安徽农业大学学报, 2017, 44(3): 429-433
Journal of Anhui Agricultural University
[DOI] 10.13610/j.cnki.1672-352x.20170524.004　网络出版时间：2017/5/24　11:16:20
[URL] http://kns.cnki.net/kcms/detail/34.1162.S.20170524.1116.008.html

基于近红外光谱技术的成年橡胶树叶片氮素含量检测

蒋灿辰，唐荣年*
(海南大学机电工程学院，海口 570228)

摘　要：为了快速并无损地检测成年橡胶树叶片的氮素含量，使用近红外光谱检测技术获取叶片的光谱数据，采用多元散射校正（MSC）对光谱数据预处理后，使用 SPA（连续投影算法）提取光谱数据的有效波长，PCA（主成分分析法）提取光谱数据主成分，然后分别将提取的光谱数据特征值输入到线性回归模型 PLS（偏最小二乘回归）、非线性回归模型 BPNN（BP 神经网络）和 LSSVM（最小二乘支持向量机）中，得到 6 个现有主流模型：PCA-BPNN、PCA-PLS、PCA-LSSVM、SPA-BPNN、SPA-PLS 和 SPA-LSSVM。用这 6 个模型去预测实验样本数据，经比较发现 SPA-LSSVM 模型对于该组实验样本的预测效果最好，其预测相关系数 *R*p 和预测残差均方根 *RMSEP* 分别为 0.9253 和 0.1190。因此对于成年橡胶树氮素含量的光谱快速检测，SPA-LSSVM 算法模型的性能更为突出，有较好的应用潜力。

关键词：成年橡胶树叶片；氮素含量；SPA；PCA；PLS；BPNN；LS-SVM
中图分类号：S794.1　　**文献标识码：**A　　**文章编号：**1672–352X (2017)03–0429–05

Detection of nitrogen in mature rubber tree leaves using near infrared spectroscopy

JIANG Canchen, TANG Rongnian
(Mechanical and Electrical Engineering College, Hainan University, Haikou 570228)

Abstract: For a quick non-destructive detection of nitrogen content in the mature rubber tree leaf, spectral data of the leaf were obtained using the near infrared spectroscopy technique in this paper. Two different methods—SPA (Successive projections algorithm) and PCA (Principal component analysis) were used to extract the effective wavelength of the spectral data and the principal component after a multiplicative scatter correction (MSC) of the data, respectively. Then, the extracted characteristic values of the spectral data were respectively input into a linear regression model—PLS (Partial least squares) and non-linear regression model—BPNN (BP neural network) and LSSVM (Least squares support vector machine) to develop six mainstream models—PCA-BPNN, PCA-PLS, PCA-LSSVM, SPA-BPNN, SPA-PLS and SPA-LSSVM. It was found that the experimental sample prediction effect using SPA - LSSVM model was the best for the group after comparison of the results with the six models, and the prediction correlation coefficient (Rp) and root mean square error (RMSEP) were 0.9253 and 0.1190, respectively. Therefore, in terms of rapid spectral detection of nitrogen content in the mature rubber tree leaf, the performance of SPA - LSSVM algorithm model is more excellent with a broad application potential.

Key words: mature rubber tree leaves; nitrogen content value; SPA; PCA; PLS; BPNN; LS-SVM

橡胶树是我国重要的天然橡胶生产源。在国防、工业、交通和医药卫生等诸多领域，橡胶的优势在于其绝缘性、弹性和可塑性强，抗拉伸，耐磨损，使用寿命长，并可以阻隔水和气体等，因此被广泛地应用。橡胶树是我国重要的经济作物，其叶片氮素含量是判断橡胶树营养状态的一个重要量化标，对于橡胶树的营养诊断和实施变量施肥等有着重要的指导意义[1]。

近红外光谱技术是一种不破坏检测材料的快速检测技术，目前已被广泛应用于农业、食品、石油工业和医药等行业[2-4]。对于农产品中一些化学成分的检测，孔汶汶等[5]针对黄瓜叶片的光谱检测进行

