Rapid Determination of Cellulose Content in Pulp using Near Infrared Modeling Technique

Chengfeng Zhou,[a] Guangting Han,[a] Shouwu Gao,[a] Meiyi Xing,[b] Yang Song,[c,]* and Wei Jiang [a,d,]*

The feasibility of using near infrared spectroscopy (NIR) to rapidly determine the cellulose content in pulp was investigated in this study. Partial least square regression analyses were performed to describe the relationships between the data sets of wet chemistry analysis and the NIR spectra. The selection of relevant wavenumbers combined with the appropriate data pre-processing methods produced satisfactory prediction models. The test statistics (R^2, RMSECV, and RPD) improved compared with the models over the wavenumber range 10000 cm^{-1} to 4000 cm^{-1}. The predicted cellulose content models, using the cross validation in the appropriate wavenumber ranges coupled with the spectral data preprocessing methods of multiplicative scattering correction (MSC), standard normal variate (SNV), and first derivative (FD) normalization, were established. The highest R^2 value was found to be 0.92 with the lowest RMSECV values 0.60 using FD 19 normalization at the wavenumber range from 7250 to 6500 and 5500 to 4000 cm^{-1}. The highest RPD value was 2.45. NIR spectroscopy, combined with multivariate statistical analysis, could predict cellulose content in the pulp with efficient accuracy.

Keywords: Pulp; Cellulose; Near infrared; Quantitative analysis

*Contact information: a: Laboratory of New Fiber Materials and Modern Textile, The Growing Base for State Key Laboratory, Qingdao University, Qingdao, 266071, P.R. China; b: College of Textiles, Donghua University, Shanghai, 200051, P.R. China; c: School of Basic Medicine, Qingdao University, Qingdao, 266071, P.R. China; d: College of Textile and Garment, Qingdao University, Qingdao, 266000, P.R. China; *Corresponding authors: qdsongyang@126.com; weijiangqd@gmail.com*

INTRODUCTION

Woody biomass is a promising material due to its renewability and abundance (Cheng 2009). Cellulose, hemicellulose, and lignin are the three main chemical components in woody biomass; cellulose has the most applicable value (Rabemanolontsoa *et al.* 2011). In textile industry, woody biomass is traditionally used for producing cellulosic fibers, which is called viscos fiber. Wood pulp is a necessary intermediate product between wood and viscos fiber. Wood pulp contains as much as 95% or higher cellulose content, while the number is very important for the following procedure to produce the viscos fiber (Costa *et al.* 2013).

Traditionally, the cellulose content in pulp is quantified by wet chemistry methods (GB/T 2677.20 1995, TAPPI Test Methods 1996, TAPPI T222 om-11 2011). This method, however, is time consuming and a cause of pollution due to the use of chemicals, which is not applicable to the modern industry with the requirement of fast and clean production

(He *et al.* 2015). Recently, the chemometric method coupled with spectroscopy was proposed to determine cellulose content in pulp/viscos/biomass. These methods include the near-infrared diffuse reflectance spectroscopic (NIRS) method (Mayr *et al.* 2015), the UV/Vis spectroscopic method (He *et al.* 2015), and the thermogravimetric (TG) method (Acquah *et al.* 2017). These methods have the advantages of being fast, free of chemicals, and mostly nondestructive.

Current research mostly focuses on the NIR method for quantitative analysis on pulp or biomass. Axrup *et al.* (2000) utilized a silicon diode array spectrometer as a tool to investigate the components of wood chips and bark samples. Fardim *et al.* (2002) reported on the prediction of chemical composition and physico-chemical characteristics of eucalypt unbleached kraft pulps with the application of NIR spectroscopy. The PLS method was established after first derivative (FD) spectra pre-processing and application of cross validation methodology. The predictive models were time-saving and recommended for process monitoring. Henriksen *et al.* (2004) also reported on the feasibility of NIR spectroscopy to predict the kappa number and viscosity of unbleached pulp. The work measured the cooking liquor instead of pulp and adopted four different preprocessing methods, as well as calibration methods. Results showed that the recommended model was partial least-squares on the first derivative spectra. Furthermore, Terdwongworakul *et al.* (2005) investigated rapid prediction of chemical composition and final pulp yield of *Eucalyptus camaldulensis* wood from Thailand, using NIR spectroscopy. Results could be used as a basis for the characteristics of pulp and paper properties in the industry. Schimleck *et al.* (2007) examined the ability of NIR spectroscopy for estimating poplar pulpwood properties. The study indicated that the NIR spectroscopy provides a rapid alternative method for prediction of a whole-tree basis.

