

基于CARS-PLSR算法的土壤有效磷高光谱反演研究*

郭鹏^{1,2}, 赵阳^{1,2}, 孙子皓^{1,2}, 陈秀万^{1,2*}

(1. 北京大学遥感与地理信息系统研究所, 北京 100871; 2. 北京大学地球观测与导航教育部工程研究中心 (CEON), 北京 100871)

摘要:【目的】剔除土壤高光谱中包含的大量冗余和无效信息, 探明土壤有效磷 (SAP) 的敏感波段, 简化SAP的高光谱估算模型并提高模型的预测精度。【方法】文章以四川省崇州市西河流域110个土壤样本为研究对象, 利用ASD Fieldspec3地物光谱仪在室内条件下测定350~2500 nm波段范围的土壤高光谱数据。对光谱数据进行预处理后, 采用连续投影算法 (SPA) 和竞争性自适应重加权算法 (CARS) 优选的波长变量作为建模参数, 运用偏最小二乘回归 (PLSR) 方法建立模型并比较其精度。【结果】结果表明, 标准正态变换预处理方法是SAP的最佳土壤光谱数据预处理方法。基于标准正态变换后的光谱数据, CARS、SPA算法可将预测SAP的关键波段变量分别压缩至54和13个, CARS-PLSR模型与SPA-PLSR模型相比, 相关系数由0.894提高到0.945, 均方根误差由5.73降低到3.56。【结论】土壤高光谱数据经标准正态变换后, 采用CARS-PLSR算法可有效提高有效磷含量预测的鲁棒性。该结果可为高光谱数据快速反演土壤有效磷含量提供理论依据。

关键词: 土壤有效磷; 高光谱; 竞争性自适应重加权算法; 连续投影算法; 偏最小二乘回归法

DOI: 10.12105/j.issn.1672-0423.20230105

0 引言

土壤有效磷对植物根系生长、光合进程及其产物的合成、运输、代谢等具有重要作用, 是农田土壤肥力评价和作物施肥的重要指标^[1-2]。自然演变及人类活动的影响使土壤有效磷分布具有高度的空间异质性^[3-4], 导致快速获取其分布信息存在一定难度。土壤主要养分含量 (如氮、磷、钾等) 的定量测定一般采用传统实验室化学测试法, 其结果相对准确, 但通常比较费时、费工、有害或有污染。而依据地统计学揭示元素的空间变异性, 则需要大量的土壤样本为前提, 难以实现真正的大田测土施肥。随着信息化技术在土壤科学领域的深入推进, 目前较多学者利用土壤组成物质对特征光谱的吸收或反射作用建立光谱值与测量值模型, 进而预测土壤有效磷含量及其分布状况^[5]。高光谱遥感技术

收稿日期: 2023-01-04

第一作者简介: 郭鹏 (1996—), 博士研究生。研究方向: 遥感图像信息提取。Email: peng_guo@pku.edu.cn

※ 通信作者简介: 陈秀万 (1964—), 博士、博士生导师。研究方向: 农业遥感、农业信息化。Email: xwchen@pku.edu.cn

* 基金项目: 国家重点研发计划项目中国和蒙古国政府间国际科技创新合作重点专项“智慧农牧关键技术集成与中蒙协同创新示范园研究” (2021YFE0102000); 北京大学中央高校基本科研业务费“地球观测与导航教育部工程研究中心——智能飞行器侦测与信息服务系统” (7100604290)

凭借其光谱分辨率高、波段连续性强和光谱信息精细等特征，具备了定量获取土壤养分含量的研究潜力^[6-9]。然而，其高分辨率性质也导致原始高光谱数据中存在大量冗余和共线性信息，降低了其建立光谱模型的精度。鉴于此，有必要探究土壤有效磷的高光谱特征、优选其敏感波长并构建较高精度的估测模型，以期为实时获取土壤养分状况和作物施肥管理提供较为精准的科学依据。

利用土壤有效磷敏感波长变量建立其含量估算模型，可以降低模型复杂度并保持甚至提升模型精度。早期研究较多采用相关分析法研究土壤磷素与土样光谱反射率（或其不同数学变换形式）的关系，将相关系数高的波长作为土壤有效磷敏感波长^[10-13]。而后，部分学者通过分析土壤去除有效磷前、后的光谱变化特征，获取土壤光谱反射率变化较大的波段，将其作为土壤有效磷的敏感波段^[14,16]。近年来，越来越多的学者采用竞争性自适应重加权算法（competitive adaptive reweighted sampling, CARS）、遗传算法（genetic algorithm, GA）、连续投影算法（successive projections algorithm, SPA）和无信息变量去除（uninformative variables elimination, UVE）等变量优选方法，从高光谱全波段数据中滤除无效变量或冗余变量，从而优选出敏感波段^[15-16]。如蔡亮红等^[15]在对土壤水分含量的预测研究中发现，CARS-SPA算法能够较为彻底地去除噪声和无关信息；朱亚星等^[16]在对土壤有机质进行高光谱反演研究中，得到UVE-CARS为最佳变量优选方法，能够在有效压缩波长数的同时提升整体的预测精度；于雷等^[17]利用CARS算法对土壤水分敏感波段进行提取，预测模型 R^2 达到0.983。综上，目前变量优选方法较多应用于土壤有机质、土壤含水量等的光谱响应机理，在土壤有效磷高光谱反演研究中，大多进行相关分析法和建模方法的研究^[18-20]，而变量优选方法应用较少。

