

徐丽华, 谢德体. 土壤有机质含量预测精度对光谱预处理和特征波段的响应[J]. 江苏农业学报, 2019, 35(6): 1340-1345.
doi:10.3969/j.issn.1000-4440.2019.06.010

土壤有机质含量预测精度对光谱预处理和特征波段的响应

徐丽华, 谢德体

(西南大学资源环境学院, 重庆 400716)

摘要: 为了提高土壤有机质含量预测的精度, 对光谱预处理方法和特征波段的选择进行了研究。分别用 Savitzky-Golay 平滑 (SGS)、多元散射校正 (MSC)、标准正态化 (SNV)、标准正态化+去趋势 (SNV_Detrend)、一阶导数 (FD)、二阶导数 (SD)、包络线去除 (CR) 和克里克滤波 (KF) 8 种方法对 33 个水稻土土壤样本进行了光谱预处理, 用分区极值法选择的特征波段进行了建模。结果表明: 经过预处理以后, 除了 SGS 和 KF 处理外, MSC、SNV、SNV_Detrend、FD、SD、CR 预处理获得的土壤光谱与土壤有机质 (SOM) 含量的相关性都得到了显著提高; CR 预处理方法获得的预测模型精度最高, 其标定集和验证集的决定系数分别是 0.728 和 0.666, 最小均方根误差 (RMSE) 分别是 2.240 g/kg 和 2.770 g/kg; 利用分区选择的 4 个特征波段建立的预测模型精度远高于利用 4 个相关系数最大绝对值对应的波段及所有相关系数绝对值大于 0.5 的 77 个波段建立的预测模型。CR 预处理方法和基于分区极值选择的特征波段能够改善土壤有机质含量的预测精度。

关键词: 克里克滤波; 光谱预处理; 分区特征波段; 土壤有机质

中图分类号: S127 TP79 **文献标识码:** A **文章编号:** 1000-4440(2019)06-1340-06

Response of soil organic matter content prediction accuracy to preprocessing of spectra and feature bands

XU Li-hua, XIE De-ti

(College of Resources and Environment, Southwest University, Chongqing 400716, China)

Abstract: In order to improve the prediction accuracy of soil organic matter (SOM) content, the spectral preprocessing methods and the selection of characteristic bands were studied. Savitzky-Golay smoothing (SGS), multiplicative scatter correction (MSC), standard normal variate (SNV), standard normal variate+detrend (SNV_Detrend), first derivative (FD), second derivative (SD), continuum removal (CR) and kriging filter (KF) were used to preprocess the original spectra of 33 soil samples from paddy fields. Feature bands selected by zonal extremum method were used to build prediction model of SOM content. The results showed that the correlations between soil organic matter content and soil spectra preprocessed by MSC, SNV, SNV_Detrend, FD, SD and CR had been significantly improved. The prediction precision of

model obtained by CR pretreatment method was the highest. The determination coefficients of calibration set and validation set were 0.728 and 0.666, and root mean square errors (RMSEs) were 2.240 g/kg and 2.770 g/kg, respectively. The prediction precision of the model based on four zonal feature bands was much higher than that based on four feature bands with maximum absolute values of correlation coefficient and based on the 77 feature bands with

收稿日期: 2019-04-01

基金项目: 中央高校基本科研业务费专项基金项目 (XDJK2016C083);
重庆市基础科学与前沿技术研究一般项目
(cstc2016jcyjA0184); 国家自然科学基金项目 (41671291)

作者简介: 徐丽华 (1976-), 女, 黑龙江绥滨人, 博士, 副教授, 主要从事土壤养分预测与制图、遥感图像信息处理研究, (E-mail) sweitlianna@126.com

通讯作者: 谢德体, (E-mail) xdt@swu.edu.cn

absolute values of correlation coefficient greater than 0.5. Therefore, CR pretreatment method and feature bands selected by zonal extremum can improve prediction precision of SOM content.

