Mechanical properties of wood materials using near-infrared spectroscopy based on correlation local embedding and partial least-squares

Lei Yu[1] · Yuliang Liang[1] · Yizhuo Zhang[1] · Jun Cao[1]

Abstract This study used near-infrared (NIR) spectroscopy to predict mechanical properties of wood. NIR spectra were collected in wavelengths 900–1700 nm, and spectra averaged by radial and tangential surface spectra were used to establish a partial least square (PLS) model based on correlation local embedding (CLE). Mongolian oak (*Quercus mongolica* Fisch. ex Ledeb.) was used to test the effectiveness of the model. The cross-validation method was used to verify the robustness of the CLE–PLS model. Ninety samples were tested as the calibration set and forty-five as the validation set. The results show that the prediction coefficient of determination (R_p^2) is 0.80 for MOR, and 0.78 for MOE. The ratio of performance to deviation is 2.23 for MOR and 2.15 for MOE.

✉ Yizhuo Zhang
  nefuzyz@163.com

✉ Jun Cao
  zdhcj@126.com

[1] Mechanical and Electrical College, Northeast Forestry University, Harbin 150040, People's Republic of China

Introduction

Wood products require accurate identification of structural performance and reliability, especially for construction material. The modulus of rupture (MOR) and modulus of elasticity (MOE) are important mechanical properties. Traditional wood-bending mechanical measurements are accurate and reliable but are destructive, time-consuming, and contribute to material waste. Near-infrared (NIR) spectroscopy, coupled with multivariate data analysis, has been shown to be a promising technique for measuring the physical, chemical and mechanical properties of wood (Via et al. 2003). NIR spectroscopy measures the absorbance of near-infrared light (780–2500 nm), and the spectra are mainly composed of absorptions based on stretching vibrations of hydrogen-containing groups (C–H, O–H and N–H) (Schimleck et al. 2001). Research using NIR spectroscopy has identified wood properties such as density, moisture content, lignin content, MOR, and MOE (Downes et al. 2014; Fujimoto et al. 2015; Tsuchikawa and Kobori 2015). For MOR or MOE determination, Kelley et al. (2004) used NIR spectroscopy on six softwoods. It was also used to predict mechanical properties of species such as *Pinus taeda* L. (Schimleck et al. 2005) and Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook. (Yu et al. 2009). Horvath et al. (2011) used transmittance NIR spectroscopy to predict the MOE of one- and two-year-old transgenic and wild-type aspen (*Populus tremuloides* Michx.). Kothiyal and Raturi (2011) reported the strong relationship between laboratory measurements and NIR predictions for five-year-old *Eucalyptus tereticornis*. Schimleck et al. (2011) illustrated MOR and MOE using NIR spectra collected from the transverse surface of pernambuco (*Caesalpinia echinata*). Xu et al. (2011) developed calibration models to predict MOE of oven-dried and green wood samples of balsam fir (*Abies balsamea*) and black spruce

(*Picea mariana*). Todorović et al. (2015) evaluated whether thermally modified beech wood (*Fagus moesiaca* (K. Maly) C.) could enhance the accuracy of prediction models. The NIR spectra are usually obtained from radial, tangential and transverse surfaces of solid wood samples. Zhao et al. (2012) investigated the effects of average NIR spectra from radial and tangential surfaces. Unfortunately, recent research shows that the coefficient of determination (R^2) and the ratio of performance to deviation (RPD) of mechanical properties are lower than other physical and chemical properties. The reason is that the partial least squares (PLS) regression model is only well adapted to linear relationships between NIR spectra and properties. In a non-linear relationship such as MOE, the accuracy will be reduced.

In order to improve the accuracy of MOE determination, different data pre-processing methods were evaluated such as, multiplicative scatter correction (MSC), first derivative (1stDer), second derivative (2ndDer), Savitzky–Golay smooth (S–G) (Andrade et al. 2010; Bächle et al. 2010). In the discipline of mathematics many improved PLS models have been reported, including moving window PLS (MWPLS), and synergy interval PLS (SiPLS) (Yang et al. 2015; Deng et al. 2016). These models commonly use partial spectra bands to select linear data. Saul and Roweis (2003) implemented a non-linear method to reduce data dimensionality and as a tool to simplify and accelerate machine learning in high dimensional spaces. Roweis and Saul (2012) used one of the manifold non-linear methods, locally linear embedding (LLE), to use the lower dimensionality within the Euclidean distance, but errors might be larger while certain neighbor points map into a lower dimensionality space. For this problem, Nguyen et al. (2015) showed that using correlation to seek neighbor points is better than using Euclidean distance. While somewhat effective, the disadvantage of these methods is that information related to mechanical properties may be missed when only relevant spectra are selected.