However, studies on fast and accurate analysis of pulp cellulose content were barely found in the literature. In this case, a rapid, non-destructive method for estimating cellulose content in pulp, such as NIR spectroscopy, may provide a useful advancement for the pulp industry.

EXPERIMENTAL

Materials

A total of 108 pulp samples were supplied by CHTC Helon CO., Ltd. (Weifang, Shandong, China). The pulp samples were prepared by dissolving wood using sodium hydroxide, which degrades lignin and hemicellulose and separates cellulose fiber from wood. Each pulp sample was the size of 10 × 10 cm and with 0.5 cm thickness; the weight of each sample was approximately 8 g. After storage in ambient conditions for 1 week, 7.5 g of the air-dried samples were ground to 40-mesh powders using a Willey mill, respectively. And then 2 g of the 40-mesh samples were further ground to 80-mesh powders. The 40-mesh samples were used during wet chemistry analysis, and the 80-mesh samples were used for near infrared reflectance (NIR) spectra collection (Zhou *et al.* 2016).

Wet Chemistry Analysis

The cellulose content was determined using wet chemistry analysis. According to the standard method (GB/T 9107 1999), the sample was treated with 17.5% (w/w) sodium hydroxide, and the solution was passed through filter paper and washed with distilled water until neutral. After drying and weighing, the residue (alpha-cellulose) was calculated. Each specimen was conducted in triplicate; the expected error was less than 0.3%.

NIR Spectra Collection

All 80-mesh samples were oven dried (105 °C ± 2 °C) for 12 h prior to the near infrared analysis. Thereafter each sample was placed in the light path, and the reflectance spectra were collected using a PerkinElmer spectrum Frontier FT-NIR spectrometer (Waltham, MA, USA). For NIR, the spectrum covered a range of 10,000 to 4000 cm^{-1} with a spectral resolution of 4 cm^{-1}. Each spectrum collected represents an average of 32 scans.

Chemometric Analysis

Chemometric modeling was performed with the Spectrum Quant + software (Waltham, MA, USA). A total of 108 NIR spectra were randomly divided into two groups. The first group of 85 spectra was for cross validation (CV), whereas the second group of 23 spectra was for the validation. The following data pre-processing algorithms were applied prior to model building:

1. Multiplicative scattering correction (MSC). First, a mean spectrum was calculated from all the spectra of the calibration data set. Then, each spectrum $X(i)$ was transformed according to $X(i)' = u + v \cdot X(i)$. The coefficients u and v were selected such that the difference between the transformed spectrum $X(i)'$ and the mean spectrum was minimized.

2. Standard normal variate (SNV). First, the spectra were centered. Then, the sum of all squares of all Y–values was calculated, and the respective spectrum was divided by the square root of the sum. The so-called vector norm of the resulting spectrum is always 1.

3. First derivative (FD). The first derivative of the spectrum was calculated.

Wavenumber selection was performed iteratively by comparing and combining wavenumber ranges, and automatically by Spectrum Quant + software defining significant wavenumber ranges. In the first step, the infrared dataset was initially applied for the cross validation. With the use of leave-one-out cross validation, only one sample is left for the test set and the other samples for the training set. In the second step, the amount of cellulose was predicted using calibration based on the spectra from the previous cross-validation. All models were calculated and the results of the coefficient of determination (R^2), the root mean square error of cross-validation (RMSECV), and the residual predictive deviation (RPD) were compared.

RESULTS AND DISCUSSION

Wet Chemical Analysis

The variability of cellulose content in pulp is summarized in Table 1. Results from the 108 samples showed that the amount of cellulose content ranged from 91.2% to 96.5% in the calibration set, and from 91.7% to 96.1% in the validation set. The standard deviation was 1.47% and 1.41% for the calibration set and the validation set, respectively. Thus, the calibration set and validation set were statistically similar, which provides a good database for the further analyses.