文章以成都平原西部西河流域土壤为研究对象，拟从倒数(1/R)一阶微分、倒数的对数($\log(1/R)$)一阶微分、对数的倒数(1/logR)一阶微分、标准正态变换(SNV)和多元散射校正(MSC)5种方法中筛选最适宜的光谱预处理方法，并采用竞争性自适应重加权算法(CARS)和连续投影算法(SPA)优选土壤有效磷的敏感波段，利用偏最小二乘回归法(PLSR)分别建立其预测模型，综合比较不同光谱变量优选算法在土壤有效磷定量预测中的性能，从而获取适用于土壤有效磷定量分析的高光谱的最佳变量选择方法。

1 材料与方法

1.1 研究区概况

四川省崇州市西河流域地处成都平原西部，介于东经 $103^{\circ}26'45''\sim 103^{\circ}40'52''$ ，北纬 $30^{\circ}38'28''\sim 30^{\circ}50'12''$ 之间（图1）。该区域属四川盆地亚热带湿润性季风气候，年平均气温 15.9°C ，年平均降水量 $1\,012.4\text{ mm}$ ，年均无霜期 285 d ，年均日照时长 $1\,161.5\text{ h}$ 。区内海拔高度 $468\sim 882\text{ m}$ ，以丘陵和平原为主。水稻土、黄壤和紫色土为其主要土壤类型，土地利用类型以耕地为主，轮作方式主要为水稻—小麦、水稻—油菜水旱轮作及玉米—油菜、园艺作物旱作。多条国道省道等交通干线以及西河贯穿研究区全境。

2023年2月

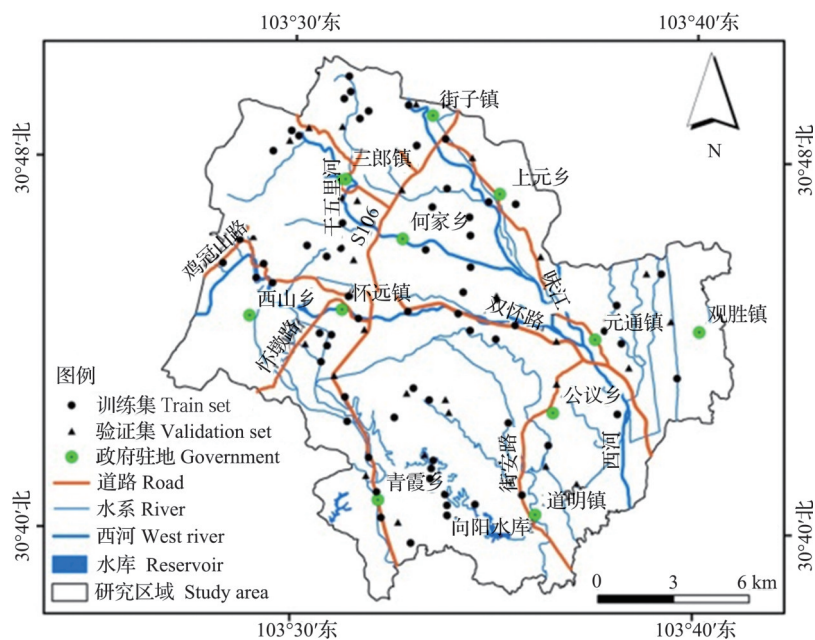


图1 采样点分布

Fig.1 Distribution of soil samples

1.2 土样采集、室内化验分析与样本划分

根据地形地貌、土地利用方式和土壤类型划分采样单元,采用格网布点与重点抽样相结合的方式设计样点,采用五点混合的方式在每个采样单元采集一个混合土样,采样深度为0~20 cm,采土1.5 kg左右,共采集混合土样110个(图1)。土样封存移送实验室,经去除土壤以外的其他侵入体后进行风干处理,随后研磨并过2 mm孔径筛子,混合均匀。采用四分法取土样200 g,平均分为两份,一份装于盛样皿(直径10 cm,深度2 cm),压平并使其表面平整,用于高光谱数据采集;另一份采用碳酸氢钠浸提-钼锑抗分光光度法测定其土壤有效磷含量^[21]。

蒙特卡洛交叉验证能够对异常样本进行筛选,通过剔除和检验光谱矩阵和有效磷含量矩阵方向上的异常点,消除异常样本数据引起的掩蔽效应,保留用于数据分析的有效样本^[22]。该文根据样本分布的离散程度,将有效磷样本中明显偏离了主体的异常样本剔除,以优化样本质量,最终确定101个有效样本进行分析。采用SPXY样本划分算法按照7:3划分建模样本集与验证集,建模集共71个,验证集30个。

