XV WORLD FORESTRY CONGRESS
Building a Green, Healthy and Resilient Future with Forests
2–6 May 2022 | Coex, Seoul, Republic of Korea

Near-infrared Analysis and Models Optimization about Main Chemical Components of Pulpwood in Hainan Province

Ting Wu[1*]

[1] Institute of Chemical Industry of Forest Products, Chinese Academy of Forestry, China - wuting@icifp.cn

Abstract

In order to improve the utilization of pulpwood in Hainan Province, alleviate the shortage of pulping materials, reduce pollution and overall costs in Chinese pulping and papermaking industry, this study aimed to use near-infrared spectroscopy to realize the rapid analysis of holocellulose and Klason lignin content in pulpwood. A holographic grating near-infrared spectrometer, with a simple structure and easy modification, was used to collect the near-infrared spectrum of 205 samples of pulpwood common in Hainan (*E.urophlla× E.tereticornis, Eucalyptus urophylla×grandis, Eucalyptus urophylla, Acacia mangium, Acacia crassicarpa Benth.*), and the content of holocellulose and Klason lignin were measured according to the traditional laboratory methods. Suitable pretreatment methods were selected in combination with partial least squares (PLS) to establish analysis models of holocellulose and Klason lignin. Then genetic algorithm was used to eliminate the irrelevant variables and clarify the feature absorption of holocellulose and Klason lignin in order to optimize the models. The holocellulose model was established by pretreatment methods of smoothing, vector normalization, first derivative of the original spectrum, with 1150.3~2362.0nm bands participated in modeling. The RMSEP value of the model was 0.55% and the absolute deviation range was -0.91~0.87%. The Klason lignin model was established by pretreatment methods of smoothing, multiplicative scatter correction(MSC), second derivative of the original spectrum, with 1137.6~1872.5nm and 2131.0~2424.1nm bands participated in modeling. The RMSEP value of the model was 0.45% and the absolute deviation range was -0.76~0.79%. The RPD values of the two models were 4.71 and 3.47, respectively, which can meet the industrial needs of online rapid analysis. At the same time, this study provides a theoretical basis for the establishment of a near-infrared characterization system for pulpwood.

Keywords: near-infrared; Genetic Algorithm; characteristic wavelengths; holocellulose; Klason lignin

Introduction, scope and main objectives

The pulping and papermaking industry is important related to the world economy and people's livelihood. From 2008 to 2018, China has become the world's largest pulping and papermaking country with an average annual growth rate of about 2.4% in the output of paper and paperboard. However, the production of paper and paperboard in 2019 and 2020 has declined compared with the previous year (Zhao, 2021). The reasons are firstly that China has stepped up restrictions on the import of waste paper, and secondly, the supply of wood pulp is insufficient. With the impact of the trade war and the novel coronavirus, Chinese restrictions on imported waste paper will only become stricter, and domestic wood pulp supply cannot increase significantly in the short term. Improving the utilization efficiency of domestic pulpwood has become a feasible solution for the crisis of pulping and papermaking materials (Shen et al. 2020). Due to the different species and growth conditions of pulpwood, there are obvious differences in the content of main chemical components. In order to ensure product quality, high power consumption input and excessive dosing are usually used, which also lead to high production costs and serious pollution. Among the pulpwood chips, holocellulose is directly related to the yield of pulp, while Klason lignin is related to the whiteness of the product. Achieving rapid

analysis of both in pulpwood chips and regulating the pulping process parameters accordingly is expected to improve the utilization efficiency of the pulpwood and reduce the pollution and overall cost of the industry (Liang et al. 2019).

In recent years, near-infrared (NIR) spectroscopy has become a commonly used rapid analysis method, and has played an important role in agriculture, biochemistry and other fields. Applying near-infrared spectroscopy to the composition analysis of pulpwood, in addition to solving practical problems in the pulping and papermaking industry, it also clarifies the characteristic absorption of complex chemical components such as holocellulose and Klason lignin, in order to further form a near-infrared characterization system for pulpwood (Wu et al. 2018). Hainan Province has superior natural conditions such as climate, soil, and precipitation, as well as abundant forest resources and excellent ecosystems. In this study, the common pulpwood in Hainan Province were used as samples, the analysis models were established by traditional partial least squares (PLS), and the models were optimized by genetic algorithm to improve accuracy and stability. Characteristic bands selected for modeling were explained from the component structure of holocellulose and Klason lignin. Finally, this study is promising to meet the practical needs of pulping and papermaking industry for rapid analysis of pulpwood, while providing a theoretical basis for the establishment of a near-infrared characterization system for pulpwood.

