Near-infrared Lignin Model Transfer: A Study Based on SWCSS-CARS Coupling Algorithm

Zhijian Liu, Honghong Wang, Zhixin Xiong,* Yunchao Hu, Haoran Huang, Ying Wang, Xianzhi Wu, and Long Liang

In NIR spectral modeling, the method of screening wavelengths with consistent stable signals (SWCSS) is based on a standard-free algorithm. However, the wavelengths selected by SWCSS may contain invalid information. In this paper, the Competitive Adaptive Reweighted Sampling (CARS) wavelength optimization algorithm was used in conjunction with SWCSS to eliminate the uninformative variables in the wavelengths selected by SWCSS. The SWCSS-CARS method was based on three near-infrared spectrometers (Lengguang 1, Lengguang 2, and Lengguang 3), with Lengguang 1 as the master and the other two instruments as the targets, using a total of 84 sample spectra of five types of pulpwood and their lignin contents as the research objects. Compared with the full spectrum, the number of wavelengths was reduced from 1601 to 24 in the model built using the coupling algorithm. For target 1, the value of RPD was improved from 1.9247 to 3.1880; for target 2, t the value of RPD was improved from 1.7415 to 3.2508. The wavelengths selected by the SWCSS-CARS coupling algorithm were able to build stable, robust models.

DOI: 10.15376/biores.19.1.245-256

Keywords: Near-infrared; Mode-transfer; Lignin; SWCSS-CARS

*Contact information: College of Light Industry and Food Engineering, Nanjing Forestry University, Longpan Road 159, Nanjing 210037 China; *Corresponding author: Leo_xzx@njfu.edu.cn*

INTRODUCTION

Wood used for pulp should have a high cellulose and low lignin content (Liang *et al.* 2020). The lignin content determines the amount of bleach, so the rapid detection of lignin in the control of the pulp production process is important. Near-infrared spectroscopy (NIR) analysis technology has the advantages of being fast, nondestructive, and green, and the method has been widely used in the fields of food (Castro *et al*. 2023), medicine (Yin *et al.* 2019) and agriculture (Cortés *et al.* 2019). In most cases, it is time-consuming and expensive to build a good multivariate calibration model, so it is desirable for a model to be stable and valid for a long time. However, changes in measurement conditions, aging or replacement of instrument components, and changes from external environments and samples may affect the accuracy and applicability of the calibration model. Model transfer, on the other hand, can adapt the calibrated model to a new instrument (target) or testing conditions through various chemometrics methods. Although the transferred model is usually not as accurate as the original model applied in the master machine, its accuracy usually still meets the requirements of the application and saves a lot of time and cost required for modifying the instrument or rebuilding the model (Dardenne 2002).

Currently, model transfer research is mostly focused on improving and developing new algorithms to realize model transfer sharing between different models of NIR spectrometers, which often requires a large number of samples to ensure the reliability of the model transfer, and this method is called scaled sample model transfer. In order to solve the problem of requiring a large number of samples for the traditional model transfer method with labeled samples, model transfer using the SWCSS method is a new strategy worthy of study (Ni *et al.* 2018; Zhang *et al.* 2020; Wang *et al.* 2022), whose transfer process is relatively simple and only requires the selection of very few representative samples for screening the consistent wavelengths, which makes the model transfer easier to realize. However, the wavelengths selected by the current SWCSS method for consistency between different spectroscopic instruments may contain no information as well as wavelengths with very little information, which leads to poor SWCSS-PLSR model transfer performance. If the invalid wavelength variables in the SWCSS results can be eliminated with other preferred algorithms, and then the PLSR model constructed with the remaining wavelengths can be transferred, it will have stronger robustness and prediction accuracy compared with the model built by the original SWCSS method. It has been shown that the use of CARS and Uninformative Variables Elimination (UVE) (Cai *et al.* 2008) as well as Successive Projections Algorithm (SPA) (Soares *et al.* 2013) algorithms are able to effectively remove the unimportant wavelengths from the spectrum. Therefore, based on the previous studies, this paper continues to explore the transfer results of the SWCSS-CARS method between three different batches of prismatic near-infrared spectrometers and compares them with the results of analyzing the target samples by the separate SWCSS, UVE, CARS, and SPA algorithms, to validate the feasibility of the concatenation of the SWCSS-CARS algorithms.

