

基于高光谱的大叶女贞叶片水分定量测定

李阳阳^a, 孙雨安^{a,b}, 王国庆^a, 李振兴^a, 于文浩^a

(郑州轻工业学院 a.材料与化学工程学院;b.环境污染治理与生态修复河南省协同创新中心,郑州 450001)

摘要:较传统水分化学检测法相比,高光谱是一种新兴的简单、快速、无损、原位的绿色检测方法.大叶女贞是一种具有经济、医药、欣赏价值的树种,研究价值较高,然而有关高光谱对其的研究鲜有报道,将其作为研究对象,利用可见近红外高光谱采集40个叶片样本的高光谱数据,利用ENVI 4.7从中提取出反射光谱数据,在MATLAB 7.0平台对光谱数据进行直接正交信号校正(DOSC)预处理后,采用共生矩阵法(SPXY)选取30个样本作为建模集,其余作为预测集,构建支持向量回归(SVR)模型,然后在此全波段建模基础上采用自适应权重采样法(CARS)、连续投影变换(SPA)等波段筛选法进行特征波段建模分析,结果表明DOSC-SPA-SVR模型效果最优,预测平方相关系数 $R_p^2=0.9743$,预测均方根误差(RMSEP)为 4.4×10^{-5} ,较全波段模型相比,精度相当甚至有所提升,模型简单、稳定,能够用于对女贞叶片水分进行快速无损的定量分析,为其他作物营养诊断和精准灌溉提供一定的参考价值,减少水资源的浪费.

关键词:高光谱;特征波长;支持向量回归;连续投影变换

中图分类号:O657.3;S123

文献标志码:A

水乃生命之源,在植物的生长过程有着不可替代的作用,是植物营养胁迫、光合能力、生长状况的指示灯.传统的水分测定方法如烘干称重法、蒸馏法、卡尔费休法等过程繁琐、费时费力、时滞性差、需要破坏样品等缺点使其无法对作物进行大范围、全面、快速的营养诊断和指导,在精细农业的生产管理上显得尤为明显^[1].在面临当前水资源短缺和严峻的环境形势下,定量测定水分对精细灌溉、节约水资源、指导作物正常生长发育具有重要意义.

高光谱是一种集光谱和成像技术于一身的无损检测方法,具有广阔的发展前景,尽管国内外利用高光谱对叶片水分测定已有许多研究^[2-4],但利用高光谱对女贞叶片进行水分测定的研究鲜有报道.高光谱数据量大、波段维数高、波段冗余信息多的特点,会使构建的全波段模型不稳定、容易产生过拟合的现象,提取特征波段替代全波段已成为当下研究的热门方向之一,已有研究表明去除冗余信息后建立的模型更加稳定、简单、准确^[5-6].此外,高光谱数据并没有通用的处理方法,有鉴于此,本文利用高光谱在全波段光谱建模的基础上,为减少数据处理量,加快模型运行速度,采用CARS,SPA等特征波长提取方法进行建模尝试,以期找到一种简单、快速、准确的大叶女贞叶片水分无损测定方法.

1 实验部分

1.1 样本采集

在野外不同的大叶女贞树东、西、南、北、中方向均匀采样40个,立即编号放入保鲜袋中.

1.2 仪器设备和软件

实验仪器采用北京安洲科技有限公司(AZUP)的可见近红外实时成像光谱仪UHD285,光谱范围从450~950 nm,采样间隔4 nm,共125波段;CCD相机(像素为1 000×1 000);两个150 W卤钨灯线光源.

收稿日期:2017-01-18;修回日期:2017-05-20.

基金项目:国家自然科学基金(21171150);河南省科技攻关项目(102102310071).

作者简介:李阳阳(1989-),男,河南洛阳人,郑州轻工业学院硕士研究生,研究方向为分析测试技术.

通信作者:孙雨安(1964-),男,河南郑州人,郑州轻工业学院教授,研究方向为光谱色谱分析,E-mail:sya371@163.com.

实验数据处理基于 ENVI 4.7 (Research System Inc, USA)、MATLAB 7.0(The MathWorks Inc, USA)软件平台.

1.3 光谱测定

有关高光谱具体操作可参考文献[7]中方法,经多次条件尝试,最终确定的光谱采集条件是:曝光时间 $t=2.55$ ms,样本与镜头垂直距离为 25 cm,每个样本重复测量 10 次取平均值.

