

IPSO-BP 木材绝干密度近红外光谱预测模型

于 雷, 陈金浩, 李龙飞, 李 超*, 张怡卓*

东北林业大学机电工程学院, 黑龙江 哈尔滨 150040

摘 要 木材密度决定着木材力学性能,是木材物理性能的重要指标之一。近年来,由于近红外光谱分析具有操作过程简单、方便、快速等优势,已有学者应用近红外光谱数据预测木材密度,但是,在实际应用中,数据集缺乏、光谱特征欠优、非线性拟合准确性低等问题还没有得到完全解决,木材密度预测模型的精度有待于进一步提升。木材密度中,木材的绝干密度相对稳定,测量结果相对精确,以柞木绝干密度预测为研究对象,通过采集不同含水率下的近红外光谱信息构建出适合不同含水率条件的绝干密度非线性预测模型。首先,选用德国 INSION 公司的近红外光纤光谱仪,运用 SPEC view 7.1 软件对不同含水率的柞木样本采集光谱信息;然后,利用 SPXY 样本划分方法,按 2:1 划分校正集与预测集,并利用多元散射校正、二阶导数光谱及 S-G 平滑方法进行光谱预处理,以减少散射光和高频噪声的影响;接着,运用连续投影算法(SPA)提取有效的波长信息;最后,运用一种非线性权重粒子群算法优化的 BP 网络(IPSO-BPNN)建立不同含水率状态下的近红外光谱与柞木绝干密度之间的关联,实现柞木绝干密度预测模型的构建。实验过程中,加工选取了 100 个柞木试件样本,在绝干条件下测量样本试件密度和光谱信息,并浸泡水中测量出不同含水率对应的的光谱信息,实验结果表明:SPXY 保证了校正集样本的均匀分布,提高了模型泛化能力;MSC、二阶导数和 S-G 卷积平滑相结合的方法抑制了原始光谱中噪声高频信号,同时使得峰值更加突出;SPA 从 117 个光谱数据中优选出 16 个特征波长,提高了建模精度;IPSO-BPNN 模型较 SPA-PLS, BP 和 PSO-BP 具有更高的相关系数,更小的均方根误差,柞木绝干密度预测相关系数为 0.938,预测均方根误差为 0.012 9,方法可以对木材密度在线无损测量提供一定的理论基础。

关键词 木材绝干密度;近红外光谱;SPA;IPSO-BP 网络

中图分类号: S781.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)09-2937-06

引 言

木材密度是木材的重要物理性能指标,与木材力学性能有着密切的联系。木材密度可分为基本密度、气干密度和绝干密度,三种密度之间可以通过公式相互转换。其中,绝干状态下的样本质量和体积相对稳定,绝干密度测试结果更精确。木材密度的检测方法有称重法、机械力密度检测法和射线密度检测法^[1]。传统称重法测量结果准确,但是操作过程复杂,耗时长;机械力密度检测法快速有效,但属于破坏性实验;射线密度检测方法能够快速无损检测木材密度,但是实验环境要求严格,射线会对操作人员产生危害。

近红外光谱分析技术具有操作过程简单、方便、时间快

且无损等优点,近年来已被应用到木材密度检测中^[2-4]。Alves 利用近红外光谱建立了海岸松和落叶松密度的偏最小二乘法回归模型^[5]。Isik 结合近红外分析与最小二乘支持向量机建立了木材密度、力学强度的快速预测^[6]。Takkaaki 运用近红外光谱分析出影响木材密度的敏感光谱波点,并运用偏最小二乘建立了不同含水率下的木材密度预测模型^[7]。李耀翔等研究枫桦木材密度时,探讨了小波压缩的近红外光谱处理方法,并利用偏最小二乘回归方法对数据进行了建模^[8]。尽管国内外学者已经开展木材密度的近红外光谱预测方法研究,但是,多数研究是在含水率一定条件下建立的木材密度预测模型,此外,在建模过程由于缺少合理的样本集优选、光谱预处理、特征优选以及非线性建模等环节,木材密度建模精度有待于进一步提升。