EXPERIMENTAL

Analysis of Samples and their Lignin Content

Five common woods (*Pinus massoniana*, *Cunninghamia lanceolata*, *Acacia*, *Eucalyptus robusta*, and *Populus*) were provided by the Institute of Forestry and Chemical Industry of the Chinese Academy of Forestry, totaling 84 log samples. The logs were chipped into wood chips and ground, and then the wood powder samples with particle size of 0.250 to 0.425 mm (40 to 60 mesh) were selected to determine the acid-insoluble lignin content according to GB/T 2677.8-1994. The statistical characteristics of the results are shown in Table 1.

Table 1. Statistical Table of Lignin Content in Pulpwood

Types of Wood Flour	Number	Min (%)	Max (%)	Mean (%)
Eucalyptus robusta	24	21.49	27.56	23.74
Cunninghamia	23	32.55	34.20	33.44
Populus	15	14.82	20.51	18.00
Acacia	12	24.62	27.15	25.69
Pinus massoniana	10	28.48	28.95	28.63
Total	84	14.82	34.20	26.43

Near-infrared Spectral Acquisition

The experiment uses three near-infrared spectrometers (three Lengguang S450 raster scanning near-infrared spectrometer of the same type), one of which is located in Zhenjiang users of the Lengguang S450 as the master, the other two are located in Shanghai users of the Lengguang S450 as targets (Target 1 and Target 2). The instrument is a grating scanning near-infrared spectrometer with an indium gallium arsenide (InGaAs) detector, wavelength range of 900 to 2500 nm, resolution of 12 nm, sampling interval of 1 nm, and total of 1601 wavelength points. The spectral data of 84 wood flour samples were collected on these three instruments respectively. Since the S450 near-infrared spectrometer is equipped with a rotary stage, the step of repeated sample loading can be omitted. The measurement of each sample was repeated 6 times, and the average spectrum was taken as the final sample spectrum. The Kennard-Stone method was used to divide the 84 specimens into 56 correction set and 28 prediction set specimens.

Wavelength Selection Method

Screening wavelengths with consistent and stable signals (Ni et al. 2019)

The standard deviation of the two spectra was calculated and analyzed using the spectral response of the samples at each wavelength as a variable to find out the wavelengths at which the spectral signals of different instruments are consistent and stable.

The Standard Deviation of Precision Detection Spectra (SDPDS) is the standard deviation of the sample spectra taken on several consecutive occasions on the main instrument. It is calculated as follows.

$$\mathrm{SDPDS}(j)=\sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij}-\bar{X}_j\right)^2}{n-1}} \quad (1)$$

where X_{ij} is the spectral signal at the j^{th} wavelength of the sample under test at the i^{th} acquisition, and n is the number of acquisitions. $\bar{X}_j$ is the average of the spectral signals at the j^{th} wavelength. SDPDS describes the level of fluctuation in the spectra of the repeated tests. The fluctuation of the spectrum is caused by noise and measurement error of the instrument in a very short period of time. The smaller the SDPDS is, the more stable the spectral signal is at this wavelength.

The Standard Deviation of Difference Spectra Between the Instruments (SDDSI) reflects the range of fluctuations in the difference spectra of the master and target. It is calculated as follows,

$$\mathrm{SDDSI}(j)=\sqrt{\frac{\sum_{i=1}^{m}\left(A_{ij}-\bar{A}_j\right)^2}{m-1}} \quad (2)$$

where m is the number of samples, while A_{ij} denotes the difference spectrum between the spectra of the master and the target $(A_{ij}=M_{ij}-S_{ij})$, and M_{ij} and S_{ij} denote the spectral response values of the i^{th} sample of the master and the target, respectively, measured at the wavelength point j. The smaller the SDDSI_j is, the better the consistency of the spectral signal of the instrument at the j^{th} wavelength.

Screening for stable and consistent wavelengths: The K-S algorithm is used to select a certain number of representative samples, calculate the standard deviation of the difference spectra between the master and target of these samples at wavelength j, SDDSI_j, and the standard deviation of the precision test spectra of the master, SDPDS_j, and define the ratio of SDDSI_j and SDPDS_j as the consistency parameter,

$$b_j = \frac{\mathrm{SDDSI}_j}{\mathrm{SDPDS}_j} (j = 1, 2, 3, \cdots, n) \quad (3)$$

where n is the number of wavelengths. Usually, SDDSI_j is larger than SDPDS_j. The closer b_j is to 1, the smaller is the standard deviation of the difference spectrum between the master and target instruments. The goal is to match the standard deviation of the master instrument's accuracy test, so the signals of the master and the target instruments have a very good consistency at that wavelength.

