

DOI: 10.5846/stxb201910172184

朱建伟, 刘玉学, 吴超凡, 靳佳, 吕豪豪, 杨生茂. 施用生物炭后土壤有机碳的近红外光谱模型研究与应用. 生态学报, 2020, 40(20): 7430-7440.

Zhu J W, Liu Y X, Wu C F, Jin J, Lü H H, Yang S M. Study on near-infrared spectroscopy model of soil organic carbon after biochar addition and its application. Acta Ecologica Sinica, 2020, 40(20): 7430-7440.

施用生物炭后土壤有机碳的近红外光谱模型研究与应用

朱建伟^{1,2}, 刘玉学^{2,3}, 吴超凡¹, 靳佳¹, 吕豪豪^{2,3}, 杨生茂^{2,3,*}

1 浙江师范大学地理与环境科学学院, 金华 321000

2 浙江省农业科学院环境资源与土壤肥料研究所, 杭州 310021

3 浙江省生物炭工程技术研究中心, 杭州 310021

摘要: 土壤有机碳是影响土壤肥力的最重要因素之一。生物炭由于具有高度芳香化碳结构和发达孔隙结构等特性, 可以作为一种土壤改良剂, 提高土壤有机碳含量, 改善土壤物理结构, 近些年成为农业环境领域研究的热点。分别采用传统方法和可见光近红外光谱(VIS-NIRS, 400—2500 nm)技术对施加不同用量生物炭的土壤有机碳含量进行检测和对比分析, 以期对含生物炭土壤的有机碳分析建立有效预测模型。通过比较不同样本选择方法(Kennard-Stone(KS), Random selection(RS)和 Sample set partitioning based on joint x-y distances(SPX))和光谱预处理方法(Savitzky-Golay平滑(SG)、倒数的对数 $\log(1/R)$ 、标准正态变量变换(SNV)、一阶导数(Der1)、二阶导数(Der2)和多元散射校正(MSC)), 以3种模型(组合间隔偏最小二乘模型(Synergy Interval Partial Least Squares, siPLS), 遗传算法-支持向量机模型(Genetic Algorithm-Support vector machine, GA-SVM)和随机森林模型(Random Forest, RF))来建立生物炭土壤有机碳预测模型。结果表明:(1)施加生物炭增加了土壤有机碳含量, 增加幅度随生物炭添加量的提高呈增加趋势;(2)土壤反射率随土壤有机碳含量的增加而降低, 在1410、1920和2200 nm光谱附近存在明显的吸收谷;(3)对比3种样本选择方法, KS方法所划分的样本集相对于RS方法和SPXY方法更适用于生物炭土壤有机碳模型的建立;(4)以SG+MSC预处理结合GA-SVM方法建立的模型精度最高, 校正集的 R_{cal}^2 和RMSECV值分别为0.9526和0.4839, 验证集的 R_{val}^2 和RMSEP值分别为0.8598和0.9987, RPD值为2.6017。该模型因具有精度高且模拟效果较好等优点, 可用于含生物炭土壤的有机碳含量的科学预测。

关键词: 生物炭; 土壤有机碳; 近红外光谱; 预测模型; 样本选择

Study on near-infrared spectroscopy model of soil organic carbon after biochar addition and its application

ZHU Jianwei^{1,2}, LIU Yuxue^{2,3}, WU Chaofan¹, JIN Jia¹, LÜ Hao hao^{2,3}, YANG Shengmao^{2,3,*}

1 College of Geography and Environmental Sciences, Zhejiang Normal University, Jinhua 321000, China

2 Institute of Environment, Resource, Soil and Fertilizer, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

3 Engineering Research Center of Biochar of Zhejiang Province, Hangzhou 310021, China

Abstract: Soil organic carbon (SOC) is one of the most important factors affecting soil fertility. Due to the characteristics of highly aromatic carbon structure and developed pore structure, biochar can be used as a soil amendment to increase SOC content and improve soil physical structure, which has become a research hotspot in the fields of agriculture and environment in recent years. In this study, both traditional method and visible near-infrared spectroscopy (VIS-NIRS,

基金项目: 国家自然科学基金项目(41701334); 浙江省重点研发计划项目(2020C02030)

收稿日期: 2019-10-17; 网络出版日期: 2020-08-27

* 通讯作者 Corresponding author. E-mail: yangshengmao@263.net

400—2500 nm) were used to detect SOC content in samples containing different amounts of biochar, in order to establish an effective prediction model for the analysis of organic carbon in soils containing biochar. An optimal prediction model was established for quantifying SOC content through three processes, including comparing different sample-selection methods (Kennard-Stone, Random selection, and SPXY), comparing various spectral pre-processing methods, and matching with three models. The pre-processing methods included Savitzky-Golay smoothing (SG), $\log(1/R)$, standard normal variate transformation (SNV), first derivative (Der1), second derivative (Der2), and multiplicative scatter correction (MSC). The three models applied in this study were Synergy Interval Partial Least Squares (siPLS), Genetic Algorithm-Support Vector Machine (GA-SVM), and Random Forests (RF). Results showed that: (1) SOC content was increased significantly by biochar addition and was affected by the amount of biochar. (2) Soil reflectance decreased with the SOC content increasing, indicated by obvious absorption valleys at the spectra nearby 1410, 1920, and 2200 nm. (3) Compared with the three sample selection methods, the sample set divided by KS method was more suitable for the SOC modeling process than those by RS and SPXY methods. (4) The model established by SG+MSC pretreatment combining with GA-SVM method had the highest accuracy, with $R_{cal}^2 = 0.9526$ and $RMSECV = 0.4839$ in the calibration set, and $R_{val}^2 = 0.8598$, $RMSEP = 0.9987$, and $RPD = 2.6017$ in the validation set. The model can be used for scientific prediction of SOC in samples containing biochar due to its advantages of high precision and good simulation effects.

