

文章编号: 1000-7032(2018)07-1016-08

土壤碱解氮含量可见/近红外光谱预测模型优化

汪六三^{1*}, 鲁翠萍¹, 王儒敬¹, 黄 伟¹, 郭红燕¹,汪玉冰¹, 林志丹², 王 键¹, 蒋 庆¹, 宋良图¹

(1. 中国科学院 合肥智能机械研究所, 安徽 合肥 230031; 2. 合肥电子工程学院, 安徽 合肥 230037)

摘要: 可见/近红外光谱技术是土壤成分检测的有效工具。波长筛选对可见/近红外模型土壤属性的预测精度有重要影响。以宁夏吴忠地区75个水稻土样为研究对象,利用可见/近红外光谱技术采集土壤样品光谱,采用SPXY(Sample set partitioning based on joint X-Y distance)方法选取了校正集和预测集样本,比较了分别采用Savitzky Golay平滑(SG smoothing)、多元散射校正(Multiple scatter correction, MSC)、标准正态变量变换(Standard normal variate, SNV)3种预处理方法对光谱数据处理后建立土壤碱解氮偏最小二乘法模型和原始光谱数据建模的效果。在此基础上,分别采用遗传算法(Genetic algorithms, GA)、连续投影算法(Successive projections algorithm, SPA)、竞争性自适应重加权算法(Competitive adaptive reweighted Sampling, CARS)、随机蛙跳(Random frog, RF)进行波长筛选,最后应用偏最小二乘法建立基于不同波长筛选方法的土壤碱解氮含量预测模型。研究表明,由于仪器性能稳定,样品的颗粒度比较小和均匀,本次实验原始光谱数据建模效果最好;各种波长筛选方法均可有效减少参与建模的波长数,且连续投影算法优于全谱建模,所选波长数仅为全谱波长数的1%,其预测决定系数(R^2)、预测均方根误差和相对分析误差值分别为0.726, 3.616, 1.906。这表明连续投影算法可以有效筛选水稻土碱解氮敏感波段,为土壤碱解氮传感器开发提供技术支持。

关键词: 土壤; 可见/近红外光谱; 碱解氮; 光谱预处理; 波长筛选

中图分类号: O235 文献标识码: A DOI: 10.3788/fjxb20183907.1016

Optimization for Vis/NIRS Prediction Model of Soil Available Nitrogen Content

WANG Liu-san^{1*}, LU Cui-ping¹, WANG Ru-jing¹, HUANG Wei¹, GUO Hong-yan¹,WANG Yu-bing¹, LIN Zhi-dan², WANG Jian¹, JIANG Qing¹, SONG Liang-tu¹

(1. Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China;

2. Electronic Engineering Institute, Hefei 230037, China)

* Corresponding Author, E-mail: lswang@iim.ac.cn

Abstract: Visible/near infrared spectroscopy (Vis/NIRS) is an effective tool for soil component detection. Wavelength selection plays an important role in predicting the soil properties of the visible/near infrared (Vis/NIR) model. Taking the 75 paddy soil samples of Ningxia Wuzhong area as the research object, the Vis/NIR spectra of soil samples were collected. The sample set partitioning based on joint X-Y distance (SPXY) method was used to divide calibration set and prediction set samples. Three different preprocessing methods, namely, Savitzky Golay smoothing (SG smoothing), multiplicative scatter correction (MSC), and standard normal variate (SNV), were applied to

收稿日期: 2017-11-15; 修订日期: 2017-12-24

基金项目: 宁夏回族自治区科技厅项目(XD-XMBW-2016001); 上海市农业委员会项目(沪农科推字(2016)第2-5-11号)资助
Supported by Project of Science and Technology Hall of Ningxia Hui Autonomous Region(XD-XMBW-2016001); Shanghai Agricultural Committee Project(Shanghai agriculture promotion word (2016) No.2-5-11)

pretreat the spectral data. Then, three soil available nitrogen content prediction models were established by partial least squares (PLS) regression. The consequences of the three models and the model using raw spectral data were compared. On this basis, the genetic algorithm (GA), successive projections algorithm (SPA), competitive adaptive reweighted algorithm (CARS) and Random Frog (RF) were used to select key wavelength variables. Finally, the soil available nitrogen content prediction models were established by partial least squares regression based on different wavelength selection methods. Research shows that, due to the stability of the instrument, relatively small and uniform of the sample particle size, raw spectral data is achieved the best consequence; various wavelength selection methods can effectively reduce the number of wavelengths in the model, and the continuous projection algorithm is better than full spectrum model, the selected wavelength number is only 1% of the number of full spectrum wavelength and its prediction coefficient of determination (R^2), root mean square prediction error and relative error analysis values were 0.726, 3.616 and 1.906. The research shows that wavelength selection method by SPA could predict the available nitrogen content in the paddy soil, provide a reference for development of soil available nitrogen sensor.