Table 1. Cellulose Content Distribution of 108 Samples

	Sample No.	Cellulose Content (%)		
		Min	Max	SD
Calibration Set	85	91.2	96.5	1.47
Validation Set	23	91.7	96.1	1.41

Notes: SD, standard deviation

Spectroscopic Characterization

The NIR spectra collected from the pulp samples are shown in Fig. 1a. The differences are difficult to discern because the spectra show mostly broad humps. Furthermore, because they are all pulp samples, their spectral characteristics are very similar except for the baseline differences among the samples. The baseline shift in the absorbance scale is quite common in the diffuse reflectance spectroscopy, part of which is due to the sample composition differences and to the inhomogeneities in the sample. In the first derivative mode, baseline differences were reduced greatly, and differences were easier to distinguish. Compared with the raw spectra, the spectra after first derivative pretreatment showed clearer results. The spectra were almost the same in the range of 10000 to 7500, and there is nearly no any characteristic peak. But the absorption exhibited dramatic differences between7500 and 4000 cm^{-1} (Fig. 1b), which caused by the stretching vibration of different groups in the cellulose.

NIR Calibrations

PLS regression was performed on the NIR spectra by applying spectral pre-processing methods and using different spectral wavenumber ranges.

First derivative (FD) is the most used and effective baseline correction method for NIR calibration model construction (Via *et al.* 2014). Based on the initial phase models, the R^2, RMSECV, and the number of principal components of the calibration models improved noticeably by selecting FD 17 than when no baseline correction method was considered. The R^2 reached to 0.76 which indicates a relatively high relationship between the spectra and data after baseline correction was applied, while no relationship appeared ($R^2 = 0.15$) with no baseline correction (Table 2). Several wavenumber ranges were considered during model building using FD pretreatment. The wavenumber ranges were selected based on the spectra image (Fig. 1b) and previous studies on cellulose in NIR (Jiang *et al.* 2014; Zhou *et al.* 2015; Jiang *et al.* 2017).

