

# 可见-短近红外成像光谱数据的土壤有机质含量估算

焦彩霞<sup>1</sup>, 郑光辉<sup>1\*</sup>, 解宪丽<sup>2</sup>, 崔雪峰<sup>3</sup>, 赏刚<sup>1</sup>

1. 南京信息工程大学地理科学学院, 江苏 南京 210044

2. 中国科学院南京土壤研究所, 江苏 南京 210008

3. 北京师范大学系统科学学院, 北京 100875

**摘要** 土壤有机质是土壤肥力的重要指标,也是全球土壤碳的重要存在形式,快速估算土壤有机质含量及其变化是保障粮食安全与评估气候变化的前提与基础。传统的土壤有机质测定方法存在周期长、成本高、有污染物排放等不足。近年来的大量研究表明,土壤反射光谱技术可以成功实现土壤有机质估算,具有时间短、成本低、无污染、无破坏等特点。但反射光谱技术仅能估算点状土壤样品有机质含量,要实现土壤有机质的空间制图还必须借助空间插值技术。成像光谱技术(也称为高光谱成像技术)为每个像元采集一条光谱曲线,实现了图谱合一,为土壤有机质空间制图提供了技术基础。目前利用成像光谱开展土壤有机质制图的研究尚处于起步阶段,利用室内可见-短近红外波段的成像光谱数据建立土壤有机质光谱指数可以探讨土壤有机质成像光谱估算的机理,从而为土壤有机质遥感制图奠定理论基础。建立 600 nm“弓曲差”光谱指数并分析其与土壤有机质的相关关系,通过 1 000 次随机划分数数据集、建立“弓曲差”非线性回归与偏最小二乘回归模型并对比结果精度,探讨成像光谱数据估算土壤有机质的可行性。结果表明,研究区土壤有机质含量偏低,变化范围较大,“弓曲差”与有机质含量呈显著的对数关系;对数函数可以较好实现土壤有机质的建模与预测,拟合结果稳定性较好,精度优于偏最小二乘回归。原因可能是由于偏最小二乘回归所使用的全谱数据中包含了部分与有机质无关的信息,影响了偏最小二乘回归结果精度。所以,使用三个波段光谱信息建立的光谱指数“弓曲差”可以用于成像光谱数据的有机质制图,从而为未来开展土壤有机质卫星遥感制图提供新的思路与方法。

**关键词** 成像光谱;土壤有机质;偏最小二乘回归;弓曲差

**中图分类号:** TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3277-05

## 引言

土壤有机质(soil organic matter, SOM)是土壤的重要组成部分,是植物营养物质的来源,它在形成土壤团聚体、提高土壤含水量、减少土壤退化和侵蚀方面具有重要作用<sup>[1]</sup>。土壤反射光谱技术成功实现了快速、低成本 SOM 估算<sup>[2-4]</sup>,特别是近年建立的全球土壤反射光谱数据库更有助于满足区域至全球范围土壤评价与监测的需求<sup>[5]</sup>。土壤反射光谱技术仅提供样点数据,无法提供面状数据以表现土壤水平方向及剖面垂直方向的连续变异。遥感技术是实现土壤有机质快速制图的重要方法<sup>[6-7]</sup>,但目前的遥感数据用于土壤有机质制图尚存在混合像元、水分、光谱分辨率等影响因素,因此提取土

壤成像光谱数据的有效信息、建立土壤有机质光谱指数,是应用现有遥感数据进行 SOM 制图及研制未来 SOM 制图专用传感器的理论前提。彭杰<sup>[8]</sup>研究结果表明,570~630 nm 波段为 SOM 敏感波段,600 nm 反射率与 SOM 相关系数最大。Zheng 等<sup>[9]</sup>与徐彬彬和戴昌达<sup>[10]</sup>分析土壤有机质反射光谱数据的吸收特征波段、建立 600 nm“弓曲差”光谱指数,成功实现了土壤有机质含量估算。因此,本工作利用风干、研磨后土壤样品的成像光谱数据,建立 600 nm“弓曲差”(deviation of arch, DOA)光谱指数并构建 SOM 估算模型,与偏最小二乘回归(partial least square regression, PLSR)结果相比较,探讨该光谱指数用于 SOM 制图的可行性,为未来 SOM 遥感制图奠定理论与方法基础。