收稿日期：2016-12-20
基金项目：国家自然科学基金（31460318）和海南省重点科技专项（ZDXM2014079）共同资助。
作者简介：蒋灿辰，硕士研究生。E-mail：jiangcanchen@126.com
* **通信作者：**唐荣年，副教授。E-mail：rongnian.tang@gmail.com

了研究，采用 SPA 算法提取光谱有效波长，然后分别代入不同的回归模型；孙光明等[6]针对油菜叶片的光谱检测进行了研究，采用 LS-SVM 算法作为回归模型，均取得了较好的预测结果。

橡胶树的光谱检测已吸引了研究者的注意，相关的研究还需要继续深入。尽管对于橡胶树幼苗生长状况的近红外光谱检测技术已经实现[7]，但由于橡胶树生长周期较长、胁迫实验不易等原因，对成年橡胶树的光谱检测还有一些问题没有解决，难以直接指导橡胶树割胶期的田间管理。为此，本研究以田间成年橡胶树叶片光谱数据为样本，分析了成年橡胶树叶片氮素的敏感波段，进一步探讨了主流的特征提取和预测模型算法在成年橡胶树叶片氮素含量光谱检测中的应用，着重比较分析了多种交叉组合建模算法对橡胶树氮素含量的预测精度，以期得出一种针对成年橡胶树叶片氮素含量的快速光谱检测算法模型。

1 材料与方法

1.1 实验仪器设备

橡胶树叶片的反射率采用美国 ASD 公司的光谱仪 FieldSpec3 进行测量。光谱仪的参数波段值范围从 350～2 500 nm，其中 350～1 050 nm 光谱采样间隔为 1.4 nm，光谱分辨率为 3 nm；1 050～2 500 nm 光谱采样间隔为 2 nm，光谱分辨率为 10 nm。分析软件为 MATLAB 2014b。

1.2 实验样本及氮素含量测定

实验样本共 176 个，全部采集自中国热带农业科学院试验场热研 7-33-97 品种的三龄橡胶树。实验设置了 N_0（不施氮），N_1（500 kg·hm^{-2} 尿素），N_2（1 000 kg·hm^{-2} 尿素）和 N_3（1 500 kg·hm^{-2} 尿素）等 4 个施氮水平。植株行距为 1 m×1 m，每个处理单株重复 20 次。采集植株上长势良好的中间小叶后，在实验室先进行近红外光谱扫描，然后做理化分析。实验测定光源由 ASD 公司的植物探头提供，近红外光谱扫描操作过程如下：先进行漫反射参考板校正，然后用外接单叶夹固定叶片，测定叶片正面光谱反射率。测量时将叶片以叶脉为界，分别选取上中下共 6 个区域，每个区域扫描 3 次，平均 18 条光谱曲线作为该叶片样本的 R 值[8]。叶片氮含量化学分析采用半微量凯氏定氮方法[9]，具体操作如下：将已采集光谱的橡胶树叶片在 105～108℃中进行杀青，烘干去主脉后粉碎，称样大约 0.07 g，加入 1.5 mL 浓硫酸消煮，消煮液定容 25 mL，冷却摇匀，吸取 10 mL 待测液于凯氏定氮仪（瑞士步琪有限公司 B339）测定氮含量，每片叶片重复分析 2 次，取平均值作为该叶片的氮含量。

当取得每个样本的氮素含量以及光谱数据后，采用浓度排序法[10]来挑选建模集样本和预测集样本，并且使得建模集数量与预测集数量之比为 3:1，即建模集样本数为 132，预测集样本数为 44。其中建模集和预测集氮素含量的均值和标准差如表 1 所示。