Key words: kriging filter; spectral preprocessing; zonal feature bands; soil organic matter

光谱分析技术由于快速、无损的特点在土壤有机质(Soil organic matter, SOM)含量预测方面取得了广泛应用^[1-4]。除了土壤类型、成土母质、人为活动等因素^[5]外,光谱采集时的背景、环境、测量仪器等往往也对土壤样本的光谱产生影响^[5-6],这可能导致所获得的原始光谱与土壤有机质含量之间并不具有明显的相关性。因此,在对 SOM 含量与光谱进行相关性分析之前进行光谱预处理是十分必要的。

常用的光谱预处理方法主要有 Savitzky-Golay 平滑(Savitzky-Golay smoothing, SGS)、多元散射校正(Multiplicative scatter correction, MSC)、标准正态化(Standard normal variate, SNV)、标准正态化+去趋势(Standard normal variate+detrend, SNV_Detrend)、一阶导数(First derivative, FD)二阶导数(Second derivative, SD)、包络线去除(Continuum removal, CR)等方法。在对土壤光谱与 SOM 含量进行灰度关联时发现,CR 处理后得到的光谱反射率与 SOM 含量之间具有较低的灰色关联度,而 SNV 处理在砂质土和黏壤土的有机质含量回归模型中都取得了最高的精度^[1]。在比较预处理方法对于医用辅料模式识别影响的研究中,利用 SNV 处理后的光谱反射率建立的模型也获得了较高的精度^[6]。于雷等的研究结果则显示,用 CR 处理后的光谱建立的 SOM 模型精度高于 FD 处理^[7]。也有研究结果显示,SGS 处理后的导数光谱对土壤有机质含量预测的精度高于 SNV 处理,基于 SD 处理的光谱建立的模型精度高于 FD 处理^[3]。Luybaert 等^[6]的研究结果表明,SD 处理比 FD 和 SNV 处理更能有效减少不同样本的方差贡献。虽然光谱预处理虽然能在一定程度上提高模型的精度和稳定性^[8],但目前还难以找到一种具有普适性的光谱预处理方法。对不同类型不同区域的有机质含量进行预测时,通过不同预处理方法的比较来选择最优的预处理方法,可在一定程度上提高土壤有机质含量的预测精度。另外,上述方法均是从光谱空间对光谱进行预处理,没有从地理空间上考虑土壤样本对彼此光谱的影响。克里克滤波(Kriging filter, KF)方法从样本点的地理空间出发,考虑了周围样本点对中心样本点光谱的影响,改

善了光谱的信噪比^[9]。但是,该方法是否能够取得比其他预处理方法更令人满意的精度,还有待进一步证明。除此之外,特征波段的选择也是影响 SOM 含量预测精度的重要因素。因此,本研究将从 SGS、KF、MSC、SNV、SNV_Detrend、FD、SD 和 CR 8 种预处理方法中选择合适的预处理方法来校正水稻土土壤光谱,并用分区极值法来确定特征波段,通过选择合适预处理方法和特征波段来提高水稻土 SOM 含量预测模型的精度。

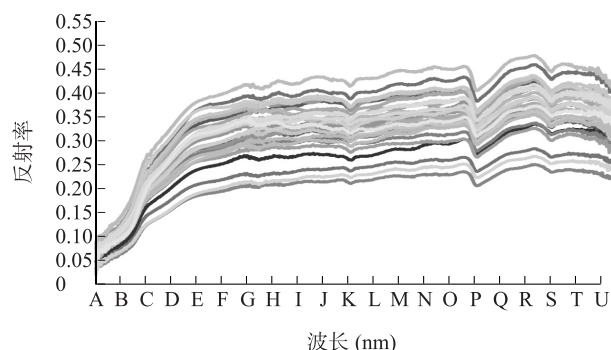
1 材料与方法

1.1 土壤样本的野外采集及实验室化学分析

33 个耕层土壤(0~20 cm)水稻土土壤样本采集于三峡库区王家沟小流域。采集的土壤样本带回实验室进行风干后,经过研磨、过筛处理后被分成 2 份,一份用于 SOM 含量的测定,一份用于光谱反射率的测定。土壤样本 SOM 含量用重铬酸钾容量法测定^[10]。水稻土土壤 SOM 含量的最小值、最大值分别为 13.20 g/kg、32.40 g/kg,均值和标准差分别为 19.24 g/kg、4.60 g/kg,偏度和峰度分别为 0.97 g/kg、0.96 g/kg,变异系数为 23.89%。