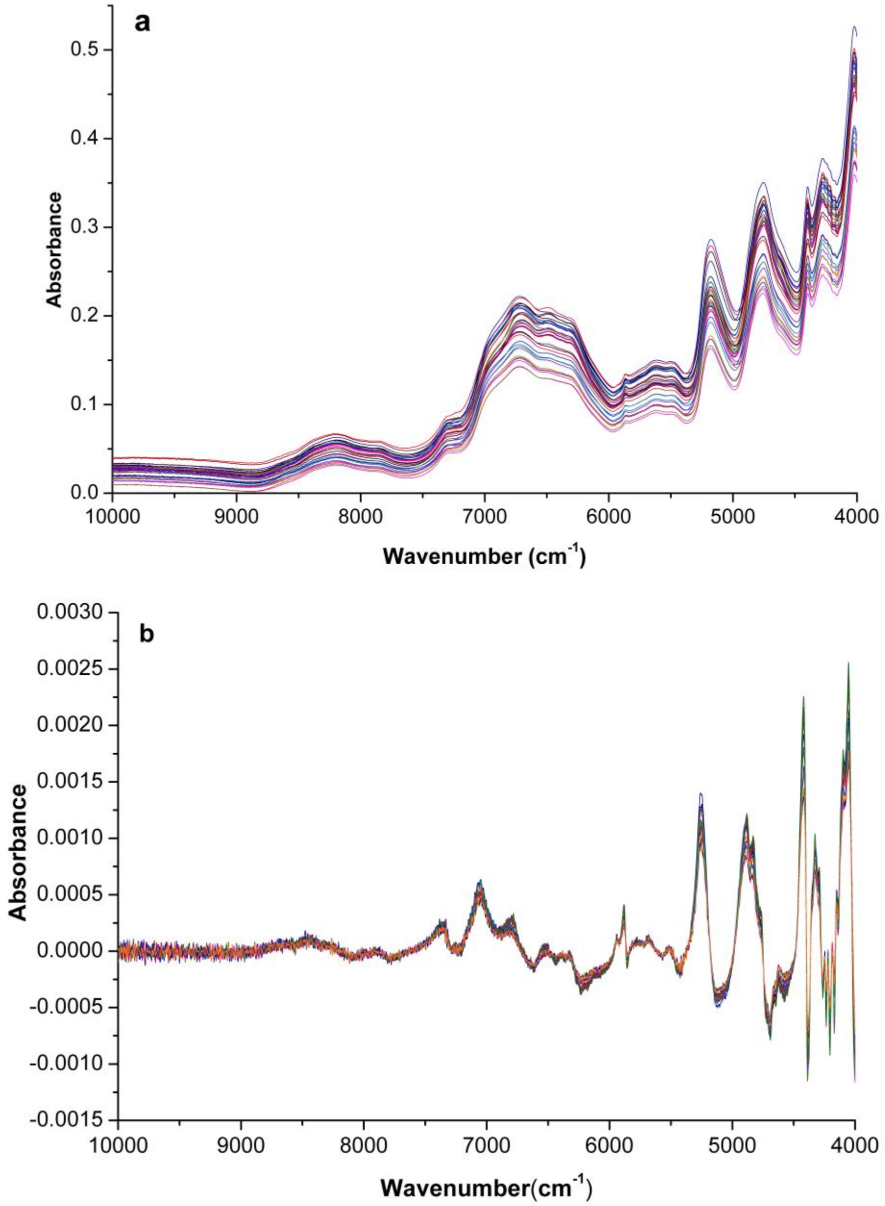

Fig. 1. Cellulose prediction (a) when raw spectra was processed and (b) when a first derivative pretreatment was processed

As is evident in Fig. 1b, the wavenumber range 7500 to 4000 cm^{-1} exhibited the main information in pulp NIR spectra, indicating that the NIR model based on this wavenumber range should have a better precision than the full wavenumber range. The results in Table 2 confirmed the above hypothesis. In comparison with the full wavenumber range, after selecting the wavenumber range of 7500 to 4000 cm^{-1}, the R^2 and RPD values were increased from 0.76, 1.60 to 0.88, 2.19, respectively; while the RMSECV decreased from 0.92 to 0.67. The higher R^2 and RPD values and lower RMSECV value represent the higher precision of the model, which indicates that the wavenumber range was selected compatibly. However, if the wavenumber range was narrowed to the range of 7250 to 6500 cm^{-1} and 5500 to 4000 cm^{-1}, the R^2, RPD, and RMSECV were improved slightly, to the values of 0.92, 2.45, and 0.60, respectively, which were the optimal results in this study.

Notably, the normalization method had barely any effect on the NIR model improvement in this research (Table 2). The NIR models constructed using no normalization, MSC, and SNV showed similar predicting ability, with R^2 values around 0.90 and RMSECV values around 0.62. Furthermore, the number of points used in FD 19 showed similar effects as FD 17.

Table 2. Results of the PLS Models Applying None, SNV, and MSC in the Different Wavenumber Regions

No.	Methods				Cross Validation Results		
	Wavenumber Range (cm^{-1})	Normalization	Baseline Correction	PC	R^2	RMSECV	RPD
1	10000-4000	None	None	1	0.15	1.47	1.00
2	10000-4000	None	FD 17	5	0.76	0.92	1.60
3	7500-4000	None	FD 17	7	0.88	0.67	2.19
4	7500-4000	SNV	FD 17	8	0.90	0.62	2.37
5	7500-4000	MSC	FD 17	8	0.89	0.64	2.30
6	7500-4000	None	FD19	8	0.91	0.61	2.41
7	5500-4000	None	FD19	8	0.92	0.60	2.45

Notes: The chemometric technique is PLS. R^2 corresponds to cross validation. RMSECV, stands for root-mean-square error of cross validation; SNV, standard normal variate; MSC, multiplicative scattering correction; FD, first derivative; PC, factor. Additionally, 17 and 19 are the number of points used in FD.