1.3 光谱数据特征变量选取方法

近红外光谱通常由大量数据点构成，建模时样本数远少于波长点数，同时近红外光谱本身的有效信息量较弱，有些波段与所测样本的化学成分之间的相关关系较低，如果采用全波段数据进行建模分析，不仅增加了建模所需要的时间，而且难以达到期望的模型预测精度，因此挑选出有效波长点的数据不仅可以简化模型，还能提高其预测精度。本研究采用主成分分析和连续投影算法来实现数据降维。

表 1 建模集和预测集氮素含量的均值和标准差

Table 1 The mean and standard deviation of nitrogen content values of modeling set and prediction set

样本集 Sample set	样本数 Sample size	含量范围（%） Contant range	平均值 Mean value	标准差 Standard deviation
训练集 Training set	132	2.84～4.33	3.62	0.3106
预测集 Prediction set	44	2.95～4.35	3.64	0.3135
全部 Total	176	2.84～4.35	3.62	0.3106

主成分分析（PCA）是一种传统的数据降维方法[11]。其基本方法是将原来相关性较高的众多数据，重新组合成一组较少个数且互不相关的综合指标来表示原来的指标。同时在尽可能保留原有信息的基础上将高维空间中的样本数据映射到较低维的主成分空间中，从而使的数据简化，降低维数，提取重要信息。每一个主成分所提取的信息量可用其方差累积贡献率来度量，其方差累积贡献率越大，表示包含的信息越多。本研究将采用主成分分析法求得光谱数据的主成分。

连续投影算法（SPA）是一种常用的特征变量选择方法[12]，由于光谱数据一般存在数据量大、数据之间存在共线性和大量冗余的问题，SPA 算法可以对光谱数据进行有效波长点信息提取，充分寻找含有最小限度的冗余信息的变量组，确保变量之间的共线性关系达到最小。并且很大程度上减少建模所需要的变量个数，提高建模的效率。本研究将采用连续投影算法（SPA）提取光谱数据有效波长（EW）。

1.4　光谱回归模型

本研究采用的线性回归模型，偏最小二乘法（PLS），是一种常用的线性回归模型，目前在近红外光谱定量分析中应用相当广泛；而采用的非线性回归模型为 BP 神经网络和最小二乘支持向量机（LSSVM）。BP 神经网络是目前应用相当广泛的神经网络模型之一，其基本特点是一种按误差逆传播算法训练的多层前馈网络。最小二乘支持向量机（LSSVM）是在支持向量机（SVM）的基础上改进的一种回归模型，相比传统的支持向量机，它降低了计算的复杂性，提高了计算的效率，充分挖掘了光谱数据的有效信息，并且能够很好地解决小样本、非线性和高维数等实际问题，已经在光谱分析领域得到广泛的应用[12]。其中采用径向基函数（RBF）为核函数，该非线性函数能够简化训练样本过程中的计算量。LSSVM 模型通过交互验证的网格搜索（Grid search）法[6]来确定模型中的 2 个参数 *gam* 和 *sig*2。以上的回归模型均采用预测集的预测相关系数（*R*p）和预测残差均方根（*RMSEP*）作为模型评价标准，预测相关系数越接近 1，*RMSEP* 越小，表示模型的预测性能越好。

2　结果与分析

2.1　经多元散射校正的橡胶树叶片反射率光谱

由于 176 个橡胶树叶片样本两端有一定量的噪声影响，并且本研究是以近红外光谱段为研究对象，故选取其中波长为 780～2 400 nm 的部分研究。如图 1 所示，每个橡胶树叶片样本的光谱曲线都有较接近的趋势，但同一个样本在不同的波段上升和下降的趋势不同。在 780～1 300 nm 波段是橡胶树叶片高反射平台（红外高台阶），反射率都高于 80%且 1 450 nm 段和 2 200 nm 段各有 1 个小波峰，而 1 250～1 400 nm、1 800～1 900 nm 和 2 200～2 400 nm 有明显下降的趋势，其叶片反射率的不同，说明其内部化学成分等可能存在差异，可能是因为氮素含量影响了其反射率。

