Detection of Knot Defects on Coniferous Wood Surface Using Near Infrared Spectroscopy and Chemometrics

Zhu Zhou,[a,b,c,*] Jianxin Yin,[a,b,c] Suyin Zhou,[a,b,c] Houkui Zhou,[a,b,c] and Yun Zhang [a,b,c]

Lumber pieces usually contain defects such as knots, which strongly affect the strength and stiffness. To develop a model for rapid, accurate grading of lumbers based on knots, Douglas fir, spruce-pine-fir (SPF), Chinese hemlock, and Dragon spruce were used. The experiments explored the effects of modelling methods and spectral preprocess methods for knot detection, and investigated the feasibility of using a model built within one species to discriminate the samples from other species, using a novel variable selection method-random frog (RF)- to select effective wavelengths. The results showed that least squares-support vector machines coupled with first derivative preprocessed spectra achieved best performance for both single and mixed models. Models built within Dragon spruce could be used to classify knot samples from SPF and Chinese hemlock but not Douglas fir, and *vice versa*. Eight effective wavelengths (1314 nm, 1358 nm, 1409 nm, 1340 nm, 1260 nm, 1586 nm, 1288 nm, and 1402 nm) were selected by RF to build effective wavelengths based models. The sensitivity, specificity, and accuracy in the validation set were 98.49%, 93.42%, and 96.30%, respectively. Good results could be obtained when using data at just eight wavelengths, as an alternative to evaluating the whole spectrum.

Keywords: Near infrared spectroscopy (NIRS); Coniferous wood; Knot detection; Random frog algorithm; Least squares-support vector machines (LS-SVM)

Contact information: a: School of Information Engineering, Zhejiang A&F University, Lin'an 311300, Zhejiang, China; b: Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Lin'an 311300, Zhejiang, China; c: Research Center for Smart Agriculture and Forestry, Zhejiang A&F University, Lin'an 311300, Zhejiang, China;
** Corresponding author: zhouzhu@zafu.edu.cn*

INTRODUCTION

Knots are serious defects that greatly reduce the strength and stiffness of lumber and affect the quality of wood products (Nagai *et al.* 2009; Fujimoto *et al.* 2010). A skilled worker performs the traditional method of knot detection (Hu *et al.* 2011). However, human inspection is tedious and can cause eye fatigue. Additionally, it is also unsatisfactory because of low efficiency, accuracy, and subjective judgment. Thus, it is important to develop rapid and accurate ways to detect knots on wood surfaces.

Recently, several nondestructive technologies have been explored for automated detection of knots such as computer vision (Takeshi *et al.* 2006; Xie and Wang 2015), microwave (Baradit *et al.* 2006), dielectrical properties (Rice *et al.* 1992), acoustic (Kodama and Akishika 1993; Karsulovic *et al.* 2000; Machado *et al.* 2004), laser scattering (Tormanen and Makynen 2009), and X-ray (Cristhian *et al.* 2008). These methods mentioned above have produced good results. However, there are still some limitations (Szymani and McDonald, 1981; Nagai *et al.* 2009;Yang *et al.* 2016).

Near infrared spectroscopy (NIRS) is a rapid, accurate, and non-destructive detection technique. It is a powerful analytical tool used in the wood industry for predicting chemical composition of wood (Poke and Raymond 2006), classifying the type of fungal decay in wood (Fackler *et al.* 2007; Yang *et al.* 2008), discriminating blue-stained wood (Via *et al.* 2008), and identifying wood species (Tsuchikawa *et al.* 2003; Shou *et al.* 2014). Recently, NIRS has been used to detect knots on wood surfaces. For instance, Fujimoto and Tsuchikawa (2010) explored the probability of detection of dead and sound knots on Japanese larch board surfaces using NIRS. Fifty spectra including dead and sound knots and 100 reference spectra of knot-free wood were collected. The wavenumber range from 6600 cm^{-1} to 5400 cm^{-1} (1515 nm to 1852 nm) was employed for the multivariate analysis, and the soft independent modelling of class analogies (SIMCA) classification technique was conducted for model building. The results showed that NIRS coupled with SIMCA could identify knots. Yang *et al*. (2012, 2015) assessed the potential of visible-NIR spectroscopy for identifying knots in *Pinus massoniana* veneer and eucalypt veneer, with good classification results. Yang *et al*. (2016) also investigated the feasibility of NIRS coupled with partial least squares discriminant analysis (PLS-DA) to discriminate knots on the surface of poplar, eucalypt (hardwood), and masson pine (softwood) veneers. The model built with eucalypt samples could predict the samples from poplar, and *vice versa*; nevertheless, the model built with masson pine samples could not predict the other two samples' species, and *vice versa*. The above research indicated that NIRS has potential to accurately identify visible knots on wood surfaces. However, there are no reports that NIRS can be used to detect knots on different coniferous wood surfaces. In addition, due to one innate disadvantage of NIR, *i.e*., the severe collinearity in spectral responses, it may be desirable to select a decreased number of wavelengths. This selection has potential to avert spectral collinearity, largely reduce computation time, and make the model more understandable (Feng *et al.* 2015). Furthermore, the selected wavelengths may be used to develop a simple detection system.