Methodology/approach

1-Samples selection

The pulpwood forests in Hainan Province are dominated by *E.urophlla× E.tereticornis*, *Eucalyptus urophylla× grandis* and *Eucalyptus urophylla*, mixed with *Acacia mangium* and *Acacia crassicarpa Benth*. 5 species of pulpwood above were collected from the pulpwood forests of Ledong, Changjiang, and Qiongzhong. The trees were 5~6 years old, and 205 discs were collected at equal distances from the trunk after bark removal. Then cut into wood chips, placed in air to fully equilibrate the moisture and ground, intercepting wood powder between 0.25~0.42 mm in particle size as samples (Liang et al. 2020). After collecting the spectrum and measuring the chemical composition, 8-9 samples of each pulpwood species were chosen as a validation set for independent verification according to the overall ratio of 4:1 (Acquah et al. 2018). The other 164 samples were used as a training set for modeling (see Table 1).

Table 1: Details of Hainan pulpwood samples

Wood species	Sources	Age	Samples	
			Training set	Validation set
E.urophlla× E.tereticornis	Ledong	6	31	8
Eucalyptus urophylla×grandis	Ledong, Changjiang, Qiongzhong	5，6	37	9
Eucalyptus urophylla	Ledong, Changjiang	6	33	8
Acacia mangium	Ledong, Qiongzhong	6	32	8
Acacia crassicarpa Benth.	Ledong, Changjiang	5	31	8
Total	——	——	164	41

2-Spectrum collection

The near-infrared spectrometer adapted to the pulping and papermaking production line has practical requirements such as easy construction, easy modification, low cost, and strong environmental adaptability. A

simple structured holographic grating spectrometer (Shanghai Fuxiang, NIR2510) was selected to collect the NIR spectrum of samples with a wavelength range of 900-2500 nm and 256 wavelength points. Each sample was sampled five times to take the average spectrum.

3-Analysis of holocellulose and Klason lignin content

After collecting near-infrared spectrum of 205 training set and validation set samples, the content of holocellulose and Klason lignin were measured according to GB/T 2677.10-1995 and GB/T 2677.8-1994.

4-Modeling and evaluation

The spectrum were pre-processed by combined methods and thereafter modeled in combination with partial least squares (PLS), respectively. The characteristic bands were screened by genetic algorithm (GA) to further optimize the holocellulose and Klason lignin analysis models. The cross validation was evaluated by R^2_{cv} and root mean square error cross validation (RMSECV). Models performance were evaluated by R^2_{val}, root mean square error of prediction (RMSEP), residual predictive deviation (RPD), and absolute deviation range (AD) as reference indicators (Hodge et al. 2018), and models systematic error was characterized by bias value (Bias).

Results

1-Distribution of measured values of holocellulose and Klason lignin content

The holocellulose and Klason lignin content of 205 samples were analyzed according to the national standard method(see Table 2). The holocellulose content of three species of *Eucalyptus* was distributed between 77.59% and 81.63%, and the holocellulose content of *Acacia* was significantly lower than that of *Eucalyptus*, distributed between 72.96% and 78.45%. The overall Klason lignin content of *Eucalyptus* and *Acacia* was not significantly different, both distributed between 21.01% and 27.96%. The samples covered the main pulpwood species in Hainan, which is expected to establish representative and applicable analytical models.

Table 2: Chemical composition content distribution of samples

Wood species	Holocellulose content(%)				Klason lignin content(%)			
	Range	Mean	Median	Standard Deviation	Range	Mean	Median	Standard Deviation
E.urophlla× E.tereticornis	77.59～81.56	79.68	79.63	1.23	22.04～27.85	25.03	25.26	1.44
Eucalyptus urophylla×grandis	78.04～81.63	79.76	79.74	1.19	21.13～27.96	24.88	25.14	1.67
Eucalyptus urophylla	77.91～81.30	79.66	79.67	1.17	21.55～27.43	24.87	25.18	1.58
Acacia mangium	73.19～78.37	75.50	75.40	1.56	21.01～27.78	24.70	24.98	1.76
Acacia crassicarpa Benth.	72.96～78.45	75.37	75.23	1.67	21.60～27.30	24.62	24.99	1.63
Total	72.96～81.63	78.06	78.42	2.57	21.01～27.96	24.82	25.07	1.82

2- Spectrum pretreatment and models establishment

The NIR spectrum of 205 pulpwood samples were collected at a temperature of 20°C and a relative humidity of 50%. The horizontal axis is the spectral wavelength, and the ordinate reflects the absorption(Fig.1). It can be seen that the original spectrum absorption bands are wide and overlap seriously, which is due to the interference of irrelevant information during the collection process, and the fact that the pulpwood contain a certain proportion of resin, inorganic salt, fat and other substances in addition to holocellulose and Klason lignin, and have a complex physicochemical structure.