Key words: soil; visible/near infrared spectroscopy (Vis-NIRS); available nitrogen; spectral pretreatment; wavelength selection

1 引 言

氮素是作物最重要的营养元素,对绝大部分作物来说,生长所需的氮素 90% 以上来源于土壤^[1]。分析测定土壤中各种形态的氮含量,对了解土壤氮的供给水平和指导施肥具有重要意义;在精准农业生产中,土壤不同形态氮素水平的测定是精准施肥或平衡施肥的重要依据。土壤碱解氮是土壤氮素的重要指标之一,碱解氮含量的高低能反映出近期土壤氮素丰缺。

传统土壤碱解氮的测定是基于湿化学法^[2],操作复杂、测量时间长、成本高,难以满足快速检测土壤碱解氮含量的需求。可见/近红外光谱技术具有快速、无损、无污染等特点,它可以在短时间内分析大量土壤样品,实现土壤参数的快速测量,在土壤养分检测中得到大量应用^[3-9]。通常可见/近红外光谱包括百千个波长变量,而这些变量中含有与待测属性无关的冗余噪声^[10]。因此,建模前有必要对波长变量进行选择。这样可以剔除不相关的变量,简化模型,更重要的是可以提高校正模型的预测能力。常用的变量选择方法包括:无信息变量消除法 (Uninformative variable elimination, UVE)^[11]、遗传算法 (Genetic algorithms, GA)^[12]、连续投影算法 (Successive projections algorithm, SPA)^[13]、竞争性自适应重加权算

法 (Competitive adaptive reweighted sampling, CARS)^[14]、随机蛙跳 (Random frog, RF) 等^[15]。刘燕德等^[16]以江西脐橙果园土壤为研究对象,通过遗传算法、连续投影算法、竞争性自适应加权算法筛选出有机质敏感波段,结果显示竞争性自适应加权算法更加有效。贾生尧等^[17]利用变量投影重要性 (Variable importance in the projection, VIP)、无信息变量消除法选择波长变量,并分别用递归偏最小二乘回归法 (Recursive partial least squares regression, RPLS)、偏最小二乘回归法 (Partial least squares regression, PLS) 建立土壤全氮和有机质含量预测模型,结果表明 VIP-RPLS 对土壤全氮和有机质含量具有更高的预测精度。高洪智等^[18]将连续投影算法与贡献值结合筛选土壤总氮的特征波长,所建模型的预测精度优于全谱偏最小二乘回归结果。林志丹等^[19]针对土壤有机质含量预测,使用连续投影算法和遗传算法进行波长优选,主成分回归构建预测模型。结果显示连续投影算法和遗传算法都能够有效减少参与建模的波长数且提高模型预测精度。

本文以宁夏吴忠地区 75 个水稻土样为研究对象,利用可见/近红外光谱技术采集土壤样品光谱,比较不同预处理方法和原始光谱对预测模型的效果,探讨 GA、SPA、CARS、RF 4 种波长选择算法对土壤碱解氮含量预测模型的影响,为土壤碱

解氮含量快速获取和仪器开发提供技术支持。

2 材料与方 法

2.1 样本采集、制备和划分

在宁夏吴忠市叶盛贡米种植基地 2 号地(954 亩)采集土壤样品 75 个。土样的采样范围为北纬 $38^{\circ}07'23.65''$ ~ 北纬 $38^{\circ}07'39.64''$,东经 $106^{\circ}11'27.66''$ ~ 东经 $106^{\circ}12'03.28''$;采集的土壤去除杂物后,让其自然风干。风干后的土壤样品先经过人工碾磨,然后过筛。根据分析项目的要求,过 60 目数的筛子。每个样品分两份,一份用于实验室化学分析,另一份用于光谱分析,用自封袋对土样进行密封保存,并统一编号。土壤样本共 75 个,使用 SPXY (Sample set partitioning based on joint x - y distance) 方法对样本按 4:1 进行划分,其中建模集 60 个,预测集 15 个样本,如表 1 所示。