Therefore, the objective of this study was to carry out a feasibility study using the NIRS techniques for detecting knots on different coniferous wood surface. The specific objectives were to: (1) employ both linear (partial least squares-linear discriminant analysis, PLS-LDA) and non-linear (least squares-support vector machine, LS-SVM) methods to establish full-wavelength models for four wood species; (2) investigate the feasibility of using a model built with one wood species to discriminate the samples from other coniferous wood species; (3) use a novel method - random frog (RF) algorithm to select effective wavelengths for knots detection; and (4) develop simplified calibration models to classify knots on different coniferous wood surface.

EXPERIMENTAL

Materials

Four types of lumber including Douglas fir (*Pseudotsuga menziesii* Mirb.), Chinese hemlock (*Tsuga chinensis* Pritz.), Dragon spruce (*Picea asperata* Mast.), and commercial SPF were collected from local wood processing factories (Lin'an, China) in December 2015. All samples were cut into the board specimens of 200 (longitudinal) × 80 (tangential) × 20 (radial) mm. A total of 459 solitary knots, including sound and dead knots, were selected, and their diameters were measured according to Chinese national standard

GB/T 4823-1995. The controls, 597 knot free samples, were also randomly selected from the neighborhood of each knot.

Spectral Acquisition

NIR spectra were collected using a miniature spectrometer named SmartEye 1700 (NIR Photonics, Hangzhou, China) at intervals of 1 nm between 1000 to 1650 nm. A fiber-optic probe (8 mm in diameter) was oriented perpendicular to the sample surface and used to collect spectra. Each spectrum was recorded at an average of 50 scans with an integration time of 11.75 ms for each sample.

Samples Partition

All samples were divided into two data sets. Two-thirds were used as a calibration set, and the remaining one-third were used as a validation set. For each species, spectra of normal samples for calibration and validation sets were chosen using the Kennard-Stone algorithm (Kennard and Stone 1969), whereas spectra of knot samples for calibration and validation were chosen according to the size of the knot. The knot samples of each species were sorted in descending order according to their area of knot affected wood, and every third sample was removed and used as a validation set. The remaining samples were used for calibration. In this way, it was ensured that both sets covered the whole range of knot samples appropriately and consistently. The numbers of samples for calibration and prediction sets for different wood species are listed in Table 1.

Table 1. Samples of Calibration and Validation Sets in the Experiment

Wood Species	Calibration Set				Prediction Set			
	Number of Normal	Knot			Number of Normal	Knot		
		Number	Diameter (mm)			Number	Diameter (mm)	
			range	mean			range	mean
F	109	74	2-55	19.6	54	37	2-56	20.0
SPF	107	123	2-75	17.1	53	61	3-75	16.9
H	89	52	1.5-62	18.2	45	26	2-45	17.6
S	93	58	2-30	12.4	47	28	2-29	11.9
Mixed	398	307	1.5-75	17.0	199	152	2-75	16.9

Note: F - Douglas fir, H - Chinese hemlock, S - Dragon spruce, SPF - spruce-pine-fir

Spectra Pretreatment

For reliable and accurate models, it is necessary to perform data pretreatment before model calibration. There are many spectral preprocessing methods such as derivative, standard normal variate transformation (SNV), multiplicative scatter correction (MSC), smoothing, and some other new methods. In this work, two classical spectral preprocessing methods, SNV and first derivative (FD), were used comparatively. SNV is a mathematical transformation method of spectra, which is used to remove slope variation and correct scatter effects (Chen *et al*. 2010). FD focuses on eliminating baseline drifts and enhancing small spectral difference.

Modelling Methods

PLS-LDA and LS-SVM were employed to develop classification models. PLS-LDA, which is the combination of PLS for dimension reduction and LDA for classification, is a supervised method for solving problems in linear classification. PLS maximizes the covariance between the class membership and the input wavelengths, whereas LDA seeks

a linear combination of the new projections minimizing the ratio of the within-class variance to between-class variance (Zhou *et al.* 2015). The optimal number of latent variables for the PLS-LDA models is chosen according to the criterion of the lowest prediction error in Monte Carlo cross validation.

The LS-SVM, which employs a set of linear equations instead of a quadratic programming problem to obtain the support vectors, is reformulated to standard SVM (Suykens and Vandewalle 1999). It has the advantage of good generalization performance as the standard SVM, with simpler structure and shorter optimization time (Zheng and Lu 2012). LS-SVM can be used both for the regression and classification problems. The LS-SVM algorithm has been described previously (Suykens and Vandewalle 1999). In the present study, RBF kernel was used as the kernel function due to its effectiveness and speed in the training process (Zheng and Lu 2012). In addition, the grid search technique and leave one out cross validation (LOOCV) were used to find the optimal parameter values: regularization parameter (γ) and the RBF kernel function parameter (σ^2).

Variable Selection

Wavelength selection is of critical significance for removing the redundant information from spectra, enhancing the stability of the model resulting from the collinearity in multivariate spectra, and reducing computation time for model development due to a reduced number of wavelengths. Thus, selecting the most effective wavelengths for knot detection is one of the most urgent procedures in chemometrics based on pattern recognition.