收稿日期: 2019-07-17, 修订日期: 2019-11-21

基金项目: 国家自然科学基金项目(41877004), 中国国家留学基金项目(201809040007)资助

作者简介: 焦彩霞, 1977 年生, 南京信息工程大学地理科学学院副教授 e-mail: JCX@nuist.edu.cn

\* 通讯联系人 e-mail: zgh@nuist.edu.cn

## 1 实验部分

### 1.1 土壤样本采集与属性测定

研究区位于江苏省东台市, 该区域东临黄海, 属北亚热带暖湿季风气候区。区内土壤以公元 1028 年完工的范公堤为界, 堤西地区为湖相沉积形成的水稻土, 堤东为近千年黄河、长江沉积物经人类耕作发育形成的潮盐土。样品分为两批采集(图 1): 2013 年采集 8 个表土(0~5 cm)及 20 个剖面土样, 每个剖面内分别采集 0~5, 5~10, 10~20, 20~30, 30~40, 40~60, 60~80 和 80~100 cm 共 8 个深度的土壤样品; 2018 年采集 11 个剖面土样, 每个剖面内分别采集 0~5, 5~10, 10~15, 15~20, 20~30, 30~40, 40~50, 50~60, 60~80 和 80~100 cm 共 10 个深度的土壤样品。共采集土壤样品 278 个。样品自然风干后研磨过 100 目筛, 分成两份分别用于成像光谱及土壤属性测定。土壤有机质由重铬酸钾氧化-外加加热法测定。

### 1.2 成像光谱数据采集与预处理

土壤样品装于铝盒盖内并用直尺刮平, 放置于移动平台之上, 在暗室内采集样品成像光谱数据, 经黑白校正得到反

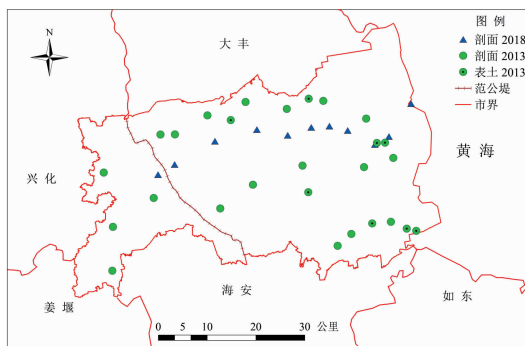


图 1 采样点位置分布图

Fig. 1 Location of samples

射率光谱数据(图 2)。仪器为加拿大 Lumenera 公司生产的 INFINITY V10E 成像光谱仪, 在 389~1 045 nm 光谱范围采集 256 个波段的数据, 人工光源为 4 盏卤素灯。采集得到的成像光谱反射率数据在 ENVI 5.3 软件中通过 ROI 提取每个土壤样品的平均反射光谱曲线, 去除两端误差较大波段、保留 425~1 015 nm 共 230 个波段数据用于后续计算与分析。

