

利用三维光谱指数定量估算土壤有机质含量：以新疆艾比湖流域为例

张子鹏, 丁建丽*, 王敬哲, 葛翔宇, 李振山

新疆大学资源与环境科学学院, 新疆 乌鲁木齐 830046

摘要 土壤光谱特征是土壤内在理化特性的外在表现, 利用可见-近红外(Vis-NIR)的漫反射光谱估测土壤有机质含量(SOMC)可为土地资源的合理开发与利用提供重要的科学依据。土壤是由多种物质组成的混合物, 其高光谱数据中存在某些成分(例如盐颗粒和矿物质)的重叠吸收, 同时波段之间存在共线性问题, 这些因素对光谱分析和建模带来了极大的挑战。光谱指数法通过迭代运算, 不但充分考虑波段之间的协同作用, 而且具有最小化无关波长影响的功能。此外该方法将光谱特征从一维扩展到多维, 能容易地检测和区分出细微的吸收峰。以新疆艾比湖流域为研究区, 采集了120个土壤样品, 在室内进行SOMC和光谱的测定。利用一阶微分(FD)和连续统去除(CR)对高光谱数据进行预处理。在已有两波段指数的基础上, 加入第三个波段, 利用最优波段算法, 构建了三种SOMC的三波段光谱指数(TBI), 并从光谱机理上讨论了TBI的合理性。最后根据支持向量机(SVM)的建模效果, 进一步比较不同维度光谱参数对SOMC估测的准确性。研究结果表明: (1)光谱预处理技术可以在一定程度上减弱反射光谱中的噪声信息, 突出更多潜在的光谱信息; (2)通过对比分析得出, SOMC的相关性随着光谱信息维度的增加而增加, 即, $TBI > 二波段指数 > 一维光谱参数$; (3)新开发的TBI在SOMC的建模和验证过程中提供了比两波段指数和一维光谱参数更好的估测效果, 其中TBI-1的估测效果最好, 建模集的决定系数(R_c^2)为0.88, 验证集的决定系数(R_v^2)为0.85, 相对分析误差(RPD)为2.43。该研究对比了不同维度光谱参数对SOMC的响应和建模精度, 发现三波段光谱指数是评价SOMC的重要参量。此外, TBI与SVM算法的结合, 可以显著弱化土壤噪声信息, 提高SOMC的预测精度, 在土壤其他生化参数的估计中具有较强的应用潜力。

关键词 高光谱; 有机质含量; 相关性分析; 三波段光谱指数; 支持向量机

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1514-09

引言

土壤有机质含量(soil organic matter content, SOMC)控制土壤环境中生物、化学和物理过程, 是评价土壤肥力和土壤质量的重要指标^[1]。传统的SOMC测定多基于大量的野外土壤采样和繁琐的室内化学分析实验, 比较费时、费力和耗资, 无法满足现代精细农业发展的需要^[2-3]。利用可见光-近红外(visible light-near infrared, Vis-NIR)的漫反射光谱分析土壤属性是当今研究的热点, 大量研究已经证明了Vis-NIR光谱在土壤属性分析中的潜力^[3-4]。

以往的研究多是在一维层面上(原始波段反射率和/或对应的数学变换)考虑光谱的敏感程度, 并没有深入研究波段之间的相互影响^[3-5]。光谱指数是由几个窄波段或宽波段通

过一定的数学变换组合而成, 不但考虑了波段之间相互作用, 而且在一定程度上提高了对待测属性的响应^[6]。已有研究多采用两波段的光谱指数进行环境建模和属性量化^[6-8]。Li等^[9]对小麦冠层N指标的估测中发现, 两波段光谱指数的量化能力和稳定性弱于三波段光谱指数。但现有研究成果中仅有很少的三波段光谱指数被用于土壤属性的估测^[10]。此外, 通过遴选得到的光谱指数可应用于遥感传感器, 以评估景观或区域尺度的地表参数, 但到目前为止基于SOMC方面的研究仍未得到深入研究与探索^[9]。

虽然Vis-NIR分析SOMC技术已经在世界某些地方使用, 但在中国西北部尚未实施有效可靠的方法来准确估测SOMC, 本研究的目的是: (1)探索不同预处理的SOMC特性及其敏感区域; (2)开发评估SOMC的三波段光谱指数, 并量化不同维度的光谱参数对SOMC的响应; (3)根据支持

收稿日期: 2019-01-14, 修订日期: 2019-06-09

基金项目: 国家自然科学基金项目(41961059, 41771470)资助

作者简介: 张子鹏, 1994年生, 新疆大学资源与科学学院博士研究生

e-mail: 13659978318@163.com

* 通讯联系人 e-mail: watarid@xju.edu.cn

向量机(SVM)的建模结果,进一步比较不同维度的光谱参数对 SOMC 的估测表现,并寻找最有效的建模方法。

1 实验部分

1.1 研究区介绍和土壤样本的制备

艾比湖流域位于亚欧大陆腹地,44°05′—45°08′N,82°35′—83°16′E,地处新疆维吾尔自治区准噶尔盆地西南部,阿拉山口的大风主通道下,研究区总面积约 2 670 km²[图 1 (a, b)]^[11]。该地区属于典型的中温带干旱大陆性气候,具有冬冷夏热(年均温 6 ℃)、降水量少(年均降水量小于 150 mm)、蒸发量大(年均蒸发量大于 1 200 mm)、风沙频繁(年均≥8 级以上大风 165 d)等特点。该地区成土母质以洪积物、

湖积物和砂质风积物为主,土壤类型主要有灰漠土、灰棕漠土、风沙土、盐土和沼泽土五大类。其中,灰漠土与灰棕漠土为主要组成部分^[12]。由于水分收支不平衡,土壤中盐分受淋洗作用的影响较小,在土壤毛细管的作用下,干涸河床和湖体周边土壤中的盐分不断地向地表聚集,造成严重的盐渍化现象[图 1(c, d, e)]^[13]。