Key Words: biochar; soil organic carbon; prediction model; pre-processing; sample selection

土壤是农业生产活动进行的基础,是极为珍贵的自然资源,而土壤质量的好坏影响着农作物的生长。土壤有机碳(Soil Organic Carbon, SOC)是生态系统中主要的碳源,可以改善土壤团聚体结构、渗透性、微生物活性^[1-2],是评价土壤质量的最重要的指标之一。SOC 是碳库的重要组成部分,其变化影响着土壤碳排放,因此,准确评估 SOC 含量是当前全球碳循环研究的重点内容。生物炭(Biochar)是农作物秸秆、生活废弃物以及动物粪便等生物质在缺氧或低氧条件下,高温裂解后产生的富碳黑色固态物质^[3-5]。生物炭由于其碳组分高度芳香化而表现出很强的热稳定性和生物化学抗分解性。此外,生物炭孔隙结构发达,比表面积大,表明含氧官能团丰富,具有较强的吸附性能,因而常被用作土壤改良剂添加到土壤中,起到增加 SOC 含量、改善土壤物理结构、持留土壤养分等作用,进而提升作物产量。Backer^[6]在加拿大魁北克南部两种类型土壤中施加生物炭,发现施炭3年后两种土壤的养分和作物产量对生物炭的响应不同,砂质粘壤土的 SOC 浓度比对照高 67%。另有研究发现,施加生物炭增加了我国南部热带和亚热带地区红壤 SOC 和土壤团聚体的碳氮比^[7]。由此可见,SOC 含量通常随着生物炭的添加而提高,且随着时间的延长而发生动态变化。SOC 含量测定方法的快速、高效和准确是开展生物炭对 SOC 影响研究的关键。但是相关研究对 SOC 的定量检测主要基于传统的化学分析方法,该分析方法存在着测定时间长、过程繁琐、样品氧化不完全引起结果存在偏差、污染环境等缺点^[8-9],因此,探索快速、准确、成本低且无污染的 SOC 测定方法十分必要^[10-11]。

近红外光谱技术具有快速准确获取待测样品属性数据等优点,被广泛应用于土壤理化性质的科学预测^[12-14]。目前国内外学者对不同类型、不同区域土壤的 SOC 预测进行了大量研究^[11, 15]。崔霞等^[16]以甘南藏族自治州高寒草地土壤为研究对象,通过比较多种光谱变化形式以及3种多元线性回归方法,建立了针对高寒草地土壤的 SOC 预测模型。Marijn 等^[17]评估了 PLSR 模型在潮沼地土壤中 SOC 预测性能,表明中红外 PLSR 模型可以很好地预测该类型土壤的 SOC 含量。Liu 等^[18]比较了不同母质土壤的光谱曲线特征以及对预测模型建立的影响。而在模型的建立中,建模方法的选择、光谱的预处理以及建模集验证集的选择影响着模型的预测性能。刘彦姝等^[19]通过对杉木林土壤氮含量的光谱预测研究,比较了不同种光谱预处理的方法,并对最小二乘-支持向量机回归模型(LS-SVR)和 PLSR 模型进行评价,结果表明 LS-SVR 比 PLSR 具有更好的预测能力。Bushang 等^[20]结合偏最小二乘法预测土壤中 SOC 含量,研究结果发现,光谱经过一阶导数处理后,模型性能明显优于原始反射率预测模型。Nawar 和 Mouazen^[21]比较了不同的土壤样本选择方法,并且在

此基础上建立 PLSR 模型来验证样本选择方法的优缺点。

然而,在众多研究中,对施加生物炭后的土壤 SOC 模型预测的研究还很少。因此,本文在室内条件测定不同处理后的生物炭土壤光谱,采用 3 种样本选择方法(KS 算法、RS 算法和 SPXY 算法)、6 种光谱预处理方法及 3 种建模方法(siPLS 模型、GA-SVM 模型和 RF 模型)分别建立含生物炭土壤的 SOC 预测模型,并对各模型的预测效果进行对比分析,探究生物炭施入土壤后 SOC 的最适模型,以期适合含生物炭土壤 SOC 含量的科学预测提供参考依据。

1 实验设计和研究方法

1.1 研究区域

研究区位于浙江省农业科学院海宁市许村镇杨渡科研基地(120°24'23"E, 30°26'07"N),属于亚热带季风气候区。其特征是暖季气候湿润,降水较多;冷季气候干燥,降水较少。平均年降水量 1187 mm,平均气温 15.9℃。供试土壤为水稻土。

1.2 实验设计及样品采集

试验设置 5 个处理:CK(无施肥)、BC0(常规施肥)、BC1(常规施肥+7.5 t/hm²生物炭)、BC2(常规施肥+15 t/hm²生物炭)、BC3(常规施肥+22.5 t/hm²生物炭)。每个处理重复 3 次,采用随机区组设计。常规施肥用量 N:P:K(尿素:过磷酸钙:氯化钾)的比例为 2:1:1.4,供试作物为油菜-水稻轮作。生物炭于 2011 年油菜季一次性施加,与 0—20 cm 表层土壤混合均匀。本研究样品采集时间为:2014 年 6 月 6 日、2015 年 12 月 25 日、2016 年 5 月 24 日、2017 年 3 月 1 日、2017 年 11 月 20 日、2018 年 11 月 20 日。采用“五点采样法”采集 0—20 cm 表层土壤样品,每个样品混匀后放入保鲜袋带回实验室进行自然风干,挑除草根石块等杂质后进行研磨后过 100 目筛。然后将每个土壤样品分为两份,一份用于传统化学分析方法,即重铬酸钾加热法,对土壤有机碳含量进行测定;另一部分用于光谱采集。

1.3 光谱采集及处理

光谱测定采用美国 ASD 公司生产的 FieldSpec 4 Hi-Res NG 光谱仪,光谱仪的波长范围为 350—2500 nm,光谱分辨率在 700 nm 为 3 nm,在 1400 nm/2500 nm 为 6 nm。光谱扫描时间为 100 ms。将土壤样品放置于培养皿内,并用玻璃片将土壤表面压平,光谱测定时培养皿内土壤厚度为 1.7 cm。光谱测量在暗室环境中进行,视角探头为 25°,每个土壤样品测定前进行白板校正。每个土壤样品测定 30 次(每测量 10 次将土壤样品旋转 90°),取 30 次反射率平均值得到该土壤样品反射率^[22]。