This study uses NIR spectroscopy to predict the MOR and MOE of Mongolian oak wood. A correlation linear embedding PLS (CLE–PLS) model was implemented to conform to the non-linear relationship between NIR spectra and MOR and MOE, which could be applied as a nonde-structive identification of wood properties.

Materials and methods

Materials

Mongolian oak (*Quercus mongolica*) was used to build the near-infrared prediction model. It is a deciduous species distributed in the northeast of China and is commonly used as structural material. Ten trees were randomly chosen from the 'Chong He' forest farm located in Wuchang, Heilongjiang province. The average diameter at breast height was 16 cm and the average height was 13 m. All the logs were taken above 1.3 m from the ground and cut into 1-m-long or 2-m-long pieces. After air-drying, each log (40 logs in total) was cut into bending mechanical samples (no pith) of 300 mm × 20 mm × 20 mm. Samples were numbered 1–4, with 1 and 4 taken from the sapwood and 2 and 3 from the heartwood (Fig. 1).

After removing samples with visible defects and deformations, 135 bending mechanical samples remained. These were numbered from 1 to 135; sapwood and heartwood samples had different serial numbers. All samples were conditioned to 12% moisture content prior to testing.
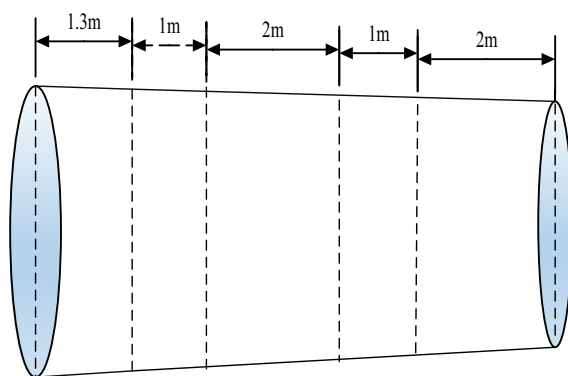
(a) Log cutting method in longitudinal section
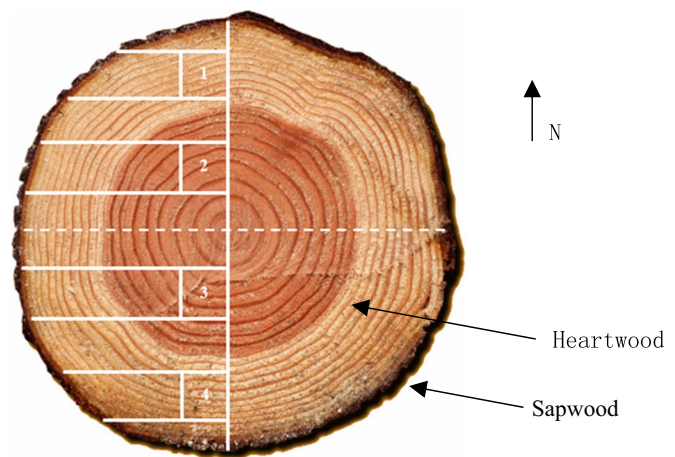
(b) Log cutting method in transverse section

Fig. 1 Process of sample production

Property tests

Mechanical test

Samples were measured in four-point bending according to Chinese national standards of the Method for Determination of the MOR in Static Bending of Wood (GB1936.1-2009) and the Method for Determination of the MOE in Static Bending of Wood (GB1936.2-2009).

NIR spectroscopy test

NIR spectra were collected with a one-chip spectrometer produced by INSION Co., Germany. This spectrometer is equipped with an optical fiber probe covering wavelengths from 900 to 1700 nm, with a 7.0 nm resolution. The NIR spectra were acquired by an optical fiber probe 5 mm in diameter (Fig. 2).

SPEC View 7.1 collected and recorded the spectra. For each scanning point, 30 scans were collected and averaged into a single spectrum. For each sample (300 mm×20 mm×20 mm), eight longitudinal spectra on the radial surface and eight longitudinal spectra on the tangential surface of the bending test samples were averaged to one spectrum. The sampling points were scanned in equal spacing 30 mm on both of the surfaces (Fig. 3). In addition, the diameter of the optical fiber probe was 5 mm.

NIR spectra pre-processing methods

Matlab 2014a software is a commonly used pre-processing method. The scattering effect of diffuse reflectance measurements can be reduced by multiplicative scatter correction (MSC), and the ground area of baseline shift can be modified by first or second derivatives, and high frequency noise can be smoothed by Savitzky–Golay (S–G).