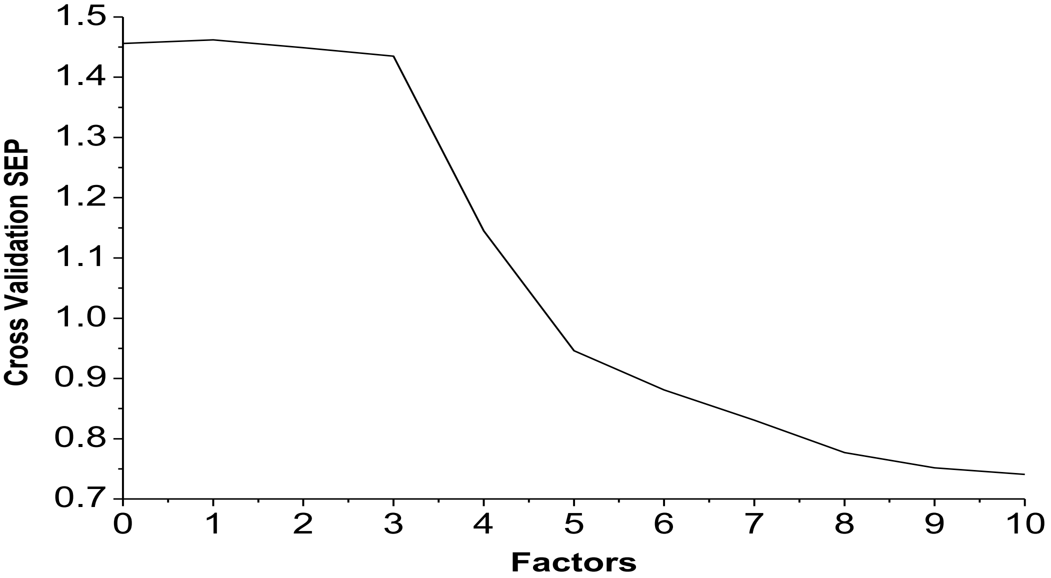

Fig. 2. The correlation between the SEP and factors

A reasonable PC value is important for a robust NIR model. As presented in Fig. 2, the relationship between the PC value and the SEP value showed a positive correlation. When the PC number is very low (1 to 3), the information extracted from the spectra is very limited, which results in the SEP value having a high level. When the PC number becomes high (3 to 8), which is the best range for NIR model construction, the SEP decreases with the increases of PCs (Dou *et al.* 2007). It decreases gently in the beginning and then sharply until the PC is at 5. The SEP becomes stable with the PC higher than 8. Obviously, 8 is the best PC for the NIR model in this study, which is in accordance with the results in Table 2.

Figure 3 depicts a spectral loading of PC1, PC2, PC3, and PC4 for cellulose analysis in pulp samples. The loading values over the absolute value of 0.001, which contribute to the PLS analysis, are shown in Fig. 3. The wavenumbers 4427 cm^{-1} and 4046 cm^{-1} showing relatively strong loading were assigned to the stretching vibration of O-H, C-O, C-H, and C-C in cellulose (Via *et al.* 2014; Zhou *et al.* 2015). The wavenumbers, which are directly assigned to the absorption bands in the wood sample, are also shown in Fig. 3 (a-d). Those results are coincident with Table 3. The loading data indicated that most of the NIR spectra represents cellulose, as previously summarized in the literature (Schwanninger *et al.* 2011), but there are still some lignin and water peaks, which is consistent with the chemical components in pulp (Froass *et al.* 1996). The loading peaks also demonstrate that wavenumbers 5500 to 4000 cm^{-1} is the most suitable wavenumber range for cellulose.