由于在土壤光谱采集过程中,受到周围环境的影响,测得的光谱常常存在随机噪声,从而影响重要信息的提取,进一步影响预测模型的准确性,所以需对采集的土壤光谱进行预处理。本文采用 Savitaky-Golay 对原始光谱数据进行平滑处理,再通过不同方法进行预处理,包括 $\log(1/R)$ 、Der1、Der2、SNV 和 MSC。

1.4 样本选择方法

校正集和验证集的划分,对模型的建立有着重要意义^[23]。本研究采用 3 种算法划分校正集和验证集:KS 算法、RS 算法和 SPXY 算法。3 种算法划分校正集和验证集数目比为 2:1,校正集样本数量为 56 个,用于建立模型;验证集样本为 28 个,用于检验模型预测效果。KS 算法是通过各样本值之间的欧式距离,选择与其他所有样本中最大距离的样本,并在此基础上,选择离此样本最远距离的样本,重复以上步骤达到初始设定的样本数后停止^[24]。RS 算法是随机选择所设定数量的样本组成校正集和验证集。SPXY 方法是 KS 方法基础上提出的基于联合 X - Y 距离的样本划分方法,充分考虑了 X 和 Y 的空间可变性^[25]。

1.5 模型及验证

1.5.1 siPLS 模型

偏最小二乘模型(Partial Least Squares, PLS)是最常用的一种多元统计数据分析方法,它是结合了多元线性回归分析、相关性分析和主成分分析的优势而成的一种统计学方法。组合间隔偏最小二乘(siPLS)是将整

个光谱波段平均分成 n 个等间隔区间,然后计算所有的两个、三个或四个区间模型组合,根据每个可能的模型组合进行 PLS 建模,当交互验证均方根误差值最小时,该组合被认定为最优波段进行建模^[26-27]。siPLS 模型通过选择合适的区间来消除无关噪声,可以提高模型的精度。

1.5.2 GA-SVM 模型

支持向量机(Support vector machine, SVM)是 Vapnik 在 1992 年发明的一种基于统计学理论的机器学习算法,它是一种二分类模型,基本原理是寻找一个最优超平面,使其分开的两类型具有最大间隔。遗传算法(Genetic Algorithm, GA)是模拟达尔文生物进化论的一种计算方法,通过模拟自然选择过程来搜索最优解^[28-29]。在 SVM 模型中,SVM 模型性能的好坏取决于惩罚参数 c 和核函数半径 g 的取值,通过使用 GA 算法进行最优参数搜索,进而提高 SVM 模型的预测精度^[30]。

1.5.3 RF 模型

随机森林(Random Forest, RF)模型是一种从分类回归树(Classification and Regression Tree, CART)算法发展而来的机器学习方法。RF 模型原理是随机有放回地选取样本并对其构建多棵决策树,随后将多棵决策树生成森林,然后通过计算各个决策树预测值求平均作为最终结果^[31]。RF 模型解决了 CART 算法的过度拟合问题,提高了模型的预测精度。与大多数的机器学习方法不同,RF 只需要设置决策树数目(N_{tree})和节点特征数(M_{try})就可以生成预测模型。

1.5.4 模型验证

本文采用决定系数(Correlation coefficient, R^2),校正集均方根误差(Root Mean Square Error of Cross Validation, RMSECV),验证集均方根误差(Root Mean Square Error of Prediction, RMSEP)和相对分析误差(Residual Predictive Deviation, RPD),对建立的模型进行性能评价。其计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \quad (3)$$

$$RPD = \frac{SD}{RMSEP} \quad (4)$$

式中, y_i 表示样本集实测值, \hat{y}_i 表示样本集预测值, \bar{y} 表示样本集平均值, N 表示样本集大小, n 表示校正集样本大小, m 表示验证集样本大小, SD 表示验证集样本标准差。 R^2_{cal} 越大, $RMSECV$ 越小, 表明模型越稳定; R^2_{val} 越大, $RMSEP$ 越小, 表明模型的预测效果越好。当 $RPD < 1.4$ 时, 表明模型效果较差, 不适用于 SOC 的预测; 当 $1.4 < RPD < 2.0$, 表明模型效果一般, 可以对 SOC 进行简单预测; 当 $RPD > 2.0$ 时, 表明模型效果极好, 适用于 SOC 的预测。

模型的建立以及验证等计算程序在 MATLAB R2018a 中进行, 图标制作与数据统计在 Origin 和 Excel 中完成。

2 结果与分析

2.1 生物炭对土壤有机碳的影响

许多研究表明, 农田中施加生物炭可以提高 SOC 的含量。不同生物炭添加量处理的土壤有机碳含量如图 1 所示。与未施加生物炭 BC0 相比, BC2 和 BC3 显著地增加了 SOC 含量, 且 BC3 处理显著高于 BC2 处理

(2018年11月20日样品除外)。BC1在各个时间段与BC0并无显著性差异,这是因为生物炭添加量比较低,不足以引起SOC含量的显著变化。本研究结果表明,添加生物炭促进土壤中SOC含量的增加,并随其添加量的增加而增加,可能因为生物炭本身含有很高的碳含量,且具有高度稳定的芳香化结构,施入土壤多年后仍然保持一定的理化和生物稳定性。这与许多研究结果相一致^[32-33]。