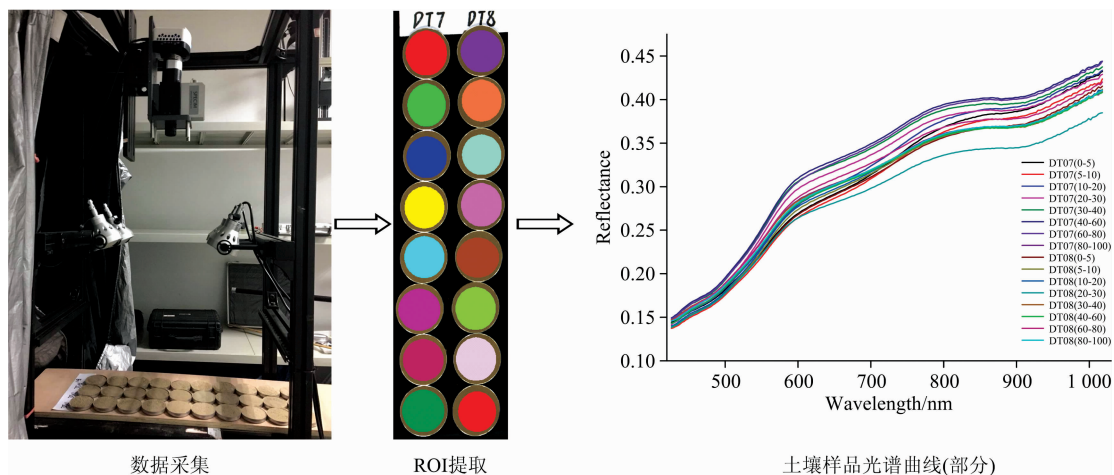


图 2 成像光谱数据采集及预处理过程图

Fig. 2 Collection and preprocessing of imaging spectroscopy

### 1.3 建模方法

使用非线性参数回归及偏最小二乘回归两种方法。非线性参数回归以光谱数据计算出的光谱指数 600 nm“弓曲差”为自变量, 可称为 DOAR(DOA Regression); 600 nm“弓曲差”是指每一条土壤光谱曲线的 600 nm 反射率与其 550 和 650 nm 光谱反射率平均值的差<sup>[10]</sup>。偏最小二乘回归以 230 个波段的光谱数据为自变量。偏最小二乘回归集典型相关分析、主成分分析与多元线性回归分析于一体, 可以较好解决自变量多重共线性问题, 允许在样本个数少于变量个数条件下进行回归建模。这种方法广泛应用于反射光谱估算土壤属性的研究<sup>[11-12]</sup>。

回归模型的精度通过均方根误差 RMSE(建模集 RMSEC、预测集 RMSEP)、决定系数  $R^2$  (建模集  $R_c^2$ 、预测集

$R_p^2$ )及 RPD(residual prediction deviation)<sup>[13]</sup> 进行评价, 计算方法详见式(1)~式(3)。

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

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

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} \quad (3)$$

式中,  $\hat{y}_i$  和  $y_i$  分别为样品  $i$  的有机质估算值和实测值;  $\bar{y}$  为样本集有机质含量的均值; SD 为预测样本集有机质实测值的标准差。RMSE 越小、 $R^2$  越接近 1、RPD 越大, 说明模型精度越高。建模及预测验证的计算过程在 Matlab R2013a 中

实现。

#### 1.4 样本集划分及建模预测

模型精度会随样本集划分结果而变化<sup>[14]</sup>。为避免样本集划分对模型结果的影响、更好地分析模型的稳定性,对样本集进行 1 000 次随机抽样,每次取 1/2 样本作为建模样本集,剩余 1/2 为预测验证集,进行 1 000 次建模与预测验证,并计算每次的 RMSE、 $R^2$  和 RPD 用于统计分析。

## 2 结果与讨论

### 2.1 土壤有机质含量统计分析

由表 1 可知,研究区内土壤有机质含量较低(平均值为  $8.72 \text{ k} \cdot \text{kg}^{-1}$ ),变异较大(最小值为  $1.39 \text{ k} \cdot \text{kg}^{-1}$ ,最大值为  $52.56 \text{ k} \cdot \text{kg}^{-1}$ ,变异系数为 87.65%)。堤东潮盐土为近

千年形成,成土年龄较短,有机质含量显著低于堤西水稻土,平均值分别为  $7.43$  和  $15.92 \text{ k} \cdot \text{kg}^{-1}$ 。