1.3 土壤有效磷高光谱测定

土壤高光谱反射率数据测定采用美国ASD Fieldspec3地物光谱仪,光谱范围为350~2 500 nm。重采样间隔设置为1 nm,光谱测定在暗室内进行,将处理好的盛样皿放在反射率几乎为0、厚3 cm的黑色橡胶上,光源为能够提供平行光的功率为50 W的标准直流钨丝石英卤素灯。采用25°视场角光纤探头,探头距土壤表面距离30 cm,光源入射角呈45°。每个土样测定4个方向(转动3次,每次转动90°),每个方向上保存5条样本线,每个样本共采集20条光谱曲线算数平均运算后得到土样反射光谱曲线。运用ViewspectPro

软件去除异常曲线并进行光滑处理，并去除边缘噪声较大的波段 350~399 nm 及 2 401~2 450 nm。

1.4 光谱预处理方法

原始光谱中通常除包含土壤样品信息外，还包含环境噪声、样本背景、仪器电噪声等无关信息，需要对原始高光谱数据做预处理以消除或减弱外界干扰因素的影响。因此，拟采用 1/R 一阶微分、log (1/R) 一阶微分、1/logR 一阶微分、SNV 和 MSC 共 5 种方法对原始光谱数据 (R) 进行预处理，以最大限度减少各类非目标因素对光谱的影响。

1.5 波长变量优选方法

1.5.1 竞争性自适应重加权算法

CARS 模仿的是达尔文进化论中的“适者生存”的原理，该算法的核心思想是利用指数衰减函数和自适应重加权采样方法，优选出每次循环所构建的 PLSR 模型中回归系数绝对值最大的变量点，再利用十折交叉验证优选出 N 个 PLSR 子集模型中十折交叉验证均方根误差 (RMSECV) 最小的子集确定为最优变量子集，其变量优选流程见图 2。

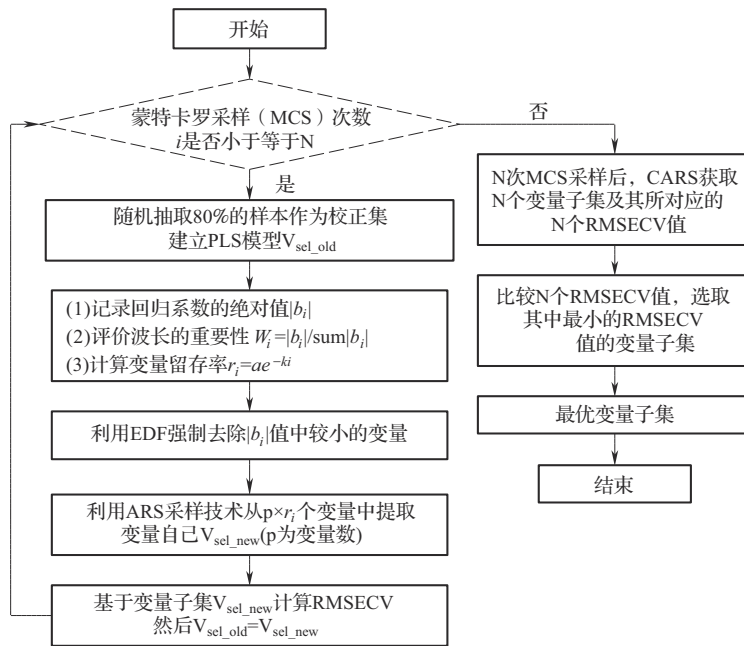


图 2 CARS 变量优选流程

Fig.2 Flowchart of key variables selection by CARS

图 2 的变量留存率式 $r_j = ae^{-ki}$ 中， a 、 k 为常数，分别代表在第一次与第 N 次蒙特卡罗采样时，样本集合中全部的 p 个变量参与建模和仅有 2 个变量参与建模，即 $r_1 = 1$ 且 $r_N = 2/p$ ，得出 a 、 k 的计算公式为：

$$a = \left[\frac{p}{2} \right]^{1/(N-1)} \tag{1}$$

$$k = \frac{\ln(p/2)}{N-1} \tag{2}$$

2023年2月

式(1)(2)中, p 为2 051, MC设置的采样次数为100, 因此, a 和 k 分别为1.07和0.07。

1.5.2 连续投影算法

连续投影算法(SPA)是一种能够使矢量空间中各变量间共线性达到最小化的前向循环选择算法。波长选择过程中将投影向量最大的波长加入到波长组合, 每个新入选的波长都与前一个线性关系最小, 因此可以最大程度优选出有用信息, 降低变量共线性, 减少高光谱冗余信息量, 提升模型运算能力^[23-24]。

1.6 模型建立与比较

偏最小二乘回归(PLSR)算法利用自变量、因变量系统中的数据信息进行分解和筛选, 从中提取对因变量解释性最强的综合变量, 能够分辨系统中的无关信息和有用信息, 消除变量多重相关性的影响, 从而建立合适的预测模型。PLSR模型的建立分析和图表制作分别在Matlab R2016a和Excel 2010中完成。

土壤有效磷预测模型的预测精度采用相关系数(包括校正集交叉验证相关系数 R_c 和预测相关系数 R_p)、均方根误差(校正集交叉验证均方根误差RMSEC和预测均方根误差RMSEP)衡量。其中, 相关系数表征模型建立和验证的稳定性, 越接近于1, 模型的稳定性越好、相应拟合程度越高; RMSE用于检验模型的预测能力, RMSE越小说明模型的估算能力越强。