表 1 土壤样品碱解氮测量值分布

Tab. 1 Measurement distribution of soil available nitrogen $\text{mg} \cdot \text{kg}^{-1}$

样品集分类	测量值分布	均值	标准偏差
校正集(60)	39.480 ~ 86.730	64.723	8.919
预测集(15)	55.300 ~ 81.830	66.871	6.892

2.2 土壤样本光谱采集

本实验采用美国 Veris 公司的可见/近红外光谱土壤检测系统。该系统将光源和漫反射收集集成于一个犁头内。犁头底部固定有蓝宝石窗片。这种窗片在近红外波段透过率大约为 90% 且足够耐用,能够承受与土壤的连续接触。系统使用卤素灯照射土壤,漫反射光通过光纤传导进入光谱仪。置于犁头内的光学快门每隔 5 min 自动获取暗光谱。光谱仪每秒大约采集 20 个光谱,同时通过 USB 接口传输给电脑保存。采集程序使用 Labview (National Instruments, Austin, TX, USA) 编写。光纤传导漫反射光进入两个分立的光谱仪。一个光谱仪 (Ocean optics, USB4000) 使用硅 CCD 作为探测器测量 342 ~ 1 050 nm 波段光强,光谱分辨率为 3 nm;另一个光谱仪 (Hamamatsu, C9914GB) 使用 InGaAs 阵列探测器测量 1 000 ~ 2 221.5 nm 波段光强,光谱分辨率为 8 nm。犁头安装于一个支架上,支架上安装有光谱仪、控制器等光谱采集与控制处理系统。整个系统包含两种测量模式——田间动态测量模式和实验室静态测

量模式。本研究中,我们将犁头从支架上拆卸下来,在实验室条件下采集土壤光谱。

为了使光源保持稳定,卤素灯光源至少预热 20 min 左右。将过筛的土壤样品装入样品杯,用样品杯盖将样品杯内的样品压实压平。样品杯贴近探头窗口,采集土壤样本光谱,转动样品杯,每个样品重复扫描 3 次,求其平均光谱曲线。

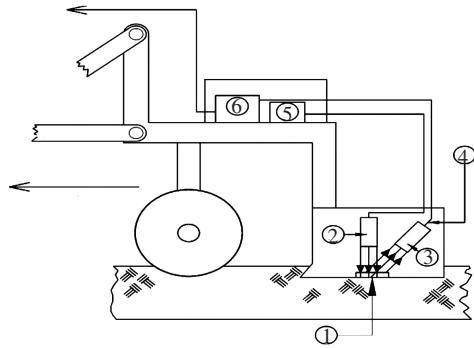


图 1 可见/近红外土壤属性测定仪结构示意图。其中:①蓝宝石窗;②卤素灯;③漫反射光收集探头;④光纤;⑤光谱仪;⑥电源。

Fig. 1 Schematic diagram of soil attribute measurement device based on Vis-NIR spectrophotometer. ① Sapphire window; ② halogen lamp; ③ collection optic; ④ fiber optic; ⑤ spectrometer; ⑥ power supply.

2.3 光谱预处理

样本表面不均引起的散射现象、装样量不一致引起的光程差、暗电流和仪器随机噪声引起的光谱曲线不重复现象和基线漂移现象,以及样本不同成分自检相互干扰引起的背景因素和多重共线性等无用信息对光谱曲线均有影响^[20]。为了达到较好的预测效果,通常需要对光谱数据进行预处理。本研究分别对土壤样本光谱数据进行了 Savitzky Golay 平滑 (SG smoothing)、多元散射校正 (Multiple scatter correction, MSC)、标准正态变量变换 (Standard normal variate, SNV) 等预处理,并对各种预处理效果和原始光谱数据效果进行了比较,具体见下文。光谱预处理使用 Uncrambler X 10. 4 (CAMO, Norway) 和 Matlab2012b (MathWorks, USA) 软件。

