

牛芳鹏, 李新国, 麦麦提吐尔逊·艾则孜, 等. 基于光谱指数的博斯腾湖西岸湖滨绿洲土壤有机碳含量估算模型[J]. 江苏农业学报, 2022, 38(2): 414-421.

doi: 10.3969/j.issn.1000-4440.2022.02.015

# 基于光谱指数的博斯腾湖西岸湖滨绿洲土壤有机碳含量估算模型

牛芳鹏<sup>1,2</sup>, 李新国<sup>1,2</sup>, 麦麦提吐尔逊·艾则孜<sup>1,2</sup>, 赵 慧<sup>1,2</sup>, 江远东<sup>1,2</sup>

(1. 新疆师范大学地理科学与旅游学院, 新疆 乌鲁木齐 830054; 2. 新疆干旱区湖泊环境与资源实验室, 新疆 乌鲁木齐 830054)

**摘要:** 以博斯腾湖西岸湖滨绿洲为研究区, 将野外原位高光谱实测数据和土壤有机碳(SOC)含量作为基础数据, 通过对原始光谱进行4种数学变换, 探索不同光谱变换形式下的弓曲差( $C$ )、差值光谱指数( $DSI$ )、简单比值土壤指数( $RSI$ )、亮度光谱指数( $BSI$ )、归一化土壤指数( $NDSI$ )与SOC含量的关系, 并建立基于随机森林法(RF)的SOC含量估算模型。结果表明: (1) 研究区SOC含量主要集中在5.25~78.76 g/kg, 平均值为21.82 g/kg, 变异系数为69.11%, 呈中等变异性; (2) 在光谱数据 $\lg R$ 下, SOC含量与 $DIS$ 指数相关系数最高, 相关系数为0.80, 最佳组合波段为(1 758 nm, 1 752 nm); (3) 基于不同光谱指数与弓曲差( $C$ )建立的模型验证集精度 $R^2$ 和 $RMSE$ 分别介于0.67~0.84和5.85~8.45 g/kg, 模型的 $RPD$ 均在1.66以上; 在基于光谱数据 $\lg(1/R)$ 变换下, 模型的验证集 $R^2=0.82$ ,  $RMSE=3.52$  g/kg,  $RPD=3.99$ , 可以较好地估算研究区SOC含量, 为干旱半干旱地区湖滨绿洲SOC含量反演提供依据和参考。

**关键词:** 土壤有机碳含量; 估算模型; 光谱指数; 随机森林; 湖滨绿洲

**中图分类号:** S153.6 **文献标识码:** A **文章编号:** 1000-4440(2022)02-0414-08

## Estimation model of soil organic carbon content in lakeside oasis on the west coast of Bosten Lake based on spectral index

NIU Fang-peng<sup>1,2</sup>, LI Xin-guo<sup>1,2</sup>, MAMATTURSUN · Eziz<sup>1,2</sup>, ZHAO Hui<sup>1,2</sup>, JIANG Yuan-dong<sup>1,2</sup>

(1. College of Geographic Sciences and Tourism, Xinjiang Normal University, Urumqi 830054, China; 2. Xinjiang Laboratory of Lake Environment and Resources in Arid Zone, Urumqi 830054, China)

**Abstract:** Taking the lakeside oasis on the west coast of Bosten Lake as the study area, and the in situ hyperspectral data and soil organic carbon (SOC) content were taken as the basic data. Four mathematical transformations were performed on the original spectrum to explore the relationships between bow curvature difference ( $C$ ), difference spectral index ( $DSI$ ), simple ratio soil index ( $RSI$ ), brightness spectral index ( $BSI$ ), normalized soil index ( $NDSI$ ) and SOC content.

A SOC content estimation model based on random forest method was established. The results indicated that the content of SOC was mainly concentrated in 5.25–78.76 g/kg, the average value was 21.82 g/kg, the coefficient of variation was 69.11%, showing moderate variability. Under the spectral data  $\lg R$ , the correlation coefficient between SOC content and  $DIS$  index was the highest, the correlation coefficient was 0.80, and the best combination band was (1 758 nm, 1 752 nm). The  $R^2$  and  $RMSE$  of the model validation set based on different spectral indices and  $C$

收稿日期: 2021-07-11

基金项目: 国家自然科学基金项目(41661047、U2003301); 新疆维吾尔自治区重点实验室开放课题(2018D04026)

作者简介: 牛芳鹏(1995-), 男, 甘肃庄浪人, 硕士研究生, 研究方向为干旱区土壤资源变化及其遥感应用研究。(E-mail) niufp0225@163.com

通讯作者: 李新国, (E-mail) onlinelxg@sina.com

were 0.67–0.84 and 5.85–8.45 g/kg, respectively, and the *RPD* of the model was above 1.66. Based on the transformation of spectral data  $\lg(1/R)$ , the  $R^2$ , *RMSE* and *RPD* of the model validation set were 0.82, 3.52 g/kg and 3.99, respectively. The model can better estimate the SOC content in the study area, and the research results can provide the basis and reference for the inversion of SOC content in lakeside oasis in arid and semi-arid regions.