收稿日期: 2019-07-25, **修订日期:** 2019-11-09

基金项目: 国家自然科学基金项目(31700643),中央高校基本科研业务费专项资金项目(2572017DB05),黑龙江省自然科学基金项目(C2017005)资助

作者简介: 于 雷,1982 年生,东北林业大学林业工程博士后 e-mail: relayjfr@163.com

* 通讯联系人 e-mail: nefuzyx@163.com; lchao820225@163.com

1 实验部分

以柞木(*Xylosma racemosum*)绝干密度预测为例,首先采集不同含水率下的柞木样本表面光谱数据,并通过 SPXY 样本划分方法得到校正集与预测集;然后,对光谱进行预处理消除基线漂移、散射影响和噪声;最后,运用 SPA 对光谱数据进行特征降维,并运用 IPSO-BP 网络建立任意含水率条件下光谱信息与绝干密度间的联系。具体建模过程如图 1 所示。

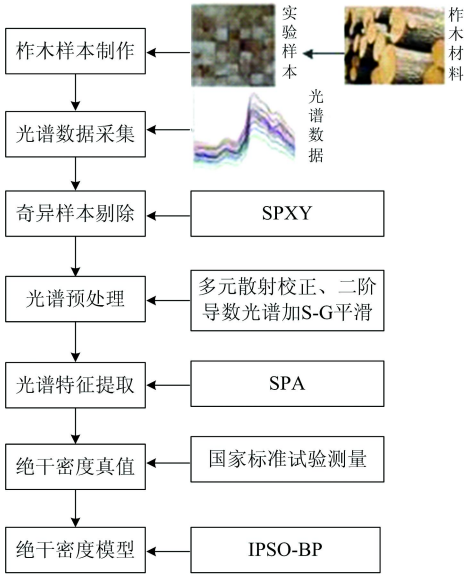


图 1 木材绝干密度建模流程

Fig. 1 Modeling process of wood absolute drying density

1.1 光谱采集与柞木绝干密度的真值获取

柞木样本来自于黑龙江省五常市林业局冲河林场,按照国家标准(GB1929—2009)制作 20 mm×20 mm×20 mm 的密度标准样本,选出其中 100 个无明显颜色差异且无缺陷的样本进行密度预测研究,图 2、图 3、图 4 分别为加工的柞木样本试件、样本干燥及称重过程。

选用德国 INSION 近红外光纤光谱仪,运用 SPEC view 7.1 软件对木材密度样本进行光谱信息采集。该光谱仪分辨率小于 16 nm,光谱范围为 900~1 700 nm,共 117 个波长点。在采集过程中,温度控制在 20 ℃,平均相对湿度为 50%。为了采集不同含水率的柞木样本近红外光谱,首先将样本充分浸泡在水中,取出后,从样本集中随机采集 10 个样本的光谱信息并记录其编号,然后按照表 1 时间间隔,从剩余样本中再随机选择 10 个样本采集光谱信息,直到最终采集全部试件光谱信息。在采集绝干密度光谱信息时,由于绝干材在空气中会很快地吸收水分而达到平衡含水率,所以在测量木材绝干密度时要足够快。实验证明,1 s 的时间差,柞木样本含水率变化非常小,可以忽略不计。实测样本绝干密度和含水率分布如表 2、表 3 所示。在光谱采集过程中,探头高度保持 1 mm 不变,分别在柞木样本的上下两个横切面采集 4 点,并取其平均值。