Based on previous experiments, when b_j is set smaller, the selected wavelengths will be very few, and much important information will be lost. However, when b_j is too large, wavelengths with large spectral differences between instruments will be included. NIR models that include information about these wavelengths have poor analytical performance for target samples. Therefore, a reasonable b_j needs to be selected based on the analytical effectiveness of the model built from the selected wavelengths on the target samples. After the wavelengths with large SDPDS values are excluded from the selected wavelengths, the set of wavelengths for which the spectral signals are consistent between the master instrument and each of the target instruments is recorded as U1,..., U_K. The intersection of these sets is Uc.

Competitive Adaptive Reweighted Sampling algorithm

Using the Competitive Adaptive Reweighted Sampling (CARS) algorithm (Jiang *et al.* 2015), based on the Monte Carlo resampling method, a fixed proportion of samples were randomly selected from the sample set as the correction set, and the importance of each variable was evaluated by the absolute value of the regression coefficients of the established PLS model. In each resampling, adaptive reweighted sampling was used to select the important spectral variables with larger absolute regression coefficients in the calibration model, and an exponential decreasing function was used to determine the number of variables to be selected. Finally, the cross-validation method was applied to select the optimal subset of variables where the root-mean-square error allowed.

Modeling and Model Evaluation Methods

The Partial Least Square Regression (PLSR) algorithm was used to establish the NIR correction model for lignin content. The size of the number of principal factors in the modeling process directly affects the effectiveness of the model. Choosing too small a number of principal factors will result in the loss of many important inter-spectral information, while choosing too large a number of principal factors introduces more redundancy into the model information, leading to overfitting (Son *et al*. 2020). In this study, the maximum number of principal factors was set to 15 and the minimum number of principal factors was set to 2 during the modeling process. The PLS component with the smallest sum of squared prediction errors (PRESS) was selected using the leave-one-out method for cross-valid

Model building, model prediction, and the effect of prediction after model delivery are evaluated, and the methods of evaluating the constructed models in this study used the Correlation Coefficient (R) (Wang *et al.* 2019), Determination Coefficient (R^2) (Morellos *et al.* 2016), and Root Mean Squared Error for Cross Validation (RMSCV) (Morellos *et al.* 2016). In addition, the Root Mean Squared Error for Prediction (RMSEP) and Ratio of Prediction to Deviation (RPD) were used (Rossel *et al.* 2006). Among them, the closer R and R^2 are to 1, the higher the RPD is, the lower the RMSE is, the better the model is.

RPD<2.5 indicates a poor model, the value of RPD is from 2.5 to 3.0 the model is fair, and RPD>3.0 the model is good. Since R^2 is associated with RPD, which represents a more straightforward measurement of the ability of an NIRS model to predict a constituent, in our following study the RPD was employed as an evaluation index.

The A_{IC} is defined as (Rossel *et al.* 2010),

$$A_{IC} = n \ln RMSEP + 2p \quad (4)$$

where n is the number of samples and p is the number of characteristics of the samples, *i.e*., the number of variables modeled. The smaller the A_{IC} value, the more streamlined the model.

Software

Spectral data preprocessing, division of correction/prediction sets, and PLS modeling and prediction were performed using NIRSA software developed in-house by the laboratory. In the test analysis, NIRSA provided similar results to the Unscrambler™ software (Xiong *et al*. 2016). The SWCSS algorithm as well as wavelength selection algorithms such as CARS were run using MATLAB 2016 software.

RESULTS AND DISCUSSION

Stable Consistent Wavelength Screening Based on Lignin

Using the standardized preprocessing method combined with the SWCSS method to screen wavelengths with consistent and stable spectral signals between master and target spectrometers, the number of selected wavelengths and wavelength points were the same, regardless of how many samples were taken to screen the wavelengths. Therefore, five representative samples were selected for use in the SWCSS algorithm to screen for consistent wavelengths (denoted as U1 and U2) between the master and target 1 and target 2 spectrometers, respectively. The wavelength sets U1 and U2 were utilized to model the host and predict the variation of RMSEP with *b*-value for the prediction set samples of target 1 and target 2, respectively (Fig. 1).