2.4 波长筛选

原始光谱数据量大,同时存在冗余噪声和大量共线信息,对光谱建模存在干扰,需要对波长进行筛选。本文对文中使用的遗传算法 (GA)、连续投影算法 (SPA)、竞争性自适应重加权算

(CARS)和随机蛙跳(RF)算法作一个简单介绍,具体的可以参见文献[12-15]。

遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种通过模拟生物进化随机寻优求解的常用算法。它利用选择、交换和突变等算子的操作,随着不断的遗传迭代,使目标函数值较优的变量被保留,较差的变量被淘汰,最终达到最优结果。

连续投影算法是一种减少变量共线性问题的前向循环选择方法。利用向量的投影分析,该方法从一个波长开始,每次循环都计算它在未选入的波长上的投影,将投影向量最大的波长引入到波长组合。每一个新选入的波长,都与前一个线性关系最小。

竞争性自适应重加权算法是一种基于回归系数进行波长选择的方法。该方法模拟达尔文进化理论的“适者生存”原则,将每个波长看作一个体,对波长进行逐步淘汰,每次采样过程中利用自适应重加权采样技术和指数衰减函数结合的方法优选出 PLS 模型中回归系数绝对值大的波长点,去除 PLS 模型中回归系数权重较小的波长,并计算变量子集 RMSECV 值,最后选择最小 RMSECV 子集作为最优变量子集。

随机蛙跳是一种迭代方式工作的波长选择方法。首先一个变量子集被随机选择,基于这个选择的变量子集按照一定的几率进行迭代,每迭代一次,变量子集更新一次。经过 N 次迭代后,每个变量选择的几率被计算出来。然后几率高的变量用来建立 PLS 模型, RMSECV 最小的 PLS 模型对应的变量为最优变量。

本研究中,GA 和 SPA 算法分别使用 GAPLS 工具箱,SPA 工具箱实现^[21];CARS 和 RF 算法使用 LibPLS 工具箱^[22]实现。GAPLS 工具箱要求初始变量小于 200。因此首先对光谱数据使用三点平均后波长数由 380 减为 127,再使用 GA 进行波长筛选。

2.5 模型验证与评价

模型分别采用完全交互验证和外部验证对其性能进行评价,由决定系数(R^2)、建模均方根误差(RMSECV)、预测均方根误差(RMSEP)和预测相对分析误差(RPD)来评价。在建模分析中, R^2 和 RPD 值高为好,同时 RMSECV 和 RMSEP 值低为好。当 RPD 值位于 1.5~2 之间表示所建立的模型对分析指标具有一定的预测能力,当位于 2~2.5 之间表示所建立的模型对分析指标定量预测是可行的,2.5~3 之间表示所建立的模型具有较好的预测精度^[3,23]。

3 结果与讨论

3.1 光谱数据预处理

75 份土壤样品的可见/近红外光谱吸收图谱如图 2 所示,在 1 400 nm 和 1 900 nm 附近有明显的水分子吸收峰。

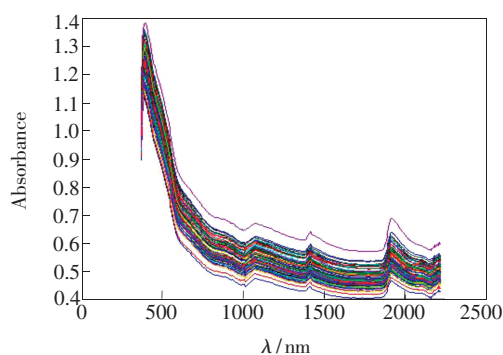


图 2 75 份土壤样品的可见/近红外光谱

Fig. 2 Visible and near infrared spectra of 75 soil samples

去除 340~366 nm 噪声大的波段,选取 366~2 221.5 nm 波段光谱数据作为后续数据处理。对土壤光谱数据分别采用 SG smoothing、SNV、MSC 等预处理方法进行光谱预处理,分别建立 PLS 模型。本文建立的所有预测模型由决定系数(R^2)、建模均方根误差(RMSECV)、均方根预测误差(RMSEP)和相对预测偏差(RPD)评价,其中 R^2 和 RPD 值越大, RMSECV、RMSEP 值越小,模型性能越好。

表 2 不同预处理方法的 PLS 建模结果

Tab. 2 Results of PLS models obtained with different spectral pretreatment methods