2 结果与分析

2.1 土壤有效磷含量及其光谱特征

土壤样本有效磷含量的基本统计特征如表1所示, 研究区内土壤有效磷的含量分布范围为2.23~49.91 mg/kg, 平均值为14.86 mg/kg。根据全国第二次土壤普查养分分级标准^[25], 研究区土壤有效磷总体处于中等含量水平。根据SPXY算法划分校正集71个和预测集30个, 校正集土壤有效磷含量分布范围能够较好地涵盖预测集样本, 此特征有助于建立较好的模型, 也验证了SPXY样本划分算法的有效性。

表1 土壤样本有效磷基本统计特征

Table 1 Basic statistical characteristics of available phosphorus in soil samples

样本设置	样本数(个)	平均值(mg/kg)	最小值(mg/kg)	最大值(mg/kg)	标准差(mg/kg)
校正集	71	15.89	3.93	49.91	12.76
预测集	30	14.43	2.23	35.25	8.09
全部	101	14.86	2.23	49.91	9.67

土壤样本有效磷的平均值及其变化范围光谱曲线如图3所示, 在400~1 110 nm范围内, 土壤有效磷样本光谱反射率增加较快; 在1 110~1 600 nm范围内, 反射率增加较慢, 并在1 145 nm处出现明显的反射吸收峰; 在1 600~2 450 nm波段范围内, 反射率急剧下降而后增加, 呈现较大波动变化, 其中在1 400 nm、2 200 nm处存在明显的土壤水分吸收谷。整体而言, 土壤有效磷含量越高, 对应光谱反射率越低。

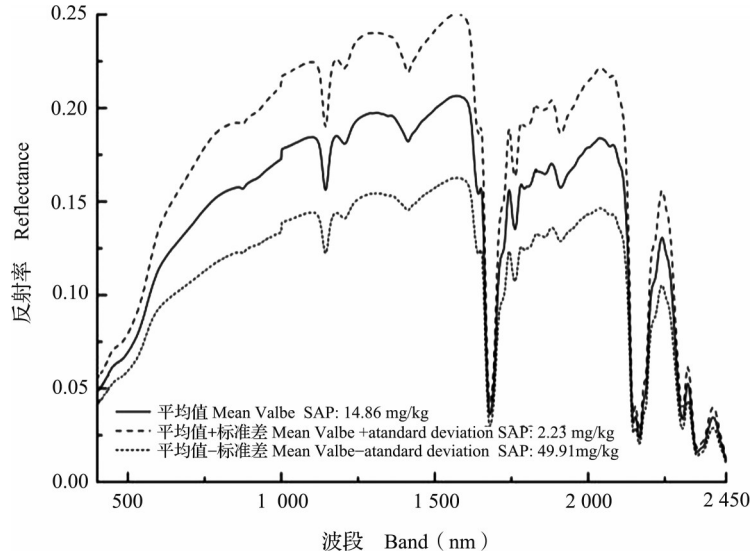


图3 土壤样本光谱平均值与标准差

Fig.3 Mean value and standard deviation of soil sample spectrum

2.2 预处理方法选择

在5种光谱预处理方法中，经过标准正态变换（SNV）处理后所建立的模型预测效果最好（表2），其预测集相关系数和最小均方根误差分别为0.924和4.02。在3种数学变换形式中，倒数一阶变换的建模效果较好，相关系数和最小均方根误差为0.950和2.51，但是存在一定的过拟合；对数的倒数一阶微分的整体建模和预测效果在三种数学变换中最佳；多元散射校正（MSC）对土壤有效磷的预测也存在一定的过拟合现象。综上，与原始光谱和其他4种预处理方法相比，标准正态变换（SNV）的建模和预测精度均较高。因此，将经标准正态变换后的光谱数据用于进一步的分析。

表2 不同光谱预处理方法的建模和预测效果

Table 2 Results of calibration and prediction with different pre-processing methods

预处理方法	因子数	校正集		预测集	
		相关系数 R_c	最小均方根误差 RMSEC	相关系数 R_p	最小均方根误差 RMSEP
R	8	0.904	3.44	0.897	5.63
1/R 一阶微分	5	0.950	2.51	0.894	6.42
log(1/R)一阶微分	2	0.925	3.08	0.889	6.18
1/logR 一阶微分	2	0.935	2.69	0.923	4.08
SNV	4	0.958	2.33	0.924	4.02
MSC	9	0.928	2.99	0.864	6.34