NIR CLE–PLS calibration models development

CLE–PLS calibrations were developed by PLS analysis. The coefficient of determination R_c^2 is for the calibration set, and R_p^2 is for the validation set. The standard error of calibration (SEC), the standard error of prediction (SEP) and the ratio of performance to deviation (RPD) were used to evaluate the performance of the calibration model. The RPD is the ratio of the standard deviation (SD) to the original data and the SEP.

Correlation local embedding is a non-liner dimension reduce method. First, neighbor points are found for each spectra by correlation. Second, a local reconstruction weight matrix is defined and the mapping invariance is used to calculate the lower dimension data by the local reconstruction weight matrix and the nearest neighbor points. Finally, lower dimension data are used and measured to establish the PLS calibration model (Fig. 4).

The detailed algorithm development process is:

Step 1 Input spectra data of the calibration set, $X = \{x_i | i = 1, 2, \ldots, n\}$ and n is the number of samples, $X \in R^N$.

Step 2 Set a value to k as the number of neighbor points.

Step 3 Calculate the correlation coefficient of two random spectra by Eq. (1):

$$\rho(x_i, x_j) = \frac{(x_i - \bar{x})^T \cdot (x_j - \bar{x})}{\sqrt{(x_i - \bar{x})^T (x_i - \bar{x})} \cdot \sqrt{(x_j - \bar{x})^T (x_i - \bar{x})}} \tag{1}$$

where, $\bar{x}$ is the average spectra of calibration, $j = 1, 2, \ldots, n$.

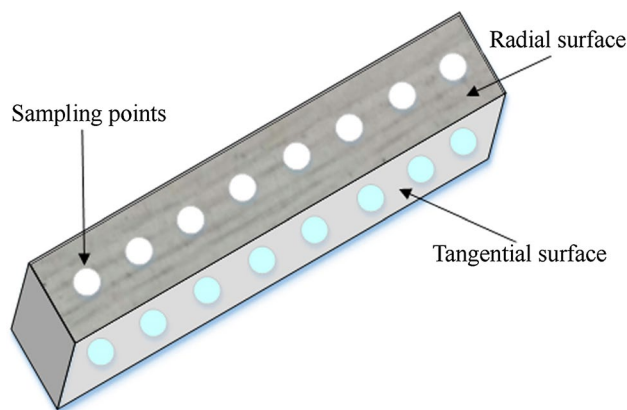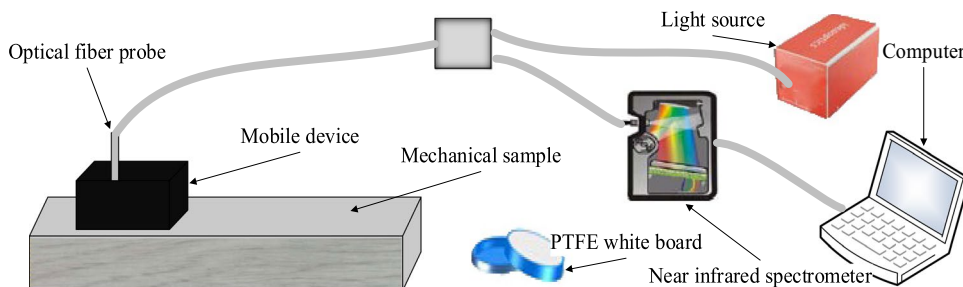*Step* 4 Define a local reconstruction weight matrix: G.

Fig. 3 Sampling points for each sample
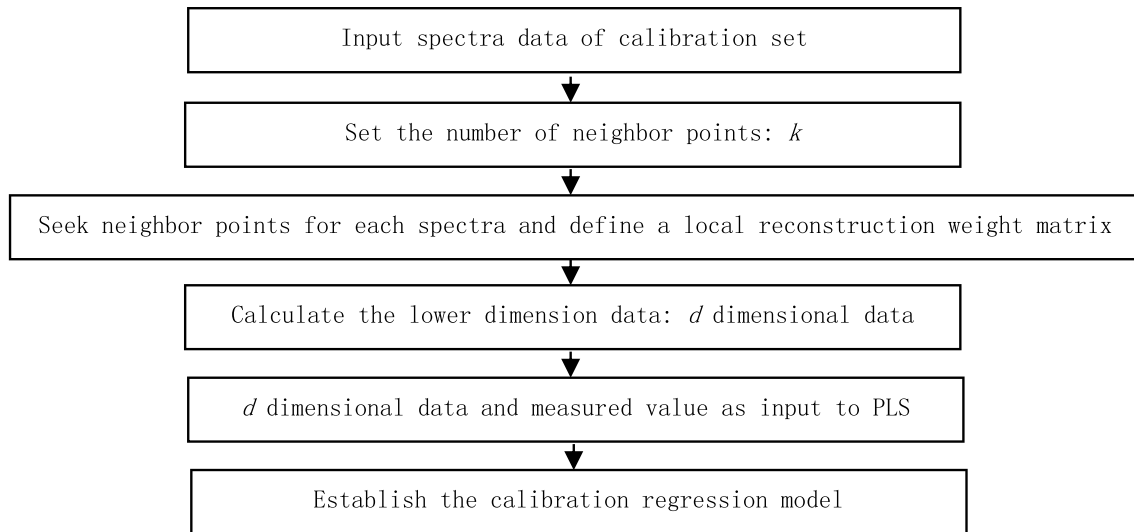

Fig. 2 NIR measurement process

Fig. 4 The block diagram of CLE–PLS

First, an error function is designed as Eq. (2):