图 2 部分样本实拍图

Fig. 2 Picture of some samples



图 3 试样烘干

Fig. 3 Specimen drying

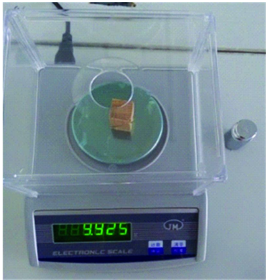


图 4 试样全干状态质量测量

Fig. 4 Quality measurement of smples in dry state

表 1 不同含水率样本制备

Table 1 Preparing process of samples with different moisture contents

步骤	干燥与浸泡时间	含水率最小值/%	含水率最大值/%
1	室温下 1 h	50.5	67.1
2	室温下 5 h	41.5	56.3
3	室温下 9 h	33.7	53.1
4	室温下 13 h	32.4	49.9
5	室温下 17 h	28.8	4.09
6	室温下 21 h	23.9	44.1
7	室温下 25 h	20.8	29.7
8	室温下 29 h	22.9	28.1
9	干燥箱 1 h	8.70	15.3

表 2 样本绝干密度分布

Table 2 Distribution of absolute dry density

样本数量	平均密度	最大密度	最小密度	标准差
100	0.753 6	0.836 4	0.699 3	0.017 1

表 3 样本含水率分布
Table 3 Sample moisture distribution

样本数量	平均含水率	最大含水率	最小含水率	标准差
100	0.336 4	0.665 8	0.087 3	0.035 2

1.2 光谱预处理与特征波长的优选

为了避免样本分布不均匀所带来的校正集不具备代表性的弊端,选用 SPXY 样本划分方法提高模型的泛化性和稳健性。由于样本光谱数据 x 和真值 y 对建模结果都有影响,在考虑样本之间的距离时,将自变量和因变量同时考虑在内;此外,为了确保 x 和 y 在空间具有同样的权重,分别除以它们自己距离的最大值。具体公式如式(1)—式(3)

$$d_x(p,q)=\sqrt{\sum_{j=1}^J[x_p(j)-x_q(j)]^2},p;q\in[1,N]$$
(1)

$$d_y(p,q)=\sqrt{(y_p-y_q)^2}=|y_p-y_q|;p,q\in[1,N]$$
(2)

$$d_{xy}(p,q)=\frac{d_x(p,q)}{\max_{p,q\in[1,N]}d_x(p,q)}+\frac{d_y(p,q)}{\max_{p,q\in[1,N]}d_y(p,q)};$$
$$p,q\in[1,N]$$
(3)

特征波长选择会影响建模的精度和速度,连续投影算法(SPA)是一种前向变量选择算法,它能够使矢量空间的共线性达到最小化^[9]。SPA 的工作原理是通过迭代对原始数据投影映射,构造新的变量集,建立回归模型评价预测效果。该算法所筛选出的敏感特征波段具有重要性顺序,能直接反映所筛选变量与因变量之间的定量关系。

1.3 基于 IPSO-BPNN 的柞木绝干密度预测模型

运用一种改进的粒子群优化(IPSO)算法优化 BP 网络,并建立柞木绝干密度近红外预测模型,以克服 BP 网络容易陷入局部优化,且训练的速度过慢的问题。传统 PSO 算法可以描述为一个 D 维的潜在解空间中,由 n 个粒子组成的种群 $x=(x_1,x_2,\cdots,x_n)$,其中第 i 个粒子表示为一个 D 维的向量 $x_i=(x_{i1},x_{i2},\cdots,x_{id})$,代表第 i 个粒子在 D 维搜索空间中的一个潜在解。根据目标函数可计算出每个粒子位置 x_i 对应的适应度值。第 i 个粒子的速度为 $v_i=(v_{i1},v_{i2},\cdots,v_{id})$,其个体极值的位置 $p_g=(p_{g1},p_{g1},p_{g2},\cdots,p_{gd})$ 。在每一次迭代过程中,粒子通过个体极值和全局极值更新自身的速度和位置,更新公式如式(4)和式(5)

$$v_{id}^{k+1}=\omega v_{id}^k+c_1r_1(p_{id}^k-x_{id}^k)+c_2r_2(p_{gd}^k-x_{id}^k)$$
(4)

$$x_{id}^{k+1}=x_{id}^k+v_{id}^{k+1}$$
(5)