2016 年 10 月,根据研究区的典型景观特征,建立了 120 个 30 m×30 m 的样方,依据五点采样法采集表层(0~20 cm)土壤样本[图 1(b, f)]。将样品带回室内经过兑水透析(去除可溶性盐)、风干、碾磨、过筛(2 mm)等操作,将制备好的土样分为两份,分别用于光谱数据的采集与 SOMC 的测定,测定依据中国标准(SL237—1999)^[2, 14]。

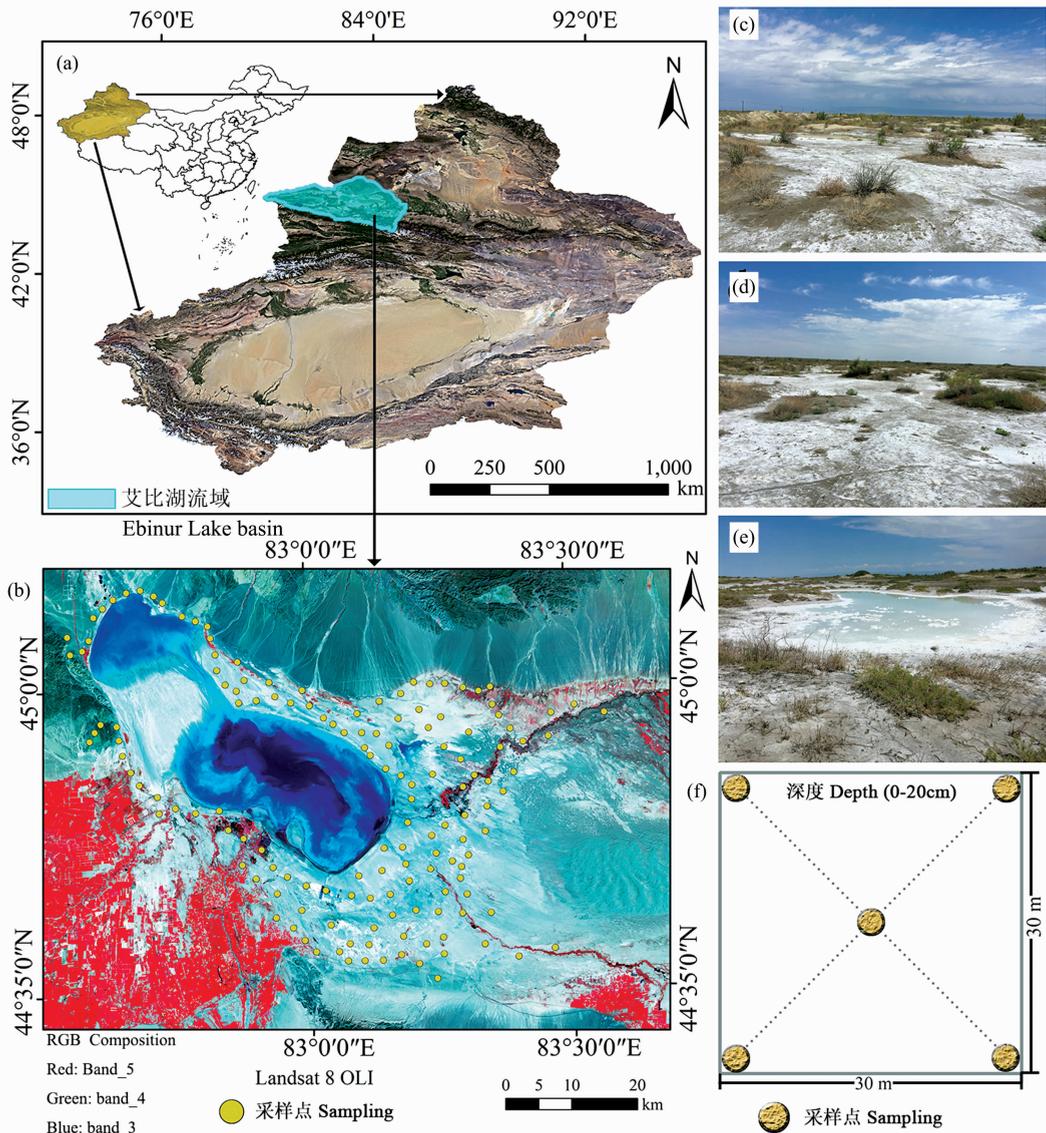


图 1 研究区位置图(a: 新疆地图; b: 艾比湖流域), (c, d, e)景观照片和方形采样方法, (f): 五点采样法
Fig. 1 Location maps of the research area (a: Xinjiang map, b: Ebinur Lake basin) and photographs of the landscape (c, d, e) and sampling method in quadrat, (f): five-point sampling method

1.2 光谱数据的采集和预处理

利用 ASDField Spec 3 光谱仪测定土壤 Vis-NIR 光谱。光谱分辨率为 3 和 10 nm(350~1 000 与 1 000~2 500 nm 区间), 重采样间隔为 1 nm, 输出波段数为 2 151 个。土壤光谱的测定在暗室中进行, 一个与光谱仪匹配的卤素灯(50 W)放置在离样品 30 cm 处, 入射角为 45°, 光谱测定开始时和每个土壤样本测定 10 次后, 利用漫反射白板定标, 每个土壤样本重复测定 10 次, 然后将其平均为一个光谱作为最终的反射光谱^[8]。

位于边缘处的光谱波段(350~399 和 2 401~2 500 nm)存在严重的机器噪声, 因此将这两个范围内的波段剔除。对 400~2 400 nm 范围内的光谱进行二阶 9 点的 Savitzky-Golay (S-G)平滑。参考 Hong 等的研究, 设置 10 nm 的重采样间隔, 并将重采样得到的 201 个波段定义为原始光谱(R)^[6]。

