

基于土壤协变量与 VIS-NIR 光谱估算土壤有机质含量的研究

马国林^{1,2,3}, 丁建丽^{1,2,3*}, 张子鹏^{1,2,3}

¹新疆大学资源与环境科学学院, 新疆 乌鲁木齐 830046;

²新疆大学绿洲生态教育部重点实验室, 新疆 乌鲁木齐 830046;

³新疆大学智慧城市与环境建模自治区普通高校重点实验室, 新疆 乌鲁木齐 830046

摘要 为研究土壤有机质(SOM)含量与土壤电导率(EC), pH 和 Fe 的相关关系, 立足于艾比湖保护区, 在 2017 年 8 月共收集了 110 个样本, 测量了土壤反射光谱、SOM 含量、土壤协变量(EC, Fe, pH)。对原始光谱进行了三种预处理: SG(Savitzky-Golay)平滑、多元散射校正(MSC)和一阶微分(FD), 并对光谱数据进行了主成分分析(PCA), 选取前 5 个主成分(PC)的特征值作为光谱变量。以使用原始光谱数据、两种预处理方法(SG-MS-C、SG-MS-C-FD)作为策略 I, 以土壤协变量(EC, Fe, pH)为预测变量作为策略 II, 以策略 I 和策略 II 组合作为策略 III, 分别利用偏最小二乘回归(PLSR)建立 SOM 的预测模型。结果表明, 基于预处理后的光谱数据的预测效果(验证集中决定系数为 $R^2=0.66\sim 0.82$) 优于以土壤协变量为预测变量的预测效果(验证集中 $R^2=0.40$), 此外将土壤协变量与光谱数据相结合可以明显改善 SOM 的光谱预测精度(最佳验证集中 $R^2=0.88$)。同时, 对光谱数据进行预处理后, 能够有效增强潜在的光谱信息, 提高模型的预测精度。综上, 将可见光-近红外光谱信息和土壤协变量相结合的策略能够有效提升 SOM 模型的预测性能。

关键词 遥感; 高光谱; 有机质; 主成分分析; 土壤协变量; 偏最小二乘

中图分类号 O433

文献标志码 A

doi: 10.3788/LOP57.192801

Soil Organic Matter Content Estimation Based on Soil Covariate and VIS-NIR Spectroscopy

Ma Guolin^{1,2,3}, Ding Jianli^{1,2,3*}, Zhang Zipeng^{1,2,3}

¹College of Resources & Environmental Science, Xinjiang University, Urumqi, Xinjiang 830046, China;

²Key Laboratory of Oasis Ecology, Ministry of Education, Xinjiang University, Urumqi, Xinjiang 830046, China;

³Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi, Xinjiang 830046, China

Abstract To investigate the relationship of the soil organic matter (SOM) content to the electrical conductivity (EC), pH, and Fe content, we collected 110 samples at the Ebinur Lake Reserve in August 2017 and measured the soil reflectance spectra, SOM content, EC, Fe content, and pH. We performed three kinds of pre-treatments, including Savitzky-Golay (SG) smoothing, multiplicative scatter correction (MSC), and first-order differentiation (FD), on the original spectrum and then performed a principal-component analysis of the spectral data. The eigenvalues of the first five principal components were selected as the spectral variables. Strategy I used the original spectrum, performed SG-MS-C and SG-MS-C-FD on it, and employed the original spectrum as a control group. Strategy II used the soil covariates (EC, Fe, pH) as the input variables. Strategy III combined strategy I and strategy II. Predictions of the SOM content were obtained for all three strategies using partial least squares regression. The results show that predictions based on the pre-processed spectral data (for the verification set, the

收稿日期: 2020-01-08; 修回日期: 2020-01-15; 录用日期: 2020-02-10

基金项目: 国家自然科学基金(41961059)

* E-mail: watarid@xju.edu.cn

coefficient of determination was $R^2 = 0.66-0.82$) were better than those based on the soil covariates as the prediction variables (for the verification set, the coefficient of determination was $R^2 = 0.40$) and that combining the soil covariates and spectral data significantly improved the spectral-prediction accuracy for SOM (for the best verification set, $R^2 = 0.88$). Pre-processing the spectral data effectively enhanced the potential spectral information and improved the predictive accuracy of the model. In summary, the combination of visible-near-infrared spectral information and soil covariates effectively improves the predictive performance of SOM models.

Key words remote sensing; hyperspectral images; organic matter; principal component analysis; soil auxiliary information; partial least squares

OCIS codes 280.4750; 300.6170; 200.4260

1 引言

土壤有机质(SOM)是土壤的重要组成部分,对土壤物理性质的改善具有重要作用,对当地生态系统的稳定与安全具有基础性的作用,对干旱区脆弱的生态系统的作用更是举足轻重^[1-2]。传统的 SOM 测定方法费时费力,对土壤具有破坏性,很难满足社会经济发展的需求。可见光和近红外光谱(VIS-NIR)技术结合化学计量学,具有快速、可重复、可低成本监测土壤属性(阳离子交换、SOM、pH、铁氧化物类型等)的优点,是实现土壤监测的重要手段^[3]。因此,研究利用高光谱分析技术定量估算 SOM 具有重要的实践意义。

目前,国内外学者利用光谱技术结合辅助变量对土壤属性的监测进行了相关研究,发现 SOM 与 Fe、土壤电导率(EC)、pH 值存在一定程度的相关性^[4-5]。Dorau 等^[6-7]研究表明:土壤中的 Fe 和 SOM 相互作用,具有极强的吸附性,对 SOM 的稳定具有关键作用;土壤 pH 是影响 SOM 分解转化重要因子,干旱区微咸地下水使土壤盐分不断积累,进而使 pH 值升高,抑制土壤矿化,导致土壤化学性质趋向恶化,对土壤肥力产生胁迫^[8-9]。Radim 等^[10]认为将其他辅助变量添加到光谱数据中时,减少光谱预测变量的数量对改善预测具有重要意义。Fernandes 等^[11]以圣保罗州为研究区,运用高光谱技术估算 SOM 时加入了 pH、钙(Ca)、镁(Mg)等辅助变量,利用人工神经网络(ANN)使建模精度得到了提升。Kweon 等^[12]采集了美国 6 个州的土壤样本,并用光谱技术估算土壤有机质,文中用 EC 作为辅助变量,利用多元线性回归模型(MLR)建模,取得了良好的效果,决定系数 R^2 为 0.80 或更高。Siebielec 等^[13]利用高光谱估算土壤有机碳(SOC)时发现,在有效利用光谱信息的基础上加入数字地形高程(DEM)为辅助变量,相对于未加入辅助变量之前,验证集精度的 R^2 提高到了 0.84。以上研究大都是在 SOM 相对较高的地区,而在 SOM 含量