1.2 土壤样本光谱测量

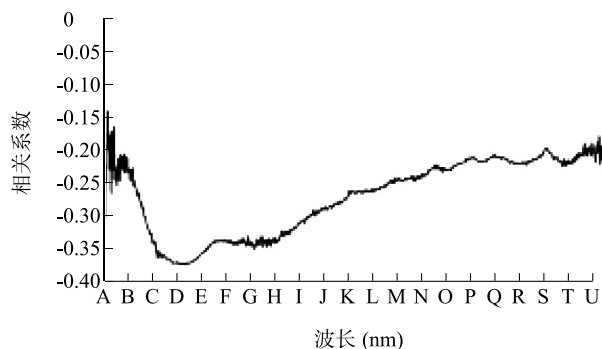
在室内使用美国 ASD (Analytical Spectral Devices) 公司生产的 FieldSpec Pro FR 野外光谱仪对风干的土壤样本反射光谱(波长范围 350~2500 nm)进行测量。测量时,将 ASD 光纤探头固定在精密三脚架上,垂直扫描土样。光纤探头距离所测土样约 3 cm,室内光源高 130 cm,以约 50°的俯角照射测量目标,测量目标位于光源光斑中心,保证视场内光源均匀,光源光谱尽量接近太阳光谱。保持标准版的清洁,并每隔 0.5 h 用标准版定标 1 次。为了保证测量的准确性,每个土壤样本重复测量 10 次。对原始光谱进行断点拟合处理来消除不同探测元件引起的光谱曲线阶跃性变化,然后剔除信噪比较低的边缘数据,使用 400~2450 nm 的光谱作为分析的原始光谱(图 1)。原始光谱与 SOM 含量在所有波段上都表现出了负相关(图 2),但在整个波长范围,SOM 含量与光谱反射率均表现为低度相关。



A:400 nm;B:500 nm;C:600 nm;D:700 nm;E:800 nm;F:900 nm;
G:1 000 nm;H:1 100 nm;I:1 200 nm;J:1 300 nm;K:1 400 nm;L:
1 500 nm;M:1 600 nm;N:1 700 nm;O:1 800 nm;P:1 900 nm;Q:
2 000 nm;R:2 100 nm;S:2 200 nm;T:2 300 nm;U:2 400 nm。

图1 土壤样本的原始光谱

Fig.1 Original spectra of soil samples



A:400 nm;B:500 nm;C:600 nm;D:700 nm;E:800 nm;F:900 nm;
G:1 000 nm;H:1 100 nm;I:1 200 nm;J:1 300 nm;K:1 400 nm;L:
1 500 nm;M:1 600 nm;N:1 700 nm;O:1 800 nm;P:1 900 nm;Q:
2 000 nm;R:2 100 nm;S:2 200 nm;T:2 300 nm;U:2 400 nm。

图2 土壤样本的原始光谱与SOM含量的相关系数

Fig.2 Correlation coefficients between original spectra of soil samples and soil organic matter (SOM) content

1.3 光谱数据预处理

分别用SGS、KF、MSC、SNV、SNV_Detrend、FD、SD和CR方法对土壤光谱进行预处理。KF用Vesper软件实现。SGS、MSC、SNV、SNV_Detrend、FD、SD和CR 7种预处理方法用R语言实现。SGS和求导处理后光谱两端的波段数会减少,为了保证各种预处理后波段数相同,先对350~2 500 nm的断点拟合光谱进行SGS和求导处理,然后截取400~2 450 nm的波段。SGS的局部多项式为2阶,窗口大小为9。设置求导的处理窗口,可解决求导产生噪声的问题。窗口大小分别设置为3、5、7、9、11、13、15、17,计算不同计算窗口得到的FD和SD光谱反射率与