**Key words:** soil organic carbon content; estimation model; spectral index; random forest; lakeside oasis

土壤有机碳(Soil organic carbon, SOC)含量是评价土壤质量的关键参数,对维护区域生态平衡有重要意义<sup>[1]</sup>。在人类活动的影响下, SOC 循环的加速变化引起了土壤环境的改变,其储量、密度和分布特征等问题受到学者们的高度关注<sup>[2-3]</sup>。传统的土壤 SOC 含量估算方法繁琐,成本高,高光谱技术可以满足大面积地区快速、准确地估算 SOC 含量的需要。国内外学者对于干旱、半干旱地区森林、草地、湿地和农田等不同生态系统类型的 SOC 估测作了大量研究<sup>[4-6]</sup>。早期学者们通过相关分析法与机器算法研究了 SOC 含量与不同变换形式下土壤光谱反射率的关系,并挑选特征波段进行 SOC 估算模型构建。叶勤等采用光谱变换与相关分析方法,基于主成分分析结果构建了土壤有机质含量(Soil organic matter, SOM)的 BP 神经网络模型,验证结果决定系数( $R^2$ )为0.893 0,均方根误差(*RMSE*)为0.118 5%<sup>[7]</sup>。AMIN 等通过对光谱数据转换分析,得到了一个在 SOC 含量3.80~6.71 g/kg范围内比较精确的模型, $R^2$ 、*RMSE* 和相对分析误差(*RPD*)分别为0.85、3.77 g/kg和2.54<sup>[8]</sup>。近年来光谱指数、弓曲差、多变量回归建模方法广泛应用于 SOC 含量定量估算模型中,基于不同预处理构建的二维光谱指数能有效地削弱噪声干扰<sup>[9-10]</sup>。赵明松等建立了光谱特征指数、弓曲差与 SOM 含量的预测模型,经检验  $R^2$  介于0.67~0.73, *RMSE* 介于3.21~3.51 g/kg<sup>[11]</sup>。张子鹏等比较了不同维度光谱指数下支持向量机(Support vector machine, SVM)对 SOM 含量估测的建模效果<sup>[12]</sup>。HONG 等研究结果表明随机森林(Random forest, RF)和连续小波变换(Continuous wavelet transform, CWT)构建的 SOC 含量预测模型具有很好的鲁棒性,交叉验证结果  $R^2 = 0.87$ <sup>[13]</sup>。张智韬等利用 RF 与 SVM 分类方法构建了荒漠区 SOC 含量预测模型,验证集  $R^2 = 0.979$ , *RMSEP* = 0.481 g/kg, *RPD* = 7.004<sup>[14]</sup>。

与以往采土后室内测量光谱相比较,基于野外原位测量土壤高光谱数据能保持土壤原有属性,建

立的光谱模型更加准确;室内光谱测量土样需要将其进行研磨和过筛<sup>[15]</sup>,破坏了土壤原有结构,尽管有利于土壤有机碳光谱特征研究,但实用性较差,难以用于野外快速测量土壤理化性质。野外原位测量时为将影响土壤光谱的因素(成土母质、土壤水分、粗糙度等)降到最低,选择土壤干燥程度、土块大小相对均一的样地进行测量,并通过光谱预处理技术进一步减弱反射光谱受到的噪声影响<sup>[12]</sup>。基于此本研究以新疆博斯腾湖西岸湖滨绿洲为靶区,以野外原位实测光谱数据为基础,探索不同光谱预处理下二维光谱指数、弓曲差与 SOC 含量间的关系,并在此基础上利用偏最小二乘回归(Partial least squares regression, PLSR)和随机森林(RF)2种机器学习算法构建研究区 SOC 含量高光谱估算模型,以期提高研究区土壤 SOC 含量的估算精度,为快速有效测定湖滨绿洲 SOC 含量提供技术支持。

## 1 材料与方法

### 1.1 研究区概况

博斯腾湖西岸湖滨绿洲(41°45'~42°10' N, 86°15'~86°55' E)是典型的山前湖泊绿洲,地处新疆焉耆盆地,隶属于新疆维吾尔自治区博湖县(图1)。地势由西北向东南倾斜,夏季月平均气温22.8℃,冬季平均气温9.0℃,无霜期176~200 d,多年平均降水量83.55 mm,年均温8.0~8.6℃,季节过渡快,属于大陆性荒漠气候。主要植被类型有山杨、柽柳和梭梭等,主要土壤类型有草甸土、绿洲潮土、沼泽土和荒漠林土等<sup>[16]</sup>。

### 1.2 材料与方法

**1.2.1 样品采集与光谱处理分析** 样品采集时间为2020年10月,均匀布点,共采集土壤样品47个,涵盖研究区主要土地利用类型,包括耕地、林地、盐渍地和非利用地。采集深度为0~20 cm,剔除杂物后,采用四分法称取200 g土壤,装袋,带回实验室进行自然干燥,研磨后过0.25 mm孔筛。SOC 含量测定采用重铬酸钾容量法(外加热法),单位为

g/kg<sup>[17]</sup>。光谱数据通过 ASD FieldSpec3 地物光谱仪于野外原位采集,在无风或风力小于 3 级,云量小于 5% 的晴朗天气,北京时间 12:00–14:00 进行测试,测量地点选择较为开阔地块,且尽量远离可能干扰土壤光谱的物体,光纤探头的视场角不大于 25°,垂直放置在土样表面 15 cm 处,使用前仪器预热 30 min,并进行白板校正,以消除暗电流影响。测试过程中,为减少外部环境带来的误差,每个采样点随机选择 2 个测试点,每个测试点测量 5 次,对 10 条光