预处理方法	主因子数	建模集		预测集	
		R^2	RMSECV	R^2	RMSEP
Raw	4	0.684	4.974	0.678	3.810
SG smoothing(3 点)	4	0.674	5.052	0.684	3.853
SNV	3	0.645	5.233	0.615	4.296
MSC	3	0.650	5.234	0.615	4.293

表 2 给出了不同预处理方法 PLS 建模结果。本次实验发现采集的光谱不经过任何预处理效果最好,是由于光谱采集时仪器性能稳定,制备的样品过了 60 目筛,样品的颗粒度比较小和均匀。

3.2 波长筛选

图 3 显示了基于 GA 算法的波长筛选结果。GA 算法操作过程中依赖于随机数发生器,每次运行的解是不同的,多次运行可以获得广泛可能

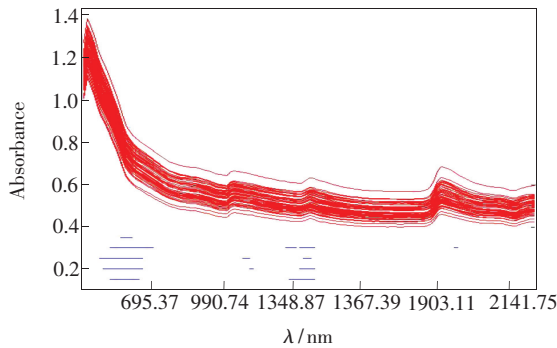


图 3 土壤光谱和 GA 变量筛选结果

Fig. 3 Plot of soil spectral and variable selection by GA

的解,对 GA 算法进行了 5 次运行。图中上半部分为建模集样品可见/近红外光谱,图中下半部分横线为 5 次运行的波长筛选结果。从图中可以看出 5 次运行筛选的波长主要集中在 400 ~ 700 nm、1 100 ~ 1 500 nm 波段。基于选择的波长建立 GA-PLS 土壤碱解氮定量预测模型,5 次预测结果的均值如表 3 所示。

图 4 表示使用 SPA 算法进行波长选择的结果。从图中可以看出选取了 4 个变量,对应波长为 535.85, 619.66, 1 091.49, 1 817.58 nm。基于这 4 个变量建立了 SPA-PLS 土壤碱解氮定量预测模型,预测结果如表 3 所示。虽然 SPA-PLS 模型参与建模的变量数仅获得 4 个变量,占原始变量的 1%, SPA-PLS 模型的性能适当, R^2 为 0.726, RPD 值为 1.906, 结果表明 SPA-PLS 模型对土壤碱解氮有一定的预测能力。

图 5 表示运行一次 CARS 算法,随着采样数的增加,采样变量数(图 5(a))、十折交叉验证 RMSECV 值(图 5(b))、各变量回归系数(图 5(c))

表 3 不同波长筛选方法的 PLS 建模结果

Tab. 3 Results of PLS modes obtained with different wavelength selection

特征波长提取方法	波长数	主成分数	R^2	RMSEP	RPD
全谱	380	7	0.678	3.81	1.809
GA	14(3 点平均后)	5	0.523	4.799	1.464
SPA	4	4	0.726	3.616	1.906
CARS	29	9	0.536	8.005	0.861
Random frog	10	7	0.622	4.425	1.557

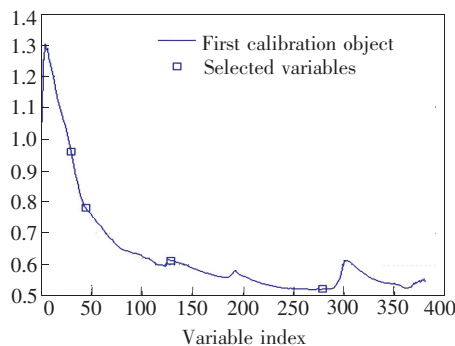


图 4 SPA 运行结果

Fig. 4 Results using SPA

的变化趋势。从图 5 中可以看出, CARS 变量选择过程分两个阶段进行。第一阶段由于指数衰减函数的作用,采样的变量数减少速度很快,称为粗选;第二阶段采样的变量数减少速度减慢,称为细