SOM的相关系数,从中选择最大相关系数对应的窗口作为FD和SD的计算窗口,其值分别为11和15。

1.4 基于分区极值法的特征波段选择

对土壤光谱与SOM含量进行相关分析,根据二者之间的相关性选择特征波段是较常用的方法^[8,11-12]。在很多情况下,与SOM含量相关性最大的波段可能集中在一个区域(例如,SOM含量通常对400~1 000 nm的波段比较敏感^[13]),在这个区域之间的相关系数绝对值可能高于其他波段,如果只从该区域选择特征波段,往往会忽略其他波段对SOM含量的影响;而且由于相邻波段之间的高度相关性,选择1个特征波段和选择2个相邻的特征波段可能对模型精度的影响并不大。另一方面,在整个波长范围内,土壤光谱与SOM含量的相关系数通常并不呈现单调递增或单调递减的趋势,而是呈现波浪形的变化,在处理的波长范围内出现多个极大值,因此,采用分区极值法来选择特征波段。具体做法是设定相关系数绝对值的阈值 T ,利用 $y=T$ 和 $y=-T$ 两条直线将相关系数曲线划分为不同的区间,将各个区间内最大相关系数绝对值对应的波段作为特征波段。阈值 T 应该满足以下条件:(1)阈值 T 的取值范围为 $0.5 \leq T \leq 1$ 。根据相关的参考文献^[14-15],当两个变量之间相关系数的一般判断标准是 $0.5 \leq |R| < 0.8$ 时,表示两个变量显著相关;当 $0.8 \leq |R| < 1$ 时,表示两个变量之间具有高度相关性;当 $R=1$ 时,表示两个变量之间具有完全的线性相关性。因此,利用分区极值法选择特征波段,其阈值 T 的取值范围为 $0.5 \leq T \leq 1$ (T 的取值保留1位小数)。(2)相关系数绝对值大于 T 的波段的反射率与SOM含量之间的相关性应通过0.01水平的显著性检验。

1.5 建模方法及模型精度评价指标

建模方法采用偏最小二乘回归(Partial least squares regression, PLSR)方法,该方法在Unscrambler软件中实现。采用留一交叉验证法作为模型验证方法。决定系数(R^2)和最小均方根误差(Root-mean-square prediction error, RMSE)作为精度评价指标。