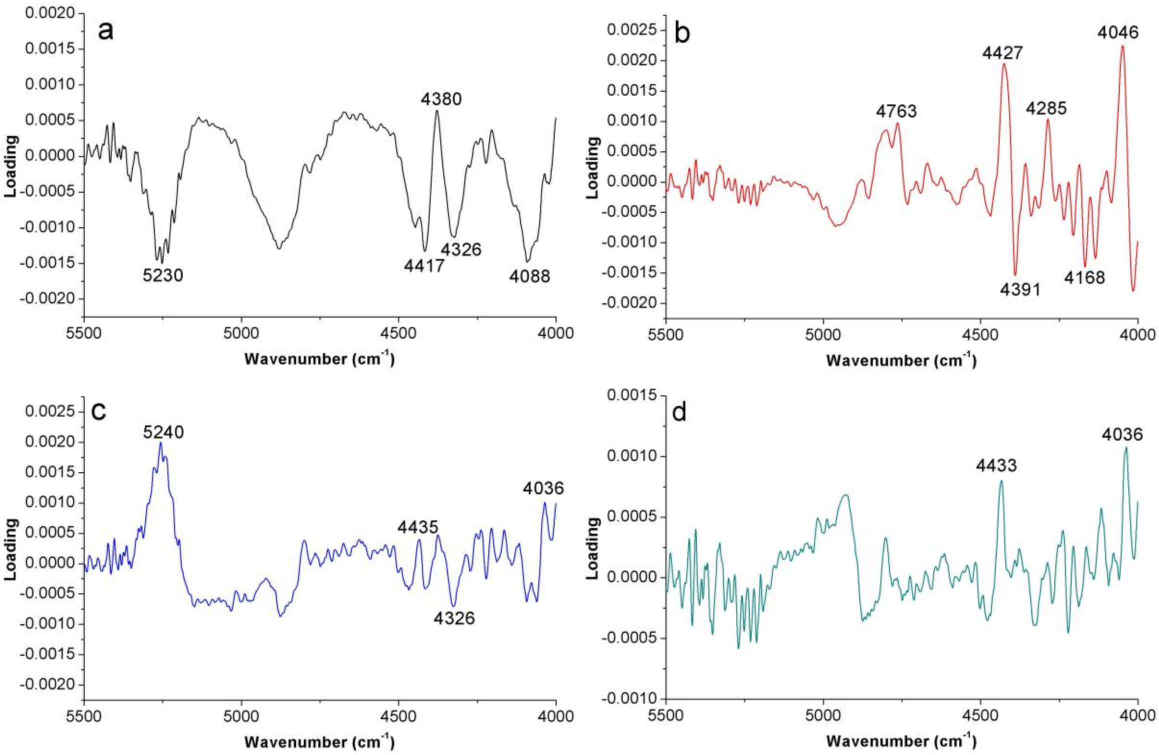

Fig. 3. A comparison of (a) PC1, (b) PC2, (c) PC3, and (d) PC4 loadings for the prediction of cellulose with NIR spectroscopy

Table 3. Band Assignments from the Literature and PC1, PC2, PC3, and PC4 Loadings for the Prediction of Cellulose

Band Location Wavenumber (cm^{-1})	PC1	PC2	PC3	PC4	Band Assignment	Components
5220-5150	5230	5240	—	—	O-H asym.str. + O-H def.	Water
4762	—	4763	—	—	O-H and C-H def. + O-H str.	Cellulose
4435	4417	4427	4435	4433	O-H str. + C-O str.	Cellulose
4392	—	4391	—	—	O-H str. + C-C str. + C-H def.	Cellulose
4365	4326	—	4326	—	C-H str. + O-H str.	Cellulose
4277	—	4285	—	—	C-H str. + C-H$_2$ def.	Cellulose
4195	—	4168	—	—	Not assigned	Lignin
4063	4088	—	—	—	C-H str. + C-C str.	Cellulose
4019	—	4046	4036	4036	C-H str. + C-C str.	Cellulose

Notes: asym.: antisymmetric, str.: stretching vibration, def.: deformation vibration (Schwanninger *et al.* 2011)