(质量比,下同)低的干旱区,其光谱特性容易受到土壤其他特性(例如 EC、Fe、pH 和阳离子交换容量等)的遮蔽^[14-15],使得 VIS-NIR 技术不能有效地解释 SOM 的空间变化。目前,针对干旱区 SOM 的研究大都是对光谱采用各种方法和技术进行变换和筛选,已实现对 SOM 的估算,但尚未有利用光谱技术结合其他土壤属性来估算 SOM 的报道。

基于此,本文以艾比湖保护区为研究靶区,在利用高光谱技术估算 SOM 的同时加入土壤 EC、pH 和 Fe 作为辅助变量,探索添加到 VIS-NIR 光谱模型中的土壤辅助信息的潜力,提高 SOM 估算模型的准确性。利用偏最小二乘回归(PLSR)建立 SOM 高光谱的预测模型,同时比较原始光谱经过不同预处理方法后的建模效果,探索反演 SOM 的优选方法,以期对干旱区 SOM 的快速监测提供参考。

2 实验数据与方法流程

2.1 研究区概况

艾比湖湿地位于新疆维吾尔自治区西北部,地理位置为 $44^{\circ}43' \sim 45^{\circ}08'N$ 、 $82^{\circ}35' \sim 83^{\circ}10'E$,见图 1。

2.2 土壤样品采集

土壤样品的采样时间为 2017 年 8 月,根据艾比湖湿地的生态景观特征,以及先前的野外经验,选取有代表性的绿洲、荒漠及交错带景观,布设 110 个采样地,以梅花采样法采集 5 个点作为样本,采样半径为 5 m,将 5 个样本混合后作为土壤样品,采样深度为 0~10 cm,利用手持 GPS 记录采样点位置。各采样点的样品用自封袋取回。在实验室进行自然风干、剔除异物、研磨过筛(>0.15 mm)等处理。完成处理后,将土壤样品分为两份,一份用于室内采集光谱数据,另一份用于土壤属性(SOM、EC、pH 和 Fe)的测定。土壤光谱的测定采用美国 ASD (Analytical Spectral Devices) 公司生产的 FieldSpec3 型光谱仪测定,其波长范围为 350~2500 nm,其中 350~1000 nm 波长的采样间隔

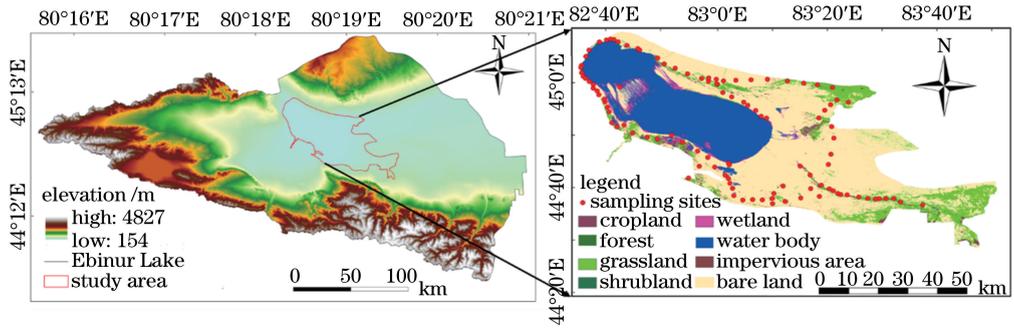


图1 研究区及采样点分布图

Fig. 1 Distribution of the study area and sampling points

为 1.4 nm, 1000 ~ 2500 nm 的波长采样间隔为 2 nm, 重采样间隔为 1.4 nm。土壤光谱的测定在暗室中进行, 选用 50 W 的卤素灯, 探头的视场角为 5°, 探头距离土壤表面 30 cm, 每个土壤样本测定前均进行白板定标, 每个土壤样品重复测定 10 次, 取算术平均值作为该样品的反射率。SOM 基于重铬酸钾进行外加热法测定^[16]。EC 采用德国 WTW 公司制造的 Cond7310 电导率仪以水土比 5 : 1 浸提, 测定浸提液的电导率。pH 采用离子选择性电极法 (ISE) 测定。Fe 元素采用 Analytick Jena 公司制造的 contrAA 测定。

2.3 数据预处理

对原始土壤的光谱数据, 去除边缘噪声较大的 350 ~ 400 nm 和 2401 ~ 2500 nm 波段的光谱^[17]。采用主成分分析 (PCA) 结合马氏距离 (Mahalanobis distance) 剔除光谱和 SOM 样本的离群点, 以降低异常值对研究的遮蔽影响^[18-19]。从 110 个样本点中确定 101 个有效样本进行后续研究, 图 2 所示为 PCA-马氏距离样本分布图。光谱采集过程中容易受到仪器噪声、土壤颗粒分布不均匀及测量随机误差等干扰, 使所测样品光谱中含有光谱噪声, 影响预测模型精度, 因此本文采用 SG (Savitzky-Golay, SG) 平滑、多元散射校正 (MSC) 和一阶微分 (FD) 对原始光谱依次进行预处理。

SG 平滑是 Savitzky 和 Golay^[20] 在 1964 年提出的, 该方法可以有效消除噪声的影响, 且不会改变光谱的位置。本次平滑采用窗口数为 7 的三次多项式拟合对原始光谱进行处理。从图中可以明显观察到, 相较于原始光谱, 经过 SG 平滑处理之后, 去除了原始光谱中存在的毛刺现象, 增强了光谱的信噪比, 得到的光谱更加光滑平稳 (图 3)。MSC 由 Isaksson 和 Naes 于 1988 年提出, 是一种多变量散射校正技术, 可以有效消除光谱散射的影响^[21]。光