1.4 水分理化测定

样本光谱测定后,立即将其放在电子天平上称湿质量,然后马上放入 $80\text{ }^{\circ}\text{C}$ 的烘箱进行恒温 12 h 处理,再称取干质量,计算其含水质量.

1.5 模型评价指标

本文建模效果采用建模平方相关系数(R_p^2),预测平方相关系数(R_p^2),建模均方根误差(RMSEC)和预测均方根误差(RMSEP)等指标进行评价, R_p^2 、 R_p^2 越大,RMSEC、RMSEP 越小,模型效果越好.

2 结果与讨论

2.1 光谱数据处理

利用 ENVI 4.7 软件从高光谱数据中提取反射光谱数据后,为减少背景和其他干扰对模型影响,需对原始反射光谱预处理.图 1 是原始反射光谱图,图 2 是直接正交信号校正处理后的光谱图.可以看出,经直接正交信号校正预处理后,光谱形状总体相似,排列更为整齐,曲线更为平滑,增强了光谱的反射特征,说明 DOSC 在有效去除部分噪音的同时,也能保留光谱中主要信息.

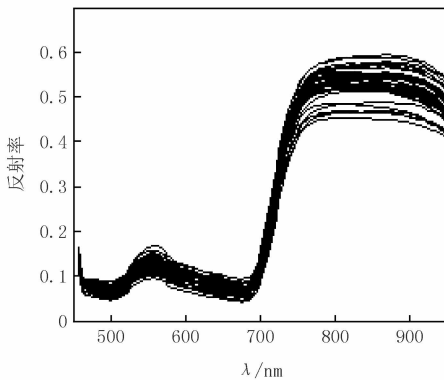


图1 原始反射光谱

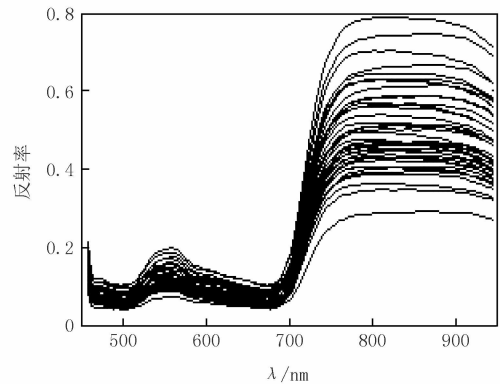


图2 直接信号校正光谱

2.2 模式识别

支持向量机(Support Vector Machine, SVM)在解决小样本、非线性及高维模式识别表现出许多特有的优势^[8-9],可避免神经网络容易陷入局部最优的问题,已在模式识别和回归分析中得到广泛应用.本文采用 Libsvm 工具箱,核函数选择“RBF”,惩罚因子 c 和核参数 g 采用网格搜索算法寻优,用法可参考文献[10].

2.3 全波段建模分析

采用 SPXY^[11]方法按 3 : 1 的比例选取 30 个样本作为建模样本,其余的作为预测样本,表 1 为样本数据分布表,可以看出预测集均匀分布在建模集中,具有一定的代表性.全波段 SVR 模型建模和预测结果如表 2 所示,从表 2 可以看出,采用 DOSC 能够取得最优的建模效果,与原始光谱相比,其能够显著提高 R_p^2 ,有效地降低 RMSEC 和 RMSEP,表明直接信号校正能够有效去除与样本水分含量无关或正交的信息,大大增强反射光谱和样本水分含量的线性相关性,削弱无关信息对模型的影响,使模型更加稳定、准确.

2.4 特征波段建模分析

SPA 是一种有效的特征波长提取方法,通过选取一组最大线性无关、最低冗余信息的变量集来实现降低数据维数和简化模型的目的.高光谱数据特点,给数据处理和仪器信息间的传输带来了困难,易产生

Hughes 效应^[12].为减少建模变量数目,降低波段间的重复信息,加快模型运行速度,本文在 DOSC 方法建模的基础上,采用 SPA,CARS 等特征提取方法筛选出的特征变量进行建模分析^[13-15],有关建模和预测效果见表 3,不难看出,几种特征波段所建模型都能够达到较好的建模和预测效果,但是从模型的简单、稳定和精度方面综合考虑,SPA 对特征变量的筛选效果最佳,筛选的光谱变量个数仅为原始光谱数据的 2.4%.