$$\min J(G) = \sum_{i=1}^{n} \left| x_i - \sum_{j=1}^{k} g_{ij} x_{ij} \right|^2 \tag{2}$$

where, x_{ij} is the neighbor points of x_i, and g_{ij} is weight. Constraint condition:

$$\sum_{j=1}^{k} g_{ij} = 1 \tag{3}$$

Second, a covariance matrix H^i is designed as Eq. (4):

$$H^i = (x_i - x_{ij})^{\mathrm{T}} \cdot (x_i - x_{im}) \tag{4}$$

where x_{ij} and x_{im} are both neighbor points of x_i

Finally, use a Lagrange multiplier algorithm to combine Eqs. (3) and (4). Calculate the g_{ij} (Eq. 5) of G.

$$g_{ij} = \frac{\sum_{m=1}^{k} (H^i)^{-1}}{\sum_{s=1}^{k} \sum_{t=1}^{k} (H_{st}^i)^{-1}} \tag{5}$$

Step 5 Preserve the mapping invariance to calculate the lower dimension data by neighbor points and G. Assuming C is d dimensional data ($d \ll N$), and the constraint of mapping is provided in Eq. (6):

$$\min J(C) = \sum_{i=1}^{n} \left| c_i - \sum_{j=1}^{k} g_{ij} c_{ij} \right|^2 \tag{6}$$

where, $J(C)$ is the objective function, c_i is the output vector of x_i, and c_{ij} is neighbor points of c_i in lower dimensional space. Equation (6) can be simplified as Eq. (7):

$$\min J(C) = \sum_{i=1}^{n} |CI - CG_i|^2 = |C(I - G)|^2 \tag{7}$$

Constraint condition:

$$\sum_{i=1}^{n} c_i = 0 \tag{8}$$

$$\frac{1}{n} \sum_{i=1}^{n} c_i c_i^{\mathrm{T}} = I \tag{9}$$

where, I is $d \times d$ dimensional unit matrix.

Therefore, Eq. (7) can be simplified continually. It is shown in Eq. (10):

$$\min J(C) = |C(I - G)|^2 = (I - G)^{\mathrm{T}}(I - G)C^{\mathrm{T}}C \tag{10}$$

Step 6 Output the lower dimensional data C. C is the vectors of d minimum non-zero eigenvalues calculated by $(I - G)^{\mathrm{T}}(I - G)$. C is $n \times d$ ($d \ll N$) dimensional matrix.

Step 7 Use C and the measured value to establish the PLS regression model.

Results and discussion

Pre-processing and PLS model

Spectra data were collected and averaged from the radial and tangential surfaces of 135 samples (Fig. 5). The original spectra are shown in Fig. 5a. The spectra were preprocessed with different preprocessing methods such as MSC, Fig. 5b, first derivative (1stDer), Fig. 5c, second derivative (2ndDer), Fig. 5d, and S–G smooth, Fig. 5e.

