a method for rapidly determining the heating value of bamboo. In this paper, the spectral data of 80 bamboo samples were obtained by Near-Infrared Spectroscopy (NIRS), and then the heating value of bamboo was measured by conventional methods. Through the analysis of bamboo sample data, the Partial Least Squares Regression (PLSR) algorithm, and the Competitive Adaptive Weighted Sampling and PLSR combination (CARS-PLSR) algorithm were used to establish the relationship between spectral data and heating value. The results show that the number of characteristic wavelengths obtained by the CARS algorithm is 24, while the wavelength of the original medium-long wavelength band is 1320, which reduces the number of wavelengths by 98.18%, and greatly reduces the data size. In addition, the selected feature wavelengths can better reflect the key information of the bamboo spectrum, which have a better model effect than the unscreened model. In this study, the heating value index (R² c, R² p, RPD) of the CARS-PLSR model was 0.97, 0.98, and 6.22, respectively. The bamboo heat generated PLSR model was established by using the characteristic wavelength selected by CARS algorithm, which greatly simplified the complexity of the model and improved the predictive ability of the model.

1. Introduction
Energy shortage has become a worldwide problem. With the rapid development of China’s economy and the continuous improvement of people’s living standards, energy issues have become increasingly prominent [1]. As the world’s fourth largest energy source, biomass energy is located behind the three fossil energy sources of coal, oil and natural gas [2], and is one of the most promising renewable energy sources [3]. At present, biomass provides only 15%~20% of global energy demand, but it is more important for developing countries, especially in rural areas [4].

Born in China, bamboo is a forestry resource with Chinese characteristics, which has valuable features such as large biomass, cluster growth, short material production period and easy breeding [5] and great advantages as a biomass energy development and utilization. China is one of the countries with the widest distribution of bamboo in the world, mainly distributed in the south, and concentrated in Fujian, Zhejiang, Jiangxi, Anhui, and Sichuan, Guangxi, Yunnan and some other provinces in the western China [1]. According to the results of the Eighth Forest Resources Survey, the existing artificially operated bamboo forests in China are over 6 million hectares and are growing at an annual...
rate of over 150,000 hectares [6]. In China, as bamboo’s planting area is large and still in rapid growth, it can be used as an ideal source of biomass. In recent decades, as the forest area on the earth's surface has been decreasing year by year but the area of bamboo forest has been expanding, and combined with that bamboo is an valuable biomass resources, it can be foreseeable that bamboo will play an increasingly important role in China's sustainable development strategy.

The heating value is an indicator of the energy content of the reaction fuel [7-8], which refers to the amount of heat generated per unit mass of fuel when it is completely burned. The heating value can be further divided into high heating value and low heating value. The high heating value refers to the heating value, which generated by the unit mass of fuel when it is completely burned, minus the heat generated by nitric acid and the heating value obtained after the sulfuric acid is corrected. The lower heating value is the heating value obtained by subtracting the heat of vaporization of water from the higher heating value. The heating value directly affects the availability of biomass energy. At present, there are two direct measurement methods and approximate analysis methods for fuel heating value [9]. The direct measurement method is based on the oxygen bomb method and is directly measured by a calorimeter. Although this method is relatively simple and inexpensive, it is a time-consuming and tedious work [10]; the approximation analysis uses a correlation between the composition and the heating value to obtain a simple prediction model to predict the heating value. The main approximation analysis methods include three methods of chemical components, elemental components and industrial component analysis. Although this method avoids dependence on lengthy experimental techniques [11], the data for mathematical models need to build relying on expensive experimental equipment, require professional technicians for experimental operation, and consume lots of labor and financial resources. Due to the widespread use of biomass materials, the traditional analysis speed has not met the needs of biomass thermal conversion process, it is necessary to explore a new rapid heating value measurement method.