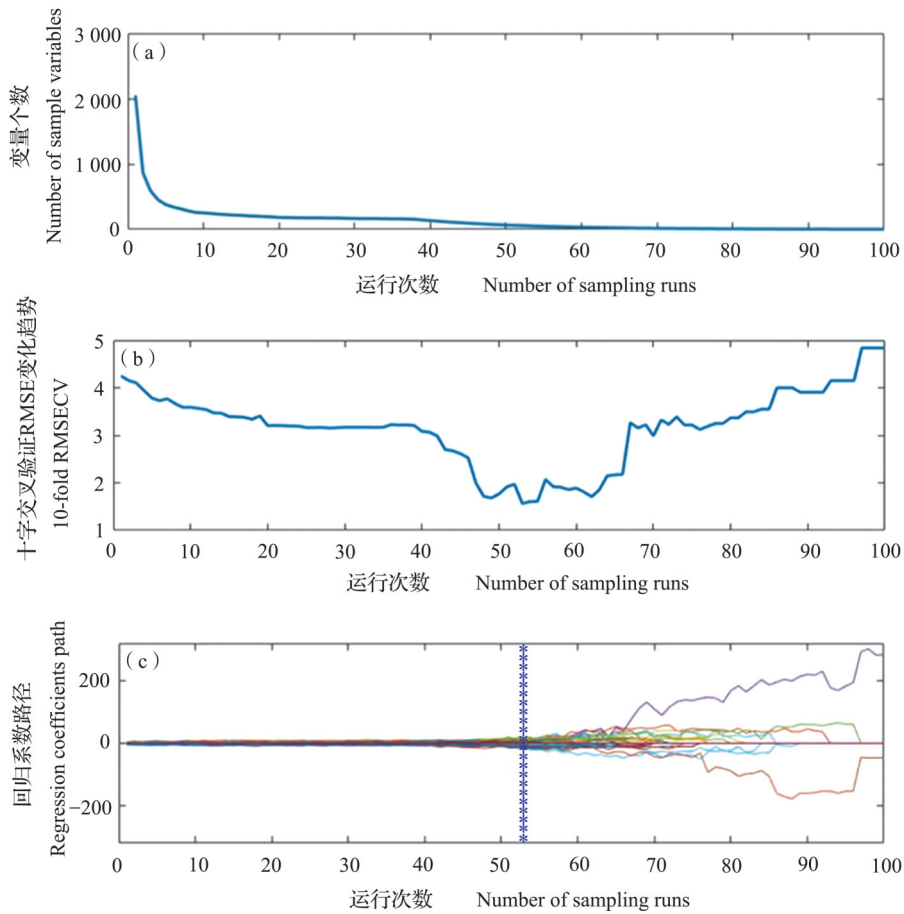
2.3 波长变量优选

2.3.1 CARS 变量优选

基于经标准正态变换后的光谱数据，采用CARS算法进行光谱特征的变量筛选。结果表明，随着运行次数的增加，一定数量的波长变量被剔除，但波长变量被剔除的速度

2023年2月

随运行次数增加逐渐降低,在运行次数为6时,仅20%变量被保留,此过程为CARS的变量粗选(图4a)。十字交叉验证均方根误差变化特征图(图4b)表明,从开始至第53次运行中,交叉验证最小均方根误差(RMSECV)随运行次数增加而降低,在第53次运行后开始增加。表明在运行第53次时,与土壤有效磷含量无关的变量已经被剔除,53次之后再运行可能导致土壤有效磷预测的关键变量被剔除,从而导致均方根误差变大。因此,以第53次运行时所保留的波长变量作为预测土壤有效磷的关键变量(54个)。



“*” 对应为第53次采样且所对应交叉验证的最小均方根误差值最小

图4 土样CARS变量优选

Fig.4 Plots of CARS variable selection for soil samples

2.3.2 SPA 变量优选

根据不同波长范围内的光谱数据信息,通过改变其最大有效波长个数,观察在最大有效波长个数不同的情况下模型的最佳参数及结果,从而确定波长优选结果(图5)。结果表明,最佳变量及所对应变量的索引为13个,对应的波长分别为401 nm、407 nm、421 nm、440 nm、490 nm、548 nm、599 nm、1 000 nm、1 007 nm、1 908 nm、1 931 nm、2 085 nm和2 211 nm。