The RMSEP values of the target samples analyzed using the master models built by U1 and U2 when the *b* value is too small were 8.8832 and 4.4979, respectively, which is a poor prediction. This indicates that the number of consistent wavelengths selected for the consistency parameter *b*=1 was too small, which will lose many important information that is beneficial to the modeling, resulting in that the transfer performance of the calibration model built by the SWCSS method will be poor. Therefore, during the experiment, U1 and U2 were screened by setting *b* to take 1 to 10. Both U1 and U2 used the minimum value of RMSEP predicted by the PLSR model of the host of lignin indicators for the 2 target samples as the criterion for selecting the appropriate *b* value, and the standard deviation was calculated between the master and the targets 1 and 2, SDDSI1 and SDDSI2. Taking the intersection of U1 and U2 yields the intersection set of Uc, containing 465 wavelength points. The consistent wavelength set Uc screened by the SWCSS method is shown in Fig. 2.

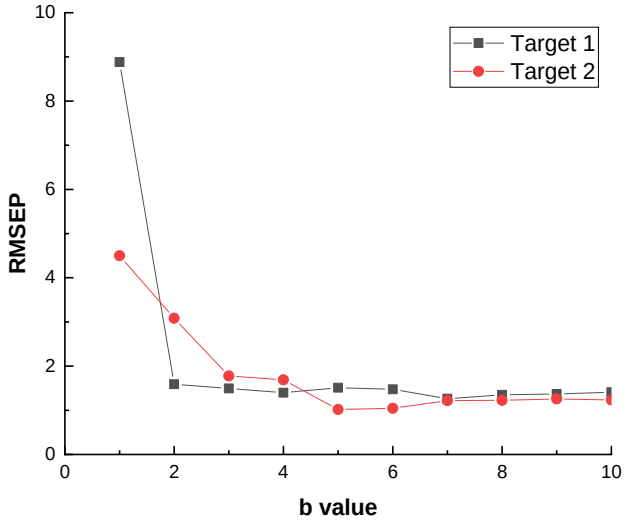

Fig. 1. The change of RMSEP with b value of two targets samples analyzed by U1 and U2 modeling

The lignin-based consistency wavelength points Uc selected using the SWCSS method are mainly located in the regions where the standard deviations SDDSI1 and SDDSI2 between the master and the 2 targets are small, and in the regions where the differences are large such as 900, 903 to 1412, 1572, 1577 to 1583, 1586 to 1593, 1611, 1614 to 1619 , 1627 to 1748, 1750 to 1942, 1974, 1976 to 1977, 1981 to 1984, 2007 to 2008, 2010, 2115 to 2116, 2130 to 2135, 2139, and 2189 to 2500 nm, then none of them can be screened by SWCSS algorithm.

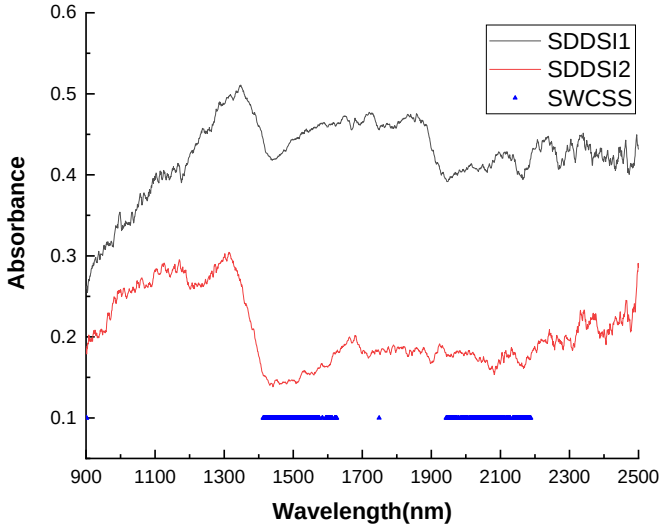

Fig. 2. The position distribution of the consistent wavelength set Uc selected based on the SWCSS method

Stable Consistent Wavelength Screening Based on Lignin

The spectral wavelength selection of SWCSS is only for wavelength points with little inter-instrumental difference, and its screening results may contain wavelengths with no or little information, which may adversely affect the calibration model, so this necessitates wavelength band optimization of the Uc wavelength set to obtain a more reliable calibration model. Based on the Uc wavelength set, the CARS algorithm is used to optimize the Uc wavelength set to obtain a new set of wavelengths containing 24

wavelength points, which accounts for 1.50% of the number of full-spectrum variables, as shown in Fig. 3. For comparison, the lignin-based wavelength sets selected by the full-spectrum-based SPA, UVE, and CARS algorithms (containing 19, 1260, and 34 wavelength points, accounting for 1.19%, 78.70%, and 2.12% of the number of full-spectrum variables, respectively) are shown in Fig. 3.