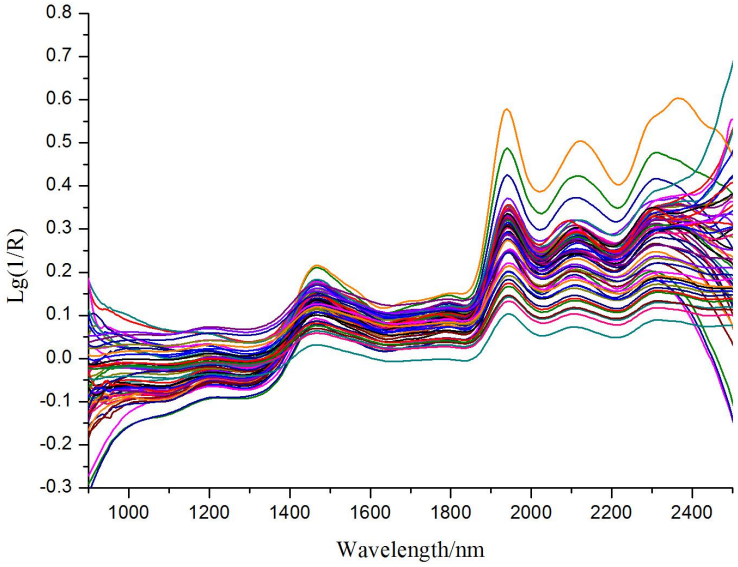

Fig. 1: Original near infrared spectrum of pulpwood samples

Among the common pretreatment methods, smoothing is used to eliminate noise, derivative method is used to deal with baseline and background interference, vector normalization is used to reduce the spectral variation due to small light range differences, and multiple scattering correction (MSC) is often used to mitigate the scattering effects caused by poor homogeneity of wood powder (Nabavi et al. 2018). After using Savitzky-Golay 13-point 3-fold smoothing in Matlab 8.0 for the original spectrum to eliminate noise, the other pretreatment methods were combined and used in conjunction with the PLS method to perform the leave-one-out cross validation modeling for each of the 164 pulpwood samples in the training set, and the effects of using different pretreatment on the modeling analysis were shown in Table 3. The performance of models were poor when the original spectrum were used to build the model directly, when both holocellulose and Klason lignin models had the lowest R^2_{cv} values and the highest RMSECV values, and model performance were somewhat improved by using the smoothing pretreatment alone. When the original spectrum were pretreated with the combination of smoothing, vector normalization, and first derivative (Fig.2-a), and the PLS optimal principal component was 8 to build the holocellulose model, the R^2_{cv} value was the highest at 0.9747; the RMSECV was the lowest at 0.43%. When the original spectrum were pretreated with the combination of smoothing, MSC, and second derivative (Fig.2-b), and the PLS optimal principal component was 11 to built the Klason lignin model, the R^2_{cv} value was the highest at 0.9441; the RMSECV was the lowest at 0.38%.

Table 3: Pretreatment methods selection and modeling analysis

Pretreatment methods	Holocellulose		Klason lignin	
	R^2_{cv}	RMSECV	R^2_{cv}	RMSECV
None	0.9603	0.54%	0.9107	0.48%
Smoothing	0.9652	0.50%	0.9187	0.46%
Smoothing, Vector normalization	0.9724	0.45%	0.9245	0.44%
Smoothing, MSC	0.9673	0.49%	0.9299	0.43%
Smoothing, First	0.9720	0.45%	0.9223	0.45%

derivative				
Smoothing, Second derivative	0.9700	0.47%	0.9325	0.42%
Smoothing, Vector normalization, First derivative	0.9747	0.43%	0.9277	0.43%
Smoothing, Vector normalization, Second derivative	0.9733	0.44%	0.9347	0.41%
Smoothing, MSC, First derivative	0.9730	0.44%	0.9373	0.40%
Smoothing, MSC, Second derivative	0.9703	0.46%	0.9441	0.38%