The random frog algorithm was carried out to select effective wavelengths. RF is a novel and efficient technique for variable selection, which borrows the framework of reversible jump Markov Chain Monte Carlo (RJMCMC) methods (Li *et al.* 2012; Yun *et al.* 2013). The details of RF algorithm can be found in literature (Li *et al.* 2012; Yun *et al.* 2013). RF works in four steps:

Step 1: Six parameters (N, Q, θ, ω, η, and V_0) should be assigned with proper values. N was the number of iterations and was sufficiently large to achieve convergence (N = 10000). Q was the number of variables in the initialized variables set (Q = 20). θ was a factor controlling the variance of a normal distribution from which the number of variables to enter a candidate variable subsets is sampled (θ = 0.3). ω was a coefficient controlling the number of variables selected to candidate variable subsets (ω = 3). η represented the upper bound of the probability for accepting a candidate variable subset (η = 0.1). V_0 represented a variable subset consisting of Q variables selected randomly from the raw dataset V.

Step 2: A random number is first generated from a normal distribution Norm (Q, θQ), then the random number is rounded to its nearest integer, denoted as Q*. Q* is the number of variables of the candidate variable subset V*. There are three possible situations between the wavelength number of candidate subset (V *) and previous subset (V) (Li *et al.* 2012): (a) if Q * = Q, let V * equal to V_0, (b) if Q * < Q, the regression coefficient of each variable in the PLS-LDA model developed by V_0 is calculated and compared. The (Q − Q*) variables associated with the smallest absolute regression coefficients are eliminated from V_0. The remaining Q* wavelengths form a candidate subset V *, and (c) if Q* > Q, another subset T with ω (Q* − Q) variables sampled from (V − V_0) is randomly generated. A PLS-LDA model is established using the combination of V_0 and T. The Q* variables with the largest absolute regression coefficients in this PLS-LDA model are retained and collected as a candidate subset V*.

Step 3: To determine whether the obtained candidate subset can be accepted, the lowest prediction error in LOOCV values using V_0 and V^* are calculated, called Err and Err*, respectively. If Err* ≤ Err, accept V^* as V_1, otherwise accept V^* as V_1 with probability ηErr/Err*. Finally, V_0 is updated using the variables in V_1 and this iteration is repeated until N loops are finished.

Step 4: Dividing the frequency of each variable by N, the selection probability for every variable can be obtained.

The larger the selection probability, the more important is the corresponding variable.

Model Performance Evaluation

Prediction performance can be assessed using three indicators—sensitivity, specificity, and accuracy—which are defined as follows,

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP+FN}}\times 100\% \quad (1)$$

$$\text{Specificity}=\frac{\text{TN}}{\text{TN+FP}}\times 100\% \quad (2)$$

$$\text{Accuracy}=\frac{\text{TN+TP}}{\text{TP+FN+TN+FP}}\times 100\% \quad (3)$$

where TP is the true positive, TN is the true negative, FN is the false negative, and FP is the false positive. Note that the positive and negative mean "knot free" and "knot" respectively. Generally, an optimal model should provide high values of sensitivity, specificity and accuracy.

Software

All calculations were implemented using MATLAB R2010a (MathWorks Inc., Natick, MA, USA) on a PC (CPU: Inter(R) Core(TM) i5-3210 @ 2.5GHz, RAM: 4.00 GB) operating with Windows 7. Spectra pretreatments were performed using a PLS extension package (PLS Toolbox v.6.5, Eigenvector Research, Wenatchee, WA, USA). Codes for the PLS-LDA and random frog algorithm were downloaded free of charge from http://www.libpls.net/. The free LSSVM toolbox (LS-SVM v.1.7, Suykens, Leuven, Belgium) was downloaded from http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

RESULTS AND DISCUSSION

Overview of Spectra

The mean raw spectra of knot samples and normal samples of the four wood species are shown in Fig. 1. The raw spectra showed some similarity in shape and several differences in absorption intensity. The knot samples were found to have higher absorption amplitude than normal samples for each wood species. These changes in spectra were consistent with previous experiments by Fujimoto and Tsuchikawa (2010) and Yang *et al*. (2015, 2016). This may be due to the difference in the orientation of the microfibrils within fibers between knots and normal wood, as the microfibrils of a knot's cells was substantially parallel to the NIR light incidence angle (Fujimoto and Tsuchikawa 2010; Yang *et al.* 2015). The mean raw spectra of knot samples and normal samples were also

different in absorption intensity for each wood species. This is possibly due to the wood species in the experiment having different in color, density, moisture content, and texture.