NIRS is considered as a fast, non-destructive, non-polluting analytical method that can be used to evaluate the properties of biomass [12]. As a faster laboratory alternative, NIRS has been successfully applied to the heating value detection of a variety of biomass energy sources. Huang [13] and others used NIRS to establish a heating value prediction model of straw, and combined with Local algorithm to explore the model optimization, which confirmed that NIRS can realize the determination of straw heating value. Xiong [14] and others based on NIRS to establish a prediction model for the high heating value of five crop straw species. The results show that NIRS can provide basic model support for the rapid detection of high-heating value of crop straw. Sheng [15] and others found that the rapid detection of biomass material characteristics is of great significance for the production of high-quality compression-molded fuel. Therefore, the NIRS technique was used to establish the heating value prediction model of pine, fir and cotton stalks. The evaluation indexes R2 and RPD are 0.87 and 2.73, respectively. The experimental results show that the NIRS technology can completely replace the traditional industrial analysis method to achieve rapid determination of heating value. In addition to the widespread use of crop waste, bamboo has been actively used in the production of fuel as a material widely used in Asia. Thailand is one of the major bamboo producing countries in the Asia-Pacific Bamboo Region. It consists of 13 genera and 60 species, accounting for 5% of the world's total. Jetsada Posom [16-17] et al. used Fourier transform NIRS instead of the traditional bomb calorimetry to determine the heating value of Thailand bamboo, and used PLSR algorithm to establish predictive model of different positions of 80 bamboo buds. The model can better predict the characteristics of bamboo samples, thus improving the efficiency of bamboo as a biomass energy conversion process.

It can be seen that the NIRS technology is expected to achieve rapid detection of heating value. However, there are more than 500 kinds of bamboo species distributed in China, and the chemical composition of different varieties is also different, resulting in different information carried by the near-infrared spectrum, and the bamboo heating value prediction model that Jetsada Posom studied only includes few bamboo species and is not applicable to China. In this paper, NIRS is used to obtain the spectral information of bamboo samples. PLSR algorithm and combination algorithm of CARS
and PLSR are used to establish a bamboo heating value prediction model for long spectral bands and characteristic bands to achieve rapid acquisition of bamboo heating value.

2. Materials and Method

2.1 Sample preparation
The bamboo samples were collected from the China Bamboo Expo Park in Anji County, Zhejiang Province. 40 bamboo varieties including Guizhu, Danzhu and Leizhu were collected. A total of 80 experimental samples including 40 bamboo leaves and 40 bamboo branches were sampled from the different growth parts of bamboo. The collected fresh bamboo samples were placed in an outdoor environment at a temperature of 37°C to be air-dried to dryness. After the sample is separated from the external moisture, it is placed in a 65°C electric heating blast drying oven (Hangzhou Blue Sky Laboratory Instrument Factory, DHG-9070A) for 6 hours, and then pulverized and dried by a pulverizer (Xulang, XL-10B). The sample is placed in a sealed plastic bag through a 60 mesh screen (0.25 mm), placed in a sealed container, stored in a desiccator at room temperature (25°C), and coated with Vaseline at the dryer ground. In this way, the sample is stored until the experiment is completed.

2.2 Heating value determination method
The commonly used method for direct determination of heating value is the oxygen bomb method. The oxygen bomb calorimeter (Shanghai Changji Geological Instrument, HRY-1B) is the main equipment for heating value determination. The experiment was carried out in accordance with the national standard GB/T 213-2008 coal heating value determination method [18]. Before the experiment, the calorimeter was first calibrated with benzoic acid, and the calibration experiment was performed at least 5 times until the relative error of the 5 results was less than 0.2%. The experiment was carried out by wrapping 0.8 g ± 0.1 g of the experimental sample with a known quality of rubbing paper. The weighing required during the experiment was carried out by using a METTLER TOLEDO electronic analytical balance. Each sample is made in three parallels and averaged. The measured result is the heating value of the cartridge of the sample. In practical applications, it needs to be converted into high heating value and low heating value. The heating value of the cartridge minus the heat of nitric acid formation and the heat of sulfuric acid is the high heating value. The value obtained by subtracting the heat of dehydration from the high heating value is the low heating value. The sulfur content and hydrogen content required to calculate the high and low heating values were determined using an EA3000 elemental analyzer (EURO, Italy). When the S content is less than 0.1%, it is negligible. The sample moisture was measured in a blast drying oven according to the national standard GB/T 28731-2012 [19]. The formula for calculating the high and low heating values is shown as follows:

\[
HHV = Q_b - 94.1S - 0.0012Q_b \\
LHV = HHV - 206H - 23M
\]

Among them, HHV is the high heating value of the sample, Qb is the bomb heating value of the sample, S is the sulfur content of the sample, 94.1 is the correction value of each 1% sulfur in the sample, 0.0012 is the heat of nitric acid formation, and LHV is the low heating value of the sample, H is the hydrogen content of the sample, and M is the moisture content of the sample.

2.3 Near infrared spectroscopy
Powdered samples usually use diffuse reflection analysis to obtain chemical information. The light signals obtained by diffuse reflection analysis mainly come from the shallow layer of the sample. The short-wave near-infrared penetrating ability is strong, the reflected back light is weak, and the short-wave diffuse reflected light is not loaded with sufficient chemical information for measurement. Therefore, when using the diffuse reflection analysis method, the near-infrared short-wave region
should not be applied to chemical analysis. Therefore, the medium-long wave (1100nm–2500nm) is selected for NIRS quantitative analysis. Savitzky-Golay Convolution Smoothing (SG) is used to smooth the spectral data to eliminate noise, and combined with Multi-Scatter Correction (MSC) to eliminate the influence of sample particle distribution and particle size.

3. Results and discussion

3.1 High and low heating value measurements
The high and low heating value and related component determination results of 80 samples are shown in Table 1. The moisture content of the sample ranged from 4.49% to 8.97%, the average value was 5.83%, and the standard deviation was 0.92. The hydrogen content ranged from 5.26% to 6.34%, the average value is 5.84%, and the standard deviation is 0.21. The sulfur element content ranges from 0% to 0.68%, the average value is 0.14%, and the standard deviation is 0.11. The high heating value range is 16077.38 J/g–19127.72 J/g, the average value is 18050.26 J/g, and the standard deviation is 670.73. The low heating value range is 16064.72 J/g–19113.79 J/g, the average value is 18036.88 J/g, and the standard deviation is 670.45. The variation range of sample moisture content is relatively wide, and the standard deviation is large, indicating that the sample source is wide, which can represent most of the bamboo energy utilization research; the average content of S is only 0.14%, which is much smaller than 1% of the standard sulfur content of coal in China, and can be reasonably developed and utilized without affecting the ecological environment and is an environmentally friendly renewable energy source. Its high and low heating value are similar with large range span, which has a high representativeness.

Table 1. Statistical results of bamboo heating value and related components

| Characteristic | Bamboo | Average value | Standard deviation |
|---------------|--------|---------------|-------------------|
| Moisture (%)  | 4.49–8.97 | 5.83          | 0.92              |
| H (%)         | 5.26–6.34 | 5.84          | 0.21              |
| S (%)         | 0–0.68   | 0.14          | 0.11              |
| High heating value (J/g) | 16077.38–19127.72 | 18050.26 | 670.73          |
| Low heating value (J/g)  | 16064.72–19113.79 | 18036.88 | 670.45           |