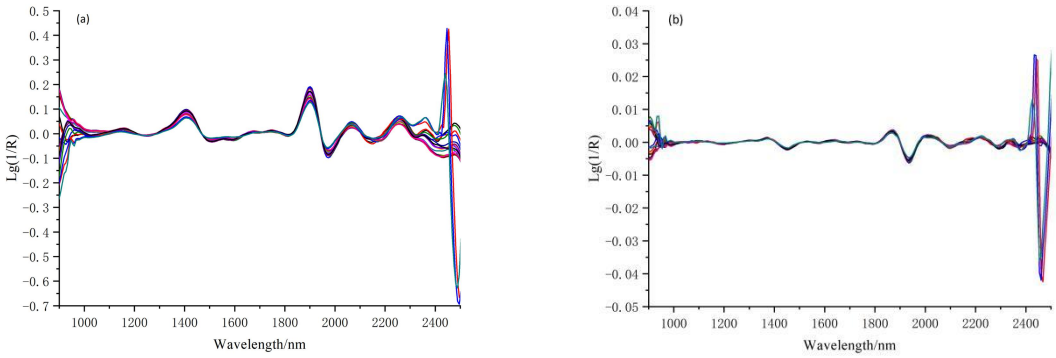

Fig.2: Pretreatment of original spectrum of samples (a. holocellulose; b. Klason lignin)

3- Wavelength selection and model optimization

The combination of pretreatment methods can effectively mitigate the interference caused by factors such as environment and sample particle size, but there is a large amount of redundant information in the wavelength region of the instrument that has little correlation with holocellulose and Klason lignin components, which has a certain impact on models performance. Therefore, after determining the specific pretreatment methods to be used in the modeling process, the genetic algorithm was selected to screen the spectral information for specific wavelengths/bands with strong correlation to holocellulose and Klason lignin, so as to optimize the analytical models performance. The 256 wavelength points in the spectral interval between 900 and 2500 nm were binary coded, labeled as 1 when selected and 0 when rejected, with artificially set number of screening wavelengths in the range of 10-200, population size of 400, evolution algebra of 150, crossover rate of 0.65, and mutation rate of 0.08. It runs in Matlab 8.0 software.

For the holocellulose content in pulpwood samples, the set of wavelength point variables screened after each evolutionary generation were modeled by PLS, and the RMSECV was used as the evaluation criterion. The final screened optimal bands were 1150.3-2362.0 nm. For the Klason lignin content in pulpwood samples, the same method was used and the optimal bands were finally screened from 1137.6 to 1872.5 nm and 2131.0 to 2424.1 nm.

Discussion

The near-infrared spectrum of samples from the validation set were analyzed with the two models for independent verification and compared with the performance of the models established using the complete wavelength range from 900 to 2500 nm (see Table 4). The RMSEP of the two models established after wavelength selection were 0.55% and 0.45%, and the AD ranges were -0.91～0.87% and -0.76～0.79%, respectively, which had smaller RMSEP and narrower AD ranges compared with the models established using the full wavelength bands. This indicates that the genetic algorithm eliminates irrelevant variables, eliminates the interference of irrelevant information, can effectively perform wavelength screening, and simplifies models while making them more predictive and more robust. The RPD values of the holocellulose and Klason lignin models are 4.71 and 3.47, which can be applied to the fast analysis and real-time control of the production line. The scatter plots of the prediction of the validation set according to the two models are shown in Fig. 3. It can be seen that the points obtained from the holocellulose model are more on the left side of the y=x line, and the Bias value is about 0.0466%, indicating that the overall predicted value is slightly higher than the actual measured value. The model has a small systematic error. The Klason lignin model has more points on the right side of the y=x line, and the Bias value is about -0.0456%, which means that the overall predicted value is slightly lower than the measured value, and the model also has a small systematic error.