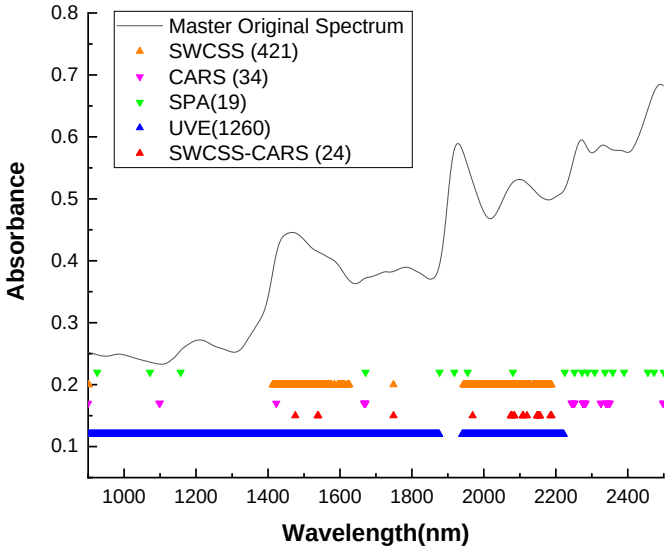

Fig. 3. The position distribution of the new wavelength group obtained by wavelength optimization of Uc and full spectrum

Lignin Model Transfer Results and Analysis

The master PLSR models were built based on SWCSS and UVE, CARS, and SPA algorithms alone or in combination, and the appropriate number of latent variables (LV) was selected by cross-validation using the leave-one-out method. The results are shown in Table 2.

Table 2. Master Models Built at Different Wavelength Sets and their Analytical Results for Master Samples

Method	LV	Correction set			Prediction set		
		R	RPD	RMSECV	R	RPD	RMSEP
Full Spectrum	9	0.9858	5.9468	0.8050	0.9843	5.6438	0.9426
SWCSS	7	0.9686	3.8169	1.2541	0.9637	3.4229	1.3985
UVE	10	0.9815	5.0194	0.9537	0.9801	5.0252	1.0587
CARS	6	0.9866	6.1199	0.8689	0.9738	4.3058	1.1118
SPA	9	0.9896	6.9338	0.7671	0.9882	6.1605	0.7770
SWCSS-CARS	5	0.9709	4.1703	1.2757	0.9641	3.6066	1.3273

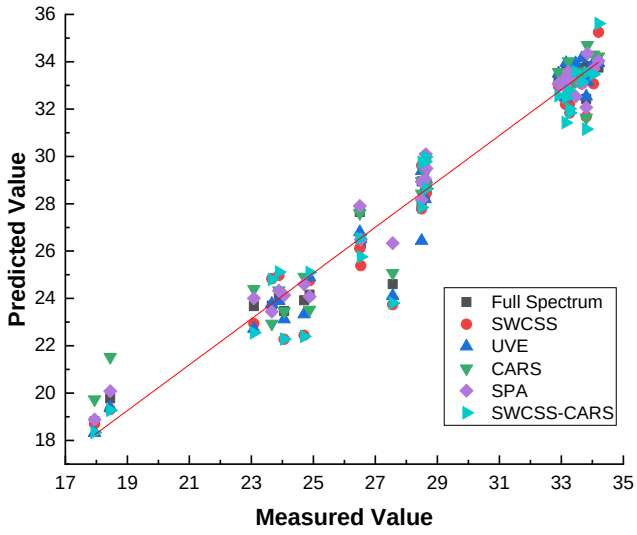

Fig. 4. The correlation between the measured value and the predicted value of the model established by different wavelength sets for the analysis of the master sample

The distribution of predicted and measured lignin values for the target samples analyzed by the eight host calibration models in Table 2 is shown in Fig. 4. The prediction results of these host correction models for target samples exceeded 3.0, indicating that the host models constructed on the basis of the above different wavelength selection methods can meet the needs of practical applications.