谱曲线取均值后作为该土样的反射光谱值<sup>[18]</sup>。为提高光谱数据的信噪比,增强光谱曲线的峰谷变化,采用 Savitaky-Golay(2 次多项式,5 个点)平滑去噪,对原始光谱反射率进行对数变换( $\lg R$ )、对数的倒数变换 $[\lg(1/R)]$ 、均方根变换( $\sqrt{R}$ )和包络线去除( $R_c$ )。统一去除受大气水汽影响较大的 1 300~1 450 nm 和 1 800~1 950 nm 波段以及噪声较大的尾部波段(2 451~2 500 nm)<sup>[19-20]</sup>。运用 MATLAB 软件进行光谱指数的构建与模型的建立。

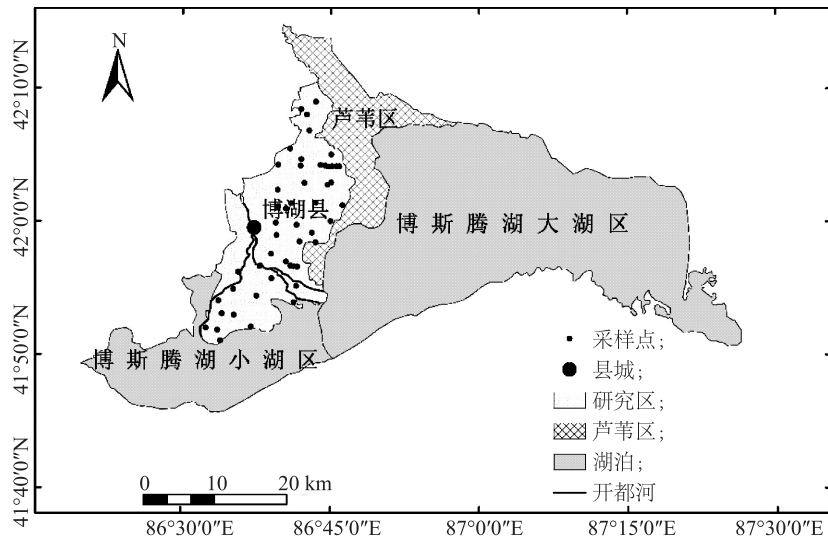


图 1 研究区地理位置图

Fig.1 Geographic location map of the study area

### 1.2.2 土壤光谱指数的构建

1.2.2.1 弓曲差(600 nm 处)计算 弓曲差( $C$ )即光谱曲线 550 nm 和 650 nm 处的光谱反射率均值与 600 nm 处的光谱反射率之差值<sup>[21]</sup>。

$$C = R_{600} - (R_{550} + R_{650}) / 2 \quad (1)$$

1.2.2.2 光谱指数构建 采用以下指数构建土壤光谱指数:差值光谱指数( $DSI$ )、简单比值土壤指数( $RSI$ )、亮度光谱指数( $BSI$ )、归一化土壤指数( $ND-SI$ )<sup>[22-23]</sup>。

$$DSI = R_j - R_i \quad (2)$$

$$RSI = R_j / R_i \quad (3)$$

$$BSI = \sqrt{R_j^2 + R_i^2} \quad (4)$$

$$NDSI = (R_j - R_i) / (R_j + R_i) \quad (5)$$

### 1.3 随机森林模型

随机森林是一种基于分类树的算法,主要利用随机森林分类器(RF)对新数据进行判别和分类,并通过总结大量的分类树来提高模型的预测

精度<sup>[24]</sup>。在处理比较大的数据时,运算速度非常快,不需要变量选择。在计算变量的非线性效应时,它对异常值不敏感,能够充分反映变量之间的相互作用。与一般分类算法相比,RF 算法结合了随机特征选择和 Bagging 算法 2 种技术,能够很好地容忍噪声和离群点的影响,能够同时处理连续和离散数据<sup>[25]</sup>。

### 1.4 估测模型构建与验证

在 47 个样本中,随机抽取 31 个作为建模集,16 个作为验证集。采用 Matlab 软件构建 PLSR 模型与 RF 模型。选取决定系数( $R^2$ )、均方根误差( $RMSE$ )和相对分析误差( $RPD$ )检验模型的准确性。 $R^2$ 取值为 0~1.00,当  $R^2$  越大、 $RMSE$  越小,模型的预测能力越好,适用性越高。 $RPD$  值可以用来解释模型的预测能力,当  $RPD < 1.40$  时,模型不能准确预测;当  $1.40 \leq RPD < 2.00$  时,模型的预测能力一般; $RPD \geq 2.00$  时,模型具有较高的预测能力<sup>[26]</sup>。

## 2 结果与分析

### 2.1 土壤有机碳含量统计分析

由表 1 可知,研究区 SOC 含量主要集中在 5.25~78.76 g/kg,平均值为 21.82 g/kg,呈中等变

表 1 土壤有机碳含量描述性统计

Table 1 Descriptive statistics of soil organic carbon content

样本集	样本数	最大值 (g/kg)	最小值 (g/kg)	平均值 (g/kg)	标准差 (g/kg)	变异系数 (%)
样本总集	47	78.76	5.25	21.82	15.08	69.11
建模集	31	78.76	5.25	22.03	15.81	71.77
验证集	16	55.30	6.39	21.42	14.03	65.50