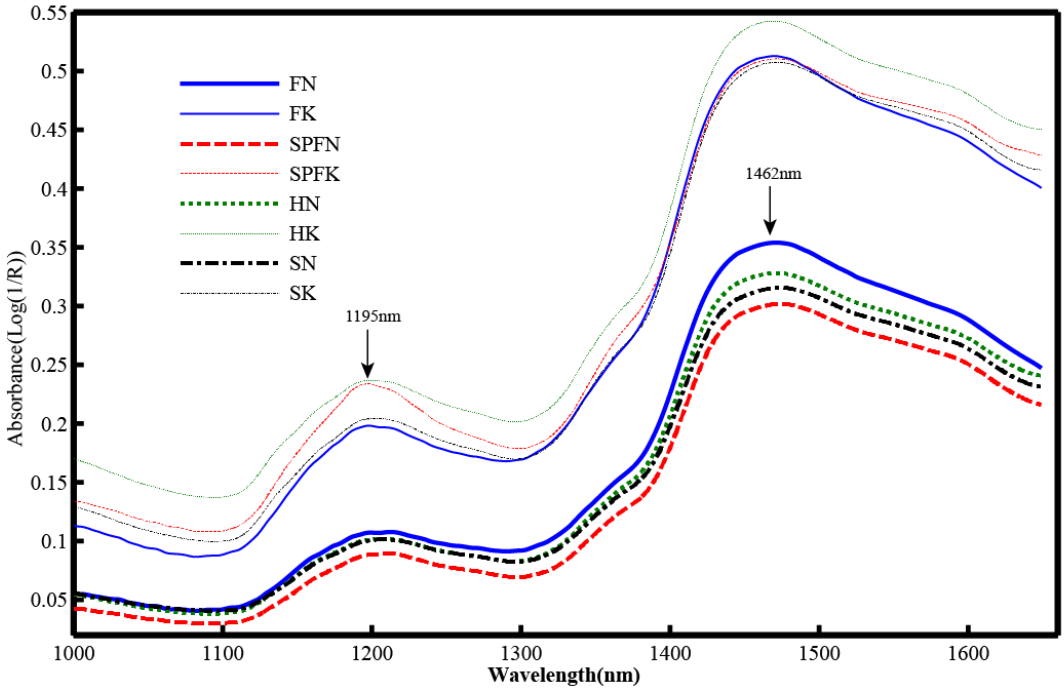

Fig. 1. The mean spectra from different types of lumbers for knot and normal samples (FN - normal samples for Douglas fir, FK - knot samples for Douglas fir, SPFN - normal samples for SPF, SPFK - knot samples for SPF, HN - normal samples for Chinese hemlock, HK - knot samples for Chinese hemlock, SN - normal samples for Dragon spruce, SK - normal samples for Dragon spruce)

All spectra showed several absorption band peaks in the wavelength region of 1000 nm to 1650 nm, including 1195 nm and 1462 nm. The absorbance near 1195 nm was bound up with the second overtone of C-H stretching vibration from lignin or cellulose (Schwanninger *et al.* 2011). The strong peak at approximately 1462 nm was primarily attributed to the O-H stretching from water or phenolic groups of lignin (Schwanninger *et al.* 2011;Yang *et al.* 2015).

Comparison with Different Modelling Methods and Spectral Preprocess Methods

To test the effects of modelling methods and spectral preprocess methods on knot detection, three types of pretreatments were used: no preprocess (NP), standard normal variate transformation (SNV), and first derivative (FD). Following, PLS-LDA and LS-SVM models for each wood species were developed based on the pretreated spectra. The performances of these models are listed in Table 2.

As can be seen in Table 2, both spectral preprocess methods and modelling methods affected the performance of knot detection. For samples from Douglas fir and Dragon spruce, FD-LS-SVM models showed superior performance to other models in both calibration and perdition sets. For samples from Chinese hemlock, the NP-LS-SVM model showed higher sensitivity, specificity, and accuracy than the SNV-LS-SVM and FD-LS-SVM models in calibration but had lower sensitivity in the perdition set. Thus, the NP-LS-SVM model may result in overfitting for hemlock samples.

Table 2. Results for Different Pre-Processing Methods and Modelling Methods

Wood Species	Modelling Method	Spectra Method	Calibration Set			Prediction Set		
			SENS (%)	SPEC (%)	ACCU (%)	SENS (%)	SPEC (%)	ACCU (%)
F	PLS-LDA	NP	99.08	93.24	96.72	100.00	89.19	95.60
		SNV	99.08	95.95	97.81	100.00	89.19	95.60
		FD	100.00	95.95	98.36	100.00	89.19	95.60
	LS-SVM	NP	100.00	94.59	97.81	100.00	89.19	95.60
		SNV	100.00	97.30	98.91	100.00	94.59	97.80
		FD	100.00	100.00	100.00	100.00	97.30	98.90
SPF	PLS-LDA	NP	99.07	95.93	97.39	100.00	95.08	97.37
		SNV	99.07	95.12	96.96	100.00	96.72	98.25
		FD	100.00	93.50	96.52	100.00	95.08	97.37
	LS-SVM	NP	100.00	97.56	98.70	100.00	95.08	97.37
		SNV	100.00	100.00	100.00	100.00	96.72	98.25
		FD	100.00	99.19	99.57	100.00	96.72	98.25
H	PLS-LDA	NP	96.63	86.54	92.91	97.78	88.46	94.37
		SNV	97.75	94.23	96.45	100.00	80.77	92.96
		FD	97.75	92.31	95.74	95.56	84.62	91.55
	LS-SVM	NP	100.00	100.00	100.00	93.33	92.31	92.96
		SNV	97.75	98.08	97.87	97.78	92.31	95.77
		FD	100.00	98.08	99.29	97.78	92.31	95.77
S	PLS-LDA	NP	100.00	89.66	96.03	100.00	89.19	96.00
		SNV	100.00	93.10	97.35	100.00	89.19	96.00
		FD	100.00	91.38	96.69	100.00	89.19	96.00
	LS-SVM	NP	100.00	98.28	99.14	100.00	96.43	98.67
		SNV	100.00	100.00	100.00	93.62	96.43	94.67
		FD	100.00	100.00	100.00	100.00	96.43	98.67

Note: NP - no preprocess; SNV - standard normal variate transformation, FD - first derivative, SENS - sensitivity, SPEC - specificity, ACCU - accuracy