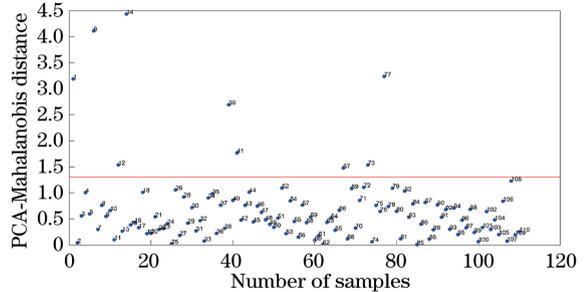


图2 PCA-马氏距离分布图

Fig. 2 PCA-Mahalanobis distance distribution

谱通过 MSC 处理后, 光谱的峰、谷、跃迁等这些细节特征更加明显 (图 3)。微分技术 (Derivative) 在土壤 VIS-NIR 光谱分析中应用广泛, 能够较好地去除基线漂移和平缓背景干扰的影响, 同时能够挖掘光谱中的隐藏信息^[14]。从图中可以看出, 经过 FD 处理后的光谱特征更加明显, 凸显了光谱细微信息差异引起的光谱变化 (图 3)。

2.4 光谱数据特征选取

PCA 是数据压缩和特征提取的基本方法, 可以处理大型、复杂的数据集, 能够揭示光谱变量的信息模式和内部结构, 其利用输入变量间的线性关系, 用少部分互不相关的主成分 (PCs) 来描述多维变量空间的绝大多数信息^[22-23]。光谱数据中存在大量的重叠吸收和自相关现象, 因此在本研究中利用 SPSS 软件对原始光谱和两种预处理后的光谱数据进行 PCA 处理。在原始光谱中, 第一个变量的贡献率为 57.78%, 前 5 个变量的累计贡献率为 98.08%。经过 SG-MSC 处理后, 第一个变量的贡献率是 61.44%, 前 5 个变量的累计贡献率为 98.16%。而经过 SG-MSC-FD 处理之后, 第一个变量的贡献率是 73.29%, 第二个变量的贡献率是 10.26%, 前 5 个表变量的累计贡献率是 91.16%。原始光谱与预处理的光谱经过 PCA 处理后, 前 5 个

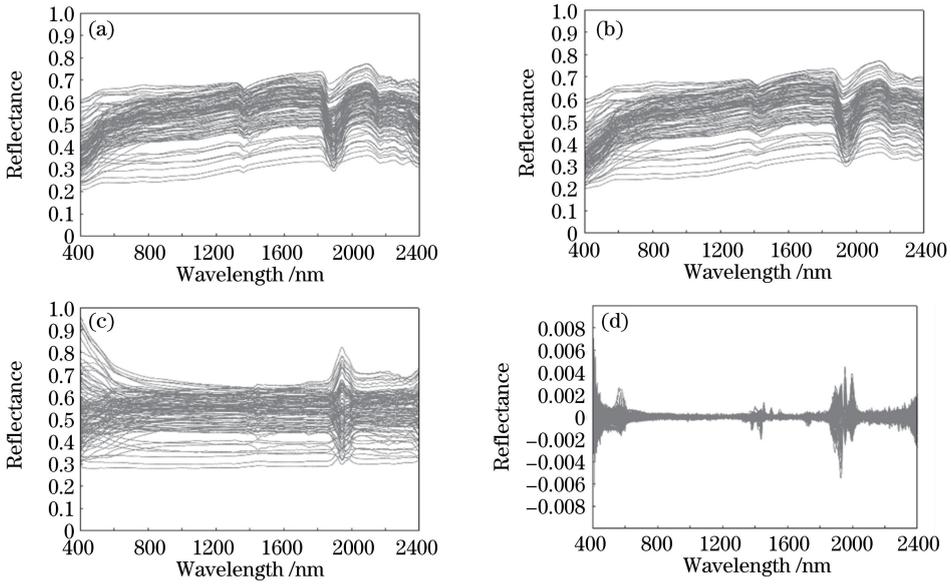


图3 原始光谱与预处理后土壤光谱反射率曲线。(a)原始光谱反射率;(b)经SG平滑处理后的光谱反射率;(c)经MSC处理后的光谱反射率;(d)经FD处理后的光谱反射率

Fig. 3 Original spectra and the pretreated soil spectral reflectance curves. (a) Original spectral reflectance; (b) spectral reflectance after SG smoothing; (c) spectral reflectance corrected for multiple scattering; (d) spectral reflectance treated with first order differentiation

变量的累计贡献率均达到90%以上,因此各选取前5个变量进行下一步分析。图4展示了原始光谱与经两种预处理后前10个变量的贡献率和累计贡献率。

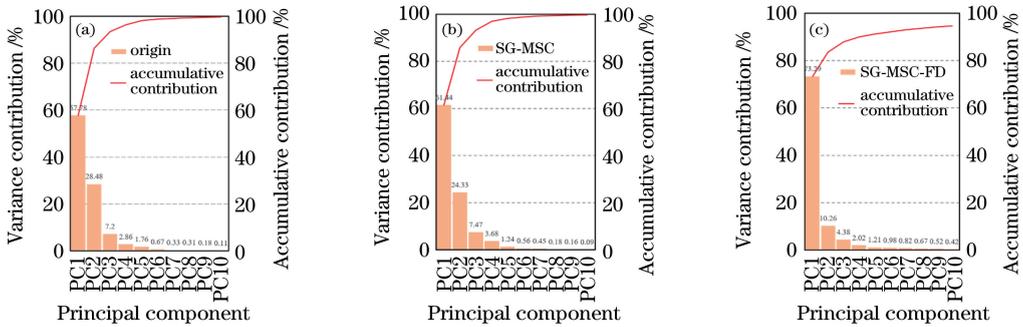


图4 前10个变量贡献图。(a)原始光谱反射率;(b)经SG-MSC处理后的光谱反射率;(c)经SG-MSC-FD处理后的光谱反射率

Fig. 4 Contribution diagram of the first 10 variables. (a) Original spectral reflectance; (b) spectral reflectance after SG-MSC treatments; (c) spectral reflectance after SG-MSC-FD treatments

2.5 精度评价

PLSR广泛应用于SOM的高光谱监测,其集成了PCA、典型相关分析(COA)和MLR三种方法的优点^[24]。其优势在于采用降维思想去除冗余噪声,精简光谱自变量,可以较好地处理强共线性和变量数量超过可用样本数量的情况^[25]。同时还考虑了因变量和自变量(SOM)之间的关系,能够有效筛选对SOM解释性最强的综合变量,使得建立的模型具有更好的鲁棒性,避免过拟合,为多元统计分析提供了极大便利^[26]。

采用分层策略将整个SOM数据集划分为建模集和验证集,它是一种有效的样本划分方法,可用于对代表性样本进行建模^[16]。具体来说,就是根据SOM含量从小到大对101个样本进行分类,然后将101个样本划分为33层,每层包含3个样本,将第二个样本划分为验证集,由于本次样本总量不能被3整除,故将其他两个样本划分为建模集,因此建模数据集包含68个样本,验证数据集包含33个样本。为了量化基于PLSR建模的效果和性能,选用决定系数(R^2)、均方根误差(RMSE)、相对分析误差