### 2.2 基于弓曲差的土壤有机碳含量建模

利用 5 种不同光谱数据的弓曲差建立研究区 SOC 含量的 PLSR 和 RF 模型。由表 2 可知,原始光谱数据( $R$ )的弓曲差构建的 PLSR 与 RF 模型效果较差,验证集  $R^2$  分别为 0.65 和 0.69,RPD 分别为 1.14 和 2.07。PLSR 线性模型中,光谱数据  $\lg(1/R)$  的弓曲差模型效果最好,模型  $R^2$  为 0.69, RMSE 为 12.75

异性,变异系数为 69.11%。随机划分建模集与验证集, SOC 含量平均值分别为 22.03 g/kg 和 21.42 g/kg,变异系数分别为 71.77% 和 65.50%,两者相差不大,能较好地反映研究区 SOC 含量平均水平。

g/kg, RPD 为 1.10; 光谱数据  $R_c$  的弓曲差模型效果最差,  $R^2$  为 0.51, RMSE 为 13.03 g/kg, RPD 为 1.08。与 PLSR 模型相比较, RF 模型预测能力均有所提升, RPD 均大于 2.00; 基于光谱数据  $\sqrt{R}$  的弓曲差构建的 RF 模型效果最好,  $R^2$  为 0.83, RMSE 为 6.14 g/kg, RPD 为 2.29。

表 2 基于弓曲差的土壤有机碳含量 PLSR 与 RF 模型

Table 2 PLSR and RF models of soil organic carbon content based on bow curvature difference

光谱数据 变换形式	PLSR 模型					RF 模型				
	建模集		验证集		$\sqrt{R}$	建模集		验证集		$\sqrt{R}$
	$R^2$	RMSE(g/kg)	$R^2$	RMSE(g/kg)		$R^2$	RMSE(g/kg)	$R^2$	RMSE(g/kg)	RPD
$R$	0.54	15.24	0.65	12.28	1.14	0.60	8.95	0.69	6.78	2.07
$\lg R$	0.52	15.36	0.58	12.75	1.10	0.66	8.39	0.70	6.73	2.08
$\lg(1/R)$	0.52	15.36	0.69	12.75	1.10	0.66	8.51	0.73	6.61	2.12
$\sqrt{R}$	0.64	15.25	0.64	12.38	1.13	0.70	8.87	0.83	6.14	2.29
$R_c$	0.51	15.45	0.51	13.03	1.08	0.61	7.97	0.79	6.32	2.22

$R$ : 原始光谱反射率;  $\lg R$ : 原始光谱反射率的对数变换;  $\lg(1/R)$ : 原始光谱反射率倒数的对数变换;  $\sqrt{R}$ : 原始光谱反射率的均方根变换;  $R_c$ : 原始光谱反射率的包络线去除。  $R^2$ : 决定系数; RMSE: 均方根误差; RPD: 相对分析误差。

### 2.3 光谱指数与土壤有机碳含量相关性分析

由图 2 可知,不同光谱数据变换下,不同光谱指数与土壤有机碳含量相关系数二维图存在明显差异。在原始光谱反射率  $R$  中, SOC 含量与  $DSI$  指数、 $RSI$  指数和  $NDSI$  指数的最佳组合波段均为 (481 nm, 480 nm)、(942 nm, 941 nm)、(999 nm, 993 nm), 最大相关系数分别为 -0.68、0.66、0.77; 与  $BSI$  指数的最大相关系数为 -0.65, 最佳组合波段为 (1 955 nm, 1 851 nm)。与  $DSI$  指数相关性较高的组合波段主要分布在 412~548 nm、586~701 nm、794~871 nm 和 2 112~2 162

nm。在  $\lg R$  光谱数据数据下, SOC 含量与  $DIS$  指数、 $RSI$  指数、 $BSI$  指数和  $NDSI$  指数最大相关系数分别为 0.80、-0.70、0.76、-0.69, 最佳组合波段分别为 (1 758 nm, 1 752 nm)、(943 nm, 941 nm)、(1 833 nm, 677 nm)、(999 nm, 993 nm); 在  $\lg(1/R)$  光谱数据下, SOC 含量与  $DIS$  指数、 $RSI$  指数、 $BSI$  指数和  $NDSI$  指数最大相关系数分别为 -0.77、-0.68、0.76、-0.70, 最佳组合波段分别为 (1 047 nm, 1 046 nm)、(942 nm, 941 nm)、(1 838 nm, 676 nm)、(943 nm, 941 nm); 在  $\sqrt{R}$  光谱数据下, SOC 含量与  $DIS$  指数、 $RSI$  指数、 $BSI$  指数



和  $NDSI$  指数最大相关系数分别为 0.77、0.77、0.76、-0.68, 最佳组合波段分别为 (1 757 nm, 1 752 nm)、(1 047 nm, 1 045 nm)、(1 838 nm, 676 nm)、(1 851 nm, 677 nm); 在  $R_c$  光谱数据下, SOC 含量与  $DIS$  指