表 1 样本数据分布

样品	样本数	质量最小值/g	质量最大值/g	质量平均值/g	质量标准差
建模集	30	0.173 9	0.470 0	0.306 8	0.073 7
预测集	10	0.242 2	0.373 2	0.297 0	0.043 9

表 2 全波段 SVR 建模和预测结果

预处理方法	建模平方相关系数(R^2)	建模均方根误差(RMSEC)/%	预测平方相关系数(R_p^2)	预测均方根误差(RMSEP)/%
原始光谱	0.990 5	0.010 0	0.268 2	0.120 0
信号校正光谱	0.983 4	0.009 0	0.981 8	0.003 2

表 3 特征波段 SVR 建模和预测结果

变量提取法	变量数目	建模平方相关系数(R^2)	建模均方根误差(RMSEC)/%	预测平方相关系数(R_p^2)	预测均方根误差(RMSEP)/%
无关变量消除法 UVE	32	0.984 5	0.008 2	0.980 0	0.004 4
自适应重加权采样法 CARS	2	0.982 5	0.000 9	0.917 9	0.001 4
连续投影 SPA	3	0.987 3	0.008 5	0.974 3	0.004 4
遗传算法 GA	46	0.983 7	0.008 7	0.978 1	0.006 2

图 3 为进入到模型中的变量数目对均方根误差(RMSE)的变化趋势图.可以看到随着进入模型变量数目的增加,RMSE 不断下降,当特征变量个数为 3 时,RMSE 变化较小,趋于稳定,与之对应的变量数目就是特征变量个数.

图 4 为 SPA 筛选的特征变量在整个变量中的分布图,曲线是经直接信号校正后的反射光谱曲线,方框代表 SPA 筛选出的特征变量即 538 nm、698 nm、806 nm 3 处波段,这与张光伟有关脐橙叶片水分特征波长提取结果基本一致^[16],可以看出这些特征波段分布在反射曲线的极大值极小值附近,从侧面也反映出 SPA 能够有效地提取与叶片水分含量有关的波长信息.

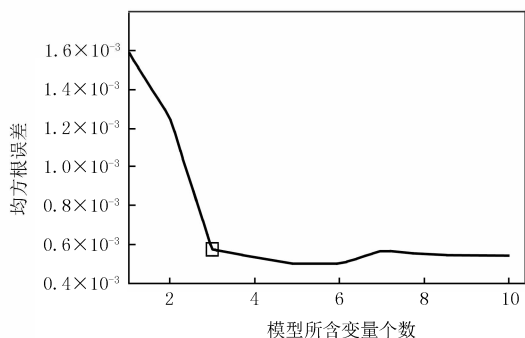


图 3 SPA 变量选择图

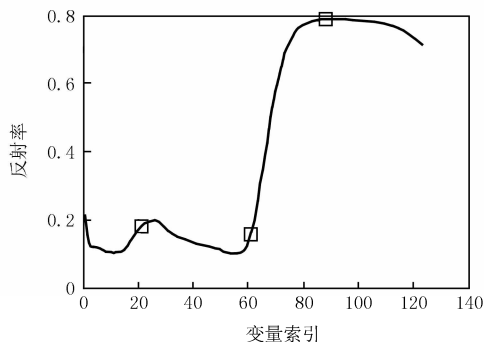


图 4 SPA 变量分布图

图 5 为采用网格搜索方法对参数选择结果 3D 视图,可以看到在选取一个最优参数后,必须要有与之相适应的另一个参数时,才能取得最低的交互验证均方根误差(RMSECV),RMSECV 越低,模型预测效果越好,本文采用 10 折交互检验选取最佳的参数 c 和 g ,参数 c 和 g 搜索范围设置为 $2^{-10} \sim 2^{10}$,步长为 1.图 6 中的线表示不同值的等高线,线中的数值代表选取不同的参数对时模型的交互验证均方根误差.实验最终选取的最优参数为 $c=84, g=16$,此时 RMSECV 达到最低 1.5×10^{-4} .