式中, ω 为惯性重, $d=1,2,\cdots,D$; $i=1,2,\cdots,n$; k 为当前迭代次数; v_{id} 为粒子的速度; c_1 和 c_2 为加速因子; r_1 和 r_2 为分布于 $[0,1]$ 上的随机数。

针对粒子群优化算法中惯性权重线性递减策略影响收敛速度等一系列的问题,弥补算法的不足,采用一种对 ω 非线性递减策略减少的群算法(IPSO)^[10],算法公式如式(6)

$$\omega=\omega_{\max}-(\omega_{\max}-\omega_{\min})k/T_{\max}$$
(6)

式中, $\omega_{\max}$ 和 $\omega_{\min}$ 分别为最大惯性权重和最小惯性权重, k 为算法当前迭代次数; $T_{\max}$ 为最大迭代次数,通常 $\omega_{\max}$ 和 $\omega_{\min}$ 取

0.9 和 0.4 时算法的性能会获得较大的提升。

2 结果与讨论

2.1 样本划分

100 个柞木样本以 2 : 1 的比例划分为校正集 67 个,预测集 33 个。为验证样本集划分的重要性,在采集柞木表面近红外光谱数据后,运用 SPXY 方法取处校正集和预测集,运用 PLS 模型来评价预测结果有效性。表 4 为运用 SPXY 和随机选择样本建模的比较,参数表明:基于 SPXY 方法的 PLS 模型的指标参数都有所改善。

表 4 SPXY 与随机样本选择建模比较
Table 4 Modeling comparison between SPXY and random selections

	校正集 R	预测集 R	校正集 RMSE	预测集 RMSE
随机选择	0.86	0.72	0.025 4	0.022 7
SPXY	0.92	0.81	0.016 7	0.018 5

2.2 光谱预处理

采用多元散射校正(MSC)、二阶导数光谱加 S-G 平滑三种方法相结合对样本光谱进行预处理,数据处理工具为 MATLAB 7.0,原始光谱如图 5 所示。经过 MSC 预处理后,原始光谱的散射光得以减弱,光谱变得集中,更能看出光谱走向和趋势,但光谱依然存在信息强度不足和吸收峰不明显现象。在此,加入二阶导数操作,并加入平滑处理,以减弱噪声信号。三种方法结合的预处理结果如图 6 所示。

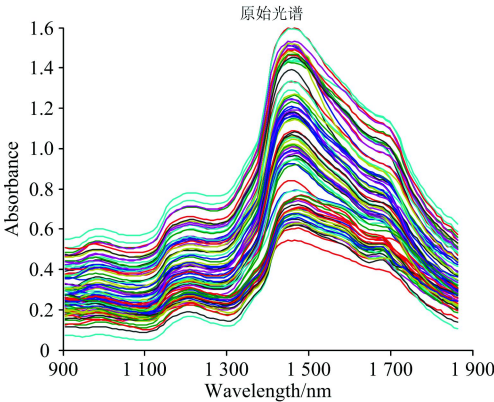


图 5 原始光谱图
Fig. 5 Original spectra

2.3 木材光谱特征选择实验

对近红外光谱的 117 个波长使用 SPA 进行筛选,运用 PLS 进行优选结果分析。在变量个数不断增加的过程中, RMSE 值大幅度降低,当变量个数为 16 时, RMSE 值达到最小,为 0.014 6。当选取变量继续增加时, SMSE 也有所上升。优选光谱中的位置如图 7 所示,不同变量数的与 RMSE 的关系如图 8 所示。