The FD-LS-SVM is suitable for knot detection for samples from hemlock because good results were archived both in calibration and perdition sets. For samples from commercial SPF, the FD-LS-SVM model showed the same performance as the SNV-LS-SVM model in perdition set. The specificity (99.19%) of the FD-LS-SVM model was only slightly worse than that (100.00%) of SNV-LS-SVM model in calibration. Therefore, the FD-LS-SVM is also suitable for knot detection on commercial SPF wood. Nearly all LS-SVM models yielded better results than PLS-LDA models for all wood species whatever spectral pretreatments were used. Also, spectral preprocess methods cannot improve the performance of PLS-LDA model. It is possible that LS-SVM could take advantage of the latent nonlinear information (color, density, moisture content, texture, *etc*.) of the spectra data, which contributed better results, and PLS-LDA only dealt with the linear relationship between the spectra and knots. Therefore, only LS-SVM was used to build models in the further study.

Effect of Single Models and Mixed Models for Knots Detection

In order to test the adaptability of a model built with samples from one wood species to detect knots from other coniferous wood species, all LS-SVM models established within one wood species in Table 2 were used to predict the samples from other three coniferous wood species. The results are shown in Table 3.

Table 3. Results of Using a Model Built Within One Species to Discriminate Samples from Other Species

Wood Species	Spectra Method	Sample Species	Prediction Set		
			SENS (%)	SPEC (%)	ACCU (%)
F	NP	SPF	100.00	75.41	86.84
		H	97.78	73.08	88.73
		S	100.00	64.29	86.67
	SNV	SPF	73.58	91.80	83.33
		H	100.00	26.92	73.24
		S	100.00	82.14	93.33
	FD	SPF	98.11	70.49	83.33
		H	100.00	38.46	77.46
		S	100.00	50.00	81.33
SPF	NP	F	57.41	94.59	72.53
		H	88.99	92.31	90.14
		S	95.74	96.43	96.00
	SNV	F	75.93	86.49	80.22
		H	100.00	61.54	85.92
		S	97.87	92.86	96.00
	FD	F	14.81	97.30	48.35
		H	91.11	92.31	91.55
		S	100.00	96.43	98.67
H	NP	F	64.81	86.49	73.63
		SPF	96.23	95.08	95.61
		S	95.74	92.86	94.67
	SNV	F	24.07	94.59	52.75
		SPF	13.21	100.00	59.65
		S	74.47	100.00	84.00
	FD	F	1.85	100.00	41.76
		SPF	69.81	96.72	84.21
		S	97.87	96.43	97.33
S	NP	F	96.30	81.08	90.11
		SPF	100.00	96.72	98.25
		H	97.78	88.46	94.37
	SNV	F	27.78	91.89	53.85
		SPF	81.13	96.72	89.47
		H	91.11	92.31	91.55
	FD	F	62.96	81.08	70.33
		SPF	100.00	95.08	97.37
		H	97.78	92.31	95.77

Most models built with the samples from Douglas fir showed high sensitivity (97.78% to 100.00%) to samples from Chinese hemlock, Dragon spruce, and commercial SPF, but the specificity and accuracy were low. Thus, regardless of the preprocess methods used, the models built with the samples from Douglas fir could not be used to discriminate a knot on the other three coniferous wood species. This result reflects the differences (color, density, *etc*.) between Douglas fir and the other three wood species. The sensitivity, specificity, and accuracy of the samples from SPF and Chinese hemlock based on the NP-LS-SVM and FD-LS-SVM models built with the samples from Dragon spruce were almost satisfactory, and *vice versa*. It seems that there is enormous potential for using a model built with one wood species to classify a knot from samples from other wood species

among Chinese hemlock, Dragon spruce, and commercial SPF. However, there is no model built with one wood species to simultaneously classify knots from samples from other coniferous species in the present study. It is necessary to establish a mixed model to classify knots from different wood species.

Table 4 shows the performance of mixed models. The mixed model of FD-LS-SVM had best performance. The sensitivity, specificity, and accuracy for the calibration and perdition sets were 100.00%, 99.02%, 99.57% and 99.50%, 95.39%, 97.72%, respectively. When the model was used to test samples of Chinese hemlock, Dragon spruce, and commercial SPF, the results were the same as those of corresponding single models. Only the classification accuracy of samples from Douglas fir based on the mixed FD-LS-SVM model was shown to be slightly worse than that of single models on the basis of samples from Douglas fir. Thus, there is enormous potential for using a mixed model built with all wood species to classify knots among wood of Douglas fir, Chinese hemlock, Dragon spruce, and SPF.

Table 4. Results for Full Wavelengths Models and Effective Wavelengths Model

Model Type	Spectra Method	Wood Species	Calibration Set			Prediction Set		
			SENS (%)	SPEC (%)	ACCU (%)	SENS (%)	SPEC (%)	ACCU (%)
FW	NP	Mixed	100.00	97.07	98.72	98.99	94.74	97.15
		F	—	—	—	98.15	91.89	95.60
		SPF	—	—	—	100.00	96.72	98.25
		H	—	—	—	97.78	92.31	95.77
		S	—	—	—	100.00	96.43	98.67
	SNV	Mixed	99.25	96.42	98.01	99.50	94.74	97.44
		F	—	—	—	100.00	94.59	97.80
		SPF	—	—	—	100.00	96.72	98.25
		H	—	—	—	97.78	92.31	95.77
		S	—	—	—	100.00	92.86	97.33
	FD	Mixed	100.00	99.02	99.57	99.50	95.39	97.72
		F	—	—	—	100.00	94.59	97.80
		SPF	—	—	—	100.00	96.72	98.25
		H	—	—	—	97.78	92.31	95.77
		S	—	—	—	100.00	96.43	98.67
EW	NP	Mixed	99.75	93.49	97.02	98.49	93.42	96.30
		F	—	—	—	98.15	91.89	95.60
		SPF	—	—	—	100.00	96.72	98.25
		H	—	—	—	95.56	88.46	92.96
		S	—	—	—	100.00	92.86	97.33