数、 $RSI$  指数、 $BSI$  指数和  $NDSI$  指数最大相关系数分别为 0.75、0.76、0.69、0.73, 最佳组合波段分别为 (942 nm, 941 nm)、(945 nm, 944 nm)、(2 494 nm, 1 861 nm)、(1 756 nm, 1 753 nm)。

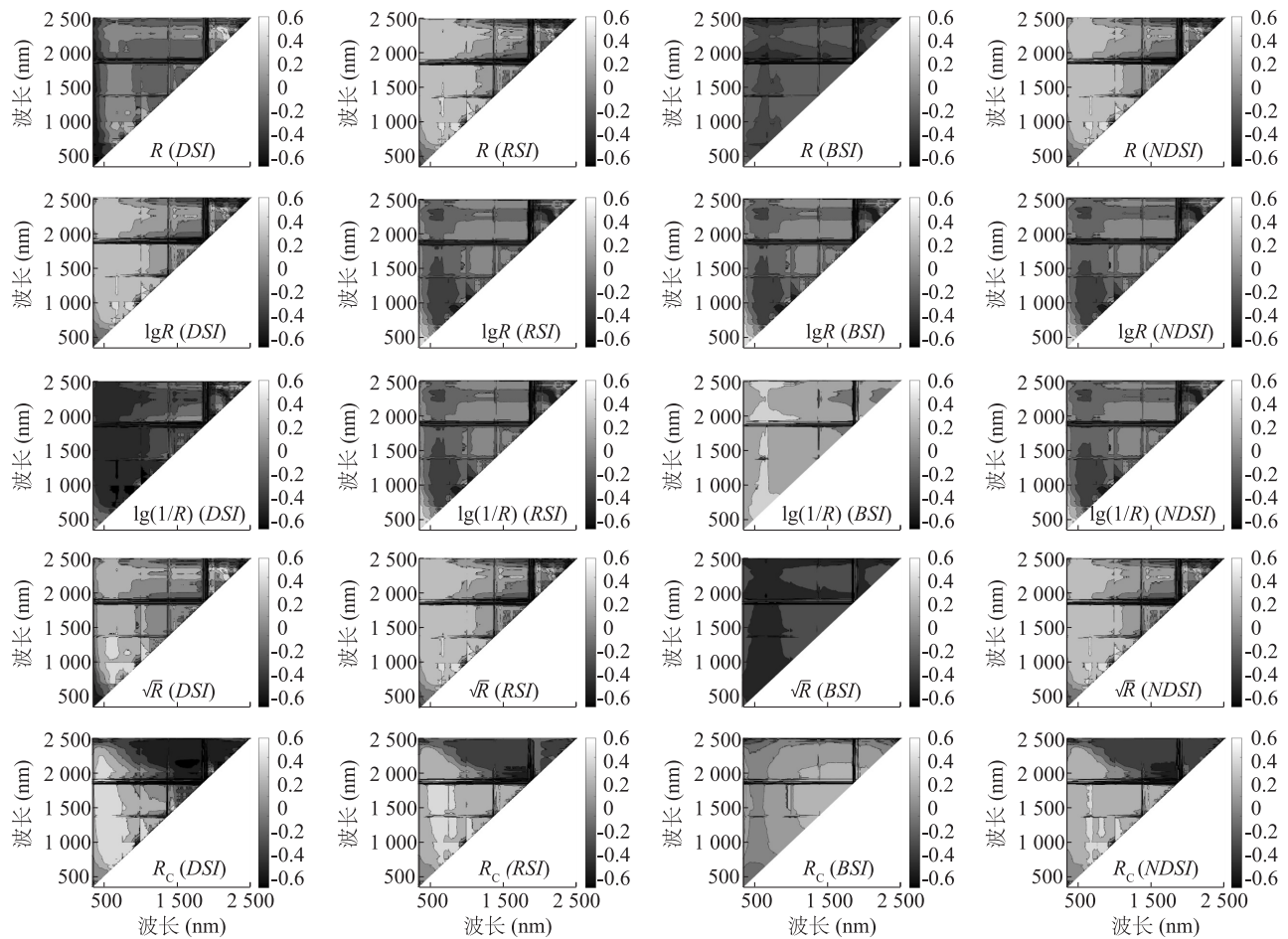


图 2 不同光谱数据变换下 4 种光谱指数与土壤有机碳含量的相关系数二维图

Fig.2 Two-dimensional diagram of correlation coefficient between four spectral indices and soil organic carbon content under different spectral data transformation

## 2.4 基于光谱指数的土壤有机碳含量建模

分别将光谱数据  $R$ 、 $\lg R$ 、 $\lg(1/R)$ 、 $\sqrt{R}$  和  $R_c$  的  $DIS$  指数、 $RSI$  指数、 $BSI$  指数和  $NDSI$  指数及其弓曲差组合作为 SOC 含量估算模型所需的独立变量, 基于 RF 方法分别构建 SOC 含量的高光谱估算模型。由表 3 可知, 在原始光谱数据  $R$  中,  $DIS$  指数与  $NDSI$  指数的 SOC 模型精度最高,  $R^2$  均为 0.75,  $RPD > 2.00$ ; 在光谱数据  $\lg R$  中,  $BSI$  指数的 SOC 模型精度最高,  $R^2$  为 0.84,  $RPD$  为 2.14,  $DIS$  指数的 SOC 模型精度最低,  $R^2$  为 0.73; 在光谱数据  $\lg(1/R)$  中,  $DIS$  指数的 SOC 模型精度最高,  $R^2$  为 0.82,  $RMSE$  为