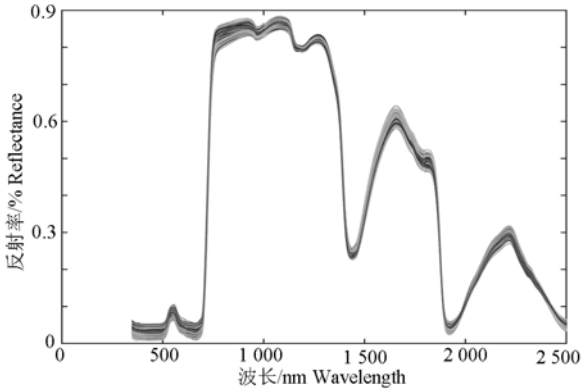

图 1　经过多元散射校正过的光谱图像

Figure 1　Spectrum images with multiplicative scatter correction

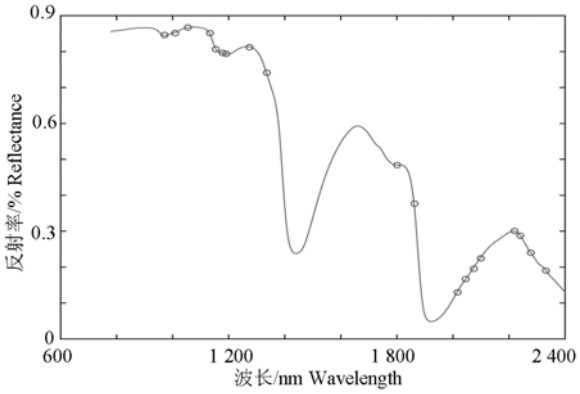

图 2　SPA 选取的有效波长点

Figure 2　Effective wavelength points extracted by SPA

2.2　几种模型预测精度的比较

对经过多元散射校正的光谱数据进行主成分分析，当主成分数为 7 时，其方差累计贡献率达到了 0.99 以上[11]。将这 7 个主成分作为输入变量分别输入到 BP 神经网络，LSSVM 以及 PLS 中，得到模型 PCA-BPNN、PCA-LSSVM 和 PCA-PLS[5-6]。模型 PCA-BPNN 中设定其隐含层节点数为 10，允许误差为 0.00001，最小训练速率为 0.1，最大迭代次数为 1 000，训练函数为“purelin”节点传递函数为“logsig”。模型 PCA-LSSVM 中 LSSVM 通过交互验证的网格搜索法分别得到最优 *gam*=988.86273，*sig*2=1 868.314。

对经过多元散射校正处理过的光谱数据进行投影计算选择光谱数据有效波长点，将提取的最大有效波长数设定为 60[12]，计算过程采用留一交互验证方法，求得有效波长个数为 19，且这 19 个有效波长点的分布如图 2 所示。将这 19 个有效波长点值作为输入变量分别输入到 BPNN、LSSVM 以及 PLS 中，得到模型 SPA-BPNN、SPA-LSSVM 和 SPA-PLS[13-14]。模型 SPA-BPNN 中建立 3 层网络，

将其隐含层节点数设定为 23，其余设置均与 PCA-BPNN 相同。SPA-LSSVM 中同样采用交互验证的网格搜索法得到最优 *gam*=1 295 343 485.6829，*sig*2=14 488 413.554424，然后分别将预测集数据输入以上 6 种模型中并且校验其精度。图 3 为以上 6 个模型的实际输出的拟合情况。

从表 2 中可以看出 SPA-LSSVM 模型的预测精度最高，其预测相关系数 *R*p 为 0.9253，预测残差均方根 *RMSEP* 为 0.1190。SPA-PLS 和 SPA-BPNN 模型的预测精度比较接近并且略低于 SPA-LSSVM 模型，而 PCA-BPNN，PCA-LSSVM 以及 PCA-PLS 模型的预测精度比起以上 4 个模型差距较大。综上所述，SPA-LSSVM 模型对于该组实验样本的预测精度最好。