土壤光谱的预处理包括: 用于分辨重叠峰、减弱地物背景干扰和修复基线漂移的一阶微分(first derivative, FD); 用于消除散射影响, 突出光谱吸收特性的连续统去除(continuum removal, CR)。这两种预处理技术被广泛应用于光谱分析中, 具有很强的普适性^[2, 6]。

1.3 二维和三维相关系数

光谱指数根据地物的光谱特性, 对波段进行简单组合, 达到对地表参量简单有效的度量^[15], 不仅将光谱特性从一维扩展到了二维, 还充分考虑了光谱之间的相互关系^[7]。在前人研究基础上选择三种光谱指数, 用于估测 SOMC 的最佳波段^[6-8]。指数的数学表达式如式(1)一式(3)

$$RI = R_{\lambda_1} / R_{\lambda_2} \quad (1)$$

$$DI = (R_{\lambda_1} - R_{\lambda_2}) \quad (2)$$

$$NDI = (R_{\lambda_1} - R_{\lambda_2}) / (R_{\lambda_1} + R_{\lambda_2}) \quad (3)$$

其中, RI 为比值指数(ratio index); DI 为差值指数(difference index); NDI 为归一化指数(normalized difference index)。

参考 Tian 等^[10]的研究成果, 本研究在所选 3 种两波段指数(λ_1 和 λ_2)的基础上增加第三个波段 λ_3 , 具体如式(4)一式(6)

$$TBI-1 = (R_{\lambda_1} - R_{\lambda_2}) + (R_{\lambda_3} - R_{\lambda_2}) \quad (4)$$

$$TBI-2 = R_{\lambda_1} / (R_{\lambda_2} + R_{\lambda_3}) \quad (5)$$

$$TBI-3 = (R_{\lambda_1} - R_{\lambda_2}) / (R_{\lambda_1} + R_{\lambda_3}) \quad (6)$$

式中, TBI-1 为三波段指数 1(three-band index 1); TBI-2 为三波段指数 2(three-band index 2); TBI-3 为三波段指数 3(three-band index 3)。

式(1)一式(6)中, R_{λ_1} , R_{λ_2} 和 R_{λ_3} 分别代表了波长在 400~2 400 nm 范围内第 λ_1 , λ_2 和 λ_3 个波段的光谱反射率。相关操作均在 MATLAB 2018b 中完成, 原始光谱(R)和预处理

的光谱(FD 和 CR)都被用来提取最佳光谱指数。

1.4 集合划分与建模预测

使用 Kennard-Stone(K-S)算法进行样本集合的划分^[3]。120 个土壤样本被分成两个部分, 其中 90 个样本用于建模, 30 个样本对建模结果进行验证。SVM 由 Vapnik 等于 1995 年提出, 该算法可最小化结构风险并提高模型泛化能力, 能够较好地解决过拟合和局部最优解等问题, 对小样本、非线性及高维数据的估测具有良好的效果, 现已被广泛应用于光谱分析与建模中^[6, 8, 16]。SVM 在 MATLAB 2018b 中编程实现核函数选择鲁棒性较好的 RBF, 利用网格搜索法遍历惩罚参数 c 和核参数 σ 的组合进而提高模型的预测精度。

1.5 建模精度检验

选择三个参数对建模的精度和效果进行评价: 决定系数(R^2), 均方根误差(RMSE), 相对分析误差(RPD)。Chang 等^[17]将 RPD 分为 3 类来评价模型的准确性: 当 $RPD < 1.4$ 时, 表明模型的拟合性和稳定性非常差, 不推荐使用; 当 $1.4 \leq RPD < 2.0$ 时, 模型可以提供较为合理的预测结果; 当 $RPD > 2.0$ 时, 表明模型的预测效果非常好。一般良好的预测模型将具有较高的 R^2 (0.8~1.0) 和 RPD, 以及较小的 RMSE。

2 结果与讨论

2.1 不同的 SOMC 和反射光谱曲线

由图 2(a)可知, 土壤光谱曲线在 1 450, 1 950 和 2 200 nm 附近都存在水分吸收谷。其中, 1 950 nm 处最为突出, 这与前人在该地区的研究结果一致^[11-12]。FD 光谱在 440, 580, 900, 1 400 和 2 300 nm 分别出现了不同的正负峰值, 可见 FD 可以凸显原始数据的局部肩峰[图 2(b)]。连续统去除使得吸收特征得到了明显的放大[图 2(c)], 如 500 和 1 100 nm 的为弱吸收带在连续统去除曲线中可以观察出来, 而在原始反射光谱中则不明显。

表 1 土壤有机质含量(SOMC)的统计特征

Table 1 Statistical characteristics of the soil organic carbon content (SOMC)

样本类型	数量	最小值 /%	最大值 /%	均值 /%	标准差 /%	变异系数/%
Whole dataset	120	0.07	4.03	1.25	0.90	72.10
Calibration dataset	90	0.07	4.03	1.18	0.89	75.76
Validation dataset	30	0.12	3.58	1.28	0.91	71.26