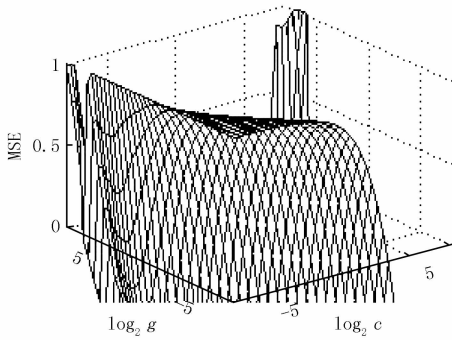


图5 SPA 参数选择三维图

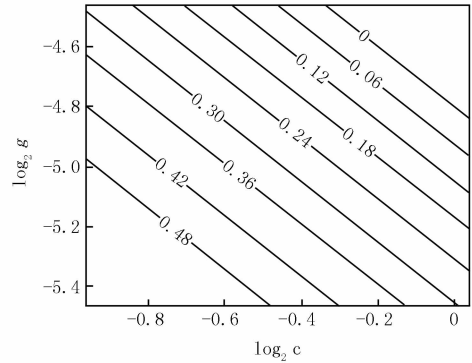


图6 SPA 参数选择二维图

2.5 水分含量建模预测图

图7是采用 DOSC-SVR 模型的建模和预测效果图,图8是采用 SPA 筛选的3个特征变量建立的 DOSC-SPA-SVR 模型效果图.蓝色和红色实线分别表示用于建模集和预测集的叶片水分含量实际值和预测值线性拟合图.从模型指标来看,后者较前者相比,虽然预测平方相关系数略微降低,但是建模平方相关系数有所提升,建模和预测均方根误差基本相当,甚至有所提升;从模型简单性来看,后者将变量个数压缩到3个,大大降低了高光谱数据的维度和数据处理量,加快了模型运行速度,模型更为稳定、简单;从预测准确性来看,后者预测的准确性并没有降低,基本相当;可以看出,SPA 在变量筛选的过程,能够有效滤除大量共线、冗余、无关的变量信息,DOSC-SPA-SVR 能够用于大叶女贞叶片水分含量的快速无损检测.

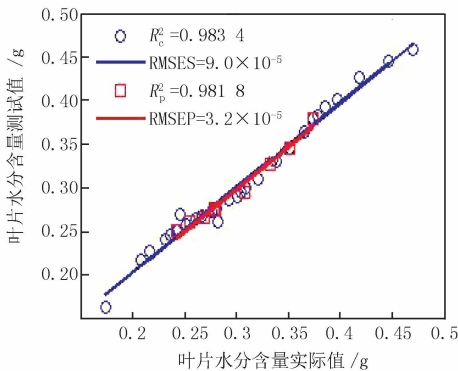


图7 全波段建模预测

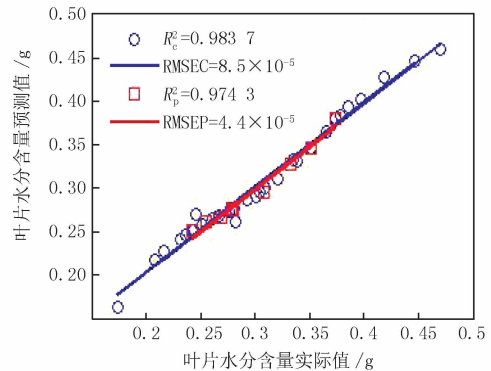


图8 特征变量建模预测

3 结 论

本研究尝试利用可见近红外高光谱技术结合化学计量学方法来快速测定女贞叶片水分含量.结果表明,全波段 DOSC-SVR 模型尽管能够对女贞叶片的水分含量进行预测,但是模型较为复杂;DOSC-SPA-SVR 模型仅仅利用全波段 2.4% 的信息就能够达到与之相当的精度,大大简化了模型,预测平方相关系数为 0.974 3,预测均方根误差为 4.4×10^{-5} ,以上说明,DOSC-SPA-SVR 模型能够进行大叶女贞叶片水分含量的快速测定,其一方面可用于指导女贞树的长势监测和施水管理,节约水资源,为其他经济作物的生长检测、产量评估提供一定的参考和借鉴价值,另一方面可为高光谱专用仪器的研发提供一定的理论依据,这也是迈向现代节水型农业发展的必由之路.

参 考 文 献

- [1] 刘燕德,姜小刚.基于高光谱成像技术对脐橙叶片的叶绿素、水分和氮素定量分析[J].中国农机化学报,2016,37(3):218-224.
- [2] 朱逢乐,何勇,邵咏妮,等.应用近红外高光谱成像预测三文鱼肉的水分含量[J].光谱学与光谱分析,2011,35(1):113-117.