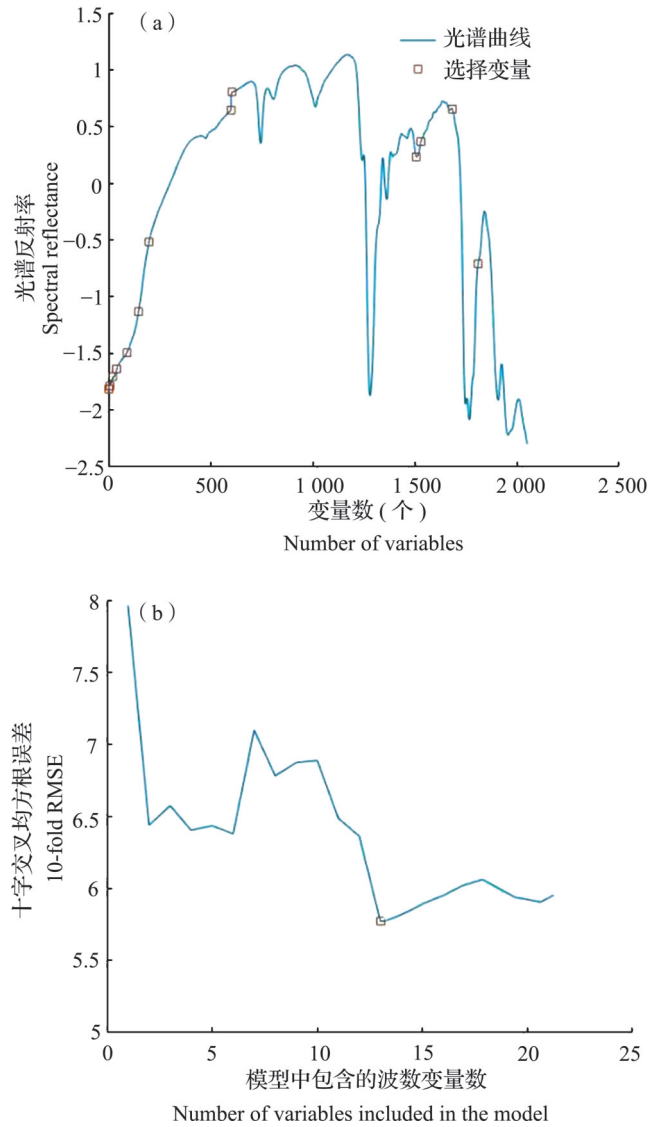


图5 基于SPA的变量筛选
Fig.5 Variables selected by SPA

2.4 PLSR模型的建立与比较

将通过CARS算法和SPA算法优选的变量集作为模型的自变量，土壤样本有效磷实测含量作为因变量，采用PLSR建立土壤有效磷预测模型。结果如表3所示，CARS算法的校正集 R_c 和RMSEC分别为0.966和2.10，预测集的 R_p 和RMSEP分别为0.945和3.56；SPA算法的校正集 R_c 和RMSEC分别为0.926和3.03，预测集 R_p 和RMSEP分别为0.894和5.73。CARS-PLSR模型的预测结果较SPA-PLSR相关系数提高了0.051，均方根误差降低了2.17；与全波段相比，变量数压缩至54个，压缩率为97.36%，极大降低了建模数据量，提高了预测精度。结果表明，CARS算法能够有效优选出预测土壤有效磷含量的波长变量。

2023年2月

表3 不同变量筛选方法的建模效果

Table 3 Results of different variable selection methods

方法	变量数	主因子数	校正集		预测集	
			相关系数 R_c	最小均方根误差 RMSEC	相关系数 R_p	最小均方根误差 RMSEP
SPA-PLS	13	12	0.926	3.03	0.894	5.73
CARS-PLS	54	4	0.966	2.10	0.945	3.56

3 讨论

不同有效磷含量的土壤高光谱曲线形态相似,即随着有效磷含量的增加,反射率呈下降趋势,这与薛丽红等^[26]和于雷等^[27]的研究结果一致,表明通过光谱特征变化规律反演和预测土壤有效磷含量是可行的。为削减和消除基线漂移、散射以及环境噪声干扰等非目标因素对光谱数据的影响,对高光谱采集的数据进行预处理^[28]。比较分析常用的5种预处理方式发现,正态标准变换(SNV)建立的模型和预测精度最高,是进行土壤有效磷建模预测的最好指标。这与王海峰等^[29]对荒漠土壤有机质含量高光谱估算,指出SNV处理能够提高模型预测能力和精度,是最佳光谱反演指标的研究结论一致。

原始高光谱数据中冗余信息的存在会削弱模型的预测能力和稳健性^[30]。采用CARS和SPA两种优选算法筛选不同土壤有效磷含量的关键光谱波段变量,两种算法筛选的变量数分别为54和13个,但CARS相较于SPA建模和预测效果,相关系数分别提高了0.040、0.051,预测偏差比分别降低了0.93、2.18。总体来看,CARS和SPA都能够有效的剔除冗余和无关波长变量,所建立的模型均能较好地预测土壤有效磷的含量。虽然SPA-PLSR参与建模的变量个数仅为原始变量的0.63%,也仅为CARS所选变量个数的24%,但是CARS算法的预测能力和稳定性较SPA更强,这与詹白勺等^[31]、温珍才等^[32]研究发现SPA、GA、CARS等算法均可有效优选出原始高光谱数据中的波长变量,其中CARS算法的优选效果最佳、预测精度最高的研究结果一致。李江波等^[33]比较了MC-UVE、GA、CARS变量优选方法后,发现CARS算法优选效果最佳,能够减小变量间的共线性的研究结果一致,其原因可能是原始变量中存在大量与土壤SAP含量不相关的无信息变量,而无信息变量与光谱有效变量存在非共线的关系,从而导致SPA获取的13个变量中含有部分无信息的变量,降低了SPA-PLSR模型的预测性能。CARS算法优选的变量中,波长大致在403~450 nm、533~626 nm、1 647~1 704 nm、2 133~2 163 nm以及2 343~2 438 nm范围内,这与徐丽华等^[34]对118个土壤样本进行土壤总磷高光谱遥感预测所提取出4个敏感波段(639 nm、1 092 nm、2 262 nm、2 342 nm)的研究结果基本一致,波长变量范围包含其中的敏感波段。比较CARS-PLSR模型与SPA-PLSR模型,两者均极大地降低了建模输入的波长变量个数,而前者的预测精度更高、预测能力更强。因此,CARS算法可以被用来优选土壤有效磷的波长变量,这也为优选其他土壤属性的敏感波段提供数理计量方法支撑。