选。在 1 ~ 25 采样区间, RMSECV 值随着采样数的增加减少快速,表明消除了与碱解氮无关变量;采样超过 25 后,随着采样次数的继续增加, RMSECV 值又逐渐递增,表明光谱中与碱解氮相关的重要信息被剔除了。图中显示采样 25 时, RMSECV 值最小(图 5(c)中星号垂线标示)。因此,第 25 次蒙特卡罗采样获得的变量为预测土壤碱解氮含量的关键变量,共计 29 个特征波长,依次为 455.18, 484.26, 501.56, 530.17, 558.47, 597.58, 608.65, 679.35, 778.83, 789.03, 995.24, 999.74, 1 013.12, 1 075.28, 1 080.7, 1 091.49, 1 096.86, 1 323.96, 1 328.95, 1 769.18, 1 890.49, 1 898.91, 1 907.31, 1 977.34, 1 981.38, 1 985.42, 2 168.37, 2 198.73, 2 221.53 nm。基于 CARS 算法获得的 29 个特征波长建立 PLS 模型对土壤碱解氮进行定量预测,预测性能

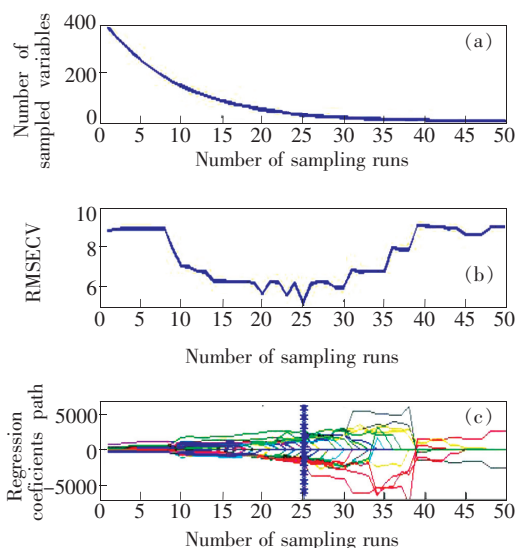


图 5 CARS 运行结果。(a) 筛选过程中采样变量数的变化趋势;(b) 10 折交叉验证所得残差 RMSECV;(c) 各变量回归系数随着采样次数增加的变化趋势。

Fig. 5 Results using CARS. (a) Changing trend of the number of sampled variables. (b) 10-fold RMSECV values. (c) Regression coefficients of each variables with the increasing of sampling runs.

如表 3 所示。

图 6 表示基于 Random frog 算法运行 10 000 次结果的均值, 由于 Random frog 算法每次运行的结果略有差异, 所以需要多次运行并且结果取平均值。越重要的变量, 被选择的概率越大。依据经验, 以 0.3 为阈值, 选择概率大于阈值的波长将作为特征波长。因此, 获得了 13 个超过点线(0.3)的波长(501.56, 789.03, 912, 949.56, 1 283.78,

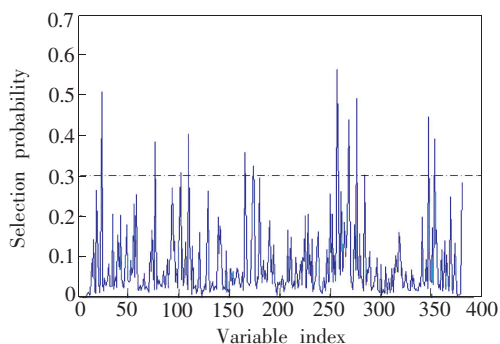


图 6 Random frog 选择波长的概率

Fig. 6 Selection probability of each wavelength by Random frog

1 323.96, 1 719.88, 1 769.18, 1 773.6, 1 839.26, 2 095.77, 2 118.83 nm) 作为特征波长, 作为 Random frog 算法运行 10 000 次后的结果。基于 13 个特征波长建立 RF-PLS 土壤碱解氮定量预测模型, 预测结果如表 3 所示。