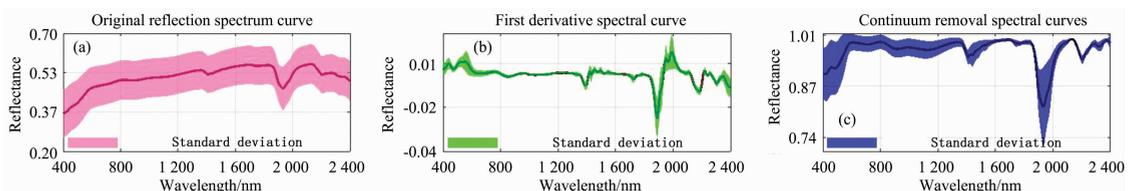


图 2 不同预处理的 SOMC 光谱曲线

Fig. 2 Different pretreated SOMC spectral curves

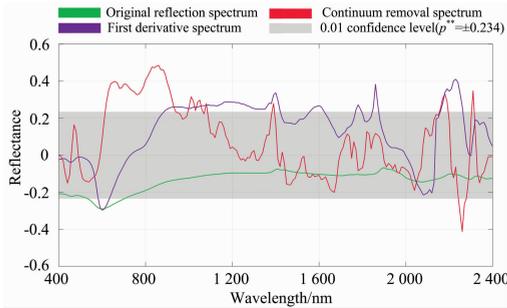


图 3 SOMC 与原始光谱和不同预处理光谱的一维相关系数
Fig. 3 One-dimensional correlation coefficients between SOMC and raw reflectance and its transformation

2.2 一维相关性分析

R 与 SOMC 的相关性曲线较为平滑, 仅有 550~710 nm 波段通过 0.01 水平上的显著性检验 ($p^{**} = \pm 0.234$), 说明 R 与 SOMC 的敏感性较低(图 3)。然而经过 FD 处理后, 显著性明显提升, 尤其是在 620~950 nm 处, 绝对值最大的相关系数(maximum absolute correlation coefficient, MACC)为 0.52 在 870 nm 处, 已有研究指出该波段附近存在 C—H 吸收带, 与 SOMC 直接相关^[6]。经过 CR 处理后, 相关系数在 610, 1 410, 1 870 和 2 240 nm 附近得到了一定程度的增加。本研究选择不同预处理中 MACC 最大的波段 (R_{610} , FD_{870} , CR_{2240}), 用于后续的对比分析和建模预测(表 2)。

表 2 一维光谱数据和二维光谱指数中最敏感的光谱参数

Table 2 Most sensitive spectral parameters in One-Dimensional spectral data and Two-Dimensional indices

一维光谱数据		二维光谱数据					
1D Data	MACC	Ratio index	MACC	Difference index	MACC	Normalized index	MACC
R_{610}	0.29	$R_{2\ 250}, R_{2\ 290}$	0.48	R_{860}, R_{870}	0.49	$R_{2\ 250}, R_{2\ 290}$	0.48
FD_{870}	0.49	$FD_{670}, FD_{2\ 250}$	0.53	$FD_{870}, FD_{2\ 270}$	0.59	$FD_{670}, FD_{1\ 890}$	0.52
$CR_{2\ 240}$	0.41	CR_{890}, CR_{900}	0.47	CR_{890}, CR_{900}	0.47	CR_{890}, CR_{900}	0.47

2.3 二维相关性分析

与图 3 相比, 二维相关系数值的阈值范围均大于一维相关系数。其中 R 的差值指数(DI)[图 4(i)]与 R 的一维数据(图 3)相比 MACC 提高了 0.2。在所有两波段指数中相关性较好的波段主要集中在可见光和短波近红外区域(400~

1 100 nm)(图 4), 这与前人的研究结果一致^[6, 8]。在每个两波段指数中, FD 与 SOMC 的敏感性最强, 其中 FD-DI[图 4(e)]得到的 MACC 最大为 0.59。提取每个指数的详细信息, 见表 2。

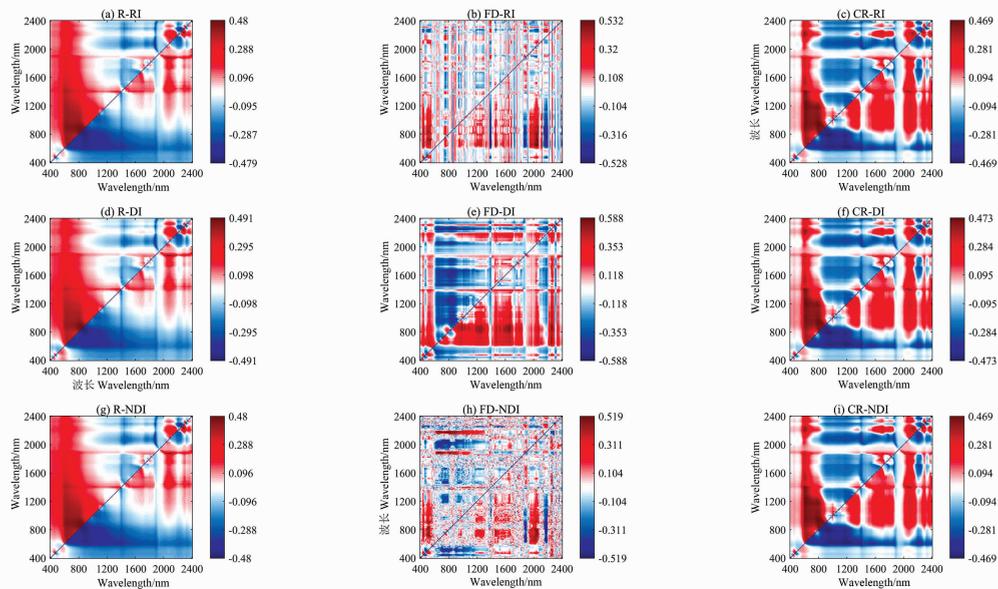


图 4 RI, DI, NDI 与 SOMC 的二维相关图

Fig. 4 Two-dimensional correlation diagram between the RI, DI, NDI and SOMC

2.4 三维相关性分析

在 R 所构建的三波段指数中(图 5, 图 6 和图 7), 最优指数的波长都在近红外短波范围内, 并且每个指数中均有一个波段靠近水分吸收带。在所有三波段指数中, R 的敏感性均

最强, 其中 TBI-1 的 MACC 最大达到 0.70。在 FD 所构建的三波段指数中(图 5, 图 6, 图 7), 均有一个波段靠近 2 300 nm, 已有研究证明 2 300 nm 附近存在一个与 SOMC 相关的 C—H 特征峰^[18]。但从切片图中来看, 基于 FD 的三波段指