Table 4: Performance evaluation of models

Components	Bands	R^2_{val}	RMSEP/%	RPD	AD range/%
Holocellulose	1150.3～2362.0nm	0.9548	0.55	4.71	-0.91～0.87
	900～2500nm	0.9455	0.60	4.29	-0.94～1.03
Klason lignin	1137.6～1872.5nm 2131.0～2424.1nm	0.9169	0.45	3.47	-0.76～0.79
	900～2500nm	0.8890	0.52	3.00	-0.80～0.85

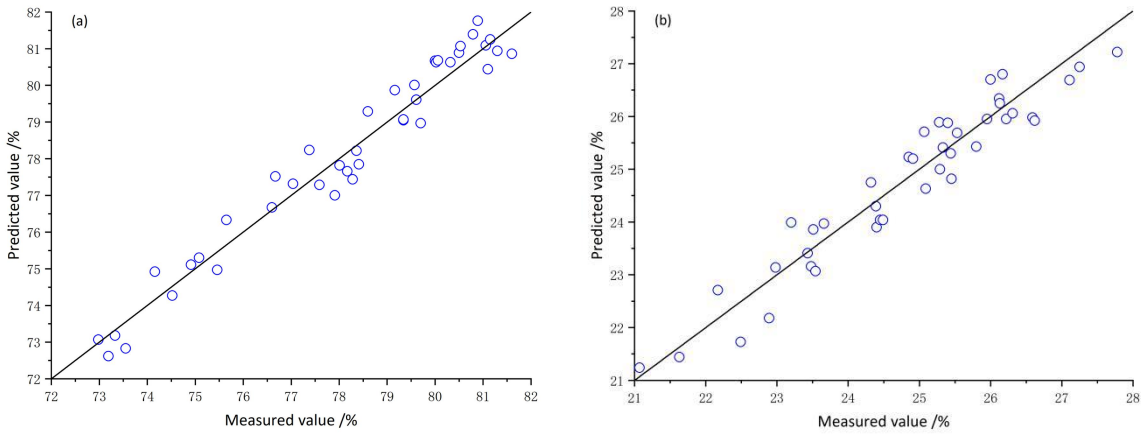

Fig.3: The distribution of scatter points in independent verification (a. holocellulose; b. Klason lignin)

The holocellulose includes cellulose and hemicellulose, and in the pulpwood species of Hainan Province, the hemicellulose is in the form of pentosan. Therefore, the NIR characteristic absorption of hemicellulose is mainly the NIR characteristic absorption of cellulose and pentosan. In the case of cellulose, the 2nd overtone of C-H stretching vibration in CH_3 exists between 1188 and 1196 nm (Satoru et al. 2003), the group frequencies absorption of the 1st overtone of C-H stretching vibration and deformation vibration in CH_3 may exist near 1366 nm, the 1st overtone of O-H stretching vibration exists in the interval of 1742-1833 nm. The absorption

of the 1st overtone of C-H stretching vibration in CH_2 is near 1703 nm (Yamamoto et al. 2007), and the absorption of the 1st overtone of C-H stretching vibration in CH_3 is concentrated near 1720 nm, the group frequencies absorption of O-H deformation vibration and O-H stretching vibration are present at 2112 nm. The group frequencies of O-H and C-C stretching vibration, the group frequencies of C-H deformation and stretching vibration are present near 2275 nm (Tsuchikawa et al. 2005). In the case of pentosan, the group frequencies of the 1st overtone of C-H stretching and deformation vibration in CH_3 exist in the spectral interval from 1350 to 1370 nm, the 1st overtone of O-H stretching vibration exists between 1470 and 1495 nm, the 1st overtone of C-H stretching vibration in CH_3 exists at 1665 nm, 1680 nm and 1721 nm. The 1st overtone of C-H stretching vibration in CH_2 exists at 1704 nm, the 2nd overtone of C=O stretching vibration absorption exist near 1906 nm and 1911 nm (Fackler et al. 2010), the group frequencies of O-H stretching and deformation vibration and C-H deformation vibration exist near 2085 nm. The group frequencies of C-H stretching and deformation vibration exist between 2325 and 2335 nm (Schwanninger et al. 2011). The characteristic absorption range of cellulose in holocellulose is about 1188-2335 nm, the characteristic absorption range of pentosan in holocellulose is about 1350-2335 nm, which are basically included in the bands screened by using genetic algorithm.

The main groups of Klason lignin are methoxy ($-OCH_3$), hydroxyl (-OH), carbonyl (C=O) and benzene ring. In NIR spectrum of samples, the 2nd overtone absorption of C-H stretching vibration in benzene ring and C-H stretching vibration in CH_3 are around 1143 nm, and the 2nd overtone absorption of C-H antisymmetric stretching vibration in HC=CH is around 1172 nm (Sandak et al. 2011). The 1st overtone of O-H stretching vibration of phenolic hydroxyl group may exist at 1408 nm and 1449 nm (He and Hu 2013), the 1st overtone of C-H stretching vibration in benzene ring exists on the interval from 1670 to 1684 nm. The 1st overtone of C-H stretching vibration exists on the interval from 1695 to 1815 nm, the group frequencies of C-H and C=O stretching vibration exist near 2205 nm (Schwanninger et al. 2011), the group frequencies of O-H and C-O stretching vibration exist near 2268 nm. Near 2337 nm there are group frequencies absorption of C-H stretching vibration and CH_2 deformation vibration. The main characteristic absorption of Klason lignin are contained in the bands selected by genetic algorithm.