(RPD, e_{RPD}) 这 3 个参数对模型进行评估。 R^2 越大, 表明建模的精度越高。 RMSE 表示模型的预测能力, 其值的大小与模型的精度成反比。 RPD 广泛用于评估土壤属性预测模型的准确性, $e_{RPD} \geq 2.5$ 表示极佳, $2.0 \leq e_{RPD} < 2.5$ 表示良好, $1.8 \leq e_{RPD} < 2.0$ 表示一般, $1.4 \leq e_{RPD} < 1.8$ 表示预测效果可以接受, 但需要进一步改进, $e_{RPD} < 1.4$ 表示模型效果差, 无法对样本实测值进行预测^[27]。

为了测试两种预处理方法, 测试加入土壤辅助信息(即 EC, Fe 和 pH)对模型预测的影响, 将原始光谱数据作为对照组, 首先分别使用两种预处理方法(SG-MSC, SG-MSC-FD)作为策略 I, 然后以土壤辅助信息为预测变量作为策略 II, 最后将策略 I 和策略 II 组合作为策略 III, 比较原始光谱、采用两种预处理方法以及加入土壤辅助信息时模型的预测效果。图 5 演示了用于 SOM 估算的不同输入变量的示意图。

3 结果与分析

3.1 有机质和其他土壤属性的统计特征

表 1 所示为 SOM 与其他土壤属性的描述性统

表 1 土壤属性统计特征

Table 1 Statistical characteristics of soil properties

Property	Dataset(CV/%)	<i>n</i>	Min	Mean	Max	Std
Content of SOM /($g \cdot kg^{-1}$)	Whole(57.91)	101	0.60	8.94	23.00	5.18
	Calibration(57.33)	68	0.60	8.90	23.00	5.10
	Validation(59.92)	33	1.3	9.02	21.72	5.41
EC /($dS \cdot cm^{-1}$)	Whole(86.52)	101	0.05	6.54	28.40	5.66
Content of Fe /($g \cdot kg^{-1}$)	Whole(56.93)	101	0.10	13.18	25.91	7.67
pH	Whole(4.12)	101	8.22	8.87	9.93	0.37

3.2 估算机制

如图 6 所示, SOM 与 EC 呈正相关(相关系数 $r=0.63$, 概率 $p < 0.001$), SOM 与 pH 也呈正相关($r=0.56$), 这表明 EC 和 pH 可能是估算土壤 SOM 的重要参数。结果表明, SOM 与 Fe 之间的相关性很弱($r=0.18$), 没有通过显著性检验, 这与 Hong 等的研究结果相近^[28]。pH 与 EC 之间没有显著相关($r=0.12$), 未通过显著性检验, 而 Fe 与 EC($r=0.58, p < 0.001$) 和 pH($r=0.69, p < 0.001$) 通过了显著性检验, 表明 Fe 与 EC 和 pH 的起源不同。

SOM 的相关系数曲线也显示出与 EC 大致相似的趋势(图 7), 这一发现进一步强调了 EC 在 SOM 光谱估算中的重要性, pH 的相关峰或谷也与 SOM 一致, 由此表明在建模的过程中不能忽略 pH 的作用。由于这些内部联系, 可以间接建立土壤

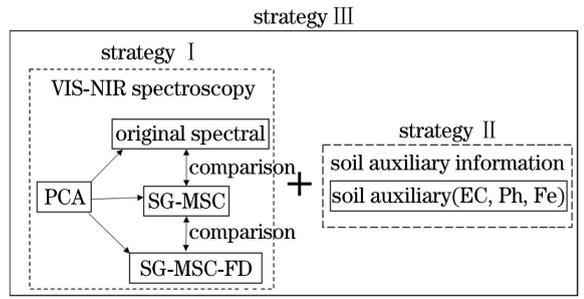


图 5 SOM 预测变量比较框架图

计特征。由表 1 可知: 研究区的土壤样本的 SOM 含量为 $0.60 \sim 23.00 g \cdot kg^{-1}$, 平均含量为 $8.94 g \cdot kg^{-1}$, 标准差(Std)为 5.12, 变异系数(CV)为 57.91%, 属于中等变异。同时, EC 为 $0.0534 \sim 28.4 dS \cdot cm^{-1}$, 均值为 $6.54 dS \cdot cm^{-1}$, 变异系数为 86.52%, 属于中等变异。Fe 的含量为 $0.10 \sim 25.91 g \cdot kg^{-1}$, 均值为 $6.54 g \cdot kg^{-1}$, 变异系数为 56.93%, 属于中等变异。而 pH 为 $8.22 \sim 9.93$, 均值为 8.74, 变异系数为 4.12%, 属于弱变异, 这表明该研究区的土壤都呈碱性。

SOM 的光谱估算模型。

3.3 有机质与光谱主成分之间的关系

图 8 展示了 SOM 与原始的, 经 SG-MSC 和 SG-MSC-FD 预处理之后的两种光谱的前 5 个主成分之间的相关系数。由图 8 可知: 对于原始光谱, 与 SOM 相关系数的最大值为 0.375(PC1); 经过 SG 平滑和 MSC 预处理之后, 相关系数的最大值为 0.398(PC3); 经过 SG 平滑、MSC 和 FD 预处理之后, 相关系数的最大值为 0.436(PC3)。由此可知, 光谱经过适当的预处理, 能够改善与有机质相关的光谱特征。

3.4 建模评价

对原始光谱以及 SG-MSC 和 SG-MSC-FD 处理后的光谱进行 PCA, 分别选取主成分后的前 5 个变量, 在上述三种策略下进行 PLSR 建模, 三种策略下 PLSR 模型结果如表 2 所示。