数其敏感区域较为分散，不集中。在 CR 所构建的三波段指数中(图 5, 图 6 和图 7), 特征波段为 890 和 900 nm(与高价铁氧化物的分子振动有关), 1 670, 1 720 和 1 780 nm(与 C—H 的伸缩振动有关), 1 900 nm(与土壤表面吸附水、粘

土矿物 O—H 晶格结构水有关)^[5-6, 18]。其中 TBI-1 的 PCC 相比两波段指数平均提升 0.18。优化后的三波段指数是估测 SOMC 较为有前景的指标, 同时也暗示两波段指数存在一定的不足。提取每个指数的详细信息, 见表 2。

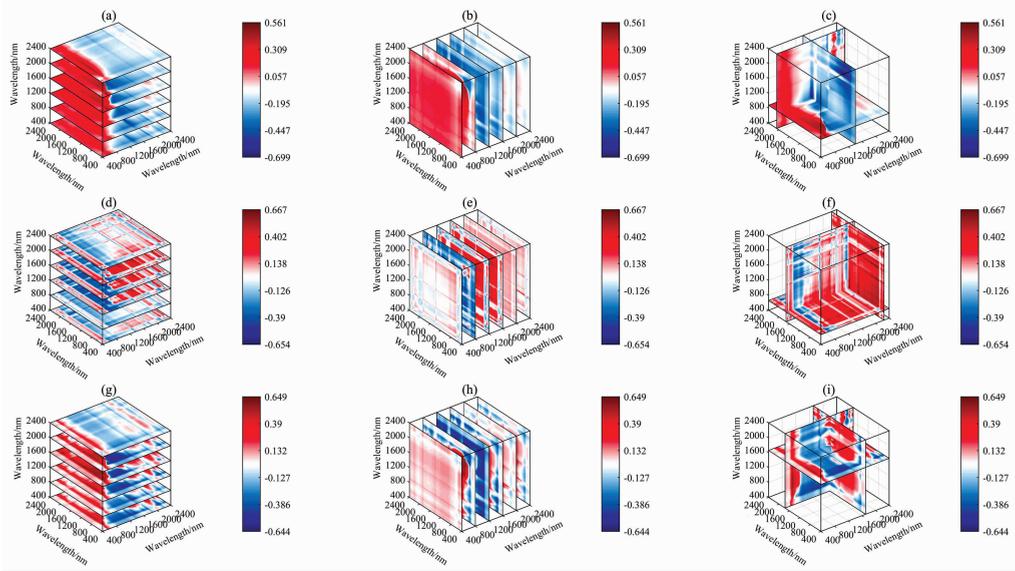


图 5 TBI-1 和 SOMC 的相关系数切片图

(a, d, g) 分别为 R, FD 和 CR 的水平切片图; (b, e, h) 分别为 R, FD 和 CR 的垂直切片图; (c, f, i) 分别为 R, FD 和 CR 的最佳切片图

Fig. 5 Correlation coefficient slice diagram of the TBI-1 and SOMC

(a, d, g): Horizontal slice map of the R, FD and CR, respectively; (b, e, h): Vertical slice map of the R, FD and CR, respectively; (c, f, i): Optimum slice map of the R, FD and CR, respectively

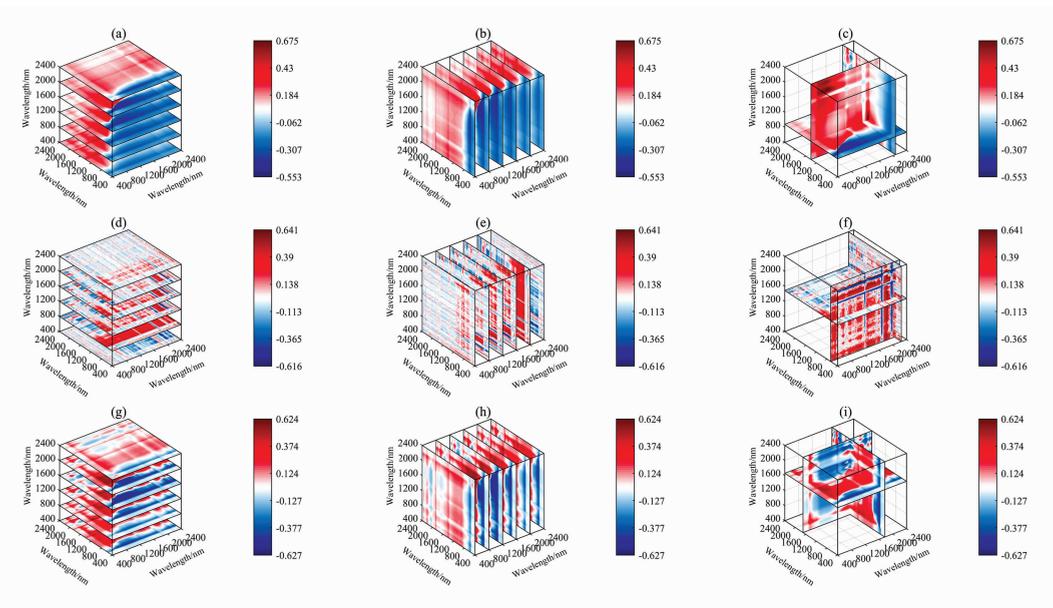


图 6 TBI-2 和 SOMC 的相关系数切片图

(a, d, g) 分别为 R, FD 和 CR 的水平切片图; (b, e, h) 分别为 R, FD 和 CR 的垂直切片图; (c, f, i) 分别为 R, FD 和 CR 的最佳切片图

Fig. 6 Correlation coefficient slice diagram of the TBI-1 and SOMC

(a, d, g): Horizontal slice map of the R, FD and CR, respectively; (b, e, h): Vertical slice map of the R, FD and CR, respectively; (c, f, i): Optimum slice map of the R, FD and CR, respectively