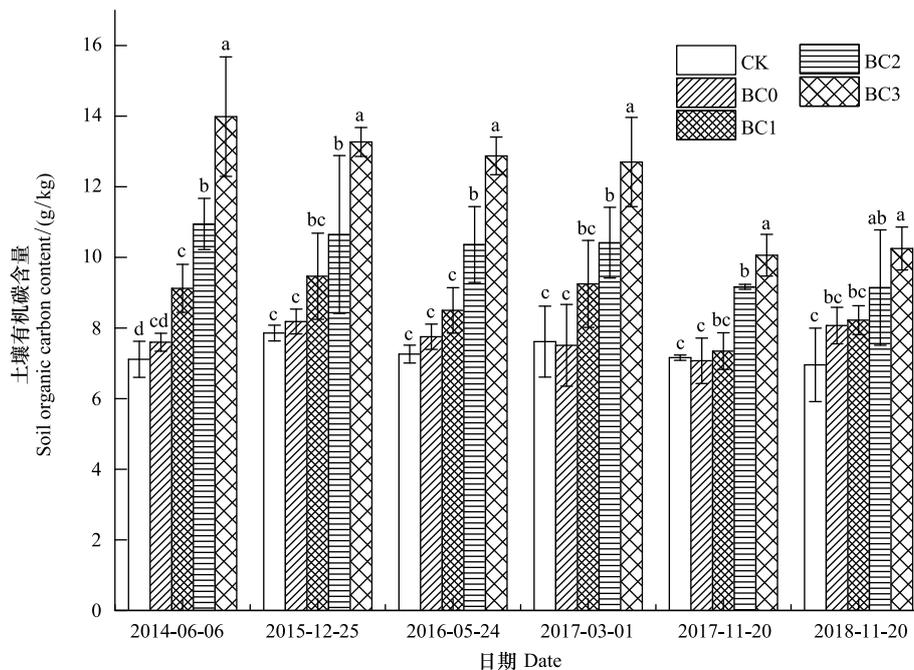


图1 不同生物炭添加量处理土壤有机碳含量

Fig.1 Soil organic carbon content with different amounts of biochar

CK: 无施肥, No fertilizer; BC0: 施肥, Chemical fertilizer; BC1: 施肥+7.5 t/hm²生物炭, Chemical fertilizer+7.5 t/hm² Biochar; BC2: 施肥+15 t/hm²生物炭, Chemical fertilizer+15 t/hm² Biochar; BC3: 施肥+22.5 t/hm²生物炭, Chemical fertilizer+22.5 t/hm² Biochar

2.2 土壤的光谱特征

通过计算各处理的土壤光谱反射率的平均值,比较各处理下土壤反射率变化。由图2可知,SOC含量越大,反射率越小,但土壤反射率总体趋势相似,基本呈上升抛物线型。在400—780 nm的斜率较陡,土壤反射率增加较快;而在780—2120 nm的斜率较缓,土壤反射率增加较慢。2120 nm之后,土壤反射率有略微下降的趋势。在1410、1920、2200 nm附近存在明显的吸收谷,这是由于在1410 nm波段和1920 nm波段主要为水分吸收带,通常认为这与OH—基团和水分子振动有关;2200 nm波段附近存在Al—OH矿物吸收带,表明高岭石和其他硅酸盐存在^[34-35]。

2.3 不同光谱预处理方法对模型性能的影响

本节以KS为样本选择方法并且基于6种光谱预处理方法,对84个土壤样品建立了siPLS、SVM、RF模型,表1—表3为3种模型的精度评价结果。在siPLS模型中,相关系数 R_{val}^2 从大到小顺序为:原始光谱>SG>SG+MSC>SG+log(1/R)>SG+SNV>SG+Der1>SG+Der2,并且其中原始光谱、SG、SG+MSC的siPLS模型的RPD值大于2,分别为2.4079、2.2345和2.1116,而SG+Der1

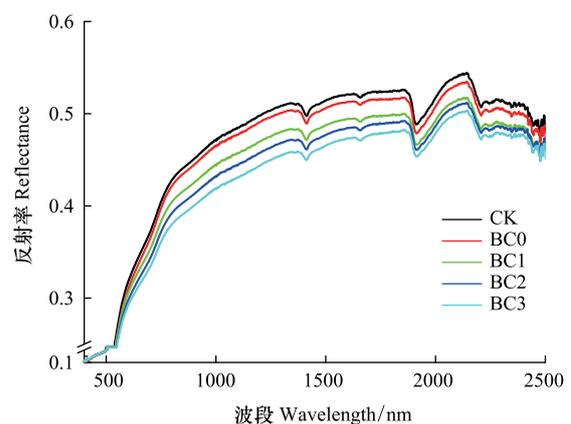


图2 不同处理下土壤光谱曲线

Fig.2 Soil spectra at different biochar amounts treatment

以及 SG+Der2 的 siPLS 模型的 RPD 值小于 1.4, 分别为 1.3000 和 0.8176, 这表明后两种组合不适合土壤 SOC 预测。因此, 未对光谱进行预处理更适用于 siPLS 模型。

在 GA-SVM 模型中, 相关系数 R_{val}^2 从大到小顺序为: SG+MSC>SG+log(1/R)>SG+SNV>SG>原始光谱>SG+Der1>SG+Der2, 其中除了 SG+Der1 和 SG+Der2 的 GA-SVM 模型的 RPD 值小于 2 外, 其余的模型均大于 2, 而 SG+Der1 和 SG+Der2 的 GA-SVM 模型的 RPD 值只有 1.7714 和 0.8621。GA-SVM 模型也适合于小样本建模, 而且避免了局部最小值^[36]。SG+Der2 预处理方法在校正集中表现良好, R_{cal}^2 值高达 0.9991, RMSECV 为 0.0883, 但是在验证集中, R_{val}^2 和 RMSEP 值只有 0.0117 和 2.2426, 表明该预处理方法所建立的 SVM 模型的稳定性较差, 出现了过拟合现象。

在 RF 模型中, 相关系数 R_{val}^2 从大到小顺序为: SG+SNV>SG+MSC>SG>原始光谱>SG+log(1/R)>SG+Der1>SG+Der2, 其中只有 SG+SNV 和 SG+MSC 的 RF 模型的 RPD 值大于 2, 分别为 2.4699 和 2.2744。在基于 RF 建立的模型中, 模拟效果普遍不佳, 相对于其他两个模型方法表现出了较差的模拟性能。RF 方法在校正集中效果表现较好, 但是在验证集模型中效果一般, 表明该方法在本研究中鲁棒性较差。

表 1 siPLS 模型精度评价

Table 1 Accuracy evaluation of siPLS model

处理 Process	校正集 Calibration set		验证集 Validation set		
	R_{cal}^2	RMSECV	R_{val}^2	RMSEP	RPD
R	0.9133	0.6563	0.8338	0.8068	2.4079
SG	0.9039	0.6949	0.8102	0.8625	2.2345
SG+log(1/R)	0.9024	0.6839	0.7399	0.9891	1.9406
SG+SNV	0.8676	0.7972	0.7178	1.0317	1.8759
SG+Der1	0.7997	0.9820	0.4884	1.4743	1.3000
SG+Der2	0.2076	1.9737	0.0042	2.3659	0.8176
SG+MSC	0.8686	0.7935	0.7762	0.9316	2.1116