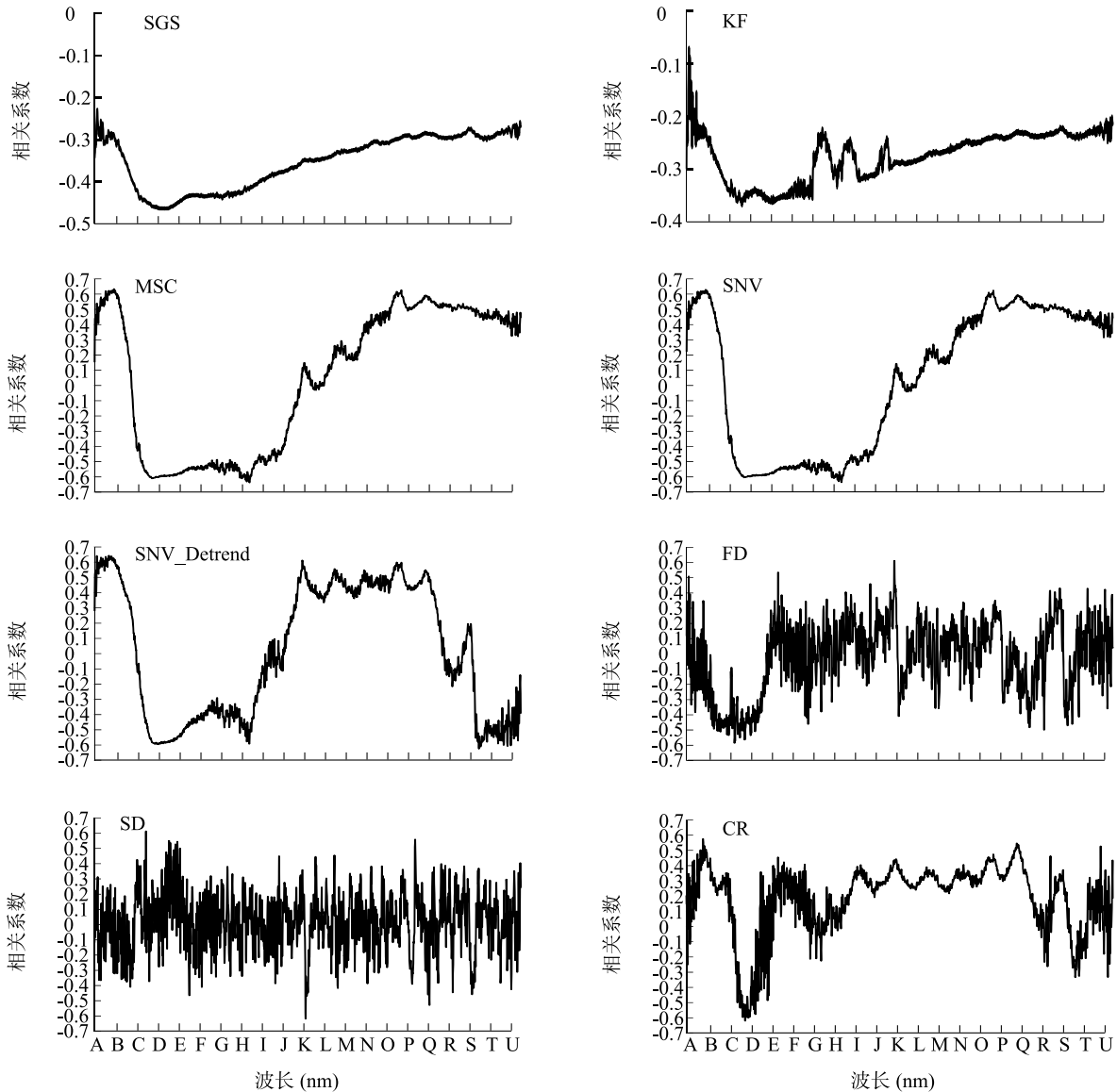
2 结果与分析

2.1 土壤样本光谱反射率与SOM含量的相关性

对SGS、KF、MSC、SNV、SNV_Detrend、FD、SD和CR处理后的土壤光谱反射率与SOM含量进行相关

性统计(图3)。SGS 光谱反射率与 SOM 含量之间的相关系数的值域区间与原始光谱基本相同(分别为 $-0.229 \sim -0.464$ 、 $-0.211 \sim -0.464$),主要原因主要是原始光谱曲线比较平滑,SGS 处理后光谱反射率变化较小。KF 光谱反射率与 SOM 含量也显示负相关,但相关性却低于原始光谱与 SOM 含量的相关性(值域为 $-0.071 \sim -0.369$),稀疏采样(采样间隔

100 m)可能是其主要影响因素。经过 MSC、SNV、SNV_Detrend、FD、SD 和 CR 处理得到的反射光谱与 SOM 含量的相关性都得到了显著的提高,出现了一些与 SOM 含量有显著相关性的波段,且其相关系数的最大绝对值都高于 0.6。因此,后续工作将只针对 MSC、SNV、SNV_Detrend、FD、SD 和 CR 处理得到的光谱进行特征波段的选择和建模。



SGS: Savitzky-Golay 平滑; KF: 克里克滤波; MSC: 多元散射校正; SNV: 标准正态化; SNV_Detrend: 标准正态化+去趋势; FD: 一阶导数; SD: 二阶导数; CR: 包络线去除。A: 400 nm; B: 500 nm; C: 600 nm; D: 700 nm; E: 800 nm; F: 900 nm; G: 1 000 nm; H: 1 100 nm; I: 1 200 nm; J: 1 300 nm; K: 1 400 nm; L: 1 500 nm; M: 1 600 nm; N: 1 700 nm; O: 1 800 nm; P: 1 900 nm; Q: 2 000 nm; R: 2 100 nm; S: 2 200 nm; T: 2 300 nm; U: 2 400 nm。

图3 不同预处理后的土壤光谱与 SOM 含量的相关关系

Fig.3 Correlation coefficients between spectral reflectance obtained by different pretreatment methods and SOM content

2.2 特征波段的选择

几种预处理得到的光谱反射率与 SOM 含量的

相关系数绝对值都未超过 0.7,且对于 FD 和 SD 处理,分别只有 2 个波段的光谱反射率与 SOM 含量的

相关系数绝对值超过 0.6,而对于 CR 处理,只有 1 个波段的反射率与 SOM 含量的相关系数绝对值大于 0.6,其他都小于 0.6。所以,为了不同预处理方法的比较,用分区极值法选择特征波段时,将阈值 T 统一设定为 0.5。而为了保证经过各种预处理后的 SOM 含量高光谱反演模型有相同数量的输入变量,选择最小分区数作为特征波段数($n=4$)。若某预处理方法的分区数大于 4,则将分区间隔窄的相邻区间合并为一个分区。各预处理方法所选择的特征波段见表 1。