4 结论

该文将近红外高光谱技术与竞争性自适应重加权算法(CARS)结合,用于土壤有效磷(SAP)含量的预测。结果表明,CARS方法能够有效优选出与土壤SAP相关的重要波长变量,优于SPA变量优选方法,变量筛选效果最佳;采用偏最小二乘回归(PLSR)建立的预测模型CARS-PLSR优于全光谱PLSR及SPA-PLSR模型,其校正集和预测集的相关系数分别为0.966和0.945,均方根误差分别为2.10和3.56。因此,近红外高光谱技术结合CARS方法可以建模并有效预测土壤SAP含量,CARS变量优选方法能够有效简化预测模型并提高模型预测精度。

参考文献

- [1] Qi H J, Paz-Kagan T, Karnieli A, et al. Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil and Tillage Research*, 2018, 175:267-275.
- [2] 李强,戴美玲,向德明,等. 湘西喀斯特区植烟土壤有效磷时空变异及风险评估. *土壤*, 2018, 50(1):181-189.
- [3] 王永壮,陈欣,史奕. 农田土壤中磷素有效性及影响因素. *应用生态学报*, 2013, 24(1):260-268.
- [4] 赵庆雷,吴修,袁守江,等. 长期不同施肥模式下稻田土壤磷吸附与解吸的动态研究. *草业学报*, 2014, 23(1):113-122.
- [5] 张东辉,赵英俊,秦凯,等. 高光谱土壤多元信息提取模型综述. *中国土壤与肥料*, 2018(2):22-28.
- [6] Onojeghuo A O, Blackburn G A, Huang J F, et al. Applications of satellite 'hyper-sensing' in Chinese agriculture: Challenges and opportunities. *International Journal of Applied Earth Observation and Geoinformation*, 2018, 64:62-86.
- [7] 王莉雯,卫亚星. 湿地土壤全氮和全磷含量高光谱模型研究. *生态学报*, 2016, 36(16):5116-5125.
- [8] Ben-Dor E, Patkin K, Banin A, et al. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-A case study over clayey soils in Israel. *International Journal of Remote Sensing*, 2002, 23(6):1043-1062.
- [9] Naveen J P Anne, Amr H Abd-Elrahman, David B Lewis, et al. Modeling soil parameters using hyperspectral image reflectance in subtropical coastal wetlands. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 33:47-56.
- [10] Bogrekcı I, Lee W S. Comparison of ultraviolet, visible, and near infrared sensing for soil phosphorus. *Biosystems Engineering*, 2007, 96(2):293-299.
- [11] 章文龙,曾从盛,高灯州,等. 闽江河口湿地土壤全磷高光谱遥感估算. *生态学报*, 2015, 35(24):8085-8093.
- [12] 徐丽华,谢德德,魏朝富,等. 紫色土土壤全氮和全磷含量的高光谱遥感预测. *光谱学与光谱分析*, 2013, 33(3):723-727.
- [13] 高会,陈红艳,刘慧涛,等. 基于高光谱的鲁西北平原土壤有效磷含量快速检测研究. *中国生态农业学报*, 2013, 21(6):752-757.
- [14] 贾生尧,杨祥龙,李光,等. 近红外光谱技术结合递归偏最小二乘算法对土壤速效磷和速效钾含量测定研究. *光谱学与光谱分析*, 2015, 35(9):2516-2520.
- [15] 蔡亮红,丁建丽. 基于变量优选和ELM算法的土壤含水量预测研究. *光谱学与光谱分析*, 2018, 38(7):2209-2214.
- [16] 朱亚星,于雷,洪永胜,等. 土壤有机质高光谱特征与波长变量优选方法. *中国农业科学*, 2017, 50(22):4325-4337.
- [17] 于雷,朱亚星,洪永胜,等. 高光谱技术结合CARS算法预测土壤水分含量. *农业工程学报*, 2016, 32(22):138-145.
- [18] 张佳佳,郭熙,赵小敏. 南方丘陵稻田土壤全磷、有效磷高光谱特征与反演模型. *江苏农业科学*, 2016, 44(7):522-525.
- [19] 吴茜,杨宇虹,徐照丽,等. 应用局部神经网络和可见/近红外光谱法估测土壤有效氮磷钾. *光谱学与光谱分析*, 2014, 34(8):2102-2105.
- [20] 齐海军,李绍稳,Arnon K,等. 基于PLS-BPNN算法的土壤速效磷高光谱回归预测方法. *农业机械学报*, 2018, 49(2):166-172.
- [21] Bray R, Kurtz L. Determination of total, organic, and available forms of phosphorus in soils. *Soil science*, 1945, 59(1):39-46.