R: 原始光谱, Reflectance; SG: SG 平滑, Savitzky-Golay smoothing; SG+log(1/R): SG 平滑+倒数的对数, Savitzky-Golay smoothing + Logarithm of the reciprocal; SG+SNV: SG 平滑+标准正态变量变换, Savitzky-Golay smoothing + Standard normal variate transformation; SG+Der1: SG 平滑+一阶导数, Savitzky-Golay smoothing + First derivative; SG+Der2: SG 平滑+二阶导数, Savitzky-Golay smoothing + Second derivative; SG+MSC: SG 平滑+多元散射校正, Savitzky-Golay smoothing + Multiplicative scatter correction; R_{cal}^2 : 校正集决定系数, Correlation coefficient of calibration set; RMSECV: 校正集均方根误差, Root mean square error of cross validation; R_{val}^2 : 验证集决定系数, Correlation coefficient of validation set; RMSEP: 验证集均方根误差, Root mean square error of prediction; RPD: 相对分析误差, Residual predictive deviation

表 2 GA-SVM 模型精度评价

Table 2 Accuracy evaluation of GA-SVM model

处理 Process	校正集 Calibration set		验证集 Validation set		
	R_{cal}^2	RMSECV	R_{val}^2	RMSEP	RPD
R	0.9984	0.0914	0.835	0.8103	2.3397
SG	0.9858	0.2626	0.8383	0.7885	2.4152
SG+log(1/R)	0.9755	0.3441	0.8477	0.7613	2.4989
SG+SNV	0.9575	0.4579	0.8414	0.7908	2.5111
SG+Der1	0.9988	0.0852	0.6817	1.0713	1.7714
SG+Der2	0.9991	0.0883	0.0117	2.2426	0.8621
SG+MSC	0.9526	0.4839	0.8598	0.9987	2.6017

在上述研究中, 最优 siPLS、GA-SVM 和 RF 模型如图 3—图 5 所示。在表 1—表 3 中, GA-SVM 模型的预测性能优于 siPLS 模型和 RF 模型, 并且在 3 种模型中, 基于 SG+MSC 方法建立的模型, RPD 值都高于 2.0, 尤其在 GA-SVM 模型中, RPD 值为 2.6017, 表明该模型具有极好的预测性能。对于 SG+Der2 方法, siPLS 校正集模型精度一般, 预测模型的效果也不好, 而 GA-SVM 模型和 RF 校正集模型建模精度很高, 但是预测模型的预

测效果并不佳,决定系数 R_{val}^2 值很低,且 RPD 值小于 1,表明该方法并不能提高模型的精度,所建立的模型不适用于土壤 SOC 的预测。在 3 个最优模型(图 3—图 5)中,预测值与测量值之间均具有较好的相关性。斜率从大到小的顺序为:siPLS>RF>GA-SVM。从模拟效果数据点分析,siPLS 斜率接近于 1,模型的总体模型能力较好,平均分布在 1:1 直线两侧;GA-SVM 模型中,在 6—10 g/kg 的土壤样品中的预测值偏大;RF 模型中,与 siPLS 模型模拟效果类似,但在较小的有机碳含量的土壤样品中出现较大偏差。总体来说,基于 SG+MSC 方法预处理光谱,以 GA-SVM 方法建立的模型的校正集和验证集的精度都较高,因此该模型更适用于生物炭土壤的 SOC 的精准预测。

表 3 RF 模型精度评价

Table 3 Accuracy evaluation of RF model

处理 Process	校正集 Calibration set		验证集 Validation set		
	R_{cal}^2	RMSECV	R_{val}^2	RMSEP	RPD
R	0.9575	0.6110	0.6851	1.1229	1.7257
SG	0.9555	0.6121	0.6760	1.1292	1.7305
SG+log(1/R)	0.9586	0.5970	0.6572	1.1500	1.6891
SG+SNV	0.9592	0.4823	0.8363	0.7975	2.4699
SG+Der1	0.9635	0.5204	0.6230	1.1901	1.6282
SG+Der2	0.9465	0.8182	0.0027	2.2032	0.8932
SG+MSC	0.9610	0.4687	0.8559	0.9040	2.2744