### 2.2 “弓曲差”非线性回归结果分析

土壤有机质含量与“弓曲差”的相关系数为  $-0.68$ ,表明有机质含量增加会使“弓曲差”减小;反之逐步减小的“弓曲差”表示样本中有机质含量的逐渐增加。这种负相关关系是“弓曲差”估算土壤有机质含量的机理。由散点图(图 3)可见,对数函数可以比线性函数更好地拟合土壤有机质含量与“弓曲差”的函数关系,调整  $R^2$  为  $0.524 2$ 。徐彬彬和戴昌达<sup>[10]</sup>的研究结果与本研究相同,但 Zheng 等<sup>[9]</sup>研究表明线性函数拟合结果最优。通过对比可以发现,样本有机质含量变化范围对土壤有机质含量与“弓曲差”的关系影响最大,即有机质含量变化范围大会导致两者呈对数关系,有机质含量变化范围小时两者呈线性关系。

表 1 土壤有机质含量描述统计

Table 1 Descriptive statistics of SOM content

	Number	Min/( $\text{k} \cdot \text{kg}^{-1}$ )	Max/( $\text{k} \cdot \text{kg}^{-1}$ )	Mean/( $\text{k} \cdot \text{kg}^{-1}$ )	Std/( $\text{k} \cdot \text{kg}^{-1}$ )	CV/%
全部样本	278	1.39	52.56	8.72	7.64	87.65
堤东样本	236	1.39	33.38	7.43	5.99	80.63
堤西样本	42	2.42	52.56	15.92	11.22	70.46

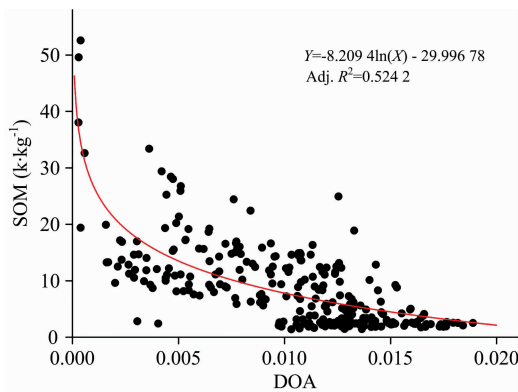


图 3 土壤有机质与“弓曲差”散点分布图

Fig. 3 Scatter plot of SOM and DOA

使用对数函数分别建立 1 000 次随机划分建模集的预测模型并使用预测集独立样本进行模型验证,1 000 次的建模及验证结果统计见表 2。建模  $R^2$  变化范围为  $0.28 \sim 0.69$ ,平均值为  $0.52$ ;预测验证  $R^2$  变化范围为  $0.02 \sim 0.64$ ,平均值为  $0.49$ ,RPD 平均值为  $1.42$ 。预测验证  $R^2$  总体小于建模  $R^2$  且最小值为  $0.02$ ,表明在一些数据集划分的情况下会产生比

表 2 DOA 回归结果统计表

Table 2 Statistics of DOA regression

	RMSEC/ ( $\text{k} \cdot \text{kg}^{-1}$ )	$R_c^2$	RMSEP/ ( $\text{k} \cdot \text{kg}^{-1}$ )	$R_p^2$	RPD
Min	4.10	0.28	4.46	0.02	1.02
Max	6.10	0.69	6.65	0.64	1.68
Mean	5.19	0.52	5.38	0.49	1.42

较差的预测结果,但大部分预测及建模  $R^2$  大于  $0.45$ (图 4),说明由成像光谱数据计算的光谱指数“弓曲差”可以较好地预测样品有机质含量。

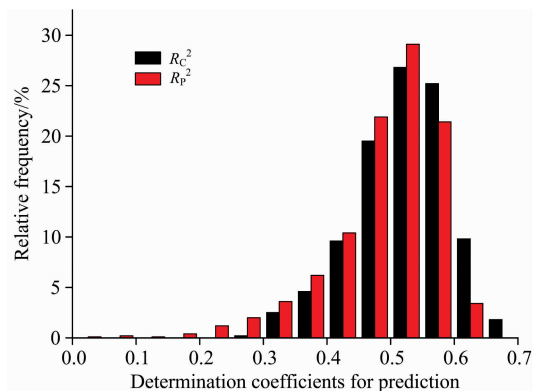


图 4 建模及预测验证决定  $R^2$  频率统计图

Fig. 4 Relative frequency for determination coefficients of calibration and prediction