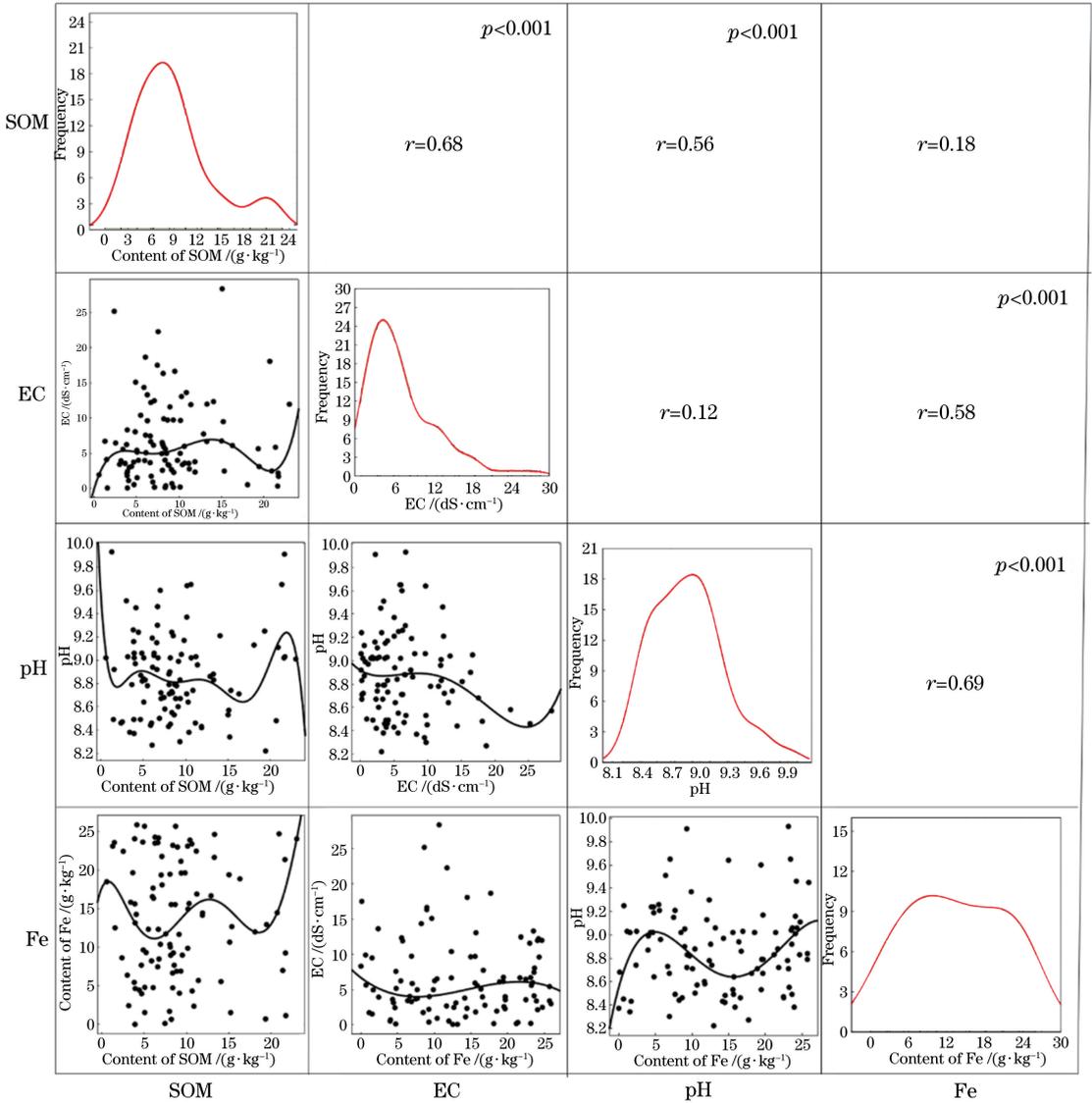


图 6 不同土壤参数之间的相关性($n=101$),图中曲线为拟合曲线

Fig. 6 Correlation between different soil parameters ($n=101$), in which the curves are fitting curves

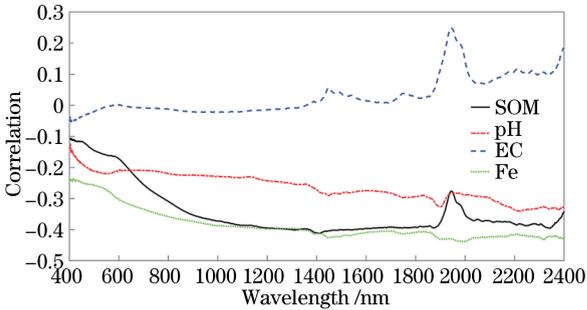


图 7 SOM,EC,Fe 和 pH 与原始光谱反射率之间的相关性($n=101$)

Fig. 7 Correlation between SOM, EC, Fe and pH and original spectral reflectance ($n=101$)

总共评估了 7 个 PLSR 模型,其中 SOM 为目标变量,光谱和土壤其他属性(EC,Fe,pH)为输入

变量,如表 2 所示。利用独立的验证数据集来验证模型的预测性能。为了便于陈述,对模型进行了编号,如表 2 所示。总体而言,校正集和验证集之间的拟合统计量(即 R^2 和 RMSE)相似,这表明已建立的模型具有稳定性(表 2)。在策略 I 中,随着对光谱预处理的增强,模型的精度得到显著提升,验证集的 R^2 从 0.66 增加到 0.82, RMSE 由 3.12 降低到 2.51, RPD 从 1.73 增加到 2.15,这表明对光谱进行相应的处理能够显著提升模型的预测性能。在策略 II 中,使用其他土壤属性信息(EC,Fe 和 Fe)作为输入变量时,建模结果表明验证集和校正集的 R^2 分别为 0.44 和 0.40, RMSE 的值为 4.19, RPD 仅为 1.21,这表明只用其他土壤属性信息建立预测模型的方法不可行。与策略 I 和 II 相比,策略 III 的

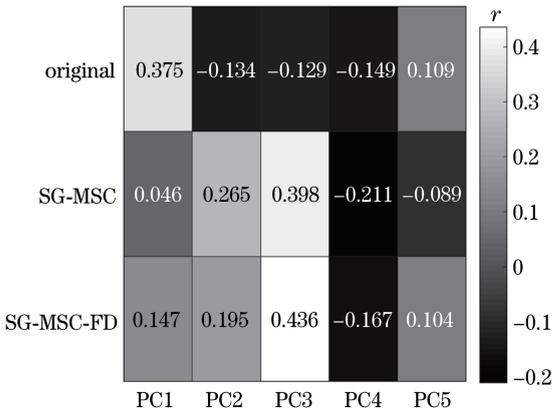


图8 原始光谱以及 SG-MSC,SG-MSC-FD 两种光谱处理下 SOM 与前 5 个主成分相关性

Fig. 8 Correlation between soil organic matter and the first five principal components for original spectrum and preprocessed spectra under two spectral treatments of SG-MSC and SG-MSC-FD