3.2 Near infrared spectroscopy acquisition results
Figure 1 is a plot of the original medium-long-wave NIRS data for 80 samples, which uses PCA combined with Mahalanobis distance to eliminate sample outliers. The wavelengths covered in the figure range from 1100 nm to 2500 nm, and the ordinate is the sample concentration corresponding to each wavelength. The chemical constituents of bamboo are mainly cellulose, hemicellulose and lignin. Cellulose’s structure is mainly composed of C-H, N-H and O-H bonds. It can be seen from the figure that the absorption peaks of the hydrogen-containing groups of the sample mainly appear near 1450 nm, 1720 nm, 1940 nm, 2150 nm, and 2350 nm. These absorption peaks reflect the constituent groups of the main active components of the sample, and the O-H double-frequency absorption band is around 1450 nm. The vicinity of 1720 nm is the C-H double frequency absorption band; the vicinity of 1940 nm is the absorption band of O-H and H2O; the vicinity of 2150 nm is the combined absorption band of N-H; and the vicinity of 2350 nm is the combined absorption band of C-H. The NIRS results of the collected samples accord with the chemical composition of the bamboo.
3.3 Model evaluation index

The criteria for judging the pros and cons of the model include the pros and cons of the model itself and its predictive ability. This paper uses the following criteria to evaluate the model:

1) Decisive factor

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

(3)

R2 represents the degree of fit between the measured and predicted values of the model. The closer R2 is to 1, the better the model effect. \( y_i \) is the measured value of the i-th sample, \( \hat{y}_i \) is the predicted value of the i-th sample, \( \bar{y} \) is the average value of all samples, and \( n \) is the number of samples.

2) Root Mean Square Error (RMSE)

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

(4)

RMSE represents the root mean square error of the training set. When nonlinearly fitting, the smaller the RMSE, the better the result.

3) Root Mean Square Error Of Prediction (RMSEP)

\[ RMSEP = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \]  

(5)

RMSEP is an important indicator to predict the prediction accuracy of the prediction set. The smaller the RMSEP value, the closer the predicted value is to the measured value.

4) Root Mean Square Error of Cross Validation (RMSECV)

\[ RMSECV = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \]  

(6)

RMSECV is an important indicator of the feasibility of modeling methods and the ability of model prediction.

5) Standard Deviation (SD) of prediction set

\[ SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2} \]  

(7)

SD represents the degree of deviation between the predicted value of the prediction set and the mean.
(6) Relative Percent Deviation

$$RPD = \frac{SD}{RMSEP}$$ \hspace{1cm} (8)

RPD is the relative analysis error of the prediction set and is an important evaluation parameter for evaluating the resolving power of the model. When RPD > 3, the model has good predictive ability and can be used for practical application; when 2.5 < RPD < 3, the model is feasible with good prediction effect; when 2 < RPD < 2.5, the model prediction effect is not good and can be used for approximate analysis.

3.4 Medium and long waveband PLSR modeling

Using SPXY partitioning, 80 samples were divided into training and prediction sets in a ratio of 3:1. The PLSR algorithm is used to establish a heating value prediction model for the long-wavelength of the sample. The RMSECV value and the model RPD value obtained by leaving the cross-validation method are used to confirm the number of hidden variables of the model. Figure 2 and Figure 3 show the relationship between the number of hidden variables and RMSECV and RPD in the PLSR modeling process. When the number of latent variables is 15, the RMSECV of the thermal value reaches the minimum and the RPD reaches the maximum.

3.5 Characteristic wavelength screening

Reducing the model wavelength by wavelength selection can better explain the relationship between the digitized spectrum and the characteristics to be studied [20]. The CARS characteristic wavelength screening algorithm is based on the principle of survival of the fittest in Darwin's theory of evolution, with the advantages of faster calculation speed and fewer characteristic wavelength need to be screen, and have been widely used in the screening of Near-Infrared characteristic variables. The characteristic wavelength screening is performed by the CARS algorithm, the number of MSCs is set to 100, and the screening process and results are shown in Figure 4. Figure 4 shows the number of variables involved in modeling during the CARS algorithm screening process, the RMSE of the adaptive weighted sampling model, and the regression coefficient of each variable as a function of the number of MSCs. From the figure, the number of MCS is 62. When the RMSE reaches a minimum value, the number of characteristic wavelengths screened is 24.
3.6 CARS-PLSR modeling
Using SPXY partitioning, 80 samples were divided into training and prediction sets in a ratio of 3:1. The PLSR algorithm is used to establish the heating value prediction model for the characteristic wavelengths selected by CARS. Figures 5 and 6 show the relationship between the number of hidden variables and RMSECV and RPD in the CARS-PLSR modeling process respectively. When the number of latent variables is 13, the RMSECV of the heating value is minimized while the RPD is maximized.