Note: FW - full wavelength, EW - effective wavelength

Effective Wavelengths Selection Based on Random Frog

RF was used to select effective wavelengths from raw data. The selected wavelengths can be used to develop a simple system for knot detection. If such a system is developed on preprocessed full spectra, the preprocessed spectral profiles from the new system will not be comparable to those from full wavelengths, owing to lack of some parameters inclusive to spectral mean and standard deviation (Feng and Sun 2013). It will not ensure the success of the new system, and make the development of a new system meaningless.

Figure 2 displays the mean selection probability of each wavelength with 100 repetitions by random frog. Though some small number of wavelengths had high selection probabilities, most of the wavelengths had low selection probability. These results showed that there existed a lot of useless information to make knot samples detected using NIRS. Based on the experience, the threshold was set as 0.72, and the wavenumbers with selection probability bigger than the threshold (0.72) were selected as the effective wavenumbers. Therefore, eight wavenumbers (1314 nm, 1358 nm, 1409 nm, 1340 nm, 1260 nm, 1586 nm, 1288 nm, and 1402 nm) above the dotted line were obtained as the final effective wavenumbers after random frog algorithm. These wavebands were ranked by descending order of selection probability.

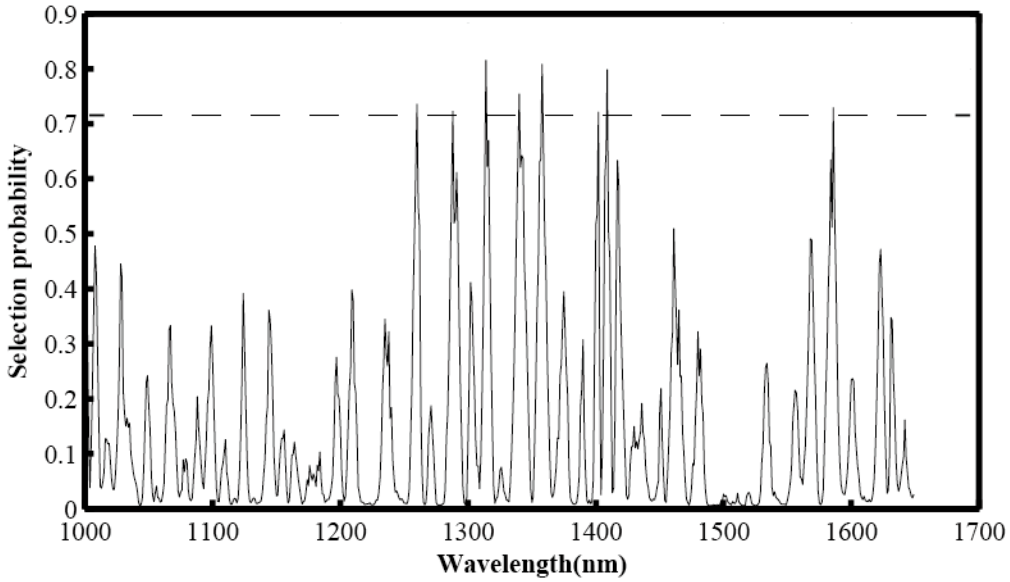

Fig. 2. Selection probability of each wavelength

Among the optimum wavelengths selected by RF, 1358 nm is primarily assigned to the first overtone of C-H stretching vibration and deformation vibration from lignin or hemicelluloses and cellulose. The peak at 1409 nm is due to the first overtone of O-H stretching vibration, which is attributed to phenolic groups of lignin, while that at 1586 nm is bound up with the first overtone of O-H stretching vibration from cellulose (Schwanninger *et al.* 2011). Knots generally contain a high percentage of compression wood, which has a relatively high content of lignin and low content of cellulose (Fujimoto and Tsuchikawa 2010). It is inexplicable that 1314 nm, 1260 nm, 1288 nm, 1340 nm, and 1402 nm were selected as effective wavelengths. However, without further investigation, the only reason to choose them is to improve the model.

By employing the above-mentioned spectral information, an effective wavelengths based model named EWs-NP-LS-SVM was developed (Table 4).

Compared with the full wavelengths (FWs) based models, the EWs-LS-SVM model was found to be slightly accurate less than the FWs-NP-LS-SVM model and the FWs-FD-LS-SVM model both in calibration and validation sets. It is worth mentioning that models with EW-LS-SVM gave an acceptance result considering 98.77% of variables were eliminated (8 *vs*. 650). Therefore, the wavelengths selected according to random frog were effective. The effective wavelengths could represent most of the features and characteristics of the raw spectra and could be applied instead of the full wavelengths to

discriminate knots for samples from different wood species. Furthermore, the effective wavelengths might be important for the development of a simple system for knot detection during industrial applications.