6.44 g/kg,  $RPD$  为 2.18; 在光谱数据  $\sqrt{R}$  与  $R_c$  中,  $RSI$  指数的 SOC 模型精度最高,  $R^2$  分别为 0.81、0.76,  $RMSE$  分别为 6.67 g/kg、5.67 g/kg,  $RPD$  分别为 2.10、2.47。20 个预测模型中, 基于光谱  $\sqrt{R}$  的  $RSI$  指数 RF 模型的预测能力最优, 能够粗略估算研究区的 SOC 含量。将 5 种光谱数据的  $DIS$  指数、 $RSI$  指数、 $BSI$  指数和  $NDSI$  指数分别与弓曲差组合进行 SOC 的 RF 建模。结果表明, 由于采用了较多的光谱特征指数, SOC 含量的建模效果整体上有不同程度的提升, 建模集的  $R^2$  和  $RMSE$  分别介于 0.69~0.81 和 7.99~9.01 g/kg, 验证集的  $R^2$  和

$RMSE$  分别介于 0.67~0.84 和 5.85~8.45 g/kg, 模型的  $RPD$  均在 1.66 以上, 其预测能力有不同程度

的提升, 基于  $\sqrt{R}$  光谱指数和弓曲差的 SOC 含量估算模型的泛化能力最高。

表 3 基于光谱指数的土壤有机碳含量建模

Table 3 Modeling of soil organic carbon content based on spectral index

光谱数据 变换形式	光谱指数	建模集		验证集		$RPD$	光谱数据 变换形式	光谱指数	建模集		验证集		$RPD$
		$R^2$	$RMSE$ (g/kg)	$R^2$	$RMSE$ (g/kg)				$R^2$	$RMSE$ (g/kg)	$R^2$	$RMSE$ (g/kg)	
$R$	$DSI$	0.68	6.99	0.75	6.21	2.26	$R$	$DSI+C$	0.69	9.01	0.70	6.49	2.16
	$RSI$	0.66	7.10	0.71	6.51	2.16		$RSI+C$	0.70	8.95	0.75	6.55	2.14
	$NDSI$	0.69	6.96	0.75	6.76	2.08		$NDSI+C$	0.70	8.82	0.76	6.95	2.02
	$BSI$	0.66	7.18	0.70	7.94	1.77		$BSI+C$	0.70	8.98	0.71	6.85	2.05
$\lg R$	$DSI$	0.68	7.10	0.73	7.62	1.84	$\lg R$	$DSI+C$	0.75	8.45	0.75	6.48	2.17
	$RSI$	0.69	6.98	0.77	6.88	2.04		$RSI+C$	0.77	8.35	0.79	6.28	2.23
	$NDSI$	0.69	6.95	0.76	7.10	1.98		$NDSI+C$	0.76	8.32	0.82	6.56	2.14
	$BSI$	0.67	7.03	0.84	6.57	2.14		$BSI+C$	0.75	8.45	0.82	6.60	2.13
$\lg(1/R)$	$DSI$	0.70	6.97	0.82	6.44	2.18	$\lg(1/R)$	$DSI+C$	0.75	8.45	0.76	8.45	1.66
	$RSI$	0.72	6.83	0.79	6.82	2.06		$RSI+C$	0.75	6.38	0.67	8.36	1.68
	$NDSI$	0.70	6.96	0.78	6.86	2.05		$NDSI+C$	0.76	8.30	0.77	5.85	2.40
	$BSI$	0.75	7.17	0.76	8.90	1.58		$BSI+C$	0.74	8.58	0.77	6.39	2.20
$\sqrt{R}$	$DSI$	0.50	6.96	0.78	6.86	2.05	$\sqrt{R}$	$DSI+C$	0.71	8.89	0.80	6.28	2.23
	$RSI$	0.50	6.86	0.81	6.67	2.10		$RSI+C$	0.71	8.85	0.83	6.01	2.33
	$NDSI$	0.69	7.40	0.70	7.65	1.83		$NDSI+C$	0.71	8.76	0.84	6.11	2.30
	$BSI$	0.74	7.13	0.77	7.86	1.78		$BSI+C$	0.72	8.75	0.77	5.97	2.35
$R_c$	$DSI$	0.70	5.36	0.77	5.65	2.48	$R_c$	$DSI+C$	0.79	8.00	0.80	6.61	2.12
	$RSI$	0.74	5.08	0.76	5.67	2.47		$RSI+C$	0.80	7.99	0.83	6.57	2.14
	$NDSI$	0.74	4.96	0.75	5.65	2.48		$NDSI+C$	0.81	8.05	0.78	6.30	2.23
	$BSI$	0.73	5.05	0.72	5.83	2.41		$BSI+C$	0.80	8.13	0.74	6.58	2.13