3.7 PLSR and CASR-PLSR modeling results
Table 2 shows the results of PLSR modeling in the medium and long-band PLSR modeling and the CARS characteristic wavelength screening. The results of the high and low heating value models are similar. Taking HHV as an example, using PLSR, the thermal index training set model index R2c of the 1320 wavelengths in the long-wavelength of the sample reaches 0.98, the prediction set model
index $R_{2p}$ reaches 0.90, and the RPD value is greater than 2.5, which is 2.89. High predictive ability, the model is feasible. The number of characteristic wavelengths obtained by the wavelength screening of CARS algorithm is 24, which is greatly reduced compared with the mid-long band spectrum, which is only 1.8% of the medium-long spectrum. The prediction ability of the model is greatly improved, and the prediction set $R_{2p}$ is 0.98, which is increased. 0.08, RMSEP is 124.37, which is 53.10% lower than that of medium-length spectrum. The RMSE after wavelength screening is increased by 12.98%, and the RPD is increased from 2.89 to 6.22, which is far beyond the practical application of models with RPD value of 3. The heating value prediction model established by using CARS algorithm combined with PLSR is better than the PLSR model without wavelength screening, which can be used in practical applications. Figure 7 and Figure 8 are the results of the high and low heating value prediction models built by the two methods.

| Modeling method | Expected target | Number of hidden variables | Number of wavelengths | $R_{2c}$ | $R_{2p}$ | RMSE | RMSEC | V RMSE | RP | D |
|-----------------|----------------|---------------------------|-----------------------|---------|---------|------|-------|-------|----|---|
| PLSR            | HHV            | 15                        | 1320                  | 0.98    | 0.90    | 265.19| 272.08| 89.87 | 2.89|   |
|                 | LHV            | 15                        | 1320                  | 0.98    | 0.90    | 265.18| 272.07| 89.86 | 2.88|   |
| CARS-PLSR       | HHV            | 13                        | 24                    | 0.97    | 0.98    | 124.37| 127.60| 103.26| 6.22|   |
|                 | LHV            | 13                        | 24                    | 0.97    | 0.98    | 124.39| 127.62| 103.25| 6.22|   |

Figure 7. PLSR model prediction diagram
4. Conclusion

In this paper, the spectrum of bamboo in Zhejiang Province was determined by NIRS. The rapid prediction models of bamboo heating value were established by using PLSR and CARS-PLSR modeling methods. Based on Darwin’s “survival of the fittest” principle, the CARS algorithm selects a combination of key wavelengths in an efficient and competitive way, which reduces redundant information. The 24 characteristic wavelengths obtained by CARS screening were involved in the PLSR modeling. The model evaluation indexes $R^2_c$, $R^2_p$ and $RPD$ were 0.97, 0.98 and 6.22, respectively. Compared with the PLSR modeled by the long-wavelength 1320 wavelengths in the sample, the data scale was greatly simplified, and the model effect has been greatly improved, which is almost a perfect model. The results show that wavelength screening is very necessary. Selecting several characteristic wavelengths with chemical significance has better prediction effect than using continuous wavelength bands. The rapid prediction model of bamboo heating value established by CARS-PLSR is outstanding which can provide reference for the research and development of heating value rapid testing equipment, and has certain application value.