The 28 prediction set samples from the 2 targets were analyzed separately using the models in Table 2, and the results are shown in Table 3. When the 2 target sample sets were directly substituted into the host model for prediction without model transfer, the RMSEP of the prediction set increased from the original 0.9426 to 2.4872 and 2.7488, and the prediction results exhibited a large deviation, which indicates that the target samples cannot be directly applied to the host model, and it is necessary to carry out model transfer for the target samples. After model transfer using different wavelength selection methods, the prediction accuracies of the established models for the target samples were different, among which the SWCSS-CARS method exhibited the highest accuracy and stability, and the transfer efficiency was significantly improved. After model transfer with the SWCSS-CARS method, the RMSEP of the two targets decreased to 1.5016 and 1.4726, respectively, and the A_{IC} value decreased from 3198.70 to 63.86. Those values are better than the prediction effectiveness and efficiency of the SWCSS method alone. This is due to the fact that the CARS method uses adaptive reweighted sampling (ARS) to select wavelengths, and the wavelength variable corresponding to the model with the smallest cross-validated root mean square error (RMSEP) is selected as the characteristic wavelength variable in PLSR modeling, which effectively eliminates the information content of the wavelengths selected by the single SWCSS method. Wavelengths with little or no information content and uninformative wavelength variables present in the wavelengths. Therefore, the wavelengths selected by the SWCSS-CARS method have better stability and transmission efficiency than the SWCSS method alone.

Tables 2 and 3 show that the models built based on the wavelength sets screened by the full-spectrum UVE, CARS, and SPA algorithms had a somewhat improved ability to analyse the host samples, but a poorer ability to analyse the lignin content of the two target samples. This is due to the fact that none of the three methods mentioned above is

based on stable wavelength screening, and most of the selected wavelengths are located in the wavelength region with large inter-instrumental differences. This difference resulted in the poor analytical ability of the host PLSR model developed by the method to analyse the spectra of the target samples.

Table 3. Model Transfer Results of Different Wavelength Selection Methods

Method	Wavelength Set	A_{IC}	Target 1			Target 2		
			R	RPD	RMSEP	R	RPD	RMSEP
Full Spectrum	1601	3198.70	0.9650	1.9247	2.4872	0.9731	1.7415	2.7488
SWCSS	421	860.78	0.9600	3.5631	1.4931	0.9596	3.1493	1.5200
UVE	1260	2523.19	0.9523	1.4334	3.3396	0.9561	1.5341	3.1204
CARS	34	73.93	0.9621	1.9276	2.4834	0.9710	2.5068	1.9096
SPA	19	23.87	0.9490	0.9748	4.9107	0.9610	1.0622	4.5065
SWCSS-CARS	24	63.86	0.9580	3.1880	1.5016	0.9597	3.2508	1.4726

Figure 5 shows the correlation plots of the measured and predicted values of lignin content of the 2 targets before and after model transfer using different methods such as SWCSS and CARS algorithms independently and in conjunction, as well as their distributions. Before the transfer, the prediction error was large when the target samples were directly substituted into the master model. The separate UVE, CARS, and SPA algorithms were directly applied to the two target machine samples with poor prediction results and large vertical deviations. After applying the master lignin model constructed by SWCSS and SWCSS-CARS methods to the two target samples, the analytical errors were reduced. Among them, the SWCSS-CARS method analyzed the target 2 samples with the smallest deviation between the predicted values and the measured values, which also indicates that the model analyzed after the delivery of the SWCSS-CARS method is good.