表2 三种策略下 PLSR 建模结果

Table 2 PLSR modeling results under three strategies

Strategy	Variable	Calibration set			Validation set		
		Model number	R^2	RMSE	R^2	RMSE	RPD
Strategy I	Original spectrum	1	0.69	3.42	0.66	3.12	1.73
Strategy I	SG-MSC	2	0.69	3.35	0.67	3.10	1.76
	SG-MSC-FD	3	0.84	2.14	0.82	2.51	2.15
Strategy II	Soil auxiliary covariates	4	0.44	4.19	0.40	4.46	1.21
Strategy III	Original spectrum combined with soil auxiliary covariates	5	0.75	3.09	0.67	3.10	1.74
Strategy III	SG-MSC combined with soil auxiliary covariates	6	0.86	1.91	0.83	2.54	2.13
	SG-MSC-FD combined with soil auxiliary covariates	7	0.91	1.54	0.88	1.20	2.70

4 讨 论

由于测量环境、测量方法、测量仪器等的影响,土壤光谱的反射率数据不可避免地存在噪声^[29]。此外,高光谱数据由于具有波段多、数据量大、数据冗余的特点,增加了数据处理与建模的工作量和复杂度,而对高光谱数据进行适当地去噪和降维能够有效地改善光谱数据的质量^[30-31]。基于策略 I 建立的 PLSR 模型的验证精度 R^2 达到了 0.82(表 2),这与沈兰芝等^[32]的研究结果相近。

对比策略 I 和策略 III(表 2)得出的模型结果,表明 SOM 中的光谱包含很多的土壤信息,对光谱数据经过一定预处理和信息提取能够提升模型的性能。验证集的 R^2 增加了 0.2。模型精度的提升主要归因于 SOM 与土壤辅助变量之间的密切关系(图 6)。当使用光谱数据和土壤辅助变量进行建模

PLSR 模型明显提高了模型在估算 SOM 方面的性能。其中模型 7 提供了最好的预测效果,验证集的 R^2 为 0.88, RMSE 为 1.20,建模效果极佳,优于所有其他 6 个模型。策略 I 和策略 III 的 R^2 比较结果还表明,经 SG-MSC-FD 处理后的预测效果优于 SG-MSC,且两者的预测效果均优于原始光谱。PLSR 模型的拟合散点图如图 9 所示。

基于不同输入源的 SOM 估算变量重要性(VIP)值,如图 10 所示(这里仅展示了模型 3、4 和 7)。对于模型 3[图 10(a)],5 个主成分的 VIP 值最大的是 PC3(2.40),其次是 PC1,PC2,PC4,PC5。对于模型 4[图 10(b)],EC 对模型的贡献值最大,pH 和 Fe 次之。对于 PCA 和土壤辅助变量[模型 7,图 10(c)],EC 是 PLSR 模型最重要的输入变量,其次是 PC3,PC1。总体而言,在模型 3、4、5 中,土壤辅助变量的重要性顺序(即 EC,pH,Fe)并未根据其相对贡献而改变。

时,后者的预测效果(验证集 $R^2=0.40$)明显低于前者的预测效果(验证集 R^2 为 0.66~0.82)。这一发现表明,尽管以土壤辅助变量建立模型的效果劣于 PCA 模型,但也突出了 SOM 光谱估算中加入辅助变量的重要性。关于土壤辅助变量对估算 SOM 的相对重要性,EC 始终显示为最重要的预测因子,其次是 pH 和 Fe。由于土壤辅助变量的测量需要大量的时间和劳动力,因此一旦光谱估算模型成功建立,便可以仅利用光谱数据来估算其他样本,节省了时间和劳动力^[33-34]。

在本研究中,尽管反射光谱法在估算 SOM 方面表现出良好的模型性能,但仍有 12% 的有机质变化信息未被捕获(根据表 2 中的最佳估算模型)。土壤通常被认为是一个多因素问题,而土壤光谱是土壤属性的全面反映,不仅包含 EC,pH,Fe 的信息,还包括其他土壤性质的信息,例如质地和黏土矿

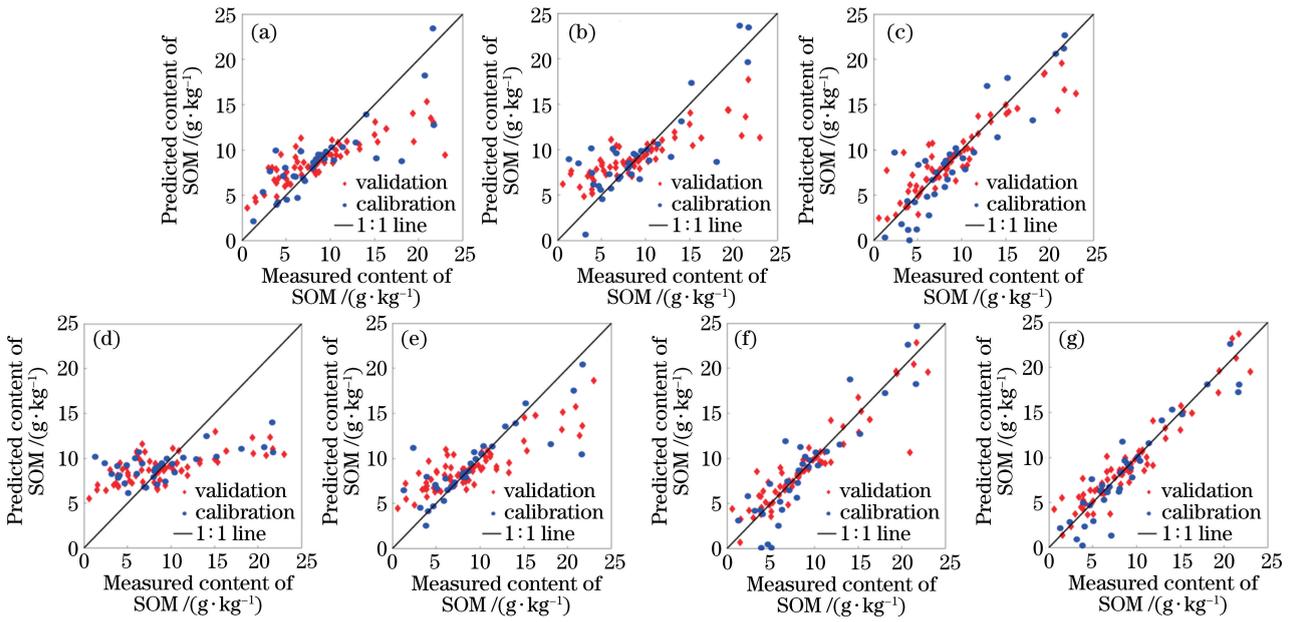


图 9 三种策略下 PLSR 模型的拟合散点图。(a)模型 1;(b)模型 2;(c)模型 3;(d)模型 4;(e)模型 5;(f)模型 6;(g)模型 7
 Fig. 9 Fitting scatter diagrams of PLSR model under three strategies. (a) Model 1; (b) model 2; (c) model 3;
 (d) model 4; (e) model 5; (f) model 6; (g) model 7