光谱数据变换形式见表 2 注。 $R^2$ : 决定系数;  $RMSE$ : 均方根误差;  $RPD$ : 相对分析误差。 $DSI$ : 差值光谱指数;  $RSI$ : 简单比值土壤指数;  $NDSI$ : 归一化土壤指数;  $BSI$ : 亮度光谱指数;  $C$ : 弓曲差。

在光谱数据变换的基础上, 联合 4 种光谱指数与弓曲差为变量, 进行模型的构建, 由表 4 可知, 光谱数据  $R_c$  建立的模型验证集  $R^2$  为 0.63,  $RMSE$  为 8.42 g/kg,  $RPD$  为 1.67, 相比于其他光谱数据建立的模型估算效果较差; 光谱数据  $\sqrt{R}$  建立的模型效果最好,  $R^2$  为 0.79, 但验证集  $R^2$  为 0.81, 估算效果较弱于  $\lg(1/R)$  模型; 光谱数据  $\lg(1/R)$  建立的模型的验证集  $R^2$  为 0.82,  $RMSE$  为 3.52 g/kg,  $RPD$  为 3.99。通过综合评估 5 种光谱数据建立的模型评价参数可知, 对 SOC 含量估算的建模效果依次为  $\lg(1/R) > \sqrt{R} > \lg R > R > R_c$ , 说明利用光谱数据  $\lg(1/R)$  建立的模型可以很好地估算研究区 SOC 含量。

表 4 基于光谱数据变换后的土壤有机碳含量建模

Table 4 Modeling of soil organic carbon content based on spectral data transformation

光谱数据 变换形式	建模集		验证集		$RPD$
	$R^2$	$RMSE$ (g/kg)	$R^2$	$RMSE$ (g/kg)	
$R$	0.71	6.88	0.71	4.51	3.11
$\lg R$	0.70	6.89	0.73	4.30	3.26
$\lg(1/R)$	0.70	6.85	0.82	3.52	3.99
$\sqrt{R}$	0.79	6.85	0.81	3.82	3.67
$R_c$	0.64	8.48	0.63	8.42	1.67

光谱数据变换形式见表 2 注。 $R^2$ : 决定系数;  $RMSE$ : 均方根误差;  $RPD$ : 相对分析误差。

### 3 讨论

现有的研究大多是基于典型特征波段或单一光谱指数建立 SOC 含量估算模型,但这些数据的光谱信息往往是饱和的,通过对光谱数据进行数学变换和二维指数提取,可以有效地压缩噪声对目标光谱的影响,充分考虑波段间的相关性。张娟娟等基于 *DSI* 指数对土壤有机质含量进行了预测,  $R^2$  为 0.79,  $RMSE$  为 4.71 g/kg,  $RPD$  为 2.30<sup>[27]</sup>。洪永胜等以江汉平原公安县为研究区,发现二维相关性分析提取的敏感光谱指数的 PLSR 建模效果优于一维相关性分析提取的敏感波段,其中 *NDSI* 指数模型的鲁棒性最好,验证集  $R^2$  为 0.82,  $RPD$  值为 2.46,模型稳定可靠<sup>[28]</sup>。焦彩霞等认为,由三波段建立的 600 nm 弓曲差光谱指数对数函数模型可以更好地实现土壤有机质含量的预测<sup>[29]</sup>。以上学者采用土壤光谱指数作为模型因子来估算土壤有机质含量,但由于没有考虑各种预处理方法对光谱指数的影响,因此模型的构建具有局限性。基于光谱数据不同变换形式,本研究以 2 种思路构建了 SOC 含量模型,第一种方法利用弓曲差与光谱指数构建 PLSR 线性与 RF 非线性的估算模型,另一种方法利用基于光谱数据变换的多种指数与弓曲差联合建模,效果显著,利用光谱数据变换形式  $\lg(1/R)$  构建的模型的验证集  $R^2$  达到 0.82,这与曹肖奕等<sup>[30]</sup>、WANG 等<sup>[31]</sup>的研究结果基本一致。在后续研究工作中将进一步探讨野外原位光谱测量中土地类型、土壤质地等因素对估算结果的影响。

### 4 结论

与 PLSR 线性模型相比较,基于光谱数据  $\sqrt{R}$  的弓曲差构建的 RF 模型效果最好,  $R^2$  为 0.83,  $RMSE$  为 6.14 g/kg,  $RPD$  为 2.29。在  $\lg R$  光谱数据下, SOC 含量与 *DIS* 指数相关系数最高,相关系数为 0.80,最佳组合波段为(1 758 nm, 1 752 nm)。

基于不同光谱指数与弓曲差的建模精度  $R^2$  和  $RMSE$  分别介于 0.67~0.84 和 5.85~8.45 g/kg,模型的  $RPD$  均在 1.66 以上。基于光谱数据  $\sqrt{R}$  和弓曲差的 SOC 预测模型的泛化能力较高。基于光谱数据  $\lg(1/R)$  变换,模型的验证集  $R^2=0.82$ ,  $RMSE=3.52$  g/kg,  $RPD=3.99$ ,可以较好地估算研究区土壤有机碳含量。