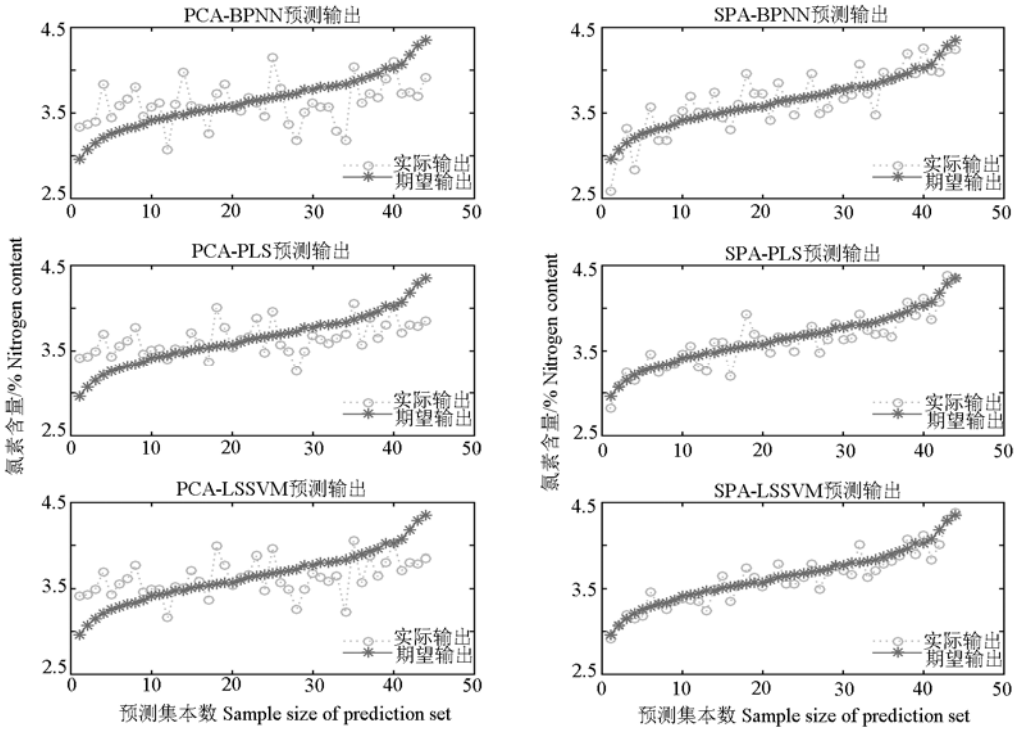

图 3 所有模型拟合结果的比较

Figure 3 Comparison of data fitting results using the six models

表 2 6 个主流模型预测精度的比较

Table 2 Comparison of predict accuracy using the six mainstream models

光谱模型 Spectrum model	主成分数/有效波长数 Principal component number/ Effective wavelength number	模型预测相关系数 *R*p Prediction correlation coefficient	模型预测残差均方根 *RMSEP* Root mean square error
PCA-BPNN	7	0.4218	0.3002
PCA-LSSVM	7	0.4311	0.2871
PCA-PLS	7	0.5011	0.2718
SPA-BPNN	19	0.8365	0.1866
SPA-LSSVM	19	0.9253	0.1190
SPA-PLS	19	0.9075	0.1392

3 讨论与结论

虽然主成分分析法（PCA）是依靠数据降维来提取特征，并且 7 个主成分的方差累计贡献率达到了 0.995，但是由于实验样本未经氮胁迫处理，光谱数据较为分散，7 个主成分可能不足以代表该组数据的大部分特征，因此使得模型的预测精度明显较低。

模型 SPA-PLS 和 SPA-LSSVM 的预测相关系数 *R*p 和预测残差均方根 *RMSEP* 十分接近，表明对于该组实验数据中使用 SPA 算法提取的有效波长点数据代入到线性和非线性回归模型中的预测精度结果较好且比较接近。