- [3] 李丹,何建国,刘贵珊,等.基于高光谱成像技术的小黄瓜水分无损检测[J].红外与激光工程,2014,43(7):2393-2397.
- [4] 方美红.基于叶片光学属性的作物叶片水分含量反演模型研究[J].光谱学与光谱分析,2015,35(1):167-171.
- [5] 孙俊,武小红,张晓东,等.基于高光谱图像的生菜叶水分预测研究[J].光谱学与光谱分析,2013,33(2):522-526.
- [6] 梁琨,杜莹莹,卢伟,等.基于高光谱成像技术的小麦籽粒赤霉病识别[J].农业机械学报,2016,47(2):309-315.
- [7] 丁希斌,刘飞,张初,等.基于高光谱成像技术的油菜叶片 SPAD 值检测[J].光谱学与光谱分析,2015,35(2):486-491.
- [8] 张学工.关于统计学习理论与支持向量机[J].自动化学报,2000,26(1):32-41.
- [9] 董进义,孟晓玲,王国庆.连续小波变换-支持向量机-紫外分光光度测定多组分 B 组维生素含量[J].河南师范大学学报(自然科学版),2006,34(1):82-86.
- [10] 高媛媛,刘强国.基于 LIBSVM 的葡萄酒品质评判模型[J].四川理工学院学报(自然科学版),2010,23(2):530-532.
- [11] 展晓日,朱向荣,史新元,等.SPXY 样本划分法及蒙特卡罗交叉验证结合近红外光谱用于橘叶中橙皮苷的含量测定[J].光谱学与光谱分析,2009,29(4):964-968.
- [12] 刘萍萍,林辉.高光谱数据的降维处理方法研究[J].中南林业科技大学学报,2011,31(11):34-38.
- [13] DENG B C, YUN Y H, LIANG Y Z. Model population analysis in chemometrics [J]. Chemometrics & Intelligent Laboratory System, 2015, 149: 166-176.
- [14] CAI W, Li Y, SHAO X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. Chemometrics and Intelligent Laboratory Systems, 2008, 90(2): 188-194.
- [15] LI H, LIANG Y, XU Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Analytica Chimica Acta, 2009, 648(1): 77-84.
- [16] 张光伟.赣南脐橙叶片生长信息光谱诊断方法[D].南昌:华东交通大学,2014.

Determination of Ligustrum Leaf Water Content Based on Hyperspectral

Li Yangyang^a, Sun Yuan^{a,b}, Wang Guoqing^a, Li Zhenxing^a, Yu Wenhao^a

(a.College of Material and Chemical Engineering; b.Ollaborative Innovation Center of Environmental Pollution Control and Ecological Restoration Hnan Province, Zhengzhou University of Light Industry, Zhengzhou 450001, China)

Abstract: Compared with traditional chemical detection methods for moisture, hyperspectral is a simple, rapid, nondestructive, new, green, and in-situ detection method. Ligustrum lucidum is a kind of tree with economy, medical and appreciation value, and it is of great value in this field, however, there are few reports about its research in hyperspectral. This paper choosed its leaves as the subject of study, 40 leaf Hyperspectral datas with visible near infrared spectrometer were collected and the reflectance spectral data were obtained by using ENVI 4.7. The direct signal correction preprocessing for hyperspectral datas were performed in the MATLAB 7.0. The co-occurrence matrix method (SPXY) was used to devide the sample set, 30 of those were chose as the modeling set and the other as prediction set, then establishing support vector regression (SVR) model, we compared and selected the optimal model. Adaptive weighted sampling method (CARS), successive projection algorithm (SPA), and the other characteristic wavelength were used to build model after screening of the whole band spectrum based on the optimal model. Results show that DOSC-SPA-SVR is optimum and the precision of model is equivalent to optimal full band model, squared correlation coefficient of prediction set is 0.974 3, root mean square error of prediction is 4.4×10^{-5} . This model is single and stable, its accuracy is equivalent and even improved compared with the whole band model. It can be used in the rapid and nondestructive quantitative analysis for ligustrum lucidum leaf moisture and can provide some reference value for other nutrition diagnosis and precision irrigation to reduce the waste of water resources.

Keywords: hyperspectral; characteristic wavelength; support vector regression; successive projection algorithm

[责任编辑 赵晓华]