Among these methods, 1stDer could eliminate the baseline drift and reduce background interference, and S–G smooth could suppress high frequency noise. In order to achieve the best prediction of mechanical properties, these methods were combined in different ways and the spectra data after preprocessing are shown in Fig. 5f. The number of windows for S–G can be set as 7, 9, 11, and 13, in which we found that 9 was the best. The weight coefficient was $[-21, 14, 39, 54, 59, 54, 39, 14, -21]$.
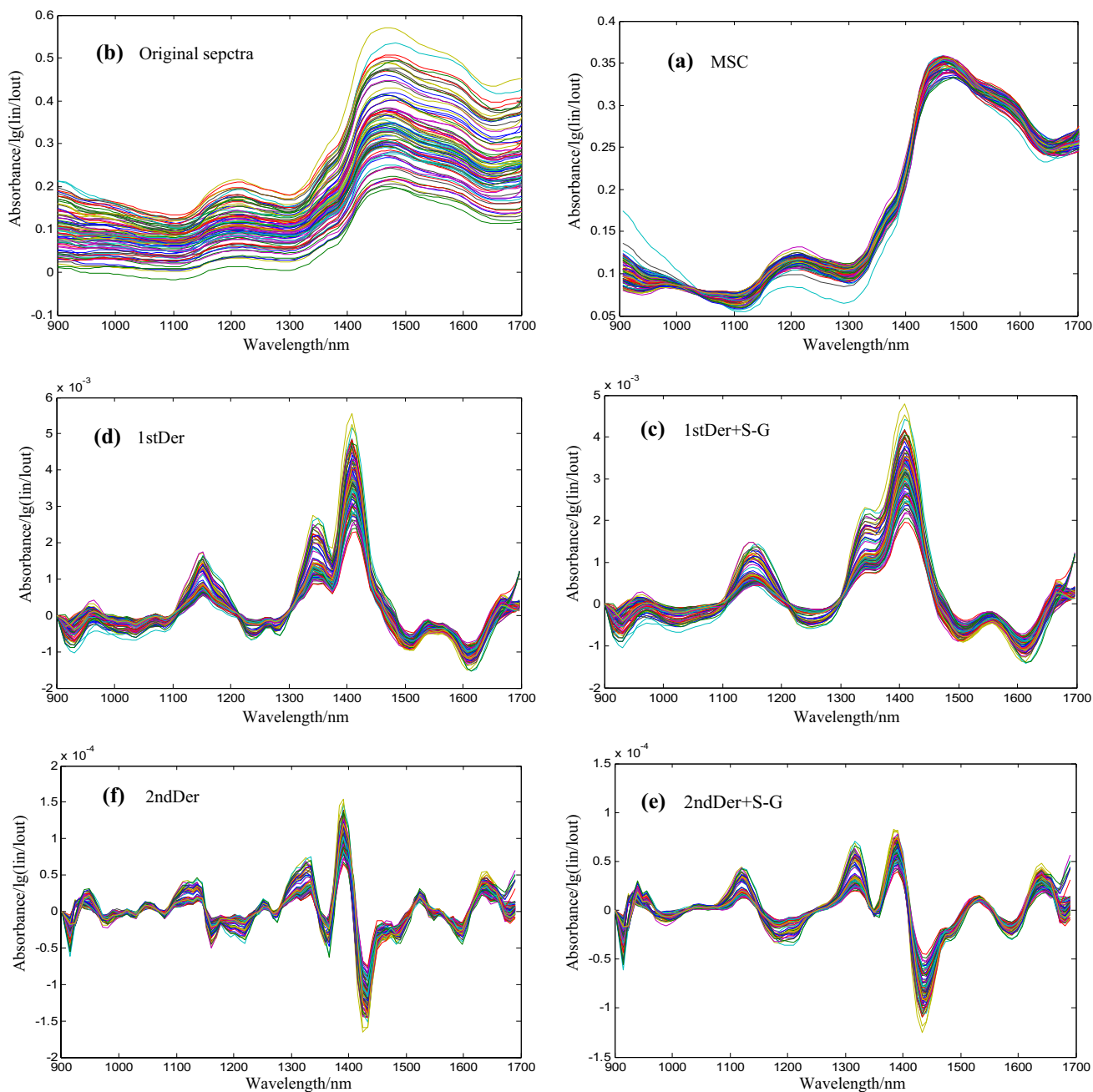

Fig. 5 Original spectrum and pre-processed spectrum

Using these pre-processing methods, the spectra are consistent by MSC, the background of gentle area can be expressed more effectively by 1stDer and 2ndDer, and the useless noise or irrelevant absorption peak can be eliminated when combined with S–G smooth (Table 1). In the PLS model, the same number of factors was set at 4.

The performance of the PLS calibration model corresponds to different pre-processing methods in a different manner. The most frequently used method is 2ndDer of the spectra. We found that using 2ndDer is better than 1stDer, a result that is consistent with the research on Eucalyptus pellita. For the spectra of Mongolian oak, 1stDer + S–G is the best of the above methods as shown by the higher R_c^2 and R_p^2, as well as the lower SEC and SEP (Table 1).

Although we used pre-processing method (1stDer + S–G) here, the R_c^2 and R_p^2 are only 0.70 and 0.62 for MOR, and

the R_c^2 and R_p^2 are only 0.67 and 0.60 for MOE. For MOR and MOE, RPD are both $1.5 < RPD < 2$, indicating that the model cannot provide an adequate prediction. Compared with other studies of wood samples with NIR spectra, the coefficient of determination (Table 1) is a slightly lower because of the linearity of the PLS model. Higher correlation could be found for MOE using various species of wood. For example, in Schimleck's paper, R_p^2 values are 0.75 (R_p is 0.87 of MOR, factors = 5), 0.70 (R_p is 0.84 of MOE, factors = 5) for six softwoods, and Zhao et al. obtained results of $R_p^2 = 0.77$ (R_p is 0.88) for MOR, $R_p^2 = 0.79$ (R_p is 0.89) for MOE by using both radial and tangential surface. The R_p^2 in these studies are still lower than 0.80, and it could be even lower if considering other wood properties such as density, moisture, and some chemical properties.