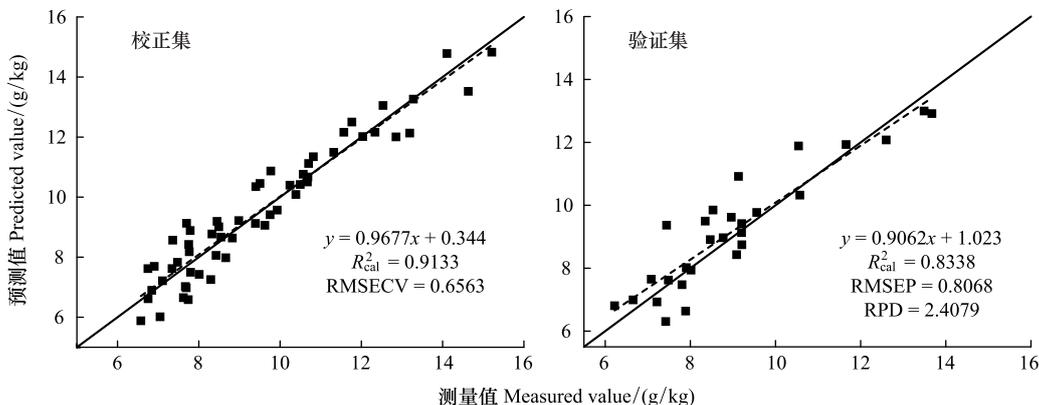


图 3 siPLS 最优模型精度评价

Fig.3 Accuracy evaluation of optimal siPLS model

2.4 不同样本选择方法对模型性能的影响

在上一节研究基础上,选择 SG+MSC 方法作为预处理方法来研究不同样本选择方法对模型性能的影响。在 siPLS 模型中,RS 样本选择方法的决定系数 R_{val}^2 值明显高于 KS 方法和 SPXY 方法,并且 RS 样本选择方法所建立模型的 RPD 值极其接近 2.5,模型预测结果较好,而 SPXY 样本选择方法建立的模型的 R_{val}^2 值较低,且相关分析误差 RPD 值仅为 1.6009。在 GA-SVM 模型和 RF 模型中,KS 样本选择方法所建立的模型明显优于其余两种样本选择方法,且 RPD 值都高于 2.0,而 SPXY 方法并不适用于两个模型,所建立的模型的鲁棒性较差,模型并不稳定,不适用于 SOC 预测。RS 方法是一种随机样本选择方法,运用该方法进行样本选择时每次的选择均有差异,每个样本的模拟效果差异较大。在表 4 所列模型中,以 KS 方法划分样本集,GA-SVM 建模方法, R_{cal}^2 和 RMSECV 值为 0.9526 和 0.4839,验证集 R_{val}^2 和 RMSEP 值为 0.8598 和 0.9987,且 RPD 值为 2.6017,表明模型的模拟效果极好。在本研究中,RS 方法更适用于 siPLS 模型,而 KS 方法更适用于 GA-SVM 模型和 RF 模型,总体而言,KS 方法相较于 RS 方法和 SPXY 方法更加适用生物炭土壤 SOC 模型。

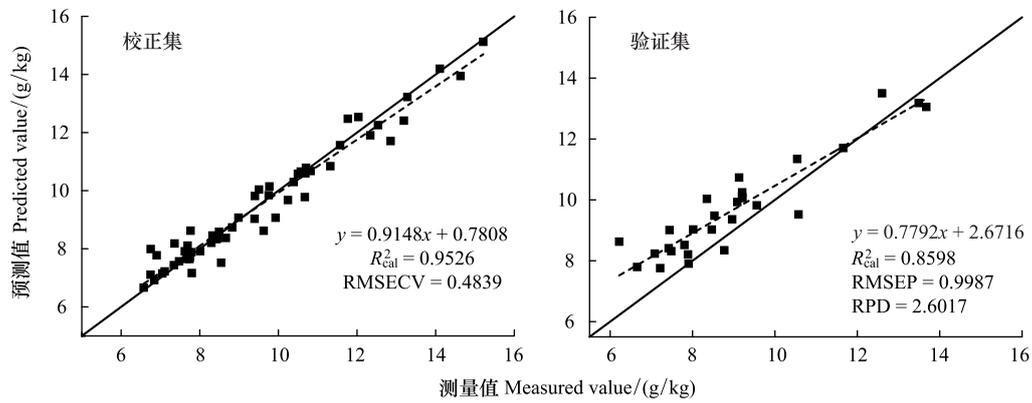


图 4 GA-SVM 最优模型精度评价

Fig.4 Accuracy evaluation of optimal GA-SVM model

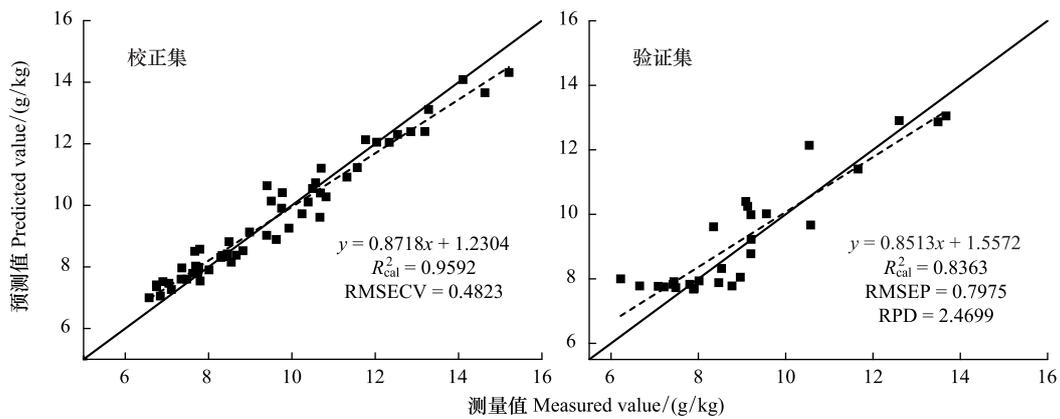


图 5 RF 最优模型精度评价

Fig.5 Accuracy evaluation of optimal RF model

表 4 不同样本选择方法的 SG+MSC 方法模型精度评价

Table 4 Accuracy evaluation of SG+MSC method model with different sample selection methods

建模方法 Modeling methods	样本选择方法 Sample selection methods	校正集 Calibration set		验证集 Validation set		
		R^2_{cal}	RMSECV	R^2_{val}	RMSEP	RPD
siPLS	KS	0.8686	0.7935	0.7762	0.9316	2.1116
	RS	0.8796	0.6950	0.8431	1.7142	2.4989
	SPXY	0.8722	0.8135	0.6623	1.0666	1.6009
GA-SVM	KS	0.9526	0.4839	0.8598	0.9987	2.6017
	RS	0.9837	0.2604	0.7775	1.5389	1.8631
	SPXY	0.9904	0.2284	0.5569	1.7150	1.2919
RF	KS	0.9610	0.4687	0.8559	0.9040	2.2744
	RS	0.9625	0.4253	0.7773	1.4563	1.5582
	SPXY	0.9644	0.4519	0.4444	1.5707	1.1620

siPLS:组合间隔偏最小二乘模型, Synergy interval partial least squares; GA-SVM:遗传算法-支持向量机模型, Genetic Algorithm-Support vector machine; RF:随机森林模型, Random Forest; KS: KS 算法, Kennard-Stone; RS: 随机选择算法, Random selection; SPXY: SPXY 算法, Sample set partitioning based on joint x-y distances