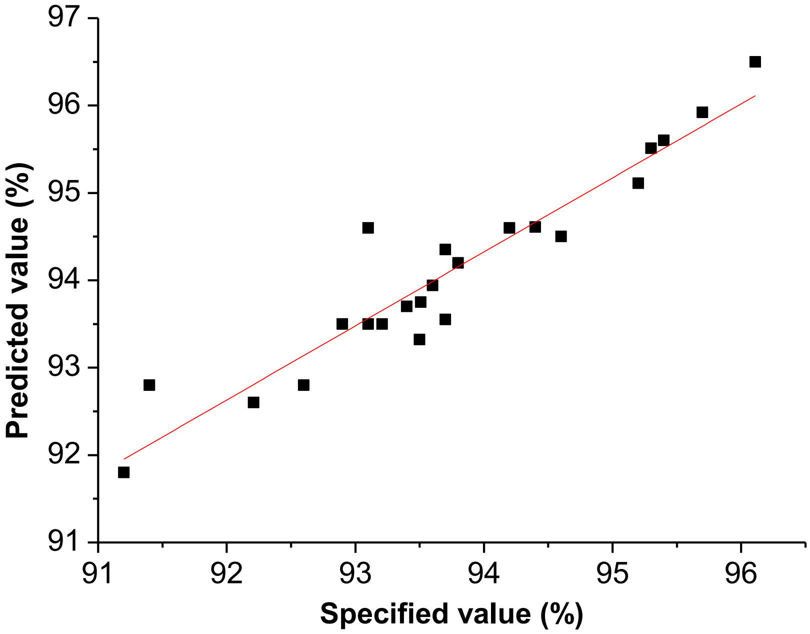

Fig. 4. Specified value and predicted value of cellulose in pulp samples

The 23 validation sets were chosen to verify the precision of the above NIR model (No. 7). The measured values using standard *vs*. predicted values are presented in Fig. 4. Using the NIR spectra to predict cellulose content produced good results. The predicted cellulose content was moderately well correlated with actual content, confirming that NIR could be used as a tool to assess this characteristic. The correlation between predicted and actual data was excellent with the Adj. R-Square (0.89), which showed that NIR spectroscopy could provide highly precise predictions.

CONCLUSIONS

1. A study of 108 pulp samples was performed with wet chemistry methods to calibrate the cellulose content of the pulp samples, and the NIR spectra combined with chemometric analysis were used to construct models. The selected appropriate wavenumber range of 7500 to 4000 cm^{-1}, 7250 to 6500 cm^{-1}, and 5500 to 4000 cm^{-1} were optimized. Good calibrations were developed for cellulose contents, with a high R^2 value of 0.92, a low RMSECV value of 0.60, and an RPD value of 2.45.
2. An accurate and precise NIR model for cellulose content prediction was obtained. Predictions were good with an Adj. R-Square of 0.89. This illustrated that the predictive model in the present work is superior and shows that it can be used for cellulose screening programs.

ACKNOWLEDGMENTS

The authors are grateful for the support of the Qingdao Postdoctoral Applied Research Project and the Natural Science Foundation of Shandong Province (ZR2017BEM045).

REFERENCES CITED

Acquah, G. E., Via, B. K., Fasina, O. O., Adhikari, S., Billor, N., and Eckhardt, L. G. (2017). “Chemometric modeling of thermogravimetric data for the compositional analysis of forest biomass,” *PLoS ONE* 12(3), e0172999. DOI: 10.1371/journal.pone.0172999

Axrup, L., Markides, K., and Nilsson, T. (2000). “Using miniature diode array NIR spectrometers for analysing wood chips and bark samples in motion,” *J. Chemom.* 14(5‐6), 561-572. DOI: 10.1002/1099-128x(200009/12)14:5/6<561::aid-cem608>3.0.co;2-2

Cheng, J. (2009). *Biomass to Renewable Energy Processes*, CRC Press, Boca Raton, FL, USA.

Costa, S. M., Mazzola, P. G., Silva, J. C. A. R., Pahl, R., Pessoa Jr., A., and Costa, S. A. (2013). “Use of sugar cane straw as a source of cellulose for textile fiber production,” *Ind. Crop. Prod.* 42, 189-194. DOI: 10.1016/j.indcrop.2012.05.028

Dou,Y., Qu, N., Wang, B., Chi, Y. Z., and Ren, Y. L. (2007). “Simultaneous determination of two active components in compound aspirin tablets using principal component artificial neural networks (PC-ANNs) on NIR spectroscopy,” *Eur. J Pharm. Sci.* 32, 193-199. DOI: 10.1016/j.ejps.2007.07.002

Fardim, P., Ferreira, M. M., and Durán, N. (2002). “Multivariate calibration for quantitative analysis of eucalypt kraft pulp by NIR spectrometry,” *J. Wood Chem. Technol.* 22(1), 67-81. DOI: 10.1081/wct-120004435

Froass, P. M., Ragauskas, A. J., and Jiang, J. (1996). “Chemical structure of residual lignin from kraft pulp," *J. Wood Chem. Technol.* 16(4), 347-365. DOI: 10.1080/02773819608545820

GB/T 2677.20 (1995). "Fibrous material - Determination of holocellulose," Standardization Administration of China, Beijing, China.