Table 1 Results of PLS Model by different pre-processing methods

Mechanical properties	Pre-processing methods	Model	R_c^2	SEC	R_p^2	SEP	RPD
MOR	Raw	PLS	0.52	18.45	0.49	18.52	1.40
	MSC	PLS	0.62	16.42	0.53	17.78	1.46
	1stDer	PLS	0.64	15.98	0.58	16.80	1.54
	1stDer + S–G	PLS	0.70	14.59	0.62	15.98	1.62
	2ndDer	PLS	0.65	15.75	0.59	16.60	1.56
	2ndDer + S–G	PLS	0.69	14.83	0.61	16.19	1.60
MOE	Raw	PLS	0.49	1.54	0.40	1.62	1.29
	MSC	PLS	0.56	1.43	0.47	1.52	1.38
	1stDer	PLS	0.60	1.37	0.55	1.40	1.49
	1stDer + S–G	PLS	0.67	1.24	0.60	1.32	1.58
	2ndDer	PLS	0.62	1.33	0.56	1.39	1.50
	2ndDer + S–G	PLS	0.66	1.26	0.57	1.37	1.53

MOR is Modulus of rupture, MOE is Modulus of elasticity, SEC is the standard error of calibration, SEP is the standard error of prediction, RPD is the ratio of performance to deviation, PLS is the model of partial least squares
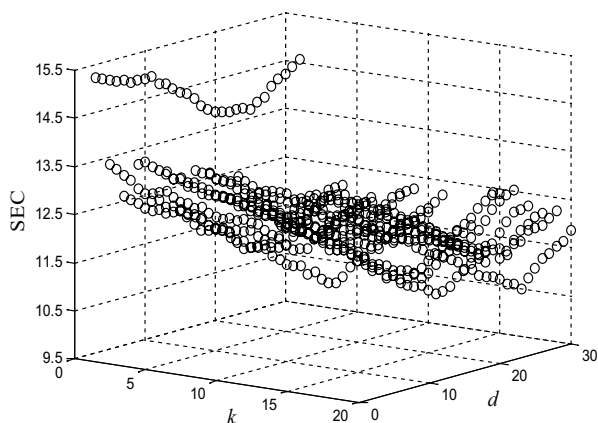
(a) MOR Calibration
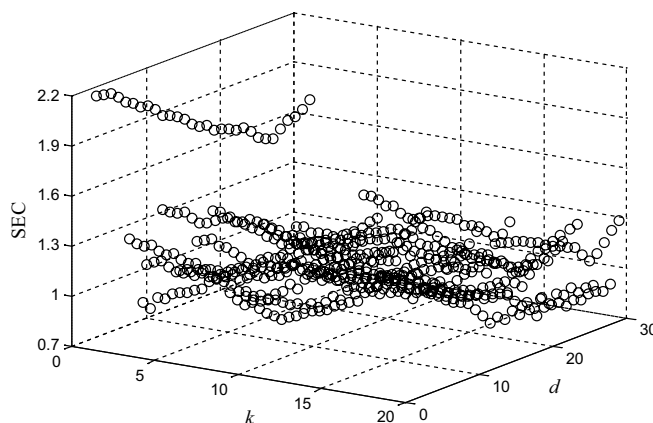
(b) MOE Calibration

Fig. 6 Effects of different d and k for calibrations

CLE–PLS model

The pre-processed spectra by 1stDer + S–G were regarded as the input data of the calibration model. As two significant parameters of CLE–PLS, the lowest dimension d and the number of neighbor points k need to be selected. The effects for SEC of different values are shown in Fig. 6.

The minimum SEC (standard error of calibration) has been determined and d and k were selected when SEC reach a minimum. For MOR, SEC (9.96) is minimum when $d = 23$ and $k = 18$. For MOE, SEC (0.81) is minimum when $d = 25$ and $k = 17$. The parameter values were calculated through 90 random samples and were used to establish the CLE–PLS

model. The number of factors of the PLS model is 4. The results are shown in Table 2 and the correlation between predicted and measured values of the CLE–PLS validation model is shown in Fig. 7.

In order to verify the robustness of the CLE–PLS model, cross-validation was applied to divide calibration and validation sets. The calibration sets, including 90 samples and the validation sets, including 45 samples, were both randomly generated 20 times. The comparisons of PLS and CLE–PLS models is shown in Table 3.