### 参考文献:

- [1] SHI Z, JI W, VISCARRA ROSSEL R A, et al. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library [J]. *European Journal of Soil Science*, 2015, 66(4): 679-687.
- [2] ALLEN R M, LAIRD D A. Quantitative prediction of biochar soil amendments by near-infrared reflectance spectroscopy [J]. *Soil Science Society of America Journal*, 2013, 77(5): 1784-1794.
- [3] WARD K J, CHARBRILLAT S, NEUMANN C, et al. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered LUCAS soil database [J]. *Geoderma*, 2019, 353: 297-307.
- [4] 林鹏达,佟志军,张继权,等. 基于 CWT 的黑土有机质含量野外高光谱反演模型[J]. *水土保持研究*, 2018, 25(2): 46-52, 57.
- [5] 赵小敏,杨梅花. 江西省红壤地区主要土壤类型的高光谱特性研究[J]. *土壤学报*, 2018, 55(1): 31-42.
- [6] GUNSAULIS F R, KOCHER M F, GRIFFIS C L. Surface structure effects on close-range reflectance as a function of soil organic matter content [J]. *American Society of Agricultural Engineer*, 1991, 34(2): 641-649.
- [7] 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.
- [8] AMIN I, FIKRAT F, MAMMADOV E, et al. Soil organic carbon prediction by Vis-NIR spectroscopy: case study the Kur-Aras plain, Azerbaijan [J]. *Communications in Soil Science and Plant Analysis*, 2020, 51(6): 726-734.
- [9] ZHENG G H, RYU D, JIAO C X, et al. Estimation of organic matter content in coastal soil using reflectance spectroscopy [J]. *Pedosphere*, 2016, 26(1): 130-136.
- [10] 王海峰,张智韬,ARNON K,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.
- [11] 赵明松,谢毅,陆龙妹,等. 基于高光谱特征指数的土壤有机质含量建模[J]. *土壤学报*, 2021, 58(1): 42-54.
- [12] 张子鹏,丁建丽,王敬哲,等. 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[J]. *光谱学与光谱分析*, 2020, 40(5): 1514-1522.
- [13] HONG Y, CHEN S, CHEN Y, et al. Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon: Feature selection coupled with random forest [J]. *Soil and Tillage Research*, 2020, 199: 104589.
- [14] 张智韬,劳聪聪,王海峰,等. 基于 FOD 和 SVM-DA-RF 的土壤有机质含量高光谱预测[J]. *农业机械学报*, 2020, 51(1): 156-167.
- [15] 周伟,谢利娟,杨哈,等. 基于高光谱的三江源区土壤有机质含量反演[J]. *土壤通报*, 2021, 52(3): 564-574.
- [16] 牛芳鹏,李新国,靳万贵,等. 博斯腾湖西岸湖滨绿洲土壤盐分特征[J]. *中国土壤与肥料*, 2020(3): 8-15.

- [17] 吴才武,夏建新,段峥嵘. 土壤有机质测定方法述评与展望[J].土壤,2015,47(3):453-460.
- [18] 何挺,程焱,王静. 野外地物光谱测量技术及方法[J]. 中国土地科学,2002(5):30-36.
- [19] 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):391-401.
- [20] 朱建伟,刘玉学,吴超凡,等. 施用生物炭后土壤有机碳的近红外光谱模型研究与应用[J].生态学报,2020,40(20):7430-7440.
- [21] 徐彬彬. 土壤剖面的反射光谱研究[J].土壤,2000(6):281-287.
- [22] ALLBED A, KUMAR L, SINHA P. Soil salinity and vegetation cover change detection from multi-temporal remotely sensed imagery in Al Hassa Oasis in Saudi Arabia[J]. Geocarto International, 2018,33(8):830-846.
- [23] 李哲,张飞,冯海宽,等. 基于波段组合的植被叶片盐离子估算研究[J].光学学报,2017,37(11):325-339.
- [24] 王李娟,孔钰如,杨小冬,等. 基于特征优选随机森林算法的农耕地土地利用分类[J].农业工程学报,2020,36(4):244-250.
- [25] 包青岭,丁建丽,王敬哲,等. 基于随机森林算法的土壤有机质含量高光谱检测[J].干旱区地理,2019,42(6):1404-1414.
- [26] 张锐,李兆富,潘剑君. 小波包-局部最相关算法提高土壤有机碳含量高光谱预测精度[J].农业工程学报,2017,33(1):175-181.
- [27] 张娟娟,田永超,姚霞,等. 同时估测土壤全氮、有机质和速效氮含量的光谱指数研究[J].土壤学报,2012,49(1):50-59.
- [28] 洪永胜,朱亚星,苏学平,等. 高光谱技术联合归一化光谱指数估算土壤有机质含量[J].光谱学与光谱分析,2017,37(11):3537-3542.
- [29] 焦彩霞,郑光辉,解宪丽,等. 可见-短近红外成像光谱数据的土壤有机质含量估算[J].光谱学与光谱分析,2020,40(10):3277-3281.
- [30] 曹肖奕,丁建丽,葛翔宇,等. 基于光谱指数与机器学习算法的土壤电导率估算研究[J].土壤学报,2020,57(4):867-877.
- [31] WANG J Z, DING J L, ABULIMITI A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared(VIS-NIR) spectroscopy, Ebinur Lake Wetland, Northwest China[J]. PeerJ,2018,6: e4703.

(责任编辑:张震林)