从表 3 可知, 采用 SPA 提取的特征波长建立的 PLS 模型的效果最好, 预测集的决定系数为 0.726, 均方根预测误差 3.616, 相对分析误差为 1.906。采用 CARS 选择的特征波长建立的模型预测效果最差, 预测集的决定系数为 0.536, 均方根预测误差 8.005, 相对分析误差为 0.861。比较基于 GA、Random frog 算法提取的特征波长建立的 PLS 模型可知, 其预测能力均低于全谱 PLS 模型, 可能是 GA、Random frog 选取的特征波长中包含无用信息, 没有达到最优作用。依据模型评价标准, SPA、GA 和 Random frog 算法及全谱 PLS 模型均对土壤碱解氮具备一定的预测能力, 相对于全谱 PLS, SPA、GA 和 Random frog 算法简化了模型; CARS 算法模型未能体现其对土壤碱解氮的预测能力, 分析可能是由于 CARS 算法在波长筛选过程中剔除了重要波长变量。

4 结 论

利用可见/近红外光谱技术对土壤碱解氮含量进行了检测。用 SG、MSC、SNV 对原始光谱数据进行了预处理, 基于全光谱建立了土壤碱解氮含量 PLS 预测模型, 比较了不同预处理方法和原始光谱的预测效果。结果表明原始光谱数据建立的模型最优, 分析是由于光谱采集时仪器性能稳定, 样品的颗粒度比较小和均匀(过 60 目筛)。在原始光谱数据的基础上, 采用 GA 算法、SPA 算法、CARS 算法和 Random frog 算法提取特征波长, 并基于特征波长建立 PLS 预测模型。结果显示, 与全谱 PLS 模型相比, SPA 算法建立的预测模型效果最好, GA 算法、CARS 算法和 Random frog 算法建立的模型的预测效果均低于全谱 PLS 模型, 说明 SPA 算法对于土壤碱解氮是一种有效的波长筛选方法。本文的研究为土壤碱解氮含量的快速检测和仪器开发奠定了基础。

参 考 文 献:

- [1] 沙丽清, 孟盈, 冯志立, 等. 西双版纳不同热带森林土壤氮矿化和硝化作用研究 [J]. 植物生态学报, 2002, 24(2):152-156.
SHA L Q, MENG Y, FENG Z L, *et al.*. Nitrification and net N mineralization rate of soils under different tropical forests in Xi shuang ban na, Southwest China [J]. *Chin. J. Plant Ecol.*, 2002, 24(2):152-156. (in Chinese)
- [2] 鲍士旦. 土壤农化分析 [M]. 北京:中国农业出版社, 2000.
BAO S D. *Soil Agro-chemical Analysis* [M]. Beijing: China Agricultural Press, 2000. (in Chinese)
- [3] CHANG C W, LAIRD D A, MAUSBACH M J, *et al.*. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties [J]. *Soil Sci. Soc. Am. J.*, 2001, 65:480-490.
- [4] 鲍一丹, 何勇, 方慧, 等. 土壤的光谱特征及氮含量的预测研究 [J]. 光谱学与光谱分析, 2007, 27(1):62-65.
BAO Y D, HE Y, FANG H, *et al.*. Spectral characterization and N content prediction of soil with different particle size and moisture content [J]. *Spectrosc. Spect. Anal.*, 2007, 27(1):62-65. (in Chinese)
- [5] 安晓飞, 李民赞, 郑立华, 等. 土壤水分对近红外光谱实时检测土壤全氮的影响研究 [J]. 光谱学与光谱分析, 2013, 33(3):677-681.
AN X F, LI M Z, ZHENG L H, *et al.*. Effect of soil moisture on prediction of soil total nitrogen using NIR spectroscopy [J]. *Spectrosc. Spect. Anal.*, 2013, 33(3):677-681. (in Chinese)
- [6] 何东健, 陈煦, 任嘉琛, 等. 土壤速效磷含量近红外光谱田间快速测定方法 [J]. 农业机械学报, 2015, 6(3):152-157.
HE D J, CHEN X, REN J S, *et al.*. Detection of soil available P content based on near infrared spectra [J]. *Trans. Chin. Soc. Agric. Mach.*, 2015, 46(3):152-157. (in Chinese)
- [7] WETTERLIND J, STENBERG B, ROSSEL R A. Soil analysis using visible and near infrared spectroscopy [J]. *Methods Mol. Bio.*, 2013, 953:95-107.
- [8] NAWAR S, BUDDENBAUM H, HILL J, *et al.*. Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy [J]. *Soil Till. Res.*, 2016, 155:510-522.
- [9] 王儒敬, 陈天娇, 汪玉冰, 等. 基于深度稀疏学习的土壤近红外光谱分析预测模型 [J]. 发光学报, 2017, 38(1):109-116.
WANG R J, CHEN T J, WANG Y B, *et al.*. Soil near infrared spectroscopy prediction model based on deep sparse learning [J]. *Chin. J. Lumin.*, 2017, 38(1):109-116. (in Chinese)
- [10] FAN S, GUO Z, ZHANG B, *et al.*. Using Vis/NIR diffuse transmittance spectroscopy and multivariate analysis to predicate soluble solids content of apple [J]. *Food Anal. Methods*, 2016, 9(5):1333-1343.
- [11] CENTNER V, MASSART D L, NOORD O E D, *et al.*. Elimination of uninformative variables for multivariate calibration [J]. *Anal. Chem.*, 1996, 68(21):3851-3858.
- [12] JOUANRIMBAUD D, MASSART D L, LEARDI R, *et al.*. Genetic algorithms as a tool for wavelength selection in multivariate calibration [J]. *Anal. Chem.*, 1995, 67(23):4295-4301.
- [13] GALVÃO R K H, ARAÚJO M C U, FRAGOSO W D, *et al.*. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm [J]. *Chemom. Intell. Lab. Syst.*, 2008, 92(1):83-91.
- [14] LI H, LIANG Y, XU Q, *et al.*. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. *Anal. Chim. Acta*, 2009, 648(1):77.
- [15] LI H D, XU Q S, LIANG Y Z. Random frog: an efficient reversible jump Markov chain Monte carlo-like approach for variable selection with applications to gene selection and disease classification [J]. *Anal. Chim. Acta*, 2012, 740:20-26.
- [16] 刘燕德, 熊松盛, 刘德力, 等. 基于 Vis-NIR 光谱的果园土壤有机质便携式检测仪研究 [J]. 中国农机化学学报, 2016, 37(2):240-244.
LIU Y D, JIANG X G, XIONG S S, *et al.*. Using CARS to predict the content of soil organic matter based on Vis-NIR spectra [J]. *J. Chin. Agricult. Mechan.*, 2016, 37(2):240-244. (in Chinese)
- [17] JIA S, LI H, WANG Y, *et al.*. Recursive variable selection to update near-infrared spectroscopy model for the determination of soil nitrogen and organic carbon [J]. *Geoderma*, 2016, 268:92-99.