CONCLUSIONS

1. Near infrared spectroscopy (NIRS) coupled with appropriate chemometrics methods seems to be an efficient technique to discriminate knots on lumber surface for four types of coniferous wood in the wavelength range of 1000 to 1650 nm.
2. The least squares-support vector machine (LS-SVM) was superior to partial least squares-linear discriminant analysis (PLS-LDA) to establish knot detection models. LS-SVM coupled with first derivative preprocessed spectra achieved best performance for both single models of each wood species and mixed models of all the four coniferous wood species. For mixed model, the sensitivity, specificity and accuracy in validation set were 99.50%, 95.39% and 97.72%, respectively.
3. Models built with the samples from spruce can be used to classify knot samples from SPF and hemlock, and *vice versa*. However, there are no models built with one wood species to classify knot samples from the other three coniferous wood species, and *vice versa*. To improve the adaptability of the model, a mixed model should be established.
4. Eight effective wavelengths (1314 nm, 1358 nm, 1409 nm, 1340 nm, 1260 nm, 1586 nm, 1288 nm, and 1402 nm) were selected by random frog algorithm to build EWs - LS-SVM model, the sensitivity, specificity and accuracy for validation set were 98.49%, 93.42%, and 96.30%. The simplified model was found to be slightly less accurate than the full wavelengths based model. However, even when using data corresponding to just eight wavelengths, acceptable results could be obtained. The selected wavelengths might be used to develop a simple system for knot detection during industrial applications.

ACKNOWLEDGMENTS

This work is funded by the Natural Science Foundation of Zhejiang Province, China (No. LQ13F050006), the Pre-research Project of the Research Center for Smart Agriculture and Forestry of Zhejiang A&F University (No. 2013ZHNL03), and the Natural Science Foundation of Zhejiang A&F University (No. 2012FR085).

REFERENCES CITED

Baradit, E., Aedo, R., and Correa, J. (2006). "Knots detection in wood using microwaves," *Wood Science and Technology* 40(2), 118-123. DOI: 10.1007/s00226-005-0027-8

Chen, Q., Jiang, P., and Zhao, J. (2010). "Measurement of total flavone content in snow lotus (*Saussurea involucrate*) using near infrared spectroscopy combined with

interval PLS and genetic algorithm," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 76(1), 50-55. DOI: 10.1016/j.saa.2010.02.045

Cristhian, A. C., Sanchez, R., and Baradit, E. (2008). "Detection of knots using x-ray tomographies and deformable contours with simulated annealing," *Wood Research* 53(2), 57-66.

Fackler, K., Schwanninger, M., Gradinger, C., Srebotnik, E., Hinterstoisser, B., and Messner, K. (2007). "Fungal decay of spruce and beech wood assessed by near-infrared spectroscopy in combination with uni-and multivariate data analysis," *Holzforschung* 61(6), 680-687. DOI: 10.1515/HF.2007.098

Feng, Y.-Z., Downey, G., Sun, D.-W., Walsh, D., and Xu, J.-L. (2015). "Towards improvement in classification of *Escherichia coli*, *Listeria innocua* and their strains in isolated systems based on chemometric analysis of visible and near-infrared spectroscopic data," *Journal of Food Engineering* 149, 87-96. DOI: 10.1016/j.jfoodeng.2014.09.016

Feng, Y.-Z., and Sun, D.-W. (2013). "Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of *Pseudomonas* loads in chicken fillets," *Talanta* 109, 74-83. DOI: 10.1016/j.talanta.2013.01.057

Fujimoto, T., Kurata, Y., Matsumoto, K., and Tsuchikawa, S. (2010). "Feasibility of near-infrared spectroscopy for online multiple trait assessment of sawn lumber," *Journal of Wood Science* 56(6), 452-459. DOI: 10.1007/s10086-010-1122-5

Fujimoto, T., and Tsuchikawa, S. (2010). "Identification of dead and sound knots by near infrared spectroscopy," *Journal of Near Infrared Spectroscopy* 18(6), 473-479. DOI: 10.1255/jnirs.887

Hu, C., Min, X., Yun, H., Wang, T., and Zhang, S. (2011). "Automatic detection of sound knots and loose knots on sugi using gray level co-occurrence matrix parameters," *Annals of Forest Science* 68(6), 1077-1083. DOI: 10.1007/s13595-011-0123-x

Karsulovic, J. T., León, L. A., and Gaete, L. (2000). "Ultrasonic detection of knots and annual ring orientation in Pinus radiata lumber," *Wood and Fiber Science* 32(3), 278-286.

Kennard, R. W., and Stone, L. A. (1969). "Computer aided design of experiments," *Technometrics* 11(1), 137-148. DOI: 10.1080/00401706.1969.10490666

Kodama, Y., and Akishika, T. (1993). "Non-destructive inspection of defects in woods by use of the pulse-echo technique of ultrasonic waves, 1: Measurements of enclosed knots," *Mokuzai Gakkaishi* 39(1), 7-12.