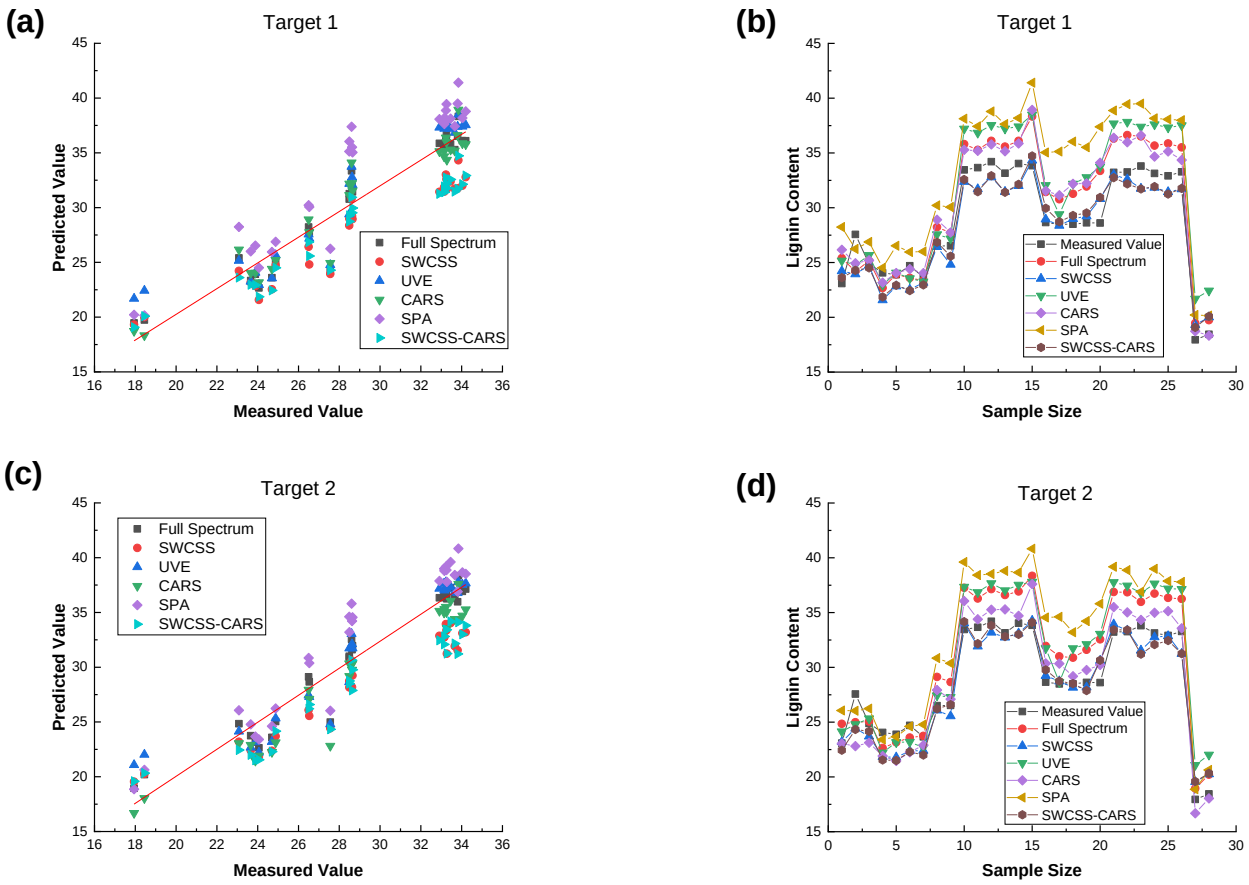

Fig. 5. Correlation plots of measured and predicted lignin content in the prediction set and their distribution

CONCLUSIONS

1. With the "screening wavelengths with consistent stable signals – competitive adaptive reweighting sampling" (SWCSS-CARS) coupling method, the wavelength was reduced from 1601 to 24, which means that the number of variables used in the modeling process was greatly reduced. For target 1, the value of ratio of prediction to deviation (RPD) was increased from 1.9247 to 3.1880; for target 2, the value of RPD was improved from 1.7415 to 3.2508, while A_{IC} decreased from 3198.70 to 63.86 for both. Comparative experiments show that the wavelengths selected using the SWCSS-CARS method can provide a more robust and simple correction model for the prediction of lignin content. It can simplify the process of correction model transfer, which is convenient for application in practice.
2. In addition, theoretically speaking, besides the CARS optimization algorithm, many wavelength selection methods can be used for the further optimization of the consistent wavelength set Uc, such as the Binary Dragonfly Algorithm (BDA), Genetic algorithm (GA), and Particle Swarm Optimization (PSO). For the dataset used in this study, not all wavelength selection methods combined with the SWCSS method give satisfactory results for the delivery of the calibration model.

3. At present, the method only realizes model transfer between the same type of NIR spectrometers produced by the same manufacturer, and model sharing between spectroscopic instruments with different wavelength intervals, different resolutions, and large differences in instrument structures requires further validation.

ACKNOWLEDGMENTS

This work was funded by the support of the Fundamental Research Funds of Research Institute of Forest New Technology, CAF(CAFYBB2019SY039).