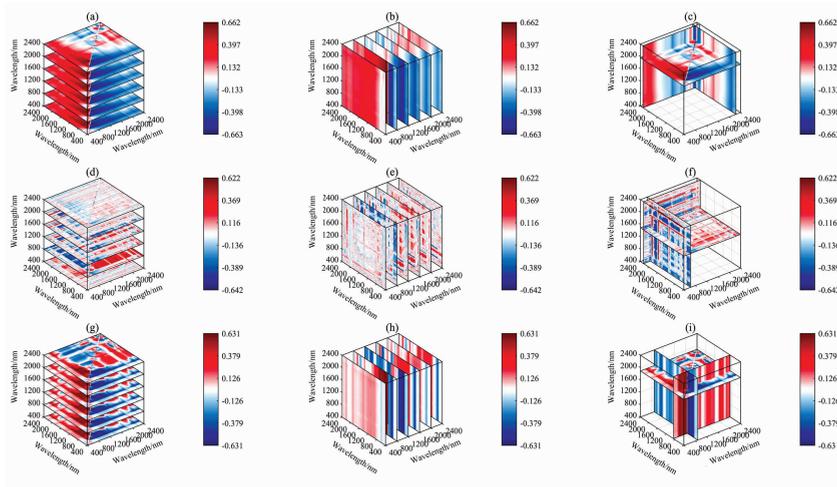


图 7 TBI-3 和 SOMC 的相关系数切片图

(a, d, g) 分别为 R, FD 和 CR 的水平切片图; (b, e, h) 分别为 R, FD 和 CR 的垂直切片图;

(c, f, i) 分别为 R, FD 和 CR 的最佳切片图

Fig. 7 Correlation coefficient slice diagram of the TBI-1 and SOMC

(a, d, g): Horizontal slice map of the R, FD and CR, respectively; (b, e, h): Vertical slice map of the R, FD and CR, respectively;

(c, f, i): Optimum slice map of the R, FD and CR, respectively

表 3 三维光谱指数中最敏感的光谱参数

Table 3 Most sensitive spectral parameters in Three-Dimensional indices

三维光谱数据					
Three band index 1	MACC	Three band index 2	MACC	Three band index 3	MACC
$R_{2100}, R_{1420}, R_{890}$	0.70	$R_{1420}, R_{2090}, R_{890}$	0.68	$R_{2290}, R_{2250}, R_{1970}$	0.66
$FD_{1740}, FD_{2260}, FD_{670}$	0.67	$FD_{670}, FD_{2260}, FD_{1500}$	0.64	$FD_{2260}, FD_{650}, FD_{1500}$	0.64
$CR_{1780}, CR_{1720}, CR_{1670}$	0.65	$CR_{1720}, CR_{1780}, CR_{1670}$	0.63	$CR_{890}, CR_{900}, CR_{1900}$	0.63

2.5 光谱参数的分析与建模

由表 2 和表 3 可知, 在一维光谱数据中, R 的 MACC 均低于 FD 和 CR; 在二维光谱指数中, R 的 MACC 略高于 CR; 而在三维光谱指数中, R 的 MACC 均大于 FD 和 CR, 研究结果说明光谱预处理效果随多维度光谱协同作用的增强而减弱。

不同维度的光谱参数(表 2 和表 3)和建模集中的 SOMC 被作为模型的训练样本(表 1), 验证集中的 SOMC 被作为模

型的预测样本(表 1), R^2 , RMSE 和 RPD 作为模型预测性能的评价指标(表 3)。总体而言, 无论在建模还是验证方面, TBI 均具有最低的 RMSE 和最大的 R^2 和 RPD, 都达到了最佳的预测性能, TBI-1 的效果最优, $R^2 = 0.85$, $RMSE_V = 3.67$, $RPD = 2.43$; 两波段指数的表现次之, 其中, NDI 的效果最优, RPD 达到 1.88; 1D 数据效果最弱, RPD 仅为 1.60。

表 4 基于 SVM 的 SOMC 建模与预测结果

Table 4 Calibration and validation results of SOMC based on SVM

光谱处理	模型	建模集			验证集			
		N	R_C^2	$RMSE_C$	N	R_V^2	$RMSE_V$	RPD
1D data	SVM	90	0.66	5.77	30	0.64	5.57	1.60
RI	SVM	90	0.78	4.62	30	0.75	5.04	1.77
DI	SVM	90	0.76	4.80	30	0.78	5.22	1.71
NDI	SVM	90	0.80	4.96	30	0.78	4.76	1.88
TBI-1	SVM	90	0.88	3.47	30	0.85	3.67	2.43
TBI-2	SVM	90	0.87	3.62	30	0.83	4.09	2.18
TBI-3	SVM	90	0.85	3.70	30	0.82	4.26	2.10

图 8 中得到的回归线斜率均低于 1, 许多研究发现, Vis-NIR 模型更倾向于高估低浓度和低估高浓度的 SOM^[19-20]。这与建模时的数据标准化有关, 数据标准化以消除不同变量样本之间量纲和数量值的差异为目的, 但数据较为集中, 则会对极值点的估测产生一定程度的偏差。

在图 8 中, 1D 数据预测值与观测值在标准 1:1 线中的偏差较大, 离群值较多, 该模型仅具有区分样本数据含量高

低的能力, 量化能力较弱。两波段指数的建模效果相比一维光谱数据有所提升, 预测值与观测值较接近于标准 1:1 线。三波段指数的估测效果均优于前两者, 其线性拟合线接近于标准 1:1 线。其中使用 (TBI-1)-SVM 的预测效果最好, ($R_c^2=0.85$, $RMSE_v=3.67$, $RPD=2.43$), 其拟合程度和准确性均优于本研究中的所有模型。