Compared with the PLS (partial least square) model, CLE–PLS performed better for MOR and MOE as it had the higher R_p^2 and RPD, as well as the lower SEP (Fig. 7).

Table 2 CLE–PLS model results

Mechanical properties	Pre-processing method	Model	d	k	R_c^2	SEC	R_p^2	SEP	RPD
MOR	1stDer + S–G	CLE–PLS	23	18	0.86	9.96	0.81	11.30	2.29
MOE	1stDer + S–G	CLE–PLS	25	17	0.86	0.81	0.79	0.96	2.18

MOR is Modulus of rupture, MOE is Modulus of elasticity, SEC is the standard error of calibration, SEP is the standard error of prediction, RPD is the ratio of performance to deviation, and CLE–PLS is the model combining correlation local embedding and partial least squares
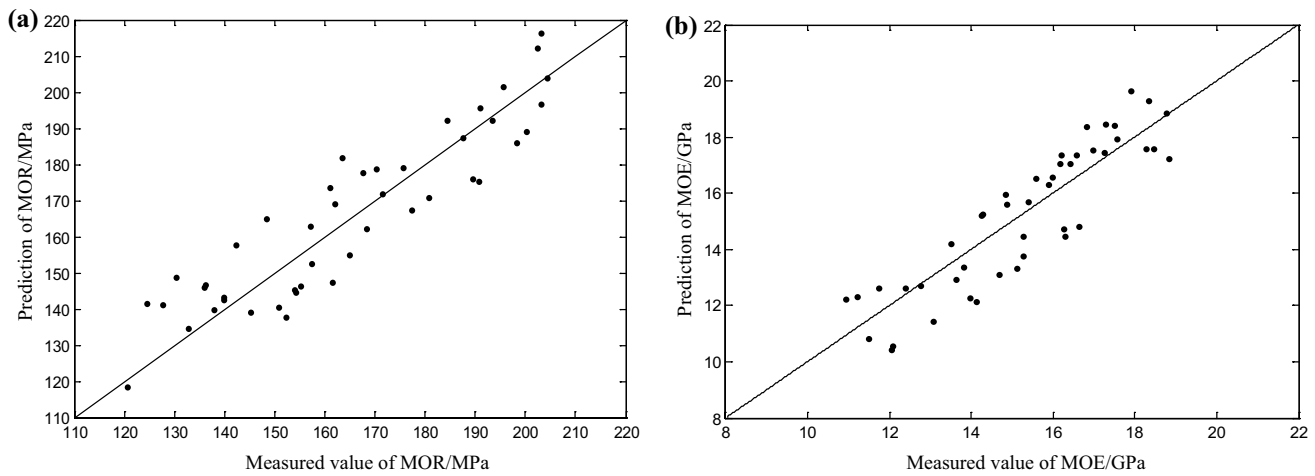

Fig. 7 CLE–PLS validation model. **a** The correlation of MOR between predicted and measured values; **b** The correlation of MOE between predicted and measured values

Table 3 Comparisons of PLS and CLE–PLS models

Mechanical properties	Pre-processing method	Model	R_c^2	SEC	R_p^2	SEP	RPD
MOR	1stDer + S–G	PLS	0.70	14.59	0.62	15.98	1.62
	1stDer + S–G	CLE–PLS	0.84	10.23	0.80	11.72	2.23
MOE	1stDer + S–G	PLS	0.67	1.24	0.60	1.32	1.58
	1stDer + S–G	CLE–PLS	0.83	0.81	0.78	1.06	2.15

MOR is Modulus of rupture, MOE is Modulus of elasticity, SEC is the standard error of calibration, SEP is the standard error of prediction and RPD is the ratio of performance to deviation, PLS is the model of partial least squares, and CLE–PLS is the model combining correlation local embedding and partial least squares

Using the CLE–PLS model, the values of R_p^2 (0.80 for MOR and 0.79 for MOE) were both more than 0.75, and the values RPD (2.23 for MOR and 2.15 for MOE) were both more than 2.0. This suggests that the CLE–PLS model could be used for preliminary screening ($2 < \text{RPD} < 3$) of Mongolia oak wood products. In the prediction of Mongolian oak mechanical bending properties, the results show that the CLE–PLS model is more effective.

Conclusions

Near-infrared (NIR) spectra, ranging from 900 to 1700 nm predicted the modulus of rupture (MOR) and modulus of elasticity (MOE) of Mongolian oak. The truth values of MOR and MOE were determined by the four-point bending method, and the spectra were collected from radial and tangential surfaces of samples. Compared with five pre-processing methods, 1stDer + S–G is better. Meanwhile, correlation-local embedding (CLE) dimensional methods were used to improve the performance of the partial least squares model (PLS). CLE–PLS [improved 0.18 determination coefficient] for both MOR and MOE in contrast with the PLS model. In addition, the ratio of performance to deviation (RPD) values for MOR and MOE were both more than 2.0 ([separately improved by 0.61 and 0.57] respectively) by using the CLE–PLS model. The CLE algorithm is effective in non-linear dimensionality, and the CLE–PLS model can be used for preliminary screening of Mongolian oak, (without visible defects and deformations), to predict MOR and MOE. Our results also indicate that non-linear dimensional methods to develop NIR spectra prediction models may be more applicable.