REFERENCES CITED

Cai, W., Li, Y., and Shao X. (2008). "A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra," *Chemometrics and Intelligent Laboratory Systems* 90(2), 188-194. DOI: 10.1016/j.chemolab.2007.10.001

Castro, R. C., Ribeiro, D. S. M., Santos, J. L.M., and Páscoa, R. N. M. J. (2023). "Authentication/discrimination, identification and quantification of cinnamon adulterants using NIR spectroscopy and different chemometric tools: A tutorial to deal with counterfeit samples," *Food Control* 109619. DOI: 10.1016/J.FOODCONT.2023.109619

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., and Talens, P. (2019). "Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review," *Trends in Food Science & Technology* 85, 138-148. DOI: 10.1016/j.tifs.2019.01.015

Dardenne, P. (2002). "Calibration transfer in near infrared spectroscopy," *NIR News* 13.4, 3-7. DOI: 10.1255/nirn.668

Jiang, H., Zhang, H., Chen, Q., Mei, C., and Liu, G. (2015). "Identification of solid state fermentation degree with FT-NIR spectroscopy: Comparison of wavelength variable selection methods of CARS and SCARS," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 149, 1-7. DOI: 10.1016/j.saa.2015.04.024

Liang, L., Wei, L., Fang, G., Xu, F., Deng, Y., Shen, K., Tian, Q., Wu, T., and Zhu, B. (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, article 117515. DOI: 10.1016/j.saa.2019.117515

Morellos, A., Pantazi, X., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., and Mouazen, A. (2016). "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy," *Biosystems Engineering* 152, 104-116. DOI: 10.1016/j.biosystemseng.2016.04.018

Ni, L., Han, M., Luan, S., and Zhang, L. (2019). "Screening wavelengths with consistent and stable signals to realize calibration model transfer of near infrared spectra," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 206, 350-358. DOI: 10.1016/j.saa.2018.08.027

Ni, L., Han, M., Zhang, L., N. M. Y., and Luan, S. (2018). "A novel calibration transfer method of near infrared spectral model without standard samples based on screening stable and consistent wavelengths," *Chinese Journal of Analytical Chemistry* 46(10), 1660-1668. DOI: 10.11895/j.issn.0253-3820.181242

Rossel, R. V., and Behrens,T. (2010). "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma* 158(1-2), 46-54. DOI: 10.1016/j.geoderma.2009.12.025

Rossel, R. V., McGlynn, R. N., and McBratney, A. B. (2006). "Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy," *Geoderma* 137(1-2), 70-82. DOI: 10.1016/j.geoderma.2006.07.004

Soares, S. F. C. S., Gomes A. A., Araujo, M. C. U., Filho, A. R. G., and Galvão, R. K. H. (2013). "The successive projections algorithm," *TrAC Trends in Analytical Chemistry* 42, 84-98. DOI: 10.1016/j.trac.2012.09.006

Son, D., Kwon, H., and Lee, S. (2020). "Visible and near-infrared image synthesis using pca fusion of multiscale layers," *Applied Sciences* 10(23), article 8702. DOI: 10.3390/app10238702

Wang, A., Yang, P., Chen, J., Wu, Z., Jia, Y., Ma, C., and Zhan, X. (2019). "A new calibration model transferring strategy maintaining the predictive abilities of NIR multivariate calibration model applied in different batches process of extraction," *Infrared Physics & Technology* 103, article 103046. DOI: 10.1016/j.infrared.2019.103046

Wang, H., Xiong, Z., Hu, Y., Liu, Z., and Liang, L., (2022). "Transfer strategy for near infrared analysis model of holocellulose and lignin based on improved slope/bias algorithm," *BioResources* 17.4, 6476. DOI: 10.15376/biores.17.4.6476-6489

Xiong, Z., Pfeifer, F., and Siesler, H. (2016). "Evaluating the molecular interaction of organic liquid mixtures using near-infrared spectroscopy," *Applied Spectroscopy* 70.4, 635-644. DOI: 10.1177/0003702816631301

Yin, L., Zhou, J., Chen, D., Han, T., Zheng, B., Younis A., and Shao, Q. (2019). "A review of the application of near-infrared spectroscopy to rare traditional Chinese medicine," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 221, article 117208. DOI: 10.1016/j.saa.2019.117208

Zhang, L., Li, Y., Huang, W., Ni, L., and Ge, J. (2020). "The method of calibration model transfer by optimizing wavelength combinations based on consistent and stable spectral signals," S*pectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 227, article 117647. DOI: 10.1016/j.saa.2019.117647

Article submitted: October 9, 2023; Peer review completed: October 28, 2023; Revised version received and accepted: November 5, 2023; Published: November 14, 2023.
DOI: 10.15376/biores.19.1.245-256