Li, H.-D., Xu, Q.-S., and Liang, Y.-Z. (2012). "Random frog: An efficient reversible jump Markov chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification," *Analytica Chimica Acta* 740, 20-26. DOI: 10.1016/j.aca.2012.06.031

Machado, J. S., Sardinha, R. A., and Cruz, H. P. (2004). "Feasibility of automatic detection of knots in maritime pine timber by acousto-ultrasonic scanning," *Wood Science and Technology* 38(4), 277-284. DOI: 10.1007/s00226-004-0224-x

Nagai, H., Murata, K., and Nakano, T. (2009). "Defect detection in lumber including knots using bending deflection curve: Comparison between experimental analysis and finite element modeling," *Journal of Wood Science* 55(3), 169-174. DOI: 10.1007/s10086-008-1016-y

Poke, F. S., and Raymond, C. A. (2006). "Predicting extractives, lignin, and cellulose contents using near infrared spectroscopy on solid wood in *Eucalyptus globulus*,"

Journal of Wood Chemistry and Technology 26(2), 187-199. DOI: 10.1080/02773810600732708

Rice, R. W., Steele, P. H., and Kumar, L. (1992). "Detecting knots and voids in lumber with dielectric sensors," *Industrial Metrology* 2(3-4), 309-315. DOI: 10.1016/0921-5956(92)80010-Q

Schwanninger, M., Rodrigues, J. C., and 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. DOI: 10.1255/jnirs.955

Shou, G., Zhang, W., Gu, Y., and Zhao, D. (2014). "Application of near infrared spectroscopy for discrimination of similar rare woods in the Chinese market," *Journal of Near Infrared Spectroscopy* 22(6), 423-432. DOI: 10.1255/jnirs.1136

Suykens, J. A. K., and Vandewalle, J. (1999). "Least squares support vector machine classifiers," *Neural Processing Letters* 9(3), 293-300. DOI: 10.1023/A:1018628609742

Szymani, R., and McDonald, K. A. (1981). "Defect detection in lumber: State of the art," *Forest Products Journal* 31(11), 34-44.

Takeshi, O., Ryo, A., Kazumi, S., and Yasuhide, M. (2006). "Development of an automatic detection system for knots appearing on the surface of sugi laminae with a two-dimensional, three-charge couple device color camera," *Journal of the Faculty of Agriculture, Kyushu University* 51(1), 29-32.

Tormanen, V.-M. O., and Makynen, A. J. (2009). "Detection of knots in veneer surface by using laser scattering based on the tracheid effect," in: *IEEE Instrumentation and Measurement Technology 2009 Conference*, pp. 1439-1443.

Tsuchikawa, S., Inoue, K., Noma, J., and Hayashi, K. (2003). "Application of near-infrared spectroscopy to wood discrimination," *Journal of Wood Science* 49(1), 29-35. DOI: 10.1007/s100860300005

Via, B. K., So, C.-L., Eckhardt, L. G., Shupe, T. F., Groom, L. H., and Stine, M. (2008). "Response of near infrared diffuse reflectance spectra to blue stain and wood age," *Journal of Near Infrared Spectroscopy* 16(1), 71-74. DOI: 10.1255/jnirs.756

Xie , Y., and Wang, J. (2015). "Study on the identification of the wood surface defects based on texture features," *Optik-International Journal for Light and Electron Optics* 126(19), 2231-2235. DOI: 10.1016/j.ijleo.2015.05.101

Yang, Z., Chen, L., Fu, Y., and Lv, B. (2012). "Rapid detection of knot defect in wood surface by near infrared spectroscopy coupled with SIMCA pattern recognition," *Journal of Northeast Forestry University* 40(8), 70-72. DOI: 10.3969/j.issn.1000-5382.2012.08.015

Yang, Z., Ren, H., and Jiang, Z. (2008). "Discrimination of wood biological decay by NIR and partial least squares discriminant analysis (PLS-DA)," *Spectroscopy and Spectral Analysis* 28(4), 793-796. DOI: 10.3964/j.issn.1000-0593.2008.04.018

Yang, Z., Zhang, M., Chen, L., and Lv, B. (2015). "Non-contact detection of surface quality of knot defects on eucalypt veneers by near infrared spectroscopy coupled with soft independent modeling of class analogy," *BioResources* 10(2), 3314-3325. DOI: 10.15376/biores.10.2.3314-3325

Yang, Z., Zhang, M., Li, K., and Chen, L. (2016). "Rapid detection of knot defects on wood surface by near infrared spectroscopy coupled with partial least squares discriminant analysis," *BioResources* 11(1), 2557-2567. DOI: 10.15376/biores.11.1.2557-2567

Yun, Y.-H., Li, H.-D., Wood, L. R. E., Fan, W., Wang, J.-J., Cao, D.-S., Xu, Q.-S., and Liang, Y.-Z. (2013). "An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 111, 31-36. DOI: 10.1016/j.saa.2013.03.083

Zheng, H., and Lu, H. (2012). "A least-squares support vector machine (LS-SVM) based on fractal analysis and CIELab parameters for the detection of browning degree on mango (*Mangifera indica* L.)," *Computers and Electronics in Agriculture* 83, 47-51. DOI: 10.1016/j.compag.2012.01.012

Zhou, Z., Zeng, S., Li, X., and Zheng, J. (2015). "Nondestructive detection of blackheart in potato by visible/near infrared transmittance spectroscopy," *Journal of Spectroscopy* 2015, 1-9. DOI: 10.1155/2015/786709

Article submitted: July 7, 2016; Peer review completed: September 4, 2015; Revised version received and accepted: September 13, 2016; Published: September 22, 2016.
DOI: 10.15376/biores.11.4.9533-9546