2023年2月

- [22] 李水芳,单杨,范伟,等. 基于MCCV 奇异样本筛选和CARS变量选择法对蜂蜜pH值和酸度的近红外光谱检测. 食品科学,2011,32(8):182-185.
- [23] Shi T Z, Liu H Z, Chen Y Y, et al. Estimation of arsenic in agricultural soils using hyperspectral vegetation indices of rice. *Journal of Hazardous Materials*, 2016, 308:243-252.
- [24] 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. *Chemometrics and Intelligent Laboratory Systems*, 2008, 92:83-91.
- [25] 中国土壤调查办公室. 全国第二次土壤普查养分分级标准. 北京:中国农业出版社,1979.
- [26] 薛利红,周鼎浩,李颖,等. 不同利用方式下土壤有机质和全磷的可见近红外高光谱反演. 土壤学报,2014,51(5):993-1002.
- [27] 于雷,洪永胜,耿雷,等. 基于偏最小二乘回归的土壤有机质含量高光谱估算. 农业工程学报,2015,31(14):103-109.
- [28] 陆婉珍. 现代近红外光谱分析技术(第2版). 北京:中国石化出版社,2005.
- [29] 王海峰,张智韬, Karnieli A, 等. 基于灰度关联-岭回归的荒漠土壤有机质含量高光谱估算. 农业工程学报,2018(14):124-131.
- [30] 杨爱霞,丁建丽. 新疆艾比湖湿地土壤有机碳含量的光谱测定方法对比. 农业工程学报,2015,31(18):162-168.
- [31] 詹白勺,倪君辉,李军. 高光谱技术结合CARS算法的库尔勒香梨可溶性固形物定量测定. 光谱学与光谱分析,2014,34(10):2752-2757.
- [32] 温珍才,孙通,许朋,等. 可见/近红外联合变量优选检测油茶籽油掺假. 江苏大学学报:自然科学版,2015,36(6):673-678.
- [33] 李江波,彭彦昆,陈立平,等. 近红外高光谱图像结合CARS算法对鸭梨SSC含量定量测定. 光谱学与光谱分析,2014,34(5):1264-1269.
- [34] 徐丽华,谢德体. 土壤总氮和总磷含量的高光谱遥感预测. 农机化研究,2012,4(4):119-122.

Prediction of soil available phosphorous using hyperspectral data based on CARS-PLSR^{*}

Guo Peng^{1, 2}, Zhao Yang^{1, 2}, Sun Zihao^{1, 2}, Chen Xiuwan^{1, 2*}

(1. Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China;

2. Ministry of Education of PRC, Engineering Research Center of Ministry of Earth Observation and Navigation (CEON), Peking University, Beijing 100871, China)

Abstract: [**Purpose**] To remove a large amount of redundancy and invalid information in soil hyperspectrum, select the sensitive wavelengths of soil available phosphorus (SAP), simplify the SAP hyperspectral estimation model and improve the prediction accuracy of the models. [**Method**] A total of 110 soil samples were collected from Xihe basin in Sichuan province. The raw hyperspectral reflectance of soil samples in the range of 350~2 500 nm wavelengths were measured by the standard procedure with a spectrometer of ASD Field Spec3 equipped with a high intensity contact probe under the laboratory conditions. Based on the pre-processing spectra, the optimized wavelength variables were selected by using the successive projections algorithm (SPA) and competitive adaptive reweighted sampling algorithm (CARS). Final model with partial least square regression (PLSR) were established. [**Result**] The results showed that SNV was the best soil spectral data pre-processing method of SAP. Based on the

pre-processing spectral data by SNV, the CARS and SPA selected 54 and 13 key wavelength variables from full-spectrum to predict SAP concentrations, respectively. Compared with the SPA-PLSR model, the correlation coefficient of the CARS-PLSR model increased from 0.894 to 0.945, and the root mean square error reduced from 5.73 to 3.56. [**Conclusion**] The utilization of CARS-PLSR algorithm based on soil hyperspectral data that is pre-processed by SNV could enhance SAP forecasting capability and reduce the model complexity, which could provide theoretical basis for the rapid inversion of SAP concentration by using hyperspectral data.

Key words: soil available phosphorous (SAP); hyperspectral; competitive adaptive reweighted sampling algorithm (CARS); successive projections algorithm (SPA); partial least square regression (PLSR)

欢迎订阅《中国农业信息》

《中国农业信息》(双月刊)由农业农村部主管,中国农学会农业信息分会、中国农业科学院农业资源与农业区划研究所共同主办,是我国目前全方位传播和刊载国内外农业遥感/农业信息科学领域的信息获取、处理、分析和应用服务的理论、技术、系统集成、标准规范等方面最新进展和成果,促进学术交流以及农业信息学科关键技术与产品的创新研发、集成推广和应用示范的综合性科学技术期刊。

主要刊登农业遥感、农业传感器、农业信息智能处理、精准农业/智慧农业、农业监测预警与信息服务系统、农业物联网、智能装备与控制、虚拟农业、人工智能、信息技术标准等方向学科热点领域的最新、最重要的理论研究和应用成果。主要栏目有:农业遥感、智慧农业、综合研究、农业信息技术、农业物联网、专题报道等。目前被中国知网(CNKI)、万方数据、中文科技期刊数据库、中国核心期刊(遴选)数据库等多家数据库收录。

《中国农业信息》为国内外公开发行的刊物,开本为16开,彩色四封,读者范围广,影响面大,全国各地邮局均有订阅。每双月25号出版,定价为25.00元/册,150元/年。

邮局汇款

收款人:《中国农业信息》编辑部

地址:北京市海淀区中关村南大街12号中国农科院资源所区划楼319

邮编:100081

银行汇款

开户行:农行北京北下关支行

行号:103100005063

账号:11050601040011896

单位名称:中国农业科学院农业资源与农业区划研究所

电话:(010)82109632

传真:(010)82109632

E-mail:nyxxbjb@caas.cn

邮发代号:2-733

投稿网址:www.cjarrp.com