Conclusions

In this study, the near-infrared spectrum of common pulpwood in Hainan Province were collected, and the analytical models of holocellulose and Klason lignin were developed by PLS method after specific pretreatment. The characteristic bands were screened by genetic algorithm, the characteristic absorptions of holocellulose and lignin were clarified, and the models were optimized. The models RMSEP values were 0.55% and 0.45%, with AD ranges from -0.91 to 0.87% and -0.76 to 0.79%, respectively. Both models can meet the practical needs of rapid analysis in pulping and papermaking industry.

Acknowledgements

The authors are grateful for the support of The Open Project of Key Laboratory of Green Chemical Engineering Process of Ministry of Education (Grants: GCP202104).

The views expressed in this information product are those of the author(s) and do not necessarily reflect the views or policies of FAO.

References

[1]Acquah G. E., Essien C., Via B. K., Billor N., Eckhardt L. G. 2018. Estimating the basic density and mechanical properties of elite loblolly pine families with near infrared spectroscopy. Forest Science, 64: 149-158.

[2]Fackler K., Schwanninger M. 2010. Polysaccharide degradation and lignin modification during brown rot of spruce wood: a polarised Fourier transform near infrared study. Journal of Near Infrared Spectroscopy, 18(6): 403~416.

[3]He W. M., Hu H. R. 2013. Prediction of hot-water-soluble extractives pentosan and cellulose content of various wood species using FT-NIR spectroscopy. Bioresource technology, 140(7): 299~305.

[4]Hodge G. R., Acosta J. J., Unda F., Woodbridge W. C. 2018. Global near infrared spectroscopy models to predict wood chemical properties of Eucalyptus. Journal of Near Infrared Spectroscopy, 26: 117-132.

[5]Liang L., Fang G. G., Deng Y. G., Xiong Z. X., Wu T. 2019. Determination of moisture content and basic density of poplar wood chips under various moisture conditions by near-infrared spectroscopy. Forest Science, 65: 548-555.

[6]Liang L., Wei L., Fang G. G., Xu F. 2020. Prediction of holocellulose and lignin content of pulp wood feedstock using near infrared spectroscopy and variable selection. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 225: 117515.

[7]Nabavi M., Dahlen J., Schimleck L., Eberhardt T. L. 2018. Regional calibration models for predicting loblolly pine tracheid properties using near-infrared spectroscopy. Wood Science and Technology, 52: 445-463.

[8]Sandak A., Sandak J., Negri M. 2011. Relationship between near-infrared (NIR) spectra and the geographical provenance of timber. Wood Science & Technology, 45(1): 35~48.

[9]Satoru T., Siesler H. W., Tsuchikawa S. 2003. Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part I: Softwood. Applied Spectroscopy, 57(6): 667~674.

[10]Schwanninger M., Rodrigues J., Fackler K. 2011. A review of band assignments in near infrared spectra of wood and wood components. Journal of Near Infrared Spectroscopy, 19(5): 287-308.

[11]Shen K. Z., Fang G. G., Li X. L. 2020. China Pulp & Paper, 39(5): 55-62.

[12]Tsuchikawa S., Yonenobu H., Siesler H. W. 2005. Near-infrared spectroscopic observation of the ageing process in archaeological wood using a deuterium exchange method. Analyst, 130(3): 379~384.

[13]Wu T., Fang G. G., Liang L., Deng Y. J., Lin Y., Xiong Z.X. 2018. Analysis of mixed pulping raw materials of Eucalyptus globulus and Acacia mangium by near infrared spectroscopy technique combined with LASSO algorithm. BioResources, 13: 1348-1359.

[14]Yamamoto H., Tsuchikawa S., Fujimoto T. 2007. Estimation of Wood Stiffness and Strength Properties of Hybrid Larch by Near-Infrared Spectroscopy, Applied Spectroscopy, 61(8): 882~888.

[15]Zhao W. 2021. China Paper Industry 2020 Annual Report. China Pulp & Paper Industry, 42(09): 11-21.