表 1 6 种预处理方法所得光谱中选出的特征波段

Table 1 The feature bands in the spectra obtained by six pre-processing methods

预处理方法	特征波段
多元散射校正(MSC)	493 676 1 145 1 875
标准正态化(SNV)	492 676 1 145 1 875
标准正态化+去趋势(SNV_Detrend)	469 708 1 144 1 399
一阶导数(FD)	631 841 1 399 2 120
二阶导数(SD)	647 758 1 415 1 942
包络线去除(CR)	481 683 1 987 2 390

2.3 SOM 含量预测精度对预处理方法的响应

利用表 1 特征波段建立偏最小二乘回归模型,得到各模型精度(表 2)。6 种预处理方法中,CR 方法得到的 SOM 含量预测模型获得了最高的精度(标定集和验证集的 R^2 分别是 0.728 和 0.666, $RMSE$ 分别是 2.420 g/kg 和 2.770 g/kg)。其主要原因可能是该处理能够很好地揭示 SOM 吸收特征。经过 MSC、SNV、SNV_Detrend 3 种预处理得到的 SOM 含量预测模型的精度比较接近,其标定集和验证集的 R^2 均低于 0.500。其中,SNV 的预测精度最低,标定集的 R^2 和 $RMSE$ 分别为 0.463、3.398 g/kg,验证集的 R^2 和 $RMSE$ 分别为 0.415、3.667 g/kg。FD 光谱获得的 SOM 含量预测精度居中,SD 光谱获得的 SOM 含量预测精度高于 FD 光谱。FD 光谱能够校正基线漂移,但是由于供试土壤样本的类型、测量条件等不同,FD 光谱所获得的模型精度并不总是高于 SNV 光谱,如王海峰等^[1]就发现,无论是砂质土还是黏壤土,SNV 方法的精度要高于 FD 方法,其标定集和验证集的决定系数均达到了 0.8 以上。

表 2 6 种预处理方法的 SOM 含量预测精度

Table 2 Prediction accuracy of SOM content based on six pre-processing methods

预处理方法	标定集		验证集	
	R^2	$RMSE$ (g/kg)	R^2	$RMSE$ (g/kg)
多元散射校正(MSC)	0.465	3.395	0.416	3.664
标准正态化(SNV)	0.463	3.398	0.415	3.667
标准正态化+去趋势(SNV_Detrend)	0.467	3.387	0.418	3.657
一阶导数(FD)	0.532	3.175	0.470	3.489
二阶导数(SD)	0.537	3.157	0.490	3.421
包络线去除(CR)	0.728	2.420	0.666	2.770

R^2 : 决定系数; $RMSE$: 最小均方根误差。

2.4 SOM 含量预测精度对特征波段的响应

利用 CR 光谱获得与 SOM 含量具有最大相关系数绝对值的 4 个波段($\lambda=683$ 、690、698、699 nm)和所有相关系数绝对值大于 0.5 的波段(共有 77 个波段,其中 48 个波段位于 658~732nm 范围内)。利用这 2 组特征波段建立 PLSR 模型,将其与分区极值法得到的 4 个分区特征波段进行比较(表 3)。模型精度从高到低排列是:分区特征波段>4 个最大相关系数绝对值对应的波段>相关系数绝对值大于 0.5 的 77 个波段。这说明基于分区极值法选取的分区特征波段比按最大相关系数绝对值选择波段更好。这可能是因为最大相关系数绝对值所对应的几个波段之间重叠信息较多。分区特征波段不仅体现了 SOM 含量在整个波段范围内的分布特征,4 个波段彼此之间的重叠信息也较少(相关系数绝对值范围为 0.053~0.614)。利用所有相关系数绝对值大于 0.5 的波段得到了最低的模型精度,其原因可能是由于光谱反射率与 SOM 含量之间的高度线性相关性^[5]。

表 3 CR 处理后不同特征波段对模型精度的影响

Table 3 The effect of different feature bands on model accuracy after CR processing