### 2.3 偏最小二乘回归结果分析

以 SOM 为因变量、230 个波段光谱数据为自变量的 1 000 次偏最小二乘回归结果统计见表 3。与表 2 中“弓曲差”拟合结果相比,PLSR 的  $R_c^2$  最大值大于 DOAR,平均值相同;但 PLSR 的  $R_p^2$  和 RPD 的最大值及平均值均小于 DOAR。综合来看,DOA 非线性拟合结果稍优于 PLSR。由光谱数据提取计算的光谱指数 DOA 虽然只使用了三个波段的光谱信息,但这三个波段位于 SOM 影响最大的波段范围<sup>[8, 15-16]</sup>,有效包含了光谱数据中关于有机质的信息。PLSR 所使用的 230 个波段光谱数据虽然信息丰富,但也包含了一

些影响有机质估算精度的噪音, 从而造成 PLSR 估算精度下降。此外, Zheng 等<sup>[10]</sup>研究认为 PLSR 结果优于 DOAR。可能与两个研究使用的光谱数据类型、波段范围及 SOM 变化范围等因素有关。

表 3 偏最小二乘回归结果统计表

Table 3 Statistics of partial least square regression

	RMSEC/ ( $\text{k} \cdot \text{kg}^{-1}$ )	$R_c^2$	RMSEP/ ( $\text{k} \cdot \text{kg}^{-1}$ )	$R_p^2$	RPD
Min	2.99	0.28	4.09	0.07	1.04
Max	7.17	0.81	7.58	0.62	1.62
Mean	5.22	0.52	5.75	0.43	1.33

### 3 结 论

通过 1 000 次随机抽样方法划分建模与预测验证数据集、利用“弓曲差”非线性回归及偏最小二乘回归两种方法建立土壤有机质估算模型, 通过精度对比探讨“弓曲差”方法估算 SOM 的有效性与稳定性。研究发现, 有机质含量与 600 nm“弓曲差”呈负相关关系, 是该光谱指数应用于 SOM 估算的机理。虽然 DOAR 仅提取了三个波段的光谱信息, 但其建模精度与 PLSR 方法相当, 预测验证精度优于 PLSR。可以得出结论, DOA 方法可以用于室内土壤成像光谱数据的 SOM 制图。此研究为高分五号及珠海一号等卫星遥感数据应用于 SOM 研究提供了新的思路。

### References

- [ 1 ] ZHAO Yong-cun, XU Sheng-xiang, WANG Mei-yan, et al(赵永存, 徐胜祥, 王美艳, 等). Bulletin of Chinese Academy of Sciences(中国科学院院刊), 2018, 33(2): 191.
- [ 2 ] Shi Z, Ji W J, Viscarra Rossel R A, et al. European Journal of Soil Science, 2015, 66(4): 679.
- [ 3 ] Ji W J, Li S, Chen S C, et al. Soil and Tillage Research, 2016, 155: 492.
- [ 4 ] ZHAO Xiao-min, YANG Mei-hua(赵小敏, 杨梅花). Acta Pedologica Sinica(土壤学报), 2018, 55(1): 1.
- [ 5 ] Viscarra Rossel R A, Behrens T, Ben-Dor E, et al. Earth-Science Reviews, 2016, 155: 198.
- [ 6 ] ZHU A-xing, YANG Lin, FAN Nai-qing, et al(朱阿兴, 杨琳, 樊乃卿, 等). Progress in Geography(地理科学进展), 2018, 37(1): 66.
- [ 7 ] ZHANG Gan-lin, ZHU A-xing, SHI Zhou, et al(张甘霖, 朱阿星, 史舟, 等). Progress in Geography(地理科学进展), 2018, 37(1): 57.
- [ 8 ] PENG Jie, ZHOU Qing, ZHANG Yang-zhu, XIANG Hong-ying(彭杰, 周清, 张杨珠, 向红英). Acta Pedologica Sinica(土壤学报), 2013, 50(3): 517.
- [ 9 ] Zheng G H, Ryu D, Jiao C X, et al. Pedosphere, 2016, 26(1): 130.
- [ 10 ] XU Bin-bin, DAI Chang-da(徐彬彬, 戴昌达). Chinese Science Bulletin(科学通报), 1980, 6: 282.
- [ 11 ] Zeng R, Rossiter D G, Yang F, et al. Geoderma, 2017, 303: 78.
- [ 12 ] Peng J, Biswas A, Jiang Q S, et al. Geoderma, 2019, 337: 1309.
- [ 13 ] Mouazen A M, Kuang B, De Baerdemaeker J, et al. Geoderma, 2010, 158: 23.
- [ 14 ] Lucà F, Conforti M, Castrignanò A, et al. Geoderma, 2017, 288: 175.
- [ 15 ] Viscarra Rossel R A, Walvoort D J J, McBratney A B, et al. Geoderma, 2006, 131: 59.
- [ 16 ] Brown D J, Shepherd K D, Walsh M G, et al. Geoderma, 2006, 132: 273.