References

Andrade CR, Trugilho PF, Napoli A, Da Silva Vieira R, Lima JT, De Sousa LC (2010) Estimation of the mechanical properties of wood from Eucalyptus urophylla using near infrared spectroscopy. Cerne 16(3):291–298

Bächle H, Zimmer B, Windeisen E, Wegener G (2010) Evaluation of thermally modified beech and spruce wood and their properties by FT-NIR spectroscopy. Wood Sci Technol 44(3):421–433

Deng BC, Yun YH, Cao DS (2016) A bootstrapping soft shrinkage approach for variable selection in chemical modeling. Anal Chim Acta 908:63–74

Downes GM, Touza M, Harwood C (2014) NIR detection of non-recoverable collapse in sawn boards of Eucalyptus globulus. Eur J Wood Wood Prod 72(5):563–570

Fujimoto T, Chiyoda K, Yamaguchi K (2015) Heritability estimates for wood stiffness and its related near-infrared spectral bands in sugi (*Cryptomeria japonica*) clones. J For Res 20(1):206–212

Horvath L, Peszlen I, Peralta P (2011) Use of transmittance near-infrared spectroscopy to predict the mechanical properties of 1- and 2-year-old transgenic aspen. Wood Sci Technol 45(2):303–314

Kelley SS, Rials TG, Groom LR, et al (2004) Use of near infrared spectroscopy to predict the mechanical properties of six softwood. Holzforschung 58(3):252–260

Kothiyal V, Raturi A (2011) Estimating mechanical properties and specific gravity for five-year-old Eucalyptus tereticornis having broad moisture content range by NIR spectroscopy. Holzforschung 65(5):757–762

Nguyen V, Hung CC, Ma X (2015) Super resolution face image based on locally linear embedding and local correlation. ACM SIGAPP Appl Comput Rev 15(1):17–25

Roweis ST, Saul LK (2012) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5):2323–2326

Saul LK, Roweis ST (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J Mach Learn Res 4(2):119–155

Schimleck LR, Evans R, Ilic J (2001) Estimation of eucalyptus delegatensis wood properties by near infrared spectroscopy. Can J For Res 31(10):1671–1675

Schimleck LR, Jones PD, Clark A, Daniels RF, Peter GF (2005) Near infrared spectroscopy for the nondestructive estimation of clear wood properties of *Pinus taeda* L. from the southern United States. For Prod J 55(12):21–28

Schimleck LR, de Matos JLM, Oliveira JTD, Muniz GIB (2011) Non-destructive estimation of Pernambuco (*Caesalpinia echinata*) clear wood properties using near infrared spectroscopy. J Near Infrared Spectrosc 19(5):411–419

Todorović N, Popović Z, Milić G (2015) Estimation of quality of thermally modified beech wood with red heartwood by FT-NIR spectroscopy. Wood Sci Technol 49(3):527–549

Tsuchikawa S, Kobori H (2015) A review of recent application of near infrared spectroscopy to wood science and technology. J Wood Sci 61:213–220

Via BK, Shupe TF, Groom LH, Stine M, So CL (2003) Multivariate modelling of density, strength and stiffness from near infrared spectra for mature, juvenile and pith wood of longleaf pine (*Pinus palustris*). J Near Infrared Spectrosc 11(1):365–378

Xu Q, Qin M, Ni Y, Defo M, Dalpke B, Sherson G (2011) Predictions of wood density and module of elasticity of balsam fir (*Abies balsamea*) and black spruce (*Picea mariana*) from near infrared spectral analyses. Can J For Res 41(2):352–358

Yang L, Guo M, Shi X, et al (2015) Online near-infrared analysis coupled with MWPLS and SiPLS models for the multi-ingredient and multi-phase extraction of licorice (Gancao). Chin Med 10(1):1–10

Yu H, Zhao RJ, Fu F, Fei BH, Jiang ZH (2009) Prediction of mechanical properties of Chinese fir wood by near infrared spectroscopy. Front For China 4(3):368–373

Zhao RJ, Xing XT, Lv JX, Zhang JZ (2012) Estimation of wood mechanical properties of *Eucalyptus pellita* by near infrared spectroscopy. Sci Silvae Sin 48(6):106–111