3 讨论

不同光谱预处理方法的选择影响着模型的预测性能,选择合适的预处理方法可以消除噪声、提取有效信息并且提高模型的性能^[37],而不合适的方法则会放大噪声并且使模型性能恶化。本文不同种预测模型的结果中显示,未经过预处理的光谱模型预测效果普遍较好,可能由于测定土壤光谱时,测定的土壤粒度较小(100目)。前期研究表明土壤粒度对反射率的影响较大,并且对土壤模型的预测精度也有很大影响,土壤粒度较大的模型其预测标准误差远高于粒度较小的模型^[38]。Kooistra 等^[39]通过研究莱茵河漫滩土壤预测锌和镉浓度,也得到了未经过预处理的模型对锌和镉的预测效果最好。而二阶导数处理过的光谱模型效果不好的原因可能是二阶导数对光谱进行预处理时,放大了噪声,未能提取较为有效的信息^[40]。但也有许多研究表明,二阶导数处理改进了 SOC 预测模型^[41-42]。因此,对于含生物炭土壤的 SOC 模型构建还仍需进一步的研究。

然而,无论使用何种预处理方法,选择具有代表性的样本集对于模型构建起着重要作用^[43],代表性样本可以提高模型的预测精度,加快模型的模拟预测并且有助于后期的改进^[44]。陈亦云等^[45]研究发现,在有机碳的预测过程中,不同的样本集构建方法对模型的精度有着不同的影响。SPXY 方法是对 KS 方法的改进^[25],但是在本研究结果中模型性能并没有得到提升,可能是由于土壤中影响因素过多,使 SOC 光谱信息被掩盖^[46]。彭杰等^[46]对不同氧化铁和有机质含量的土壤进行研究时,发现当氧化铁与有机质的比例大于 2.21 时,土壤中的氧化铁会完全遮盖有机质光谱特征。而由于 RS 方法选取的样本具有随机性,并不能代表整个样本集,所以构建的模型并不准确。也有学者认为样本集的大小也起着至关重要的作用——当样本量较少时,样本选择方法起着主要作用;当样本量较大时,样本集的大小相较于方法更为关键^[47]。未来可以对不同比例样本集进行研究,提升模型的预测精度。

在本研究表明,土壤光谱最优模型能够较好地预测施用生物炭后土壤 SOC 含量,但土壤光谱数据是基于实验室暗室环境中测量获得的,而在实际野外应用中,影响土壤光谱的环境因素复杂繁多,因此需要考虑各种环境因素对土壤光谱及预测模型的综合影响。

4 结论

本文在比较不同的样本选择方法和不同光谱预处理方法基础上,探究了 siPLS 模型、GA-SVM 模型和 RF 模型预测经生物炭处理后 SOC 的性能差异。施加生物炭增加了 SOC 含量,增加幅度随生物炭添加量的提高呈增加趋势。土壤反射率随 SOC 含量的增加而降低,在 1410、1920 和 2200 nm 附近存在明显的吸收谷。在 3 种样本选择方法中,KS 方法所划分的样本集相对于 RS 方法和 SPXY 方法更适用于生物炭 SOC 模型的建立。在各种不同组合中,以 KS 方法划分的样本集,SG+MSC 作为光谱预处理方法,且 GA-SVM 作为建模方法时,建立的模型预测性能最好。其中校正集的 R_{cal}^2 和 RMSECV 值分别为 0.9526 和 0.4839,验证集的 R_{val}^2 和 RMSEP 值分别为 0.8598 和 0.9987,RPD 值为 2.6017。应用近红外光谱技术可以很好预测施加过生物炭后的 SOC,为以后对经过生物炭处理后的土壤有机碳预测研究提供一定基础。

参考文献 (References):

- [1] 苑亚茹,李娜,邹文秀,尤孟阳,韩晓增,马大龙. 典型黑土区不同生态系统土壤团聚体有机碳分布特征. 生态学报, 2018, 38(17): 6025-6032.
- [2] Baldock J A, Beare M H, Curtin D, Hawke B. Stocks, composition and vulnerability to loss of soil organic carbon predicted using mid-infrared spectroscopy. Soil Research, 2018, 56(5): 468-480.
- [3] 关连珠,周景景,张昀,张广才,张金海,禅忠祥. 不同来源生物炭对砷在土壤中吸附与解吸的影响. 应用生态学报, 2013, 24(10): 2941-2946.
- [4] 刘玉学,刘微,吴伟祥,钟哲科,陈英旭. 土壤生物质炭环境行为与环境效应. 应用生态学报, 2009, 20(4): 977-982.
- [5] Lehmann J. A handful of carbon. Nature, 2007, 447: 143-144.
- [6] Backer R G M, Schwinghamer T D, Whalen J K, Seguin P, Smith D L. Crop yield and SOC responses to biochar application were dependent on