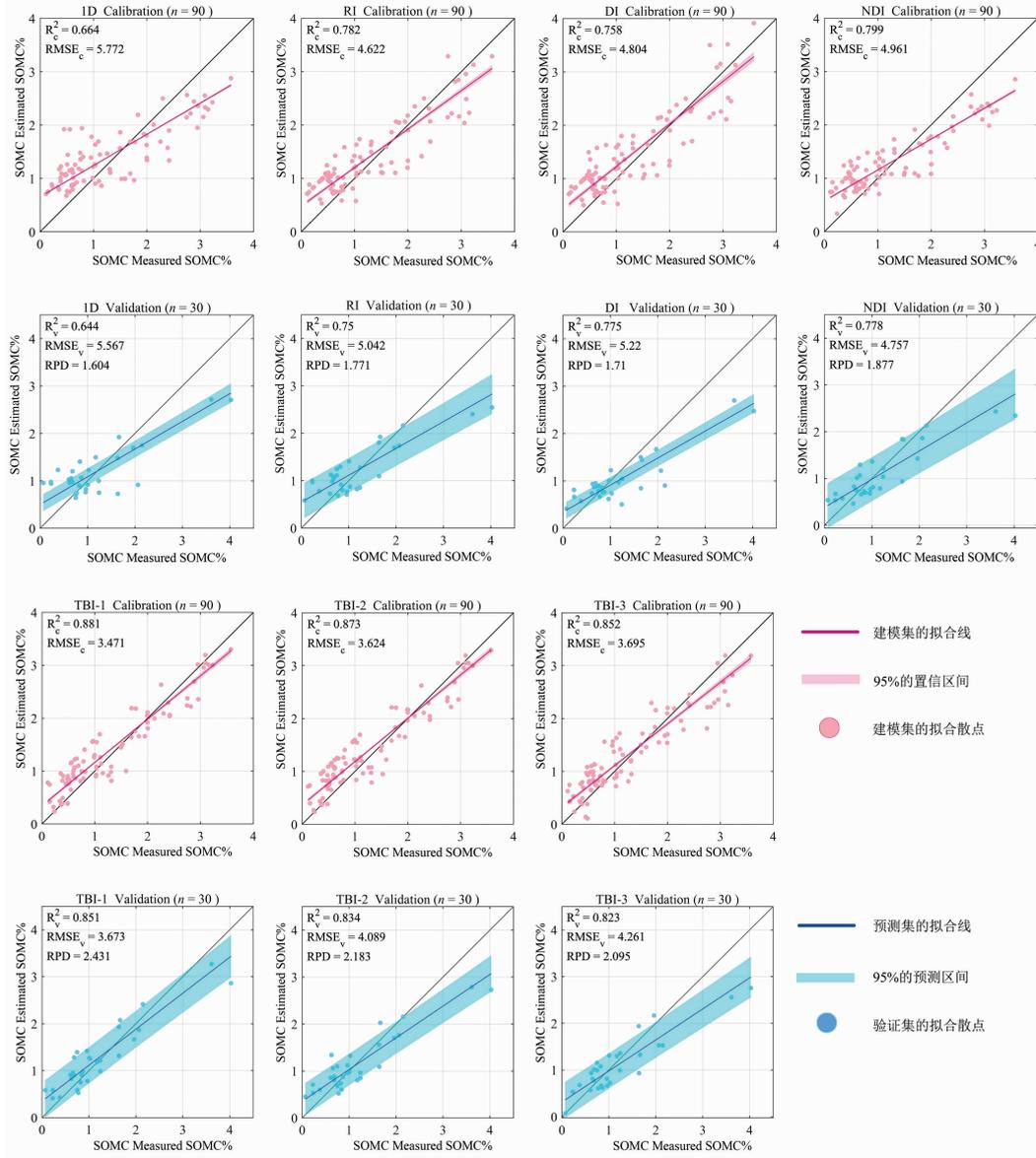


图 8 基于 SVM 的 SOMC 实测值和预测值的比较

Fig. 8 Comparison of SOMC between measured and the predicted values by SVM

3 结论

发现光谱指数法可以在一维线性数据的层面上更为深入的挖掘光谱信息, 充分考虑波段间的相互作用。在已有的两波段指数中增加特定敏感区域的第三个波段, 不仅有助于提

高光谱变量的敏感性, 而且在 SOMC 的建模和验证过程中提供了比两波段光谱指数和一维光谱参数更好的预测效果。其中基于 R 的 TBI-1 效果最优, MACC 为 0.70。利用 SVM 对不同维度的最优光谱参数分别建立预测模型, 发现预测精度随光谱参数维度的增加而增加, 在三波段指数中, (TBI-1)-SVM 的预测能力最优 ($R_c^2=0.88$, $R_v^2=0.85$ 和 $RPD=$