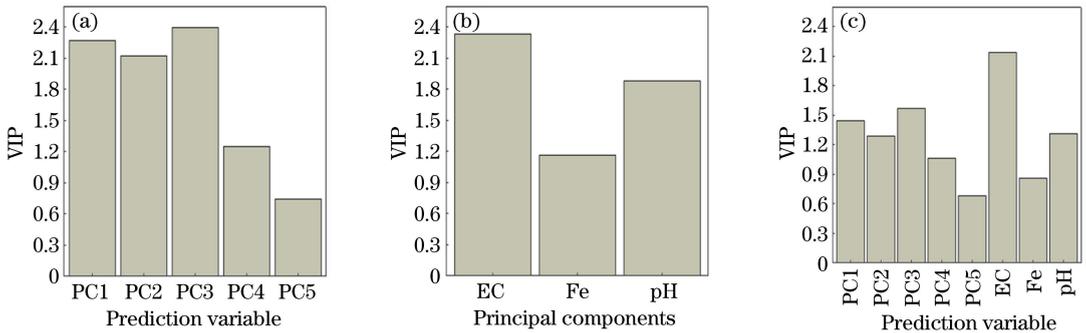


图 10 在不同 PLSR 模型中预测变量的 VIP 值。(a)模型 3;(b)模型 4;(c)模型 5

Fig. 10 VIP values of prediction variables in different PLSR models. (a) Model 3; (b) model 4; (c) model 5

物^[35]。同时,不同地区的土壤具有独特的区域性和地域性。故以艾比湖保护区的土壤为供试样本,采用高光谱技术,在室内进行样品的光谱采集与分析,加入其他土壤属性为辅助变量建立的反演模型对其他地区是否适用,仍有待进一步深入分析和验证。

土壤的属性因其包含有丰富的土壤信息,须使用多个分析工具或多个测量平台同时检测和测量相同的样品来获得^[36]。随着传感器技术的不断进步,便携式 X 射线荧光(pXRF)光谱可快速有效地进行土壤属性的实地测量^[37]。同时,多传感器数据融合是一个发展的方向,可以融合土壤变化的不同方面^[38]。反射光谱法和 pXRF 技术的联合使用具有提高估算 SOM 准确性的发展潜力。

5 结 论

结合土壤 VIS-NIR 光谱和辅助变量信息,估算艾比湖保护区 SOM 的含量时发现,与其他光谱活性成分相似(例如 EC),反射光谱随 SOM 含量的增加而降低。光谱数据的预处理对模型的改善具有重要的作用,随着对光谱预处理的增加,模型精度得到明显提升, R^2 由 0.66 提高到了 0.82, RMSE 逐渐降低(表 2)。通过提出的结合光谱数据和土壤辅助变量的策略(验证集 R^2 , 0.67~0.88), EC 被确定为重要的预测变量。因此可以通过增加土壤辅助变量作为辅助信息的来源改善 SOM 的光谱估计,但是由于光谱技术不需要其他辅助信息,故光谱技术在估算 SOM 方面仍然是值得推荐的。

参 考 文 献

- [1] Alex M B, Damien J F, Andrea K. The dimensions of soil security[J]. *Geoderma*, 2014, 213: 203-213.
- [2] Ding J L, Yu D L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments [J]. *Geoderma*, 2014, 235/236: 316-322.
- [3] Gholizadeh A, Saberioon M, Ben-Dor E, et al. Monitoring of selected soil contaminants using proximal and remote sensing techniques: background, state-of-the-art and future perspectives [J]. *Critical Reviews in Environmental Science and Technology*, 2018, 48(3): 243-278.
- [4] Chen T, Chang Q R, Clevers J G P W, et al. Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy [J]. *Environmental Pollution*, 2015, 206: 217-226.
- [5] Gholizadeh A, Boruvka L, Saberioon M M, et al. Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features[J]. *Soil and Water Research*, 2016, 10(4): 218-227.
- [6] Dorau K, Pohl L, Just C, et al. Soil organic matter and phosphate sorption on natural and synthetic Fe oxides under in situ conditions [J]. *Environmental Science & Technology*, 2019, 53(22): 13081-13087.
- [7] Wang D, Tian X P, Zhang Z Y. Distribution of various forms iron, manganese, copper, zinc and relationship between organic matter and various forms on lessive[J]. *Journal of Tianjin Agricultural University*, 2016, 23(1): 14-17, 22.
王丹, 田秀平, 张之一. 白浆土各形态 Fe、Mn、Cu 和 Zn 分布及其与有机质间的关系 [J]. *天津农学院学报*, 2016, 23(1): 14-17, 22.
- [8] Rietz D N, Haynes R J. Effects of irrigation-induced salinity and sodicity on soil microbial activity[J]. *Soil Biology and Biochemistry*, 2003, 35(6): 845-854.
- [9] Wang G D, Chu G X, Liu Y, et al. Effects of long-term irrigation with brackish groundwater on soil microbial biomass in cotton field in arid oasis [J]. *Transactions of the CSAE*, 2009, 25(11): 44-48.
王国栋, 褚贵新, 刘瑜, 等. 干旱绿洲长期微咸地下水灌溉对棉田土壤微生物量影响 [J]. *农业工程学报*, 2009, 25(11): 44-48.
- [10] Radim V, Radka K, Lubo B, et al. Combining reflectance spectroscopy and the digital elevation model for soil oxidizable carbon estimation [J]. *Geoderma*, 2017, 303: 133-142.
- [11] Fernandes M M H, Coelho A P, Fernandes C, et al. Estimation of soil organic matter content by modeling with artificial neural networks[J]. *Geoderma*, 2019, 350: 46-51.
- [12] Kweon G, Lund E, Maxton C. Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors [J]. *Geoderma*, 2013, 199: 80-89.
- [13] Siebielec G, McCarty G W, Stuczynski T I, et al. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content [J]. *Journal of Environmental Quality*, 2004, 33(6): 2056-2069.
- [14] Yu L, Hong Y S, Zhou Y, et al. Wavelength variable selection methods for estimation of soil organic matter content using hyperspectral technique [J]. *Transactions of the CSAE*, 2016, 32(13): 95-102.
于雷, 洪永胜, 周勇, 等. 高光谱估算土壤有机质含量的波长变量筛选方法 [J]. *农业工程学报*, 2016, 32(13): 95-102.
- [15] Tian M L, Ge X Y, Ding J L, et al. Coupled machine learning and unmanned aerial vehicle based hyperspectral data for soil moisture content estimation [J]. *Laser & Optoelectronics Progress*, 2020, 57(9): 093002.
田美玲, 葛翔宇, 丁建丽, 等. 耦合机器学习和机载高光谱数据的土壤含水量估算 [J]. *激光与光电子学进展*, 2020, 57(9): 093002.
- [16] Cheng H, Shen R L, Chen Y Y, et al. Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy [J]. *Geoderma*, 2019, 336: 59-67.
- [17] van der Meer F. Acknowledgement of reviewer services to the International Journal Applied Earth Observation and Geoinformation [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 58: A1.
- [18] Chen B, Zou X Y, Zhu W J. Eliminating outlier samples in near-infrared model by method of PCA-mahalanobis distance [J]. *Journal of Jiangsu University (Natural Science Edition)*, 2008, 29(4): 277-279, 292.
陈斌, 邹贤勇, 朱文静. PCA 结合马氏距离法剔除近红外异常样品 [J]. *江苏大学学报(自然科学版)*,