在模型 SPA-BPNN 和 SPA-LSSVM、PCA-BPNN

和 PCA-LSSVM 中，LSSVM 的预测精度都要明显优于 BP 神经网络，并且 BP 神经网络在某些波长点处的预测值的误差相对较大，这表明同为非线性回归模型的 BP 神经网络对于该组数据的预测能力要低于 LSSVM，同时也体现出最小二乘支持向量机有较好的预测精度和稳定性。

本研究利用所采集的成年橡胶树叶片光谱数据，分析了成年橡胶树叶片氮素的敏感波段，对现有文献中较为主流的特征提取和预测模型算法在成年橡胶树叶片氮素含量光谱快速检测中的性能进行了进一步的探讨，着重比较分析了以下 6 个模型：PCA-BPNN、PCA-PLS、PCA-LSSVM、SPA-BPNN、SPA-PLS 和 SPA-LSSVM。实验数据表明，SPA-LSSVM 模型的预测精度最好，其预测相关系数 R_p =0.9253，预测残差均方根 *RMSEP*=0.1190。因此，SPA-LSSVM 算法模型能够更好的满足橡胶树氮素营养的光谱快速诊断的应用需要。

参考文献：

[1] 陈贻钊, 林清火, 罗微, 等. 橡胶树叶片氮素光谱模型的研究[J]. 现代科学仪器, 2010 (2): 126-129.

[2] 刘贤, 韩鲁佳, 杨增玲, 等. 近红外光谱快速分析青贮饲料 pH 值和发酵产物[J]. 分析化学, 2007, 35(9): 1285-1289.

[3] 肖彦春, 窦森. 土壤腐殖质各组分红外光谱研究[J]. 分析化学, 2007, 35(11): 1596-1600.

[4] KULMYRZAEV A A, DUFOUR E.Relations between spectral and physicochemical properties of cheese, milk, and whey examined using multidimensional analysis[J]. Food Bioprocess Tech, 2010, 3(2): 247-256.

[5] 孔汶汶, 刘飞, 邹强, 等. 基于近红外光谱技术的油菜叶片丙二醛含量快速检测方法研究[J]. 光谱学与光谱分析, 2011, 31(4): 988-991.

[6] 孙光明, 刘飞, 张帆, 等. 基于近红外光谱技术检测除草剂胁迫下油菜叶片中脯氨酸含量的方法[J]. 光学学报, 2010, 30(4): 1192-1196.

[7] 吴敏. 巴西橡胶树幼苗对低钾胁迫的生理响应及差异表达基因分析[D]. 海口: 海南大学, 2011: 23-45.

[8] 陈贻钊. 基于高光谱信息的橡胶树叶片氮素营养快速检测技术研究[D]. 海口: 海南大学, 2010: 7-11.

[9] 鲍士旦. 土壤农化分析[M]. 北京: 中国农业出版社, 2000: 42-49.

[10] 吴静珠, 王一鸣, 张小超, 等. 近红外光谱分析中定标集样品挑选方法研究[J]. 农业机械学报, 2006, 37(4): 80-82.

[11] 冯雷, 陈双双, 冯斌, 等. 基于光谱技术的大豆豆荚炭疽病早期鉴别方法[J]. 农业工程学报, 2012, 28(1): 139-144.

[12] 张晓羽, 李庆波, 张广军. 基于稳定竞争自适应重加权采样的光谱分析无标模型传递方法[J]. 光谱学与光谱分析, 2014, 34(5): 1429-1433.

[13] 刘飞, 张帆, 方慧, 等. 连续投影算法在油菜叶片氨基酸总量无损检测中的应用[J]. 光谱学与光谱分析, 2009, 29(11): 3079-3083.

[14] 刘飞,王莉,何勇,等.基于可见/近红外光谱技术的黄瓜叶片 SPAD 值检测[J].红外与毫米波学报, 2009, 28(4): 272-276.