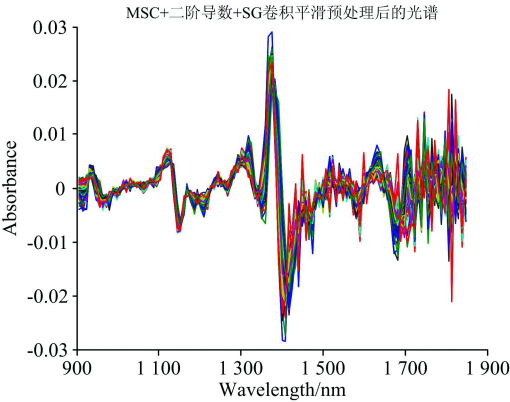


图 6 预处理后的光谱
Fig. 6 Pretreated spectra

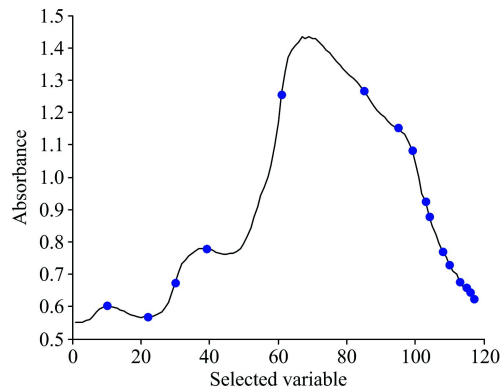


图 7 选择变量在光谱中的位置
Fig. 7 The position of selected variables

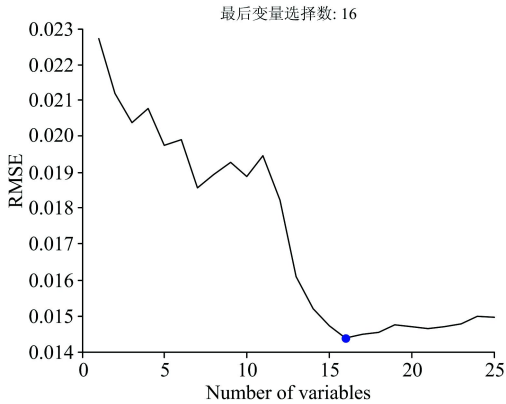


图 8 选择变量与 RMSE 之间的关系
Fig. 8 The relation between selected variables and RMSE

2.4 模型结果

利用 SPA 提取的 16 个特征, 分别建立 PLS 模型、BP 神经网络、PSO-BP 神经网络以及 IPSO-BP 神经网络模型, 并对 33 组数据进行验证, 预测结果如表 5 所示。BP, PSO-BP 和 IPSO-BP 均采用三层结构, 光谱优选波长为输入, 隐层节点数为 10, 柞木绝干密度为输出, 隐含层为 transing S 型传输函数, 输出层为 purelin 线性函数, 训练函数为 traingd 梯

度下降型函数, 学习速率设置为 0.01, 学习训练次数为 10 000 次。图 9 为 PSO-BP 模型预测值与真值散点分布图、图 10 为 IPSO-BP 预测与真值散点分布图。

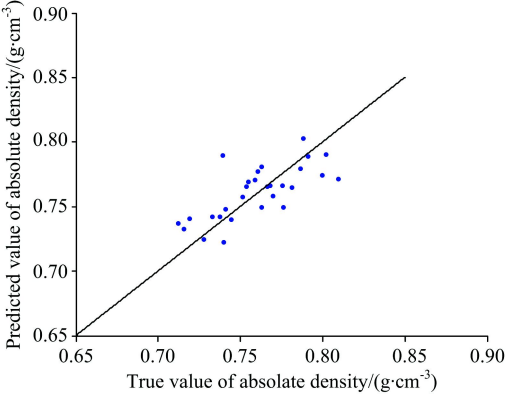


图 9 PSO-BP 预测值与真值散点分布图
Fig. 9 Scatter map of predicted value and true value based on PSO-BP