特征波段	标定集		验证集	
	R^2	$RMSE$ (g/kg)	R^2	$RMSE$ (g/kg)
4 个最大相关系数绝对值对应的波段	0.439	3.476	0.361	3.832
相关系数绝对值大于 0.5 的 77 个波段	0.356	3.724	0.284	4.055
4 个分区特征波段	0.728	2.420	0.666	2.770

R^2 : 决定系数; $RMSE$: 最小均方根误差。

3 结 论

SGS 处理并不对 SOM 含量与光谱之间的相关性有明显的改善,KF 方法降低了 SOM 含量与土壤光谱之间相关性(KF 光谱与 SOM 含量的相关系数最大绝对值低于 0.4),这说明 SGS 和 KF 2 种方法对于试验样本并不是有效的光谱预处理方法。

MSC、SNV、SNV_Detrend、FD、SD 和 CR 处理后,SOM 含量与土壤光谱的相关性都得到了显著提高,其相关系数最大绝对值都超过了 0.6。利用预处理后的光谱建立 SOM 含量预测 PLSR 模型的精度从高到低分别是:CR>SD>FD>SNV_Detredn>MSC>SNV。CR 处理后得到的光谱曲线具有明显的吸收和反射特征,该方法取得了最高的预测精度(其标定集和验证集的决定系数(R^2)分别是 0.728、0.666, RMSE 分别是 2.240 g/kg 和 2.770 g/kg)。

对 CR 处理后的光谱分别利用 4 个分区特征波段、4 个最大相关系数绝对值对应的波段和相关系数绝对值大于 0.5 的 77 个波段进行 PLSR 建模,分区特征波段获得了最高的预测精度。这说明用分区极值法选择特征波段,能更多考虑不同区间分布的光谱对 SOM 含量的影响,对提高 SOM 含量预测精度更有效。

参考文献:

- [1] 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[J]. 农业工程学报, 2018, 34(14): 124-131.
- [2] 郑曼迪,熊黑钢,乔娟峰,等. 基于综合光谱指数的不同程度人类干扰下土壤有机质含量预测[J]. 江苏农业学报, 2018, 34(5): 1048-1056.
- [3] QIAO X X, WANG C, FENG M C, et al. Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques[J]. Spectroscopy Letters, 2017, 50(3): 156-163.
- [4] WANG C, FENG M C, YANG W D, et al. Use of spectral character to evaluate soil organic matter[J]. Soil Science Society of America Journal, 2016, 80(4): 1078-1088.
- [5] 南 锋,朱洪芬,毕如田. 黄土高原煤矿区复垦农田土壤有机质含量的高光谱预测[J]. 中国农业科学, 2016, 49(11): 2126-2135.
- [6] LUYPAERT J, HEUERDING S, VANDER H Y, et al. The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams[J]. Journal of Pharmaceutical and Biomedical Analysis, 2004, 36(3): 495-503.
- [7] 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[J]. 农业工程学报, 2015, 31(14): 103-109.
- [8] 张东辉,赵英俊,秦 凯,等. 光谱变换方法对黑土养分含量高光谱遥感反演精度的影响[J]. 农业工程学报, 2018, 34(20): 141-147.
- [9] VISCARRA R R, TAYLOR H, MCBRATNEY A. Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing[J]. European Journal of Soil Science, 2007, 58(1): 343-353.
- [10] 杨剑虹,王成林,代亨林. 土壤农化分析与环境监测[M]. 北京:中国大地出版社, 2008.
- [11] 尼加提·卡斯木,茹克亚·萨吾提,师庆东,等. 基于优化光谱指数的土壤有机质含量估算[J]. 农业机械学报, 2018, 49(11): 155-163.
- [12] 侯艳军,塔西甫拉提·特依拜,买买提·沙吾提,等. 荒漠土壤有机质含量高光谱估算模型[J]. 农业工程学报, 2014, 30(16): 113-120.
- [13] 刘 炜,常庆瑞,郭 曼,等. 土壤导数光谱小波去噪与有机质吸收特征提取[J]. 光谱学与光谱分析, 2011, 31(1): 100-104.
- [14] 吴可杰. 统计学原理[M]. 南京:南京大学出版社, 1986.
- [15] 朱晓青. 统计学原理[M]. 北京:国防工业出版社, 1994.

(责任编辑:张震林)