# Prediction of Soil Organic Matter Using Visible-Short Near-Infrared Imaging Spectroscopy

JIAO Cai-xia<sup>1</sup>, ZHENG Guang-hui<sup>1\*</sup>, XIE Xian-li<sup>2</sup>, CUI Xue-feng<sup>3</sup>, SHANG Gang<sup>1</sup>

1. School of Geographic Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China

2. Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

3. School of Systems Science, Beijing Normal University, Beijing 100875, China

**Abstract** Soil organic matter (SOM) is a crucial indicator of soil fertility and an important form of global soil carbon. It is the premise and basis of ensuring food security and assessing climate change to estimate SOM content and its changes rapidly. The traditional method of SOM determination is time-consuming with the high cost and environmental risks. Soil reflectance spectroscopy can be faster and cheaper than the conventional method, do not generate chemical residues and are non-destructive to the samples. However, spatial interpolation technique was still needed to map SOM after estimation of SOM in a point soil sample by reflectance spectroscopy. Imaging spectroscopy (also known as hyperspectral imaging) collects a spectral curve for each pixel, which enlarges the envelope of point spectrometry into a spatial domain and provides a technical basis for spatial mapping of SOM. This novel technique has not yet been fully utilized for SOM mapping. Therefore, the spectral index established by laboratory visible-short near-infrared imaging spectroscopy data can be used to estimate SOM and explore the mechanism, which lays a theoretical foundation for SOM mapping of remote sensing. In this study, a spectral index, named deviation of an arch (DOA), was established using the information of three wavelengths. The correlation between DOA and SOM was analyzed by the scatter diagram. Then, the samples were randomly split into training and validation data sets for 1 000 times. Nonlinear regression and partial least square regression (PLSR) were used to calibrate DOA or spectroscopy to SOM, respectively. The performances were compared to explore the feasibility of SOM estimation using imaging spectroscopy. The results indicate that the SOM content in the study area was relatively low, and its variation range was large. There was a significant logarithmic relationship between DOA and SOM. Logarithm function can be used to model DOA and SOM and provide reasonable and stable results. The performance of DOA regression is better than PLSR. The possible reason is that the spectral data used by PLSR contains some information unrelated to SOM, which affects the accuracy of PLSR. We can conclude that this spectral index, DOA, can be used for SOM mapping, although it is deduced from three wavelengths. It provides a new idea and methods for SOM mapping based on satellite remote sensing data in the future.

**Keywords** Imaging spectroscopy; Soil organic matter (SOM); Partial least squares regression (PLSR); Deviation of arch (DOA)

(Received Jul. 17, 2019; accepted Nov. 21, 2019)

\* Corresponding author