Acknowledgments

Authors wishing to acknowledge assistance or encouragement from colleagues, special work by technical staff or financial support from organizations should do so in an unnumbered Acknowledgments section immediately following the last numbered section of the paper. This work was funded by the Development and Application of High Throughput Organic Elemental Analyzer (National Key R & D Program 2018YFF01014002) and the title of the project is the application development and method verification of organic matter element analysis in biomass fuel.

References

[1] Zhou, B.Z., Fu, M.Y., Yang, X.S., et al. (2006) Energy bamboo plant resources and its development potential in China. World Forestry Research., 19(6): 49-52.
[2] Liu H.X., Feng Y.M. (2015) The current situation and future development trend of biomass energy in the world. World Agriculture., 05: 117-120.
[3] China electricity council. (2016) Biomass energy: the most potential renewable energy. http://www.cec.org.cn/xinwenpingxi/2016-02-25/149454.html.
[4] Vargas-Moreno J.M., Callejón-Ferre A.J., Pérez-Alonso J., et al. (2012) A review of the mathematical models for predicting the heating value of biomass materials. Renewable and Sustainable Energy Reviews., 16(5): 3065-3083.
[5] Chen B.K., Yang Y.M., Zhang G.X., et al. (2007) Study on the cultivation technology and comprehensive utilization of large clumped bamboo. Journal of West China Forestry Science., 36(2): 1-9.

[6] Lin M.Y. (2019) Present situation of bamboo development and utilization in China. China Southern Agricultural Machinery., 50(02): 105.

[7] Erol M., Haykiri-Acma H., Kucukbayrak S. (2010) heating value estimation of biomass from their proximate analyses data. Renew Energy., 35: 170-173.

[8] Sheng C., Azevedo J.L.T. (2005) Estimating the higher heating value of biomass fuels from basic analysis data. Biomass Bioenergy., 28: 499-507.

[9] Posom J., Sirisomboon P. (2017) Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. Journal of Near Infrared Spectroscopy., 25(5): 301-310.

[10] Hosseinipour S., Aghbashlo M., Tabatabaei M. (2018) iomass higher heating value (HHV) modeling on the basis of proximate analysis using iterative network-based fuzzy partial least squares coupled with principle component analysis (PCA-INFPLS). Fuel., 222: 1-10.

[11] Kathiravale S., Yunus M.N.M., Sopian K., et al. (2003) Modeling the heating value of Municipal Solid Waste. Fuel., 82(9): 1119-1125.

[12] Lestander T.A., Johnsson B., Grothage M. (2009) NIR techniques create added values for the pellet and biofuel industry. Bioresource Technology., 100(4): 1589-1594.

[13] Huang C.J., Han L.J., Liu X., et al. (2009) Fast analysis of straw moisture and heating value by near infrared spectroscopy based on LOCAL algorithm. Journal of Infrared and Millimeter Wave., 28(3): 301-306.

[14] Xiong X.Q., Qian S.P., Sheng K.C., et al. (2017) High heating value of crop straw was predicted based on industrial analysis/elemental analysis and visual-near-infrared spectroscopy. Spectroscopy and Spectral Analysis., (05): 301-306.

[15] Sheng K.C., Sheng Y.Y., Yang H.Q., et al. (2012) Rapid determination of components and heating value of biomass based on spectroscopic technique. 32(10): 2805-2809.

[16] Posom J., Sirisomboon P. (2017) Evaluation of lower heating value and elemental composition of bamboo using near infrared spectroscopy. Energy., 121: 147-158.

[17] Posom J., Sirisomboon P. (2017) Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. Journal of Near Infrared Spectroscopy., 25(5): 301-310.

[18] Standardization Administration of the People's Republic of China. (2008) GB/T 213-2008 Methods for determination of heating value of coal. Standards Press of China, Bei Jing.

[19] Standardization Administration of the People's Republic of China. (2012) GB/T 28731- 2012 Industrial analysis methods for solid biomass fuels. Standards Press of China, Bei Jing.

[20] Li H., Liang., Xu Q., et al. (2009) Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Analytica Chimica Acta., 648(1): 77-84.