- [18] 高洪智, 卢名鹏, 丁海泉, 等. 基于连续投影算法的土壤总氮近红外特征波长的选取 [J]. 光谱学与光谱分析, 2009, 29(11):2951-2954.
GAO H Z, LU Q P, DING H Q, *et al.*. Choice of characteristic near infrared wavelengths for soil total nitrogen based on successive projection algorithm [J]. *Spectrosc. Spect. Anal.*, 2009, 29(11):2951-2954. (in Chinese)
- [19] 林志丹, 汪玉冰, 王儒敬, 等. 波长优选对土壤有机质含量可见光/近红外光谱模型的优化 [J]. 发光学报, 2016, 37(11):1428-1435.
LIN Z D, WANG Y B, WANG R J, *et al.*. Improvements of Vis-NIRS model in the prediction of soil organic matter content using wavelength optimization [J]. *Chin. J. Lumin.*, 2016, 37(11):1428-1435. (in Chinese)
- [20] ÅSMUND RINNAN, BERG F V D, ENGELSEN S B. Review of the most common pre-processing techniques for near-infrared spectra [J]. *Trac. Trends Anal. Chem.*, 2009, 28(10):1201-1222.
- [21] PAIVA H M. A graphical user interface for variable selection employing the successive projections algorithm [J]. *Chemom. Intell. Lab Syst.*, 2012, 118(9):260-266.
- [22] LI H, XU Q, LIANG Y. LIBPLS: an integrated library for partial least squares regression and discriminant analysis [J]. *Chemom. Intell. Lab Syst.*, 2018, 176:34-43.
- [23] SAEYS W, MOUAZEN A M, RAMON H. Potential for onsite and online analysis of pig manure using visible and near infrared spectroscopy [J]. *Biosyst. Eng.*, 2005, 91:393-402.



汪六三(1982-),男,安徽宿松人,博士,副研究员,2011年于中国科学院合肥物质科学研究院获得博士学位,主要从事光谱技术及应用方面的研究。

E-mail: lswang@iim.ac.cn