- soil texture and crop type in southern Quebec, Canada. *Journal of Plant Nutrition and Soil Science*, 2016, 179(3): 399-408.
- [7] Liu Z X, Chen X M, Jing Y, Li Q X, Zhang J B, Huang Q R. Effects of biochar amendment on rapeseed and sweet potato yields and water stable aggregate in upland red soil. *CATENA*, 2014, 123: 45-51.
- [8] Zhang L, Yang X M, Drury C, Chantigny M, Gregorich E, Miller J, Bittman S, Reynolds D, Yang J Y. Infrared spectroscopy prediction of organic carbon and total nitrogen in soil and particulate organic matter from diverse Canadian agricultural regions. *Canadian Journal of Soil Science*, 2018, 98(1): 77-90.
- [9] Minu S, Shetty A. Prediction accuracy of soil organic carbon from ground based visible near-infrared reflectance spectroscopy. *Journal of the Indian Society of Remote Sensing*, 2018, 46(5): 697-703.
- [10] McDowell M L, Bruland G L, Deenik J L, Grunwald S, Knox N M. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma*, 2012, 189-190: 312-320.
- [11] Nawar S, Mouazen A M. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. *Soil and Tillage Research*, 2019, 190: 120-127.
- [12] O'Rourke S M, Holden N M. Optical sensing and chemometric analysis of soil organic carbon - a cost effective alternative to conventional laboratory methods? *Soil Use and Management*, 2011, 27(2): 143-155.
- [13] 章文龙, 曾从盛, 高灯州, 陈晓艳, 林伟. 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.
- [14] 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.
- [15] Gupta A, Vasava H B, Das B S, Choubey A K. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. *Geoderma*, 2018, 325: 59-71.
- [16] 崔霞, 宋清洁, 张瑶瑶, 胥刚, 孟宝平, 高金龙. 基于高光谱数据的高寒草地土壤有机碳预测模型研究. *草业学报*, 2017, 26(10): 20-29.
- [17] Van De Broek M, Govers M. Quantification of organic carbon concentrations and stocks of tidal marsh sediments via mid-infrared spectroscopy. *Geoderma*, 2019, 337: 555-564.
- [18] Liu J B, Han J C, Zhang Y, Wang H Y, Kong H, Shi L. Prediction of soil organic carbon with different parent materials development using visible-near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2018, 204: 33-39.
- [19] 刘彦姝, 潘勇. 基于 SVR 算法的林地土壤氮含量高光谱测定. *生态科学*, 2013, 32(1): 84-89.
- [20] Bushong J T, Norman R J, Slaton N A. Near-infrared reflectance spectroscopy as a method for determining organic carbon concentrations in soil. *Communications in Soil Science and Plant Analysis*, 2015, 46(14): 1791-1801.
- [21] Nawar S, Mouazen A M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Computers and Electronics in Agriculture*, 2018, 151: 469-477.
- [22] 刘华, 张利权. 崇明东滩盐沼土壤重金属含量的高光谱估算模型. *生态学报*, 2007, 27(8): 3427-3434.
- [23] Saptoro A, Tadé M O, Vuthaluru H. A modified Kennard-stone algorithm for optimal division of data for developing artificial neural network models. *Chemical Product and Process Modeling*, 2012, 7(1), doi: 10.1515/1934-2659.1645.
- [24] Kennard R W, Stone L A. Computer aided design of experiments. *Technometrics*, 1969, 11(1): 137-148.
- [25] Galvão R K H, Araujo M C U, José G E, Pontes M J C, Silva E C, Saldanha T C B. A method for calibration and validation subset partitioning. *Talanta*, 2005, 67(4): 736-740.
- [26] Guo Z M, Huang W Q, Chen L P, Wang X, Peng Y K. Nondestructive evaluation of soluble solid content in strawberry by near infrared spectroscopy//*Proceedings of International Conference on Photonics and Image in Agriculture Engineering*. Sanya: SPIE, 2013.
- [27] Nørgaard L, Saudland A, Wagner J, Nielsen J P, Munck L, Engelsen S B. Interval Partial Least-Squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy*, 2000, 54(3): 413-419.
- [28] Wang B, Waters C, Orgill S, Cowie A, Clark A, Liu D L, Simpson M, McGowen I, Sides T. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecological Indicators*, 2018, 88: 425-438.
- [29] Ines A V M, Mohanty B P. Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithm: 1. Conceptual modeling. *Water Resources Research*, 2008, 44(6): W06422.
- [30] Saini L M, Aggarwal S K, Kumar A. Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market. *IET Generation, Transmission & Distribution*, 2010, 4(1): 36-49.
- [31] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5-32.
- [32] 廖敏, 彭英, 陈义, 谢晓梅, 吴春艳, 唐旭, 刘玉学, 杨生茂. 长期不同施肥管理对稻田土壤有机碳库特征的影响. *水土保持学报*, 2011, 25(6): 129-133, 138-138.
- [33] 曾爱, 廖允成, 张俊丽, 眭彦伟, 温晓霞. 生物炭对壤土土壤含水量、有机碳及速效养分含量的影响. *农业环境科学学报*, 2013, 32(5): 1009-1015.

- [34] 史舟, 王乾龙, 彭杰, 纪文君, 刘焕军, 李曦, Viscarra Rossel R A. 中国主要土壤高光谱反射特性分类与有机质光谱预测模型. 中国科学: 地球科学, 2014, 44(5): 978-988.
- [35] Islam K, Singh B, McBratney A. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 2003, 41(6): 1101-1114.
- [36] Liao K H, Xu S H, Wu J C, Zhu Q, An L S. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. Journal of Plant Nutrition and Soil Science, 2014, 177(5): 775-782.
- [37] Dotto A C, Dalmolin R S D, Caten A T, Grunwald S. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma, 2018, 314: 262-274.
- [38] 李硕. 基于 Vis/NIR 光谱不同粒径下土壤碳氮的预测研究[D]. 武汉: 华中农业大学, 2010.
- [39] Kooistra L, Wehrens R, Leuven R S E W, Buydens L M C. Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains. Analytica Chimica Acta, 2001, 446(1/2): 97-105.
- [40] Peng X T, Shi T Z, Song A H, Chen Y Y, Gao W X. Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods. Remote Sensing, 2014, 6(4): 2699-2717.
- [41] Vasques G M, Grunwald S, Sickman J O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma, 2008, 146(1/2): 14-25.
- [42] 刘晓旭. 基于不同预处理方法的小麦叶片氮素含量的高光谱估测[D]. 泰安: 山东农业大学, 2018.
- [43] Lucà F, Conforti M, Castrignanò A, Matteucci G, Buttafuoco G. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma, 2017, 288: 175-183.
- [44] Lin Z D, Wang Y B, Wang R J, Wang L S, Lu C P, Zhang Z Y, Song L T, Liu Y. Improvements of the Vis-NIRS model in the prediction of soil organic matter content using spectral pretreatments, sample selection, and wavelength optimization. Journal of Applied Spectroscopy, 2017, 84(3): 529-534.
- [45] 陈奕云, 齐天赐, 黄颖菁, 万远, 赵瑞瑛, 亓林, 张超, 费腾. 土壤有机质含量可见-近红外光谱反演模型校正集优选方法. 农业工程学报, 2017, 33(6): 107-114.
- [46] 彭杰, 李曦, 周清, 史舟, 纪文君, 王家强. 氧化铁对有机质光谱特性的影响分析. 遥感学报, 2013, 17(6): 1396-1412.
- [47] Ramirez-Lopez L, Schmidt K, Behrens T, Van Wesemael B, Demattè J A M, Scholten T. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma, 2014, 226-227: 140-150.