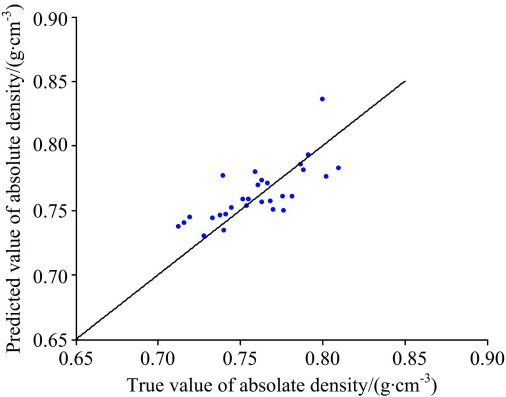


图 10 IPSO-BP 预测值与真值散点分布图
Fig. 10 Scatter map of predicted value and true value based on IPSO-BP

表 5 四种建模方法的比较				
Table 5 The comparison of 4 modeling methods				
	PLS	BP	PSO-BP	IPSO-BP
RMSEP	0.022 6	0.018 9	0.013 5	0.012 9
R	0.829	0.079 8	0.895	0.938

从表 5 可以看出, PLS 模型校正集与预测集的结果较稳定, 说明近红外光谱与绝干密度间存在一定的线性关系; BP 网络预测结果相对较低, 可能因为 BP 网络陷入了局部寻优; PSO-BP 网络相对 BP 的结果较好, 但 IPSO-BP 网络得出的结果最好, 通过图 11 和图 12 可知, 相比较 PSO-BP, IPSO-BP 达到全局最优的速度更快, 预测集相关系数达到 0.938, 预测均方根误差为 0.012 9。

3 结 论

从提高木材预测密度精准度入手, 分别就样本优选、光

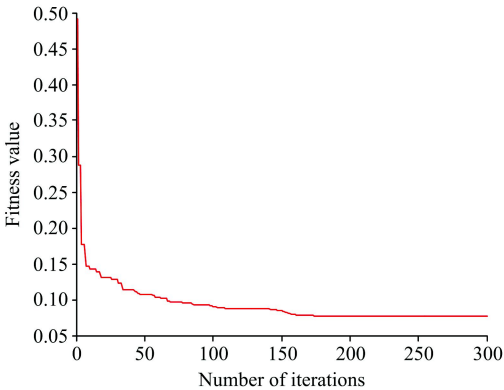


图 11 PSO 迭代次数曲线
Fig. 11 PSO iteration curve

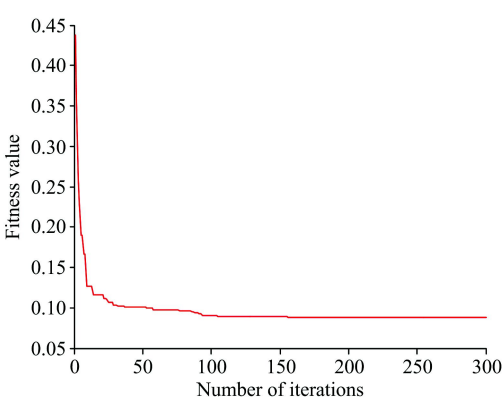


图 12 IPSO 迭代次数曲线
Fig. 12 IPSO iteration curve

谱预处理、特征优选以及光谱数据特征的非线性建模等部分环节进行了深入研究。研究结果表明：SPXY 保证了校正集样本的均匀分布，提高了模型泛化能力；MSC、二阶导数和 S-G 卷积平滑相结合的方法能够很好的抑制原始光谱中噪声高频信号，同时使得峰值更加突出，消除了基线漂移、消弱

了散射光的影响；SPA 可以有效优选出光谱波段，进而提高预测精度缩短预测时间；IPSO-BP 网络在近红外光谱与柞木绝干密度建模上具有更好的结果，对近红外光谱预测木材密度提供了新的思路。

References

[1] SUN Yan-liang, ZHANG Hou-jiang, ZHU Lei, et al(孙燕良, 张厚江, 朱磊, 等). Forest Engineering(森林工程), 2011, 27(1): 23.