- 2008, 29(4): 277-279, 292.
- [19] Ge X Y, Ding J L, Wang J Z, et al. Estimation of soil moisture content based on competitive adaptive reweighted sampling algorithm coupled with machine learning[J]. *Acta Optica Sinica*, 2018, 38(10): 1030001.
葛翔宇, 丁建丽, 王敬哲, 等. 基于竞争适应重加权采样算法耦合机器学习的土壤含水量估算[J]. *光学学报*, 2018, 38(10): 1030001.
- [20] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [21] Isaksson T, Næs T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy [J]. *Applied Spectroscopy*, 1988, 42(7): 1273-1284.
- [22] Viscarra Rossel R A, Behrens T, Ben-Dor E, et al. A global spectral library to characterize the world's soil [J]. *Earth-Science Reviews*, 2016, 155: 198-230.
- [23] Zhang Z P, Ding J L, Wang J Z. Spectral characteristics of oasis soil in arid area based on harmonic analysis algorithm[J]. *Acta Optica Sinica*, 2019, 39(2): 0228003.
张子鹏, 丁建丽, 王敬哲. 基于谐波分析算法的干旱区绿洲土壤光谱特性研究[J]. *光学学报*, 2019, 39(2): 0228003.
- [24] Weng Y L, Gong P, Zhu Z L. A spectral index for estimating soil salinity in the Yellow River Delta region of China using EO-1 Hyperion data [J]. *Pedosphere*, 2010, 20(3): 378-388.
- [25] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 58(2): 109-130.
- [26] Xue L H, Zhou D H, Li Y, et al. Prediction of soil organic matter and total phosphorus with VIS-NIR hyperspectral inversion relative to land use[J]. *Acta Pedologica Sinica*, 2014, 51(5): 993-1002.
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[J]. *土壤学报*, 2014, 51(5): 993-1002.
- [27] Ehrentraut D, Pollnau M. On the potential of BaSO₄:Mn⁶⁺ for broadly tunable laser emission in the near infrared spectral region [C] // Conference on Lasers and Electro-Optics Europe, June 22-27, 2003, Munich, Germany. New York: IEEE, 2003: 343.
- [28] Hong Y S, Shen R L, Cheng H, et al. Cadmium concentration estimation in peri-urban agricultural soils: Using reflectance spectroscopy, soil auxiliary information, or a combination of both? [J]. *Geoderma*, 2019, 354: 113875.
- [29] Hong Y S, Liu Y L, Chen Y Y, et al. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy[J]. *Geoderma*, 2019, 337: 758-769.
- [30] Jiang X Q, Ye Q, Lin Y, et al. Inverting study on soil water content based on harmonic analysis and hyperspectral remote sensing[J]. *Acta Optica Sinica*, 2017, 37(10): 1028001.
姜雪芹, 叶勤, 林怡, 等. 基于谐波分析和高光谱遥感的土壤含水量反演研究[J]. *光学学报*, 2017, 37(10): 1028001.
- [31] Ye Q, Jiang X Q, Li X C, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(3): 164-172.
叶勤, 姜雪芹, 李西灿, 等. 基于高光谱数据的土壤有机质含量反演模型比较[J]. *农业机械学报*, 2017, 48(3): 164-172.
- [32] Shen L Z, Gao M F, Yan J W, et al. Estimation model of soil organic matter based on SVR and PLSR [J]. *China Agricultural Informatics*, 2019(1): 58-71.
沈兰芝, 高懋芳, 闫敬文, 等. 基于 SVR 和 PLSR 的土壤有机质高光谱估测模型研究[J]. *中国农业信息*, 2019(1): 58-71.
- [33] Li G W, Gao X H, Xiao N W, et al. Estimation of soil organic matter content based on characteristic variable selection and regression methods [J]. *Acta Optica Sinica*, 2019, 39(9): 0930002.
李冠稳, 高小红, 肖能文, 等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. *光学学报*, 2019, 39(9): 0930002.
- [34] Rossel R A V, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra [J]. *Geoderma*, 2010, 158(1/2): 46-54.
- [35] Horta A, Malone B, Stockmann U, et al. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: a prospective review [J]. *Geoderma*, 2015, 241/242: 180-209.
- [36] Zhao Q D, Ge X Y, Ding J L, et al. Combination of

- fractional order derivative and machine learning algorithms for the spectral estimation of soil organic carbon content [J]. *Laser & Optoelectronics Progress*, 2020, 57(15): 153001.
- 赵启东, 葛翔宇, 丁建丽, 等. 结合分数阶微分技术与机器学习算法的土壤有机碳含量光谱估测[J]. *激光与光电子学进展*, 2020, 57(15): 153001.
- [37] Chakraborty S, Man T, Paulette L, et al. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging [J]. *Geoderma*, 2017, 306: 108-119.
- [38] O'Rourke S M, Stockmann U, Holden N M, et al. An assessment of model averaging to improve predictive power of portable VIS-NIR and XRF for the determination of agronomic soil properties [J]. *Geoderma*, 2016, 279: 31-44.