2.43)。该研究结果对减少土壤中其他信息噪声的干扰和土壤属性提供了新的思路。土壤属性的定量估测具有一定的参考价值,为估测土壤中其他

References

- [1] Schmidt M W I, Torn M S, Abiven S, et al. *Nature*, 2011, 478(7367): 49.
- [2] Nawar S, Buddenbaum H, Hill J, et al. *Soil and Tillage Research*, 2016, 155: 510.
- [3] ZHANG Zi-peng, DING Jian-li, WANG Jing-zhe(张子鹏, 丁建丽, 王敬哲). *Acta Optica Sinica(光学学报)*, 2019, 39(2): 0228003.
- [4] Minasny B, Mcbratney A B, Bellon-Maurel V, et al. *Geoderma*, 2011, 167(167): 118.
- [5] GE Xiang-yu, DING Jian-li, WANG Jing-zhe, et al(葛翔宇, 丁建丽, 王敬哲, 等). *Acta Optica Sinica(光学学报)*, 2018, 38(10): 1030001.
- [6] Hong Y S, Liu Y L, Chen Y Y, et al. *Geoderma*, 2019, 337(337): 758.
- [7] Jin X, Song K, Jia D, et al. *Agricultural and Forest Meteorology*, 2017, 244-245: 57.
- [8] Bao N, Wu L, Ye B, et al. *Geoderma*, 2017, 288: 47.
- [9] Fei Li, Mistele B, Hu Y C, et al. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 92(2): 112.
- [10] Tian Y C, Yao X, Yang J, et al. *Field Crops Research*, 2011, 120(2): 299.
- [11] Wang J, Ding J, Abulimiti A, et al. *PeerJ*, 2018, 6: e4703.
- [12] Wang X, Zhang F, Johnson V C, et al. *Remote Sensing of Environment*, 2018, 218: 104.
- [13] Peng J, Biswas A, Jiang Q, et al. *Geoderma*, 2019, 337: 1309.
- [14] Shang X Y, Zhou G Q, Kuang L F, et al. *Canadian Geotechnical Journal*, 2015, 52(2): 244.
- [15] Thenkabail P S. *Remote Sensing Reviews*, 2001, 20(4): 257.
- [16] Cherkassky V, Ma Y. *Neural Networks*, 2004, 17(1): 113.
- [17] Chang C W, Laird D A, Mausbach M J, et al. *Soil Science Society of America Journal*, 2001, 65(2): 480.
- [18] Shi Z, Wang Q L, Peng J, et al. *Science China Earth Sciences*, 2014, 57(7): 1671.
- [19] Nocita M, Stevens A, Noon C, et al. *Geoderma*, 2013, 199: 37.
- [20] Douglas R K, Nawar S, Alamar M C, et al. *Science of the Total Environment*, 2018, 616-617: 147.

Quantitative Estimation of Soil Organic Matter Content Using Three-Dimensional Spectral Index: A Case Study of the Ebinur Lake Basin in Xinjiang

ZHANG Zi-peng, DING Jian-li*, WANG Jing-zhe, GE Xiang-yu, LI Zhen-shan

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

Abstract The spectral characteristics of soil are the external manifestation of physical and chemical properties in soil. Estimating soil organic matter content (SOMC) by visible-near infrared (VIS-NIR) diffuse reflectance spectroscopy could provide an important scientific basis for the rational development and utilization of land resources. However, the soil is a mixture of many substances, and its hyperspectral data have overlapping absorption of certain components (such as salt particles and minerals), and there are collinear problems between the bands, which bring great challenges for spectral analysis and modeling. Through the iterative operation, the spectral index method not only fully consider the synergy between the bands, but also has the function of minimizing the influence of irrelevant wavelengths. In addition, the method extends the spectral features from one dimension to multidimensional, and can easily detect and distinguish subtle absorption peak. In this study, 120 soil samples were collected from the Ebinur Lake Basin in Xinjiang, and SOMC and spectra were measured indoors. Hyperspectral data were preprocessed using first derivative (FD) and continuum removal (CR). Based on the existing two-band index, the third band was added, and the three-band spectral index (TBI) of three SOMCs was constructed by using the optimal band algorithm. The rationality of TBI was discussed from the spectral mechanism. Finally, according to the modeling effect of support vector machine (SVM), the accuracy of SOMC estimation by different dimensional spectral parameters was further compared. The research results showed that: (1) Spectral pretreatment technology could weaken the noise information in the reflection spectrum to some extent and highlighted more potential spectral information; (2) Through comparative analysis, the correlation of SOMC increased with the increase of the spectral information dimension, that was, TBI > two-band index > one-dimensional spectral pa-

rameters; (3) The newly developed TBI provided better estimation results than the two-band index and one-dimensional spectral parameters in the modeling and verification process of SOMC. The TBI-1 had the best estimation effect and the determination coefficient of the modeling set. (R_c^2) was 0.88, the decision coefficient (R_v^2) of the verification set was 0.85, and the relative analysis error (RPD) was 2.43. In summary, this study compared the response and modeling accuracy of different dimensional spectral parameters to SOMC. It was found that the three-band spectral index was an important parameter for evaluating SOMC and had good performance. In addition, the combination of TBI and SVM algorithm could weaken soil noise information, improved the prediction accuracy of SOMC, and had strong application potential in the estimation of other biochemical parameters of soil.

Keywords Hyperspectral; Organic matter content; Correlation analysis; Three-band spectral index; Support vector machine

(Received Jan. 14, 2019; accepted Jun. 9, 2019)

* Corresponding author

《光谱学与光谱分析》对来稿英文摘要的要求

来稿英文摘要不符合下列要求者，本刊要求作者重写，这可能要推迟论文发表的时间。

1. 请用符合语法的英文，要求言简意明、确切地论述文章的主要内容，突出创新之处。

2. 应拥有与论文同等量的主要信息，包括四个要素，即研究目的、方法、结果、结论。其中后两个要素最重要。有时一个句子即可包含前两个要素，例如“用某种改进的 ICP-AES 测量了鱼池水样的痕量铅”。但有些情况下，英文摘要可包括研究工作的主要对象和范围，以及具有情报价值的其他重要信息。在结果部分最好有定量数据，如检测限、相对标准偏差等；结论部分最好指出方法或结果的优点和意义。

3. 句型力求简单，尽量采用被动式，建议经专业英语翻译机构润色，与中文摘要相对应。用 A4 复印纸单面打印。

4. 摘要不应有引言中出现的内容，换言之，摘要中必须写进的内容应尽量避免在引言中出现。摘要也不要对论文内容作解释和评论，不得简单重复题名中已有的信息；不用非公知公用的符号和术语；不用引文，除非该论文证实或否定了他人已发表的论文。缩略语、略称、代号，除相邻专业的读者也能清楚地理解外，在首次出现时必须加以说明，例如用括号写出全称。