[2] TAN Nian, WANG Xue-shun, HUANG An-min, et al(谭念, 王学顺, 黄安民, 等). Scientia Silvae Sinicae(林业科学), 2018, 54(12): 137.

[3] Vu Thi Hong Tham, Tetsuya Inagaki, Satoru Tsuchikawa. Wood Science and Technology, 2018, 52(1): 115.

[4] Haddadi A, Leblon B, Burger J, et al. Wood Material Science & Engineering, 2015, 10(1): 27.

[5] Alves A, Santos A, Rozenberg P, et al. Wood Science & Technology, 2012, 46(1-3): 157.

[6] Isik F, Mora C R, Schimleck L R. Annals of Forest Science, 2011, 68(2): 283.

[7] Fujimoto T, Kobori H, Tsuchikawa S, et al. Journal of Near Infrared Spectroscopy, 2012, 20(3): 353.

[8] LI Yao-xiang, LI Ying, JIANG Li-chun(李耀翔, 李颖, 姜立春). Journal of Beijing Forestry University(北京林业大学学报), 2016, 38(3): 89.

[9] Yu Huiling, Liang Hao, Lin Xue, et al. Optical Engineering, 2018, 57(4): 046103.

[10] HOU Yuan-hang, LIU Fei, LIANG Xiao(侯远行, 刘飞, 梁霄). Journal of Shanghai Jiao Tong University(上海交通大学学报), 2016, 50(8): 1193.

Prediction Model of Wood Absolute Dry Density by Near-Infrared Spectroscopy Based on IPSO-BP

YU Lei, CHEN Jin-hao, LI Long-fei, LI Chao*, ZHANG Yi-zhuo*

College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China

Abstract Wood density is an important physical property, which determines the mechanical properties of wood. In recent years, as NIR has the advantages of simple, convenient and fast operation, it has already been used in terms of wood density prediction. However, in practical application, the sample sets shortage, spectral characteristics selection and non-linear fitting inaccuracy still not been solved definitely, and the accuracy of wood density prediction model needs to be further improved. Among all the wood density parameters, the absolute dry density of wood is relatively stable, and the measurement results are relatively accurate. In this paper, the prediction of absolute dry density of oak is studied. By collecting spectroscopy information under different moisture content, a non-linear prediction model of absolute drying density suitable for arbitrary moisture content is constructed. The near-infrared optical fiber spectrometer of INSION Company in Germany was selected, and the spectral information of oak samples with different moisture content was collected by SPEC view 7.1 software. Then, the calibration set and prediction set were divided according to 2 : 1 using SPXY sample partition method, and multivariate scattering correction, second derivative spectroscopy and S-G smoothing method were used to reduce the influence of scattered light and high-frequency noise; After that, continuous projection algorithm SPA was used to extract effective wavelength information; finally, a BP network (IPSO-BPNN) was used to establish the correlation between near-infrared spectra and oak absolute dry density under different moisture content, which was optimized by a non-linear weighted particle swarm optimization algorithm here. The density and spectral information of 100 samples of oak wood was obtained under absolute drying condition, and the spectral information was collected corresponding to different moisture content. The experimental results show that SPXY guarantees the uniform distribution of calibration samples and improves the generalization ability of the model; Using a combination of MSC, second derivative and S-G convolution can smoothly suppress the high-frequency noise signal in the original spectrum and make the peak value more prominent; 16 characteristic wavelengths were selected by SPA from 117 spectral data. Generally, IPSO-BPNN model has a higher correlation coefficient than SPA-PLS, BP and PSO-BP, own smaller root mean square error. The correlation coefficient of the absolute dry density of oak is 0.938, and the root mean square error is 0.012 9.

Keywords Wood absolute dry density; Near-infrared spectrum; SPA; IPSO-BP network

(Received Jul. 25, 2019; accepted Nov. 9, 2019)

* Corresponding authors