GB/T 9107 (1999). "Purified cotton," Standardization Administration of China, Beijing, China.

He, L., Xin, L. P., Chai, X. S., and Li, J. (2015). "A novel method for rapid determination of alpha-cellulose content in dissolving pulps by visible spectroscopy," *Cellulose* 22(4), 2149-2156. DOI: 10.1007/s10570-015-0652-9

Henriksen, H. C., Næs, T., Rødbotten, R., and Aastveit, A. (2004). "Prediction of important sulphite pulp properties from near infrared spectra: A feasibility study and comparison of methods," *J. Near Infrared Spectrosc.* 12(5), 279-285. DOI: 10.1255/jnirs.436

Jiang, W., Han, G., Via, B., Tu, M., Liu, W., and Fasina, O. (2014). "Rapid assessment of coniferous biomass lignin–carbohydrates with near-infrared spectroscopy," *Wood Sci. Technol.* 48(1), 109-122. DOI: 10.1007/s00226-013-0590-3

Jiang, W., Zhou, C., Han, G., Via, B., Swain, T., Fan, Z., and Liu, S. (2017). "Classification and identification of plant fibrous material with different species using near infrared technique—A new way to approach determining biomass properties accurately within different species," *Front. Plant Sci.* 7, 2000. DOI: 10.3389/fpls.2016.02000

Mayr, G., Hintenaus, P., Zeppetzauer, F., and Röder, T. (2015). "A fast and accurate near infrared spectroscopy method for the determination of cellulose content of alkali cellulose applicable for process control," *J. Near Infrared Spectrosc.* 23(6), 369-379. DOI: 10.1255/jnirs.1185

Rabemanolontsoa, H., Ayada, S., and Saka, S. (2011). "Quantitative method applicable for various biomass species to determine their chemical composition," *Biomass Bioenerg.* 35(11), 4630-4635. DOI: 10.1016/j.biombioe.2011.09.014

Schimleck, L. R., Payne, P., and Wearne, R. H. (2007). "Determination of important pulp properties of hybrid poplar by near infrared spectroscopy," *Wood Fiber Sci.* 37(3), 462-471.

Schwanninger, M., Rodrigues, J.C., and Fackler, K. (2011). "A review of band assignments in near infrared spectra of wood and wood components," *J. Near Infrared Spectrosc.* 19, 287-308. DOI: 10.1255/jnirs.955

TAPPI Test Methods 1996-1997/0104 Tm96 (1996). "Technical Association of the Pulp and Paper Industry," TAPPI Press, Mumbai, India.

TAPPI T222 om-11 (2011). "Acid-insoluble lignin in wood and pulp," TAPPI Press, Atlanta, GA.

Terdwongworakul, A., Punsuwan, V., Thanapase, W., and Tsuchikawa, S. (2005). "Rapid assessment of wood chemical properties and pulp yield of Eucalyptus camaldulensis in Thailand tree plantations by near infrared spectroscopy for improving wood selection for high quality pulp," *J. Wood Sci.* 51(2), 167-171. DOI: 10.1007/s10086-004-0633-3

Via, B. K., Zhou, C., Acquah, G., Jiang, W., and Eckhardt, L. (2014). "Near infrared spectroscopy calibration for wood chemistry: Which chemometric technique is best for prediction and interpretation?" *Sensors* 14(8), 13532-13547. DOI: 10.3390/s140813532

Zhou, C., Jiang, W., Via, B. K., Fasina, O., and Han, G. (2015). "Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR," *Carbohydr. Polym.* 121, 336-341. DOI: 10.1016/j.carbpol.2014.11.062

Zhou, C., Jiang, W., Via, B. K., Chetty, P. M., and Swain, T. (2016). "Monitoring the chemistry and monosaccharide ratio of Eucalyptus dunnii wood by near infrared spectroscopy," *J. Near Infrared Spectrosc.* 24, 537-548. DOI: 10.1255/jnirs.1254

Article submitted: April 2, 2018; Peer review completed: May 13, 2018; Revised version received and accepted; May 16, 2018; Published: June 21, 2018.
DOI: 10.15376/biores.13.3.6122-6132