

# 高光谱小波能量特征估测土壤有机质含量

章涛<sup>1,2,3</sup>, 于雷<sup>1,2,3\*</sup>, 易军<sup>1,2,3</sup>, 聂艳<sup>1,2,3</sup>, 周勇<sup>1,2,3</sup>

1. 华中师范大学地理过程分析与模拟湖北省重点实验室, 湖北 武汉 430079
2. 华中师范大学城市与环境科学学院, 湖北 武汉 430079
3. 华中师范大学可持续发展研究中心, 湖北 武汉 430079

**摘要** 土壤高光谱在采集过程中难以避免噪声干扰, 造成高光谱数据信噪比较低, 影响土壤有机质含量估测精度。尝试探究小波能量特征方法, 降低高光谱噪声, 提升土壤有机质含量高光谱估测模型性能。选取湖北省潜江市运粮湖管理区为试验区, 于2016年9月采集80份深度为0~20 cm的水稻土样本; 土壤样本经风干、碾磨、过筛等一系列处理后, 在实验室内采集样本光谱, 并通过重铬酸钾-外加加热法测定土壤有机质含量; 利用浓度梯度法, 将总体样本集(80个样本)划分为建模集(54个样本)和验证集(26个样本); 以mexh为小波基函数进行连续小波变换(continuous wavelet transformation), 将土壤高光谱转换为10个分解尺度的小波系数(wavelet coefficients); 逐尺度计算小波系数的均方根作为小波能量特征(energy features), 将10个尺度的小波能量特征组成小波能量特征向量(energy features vector); 逐尺度逐波长计算小波系数与有机质含量的相关系数, 将达到极显著水平( $p < 0.01$ )的小波系数作为敏感小波系数(sensitive wavelet coefficients); 利用主成分分析法(principal component analysis)分别计算土壤高光谱和小波能量特征向量的各主成分载荷, 通过比较两者第一主成分贡献率的高低和两者前三个主成分得分的空间离散程度, 判断小波能量特征转换前后建模自变量的主成分信息变化趋势; 基于小波能量特征向量和敏感小波系数分别建立多元线性回归和偏最小二乘回归土壤有机质含量估测模型。结果表明, 土壤有机质含量越高, 全波段反射率越低, 但不同土样的光谱反射率曲线特征相似, 近红外部分的反射率(780~2400 nm)高于可见光部分(400~780 nm); 敏感小波系数对应的波长为494, 508, 672, 752, 1838和2302 nm; 土壤高光谱与小波能量特征向量的第一主成分贡献率分别为96.28%和99.13%, 小波能量特征向量的前三个主成分散点较土壤高光谱的主成分散点在空间上更为聚集, 表明小波能量特征方法有效减少了噪声影响; 比较全部土壤有机质含量估测模型, 以小波能量特征向量为自变量的多元线性回归模型具有最佳估测精度, 其验证集决定系数( $R^2$ )、相对估测误差(RPD)和均方根误差(RMSE)分别为0.77, 1.82和0.82。因此, 小波能量特征方法既能够提高数据的信噪比, 提升土壤有机质含量的估测精度, 又实现了土壤高光谱数据降维, 降低了模型复杂度, 可用于土壤有机质含量快速测定和土壤肥力动态监测等研究。

**关键词** 土壤高光谱; 小波系数; 小波能量特征; 土壤有机质; 水稻土

**中图分类号:** S127 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3217-06

## 引言

土壤有机质是评价土壤质量的关键因素, 在调控农田肥力和作物生长方面发挥着重要作用<sup>[1]</sup>。采用农化分析方法测定土壤有机质含量步骤繁琐、耗时耗力, 不利于土壤有机质

的实时高效测定<sup>[2]</sup>。土壤高光谱包含了丰富的土壤信息, 近年来, 被广泛应用于估测土壤理化性质(如土壤有机质、土壤水分及土壤全氮等), 取得了卓有成效的研究成果<sup>[3-4]</sup>。然而, 高光谱采集易受测试环境、样品质量和仪器本身等因素影响, 无法避免会产生噪声, 因此, 探索降噪方法对提升土壤有机质含量估测精度具有重要意义。

**收稿日期:** 2018-08-27, **修订日期:** 2018-12-21

**基金项目:** 国家自然科学基金项目(41401232), 中央高校基本科研业务费项目(CCNU19QN050), 和湖北省自然科学基金创新群体项目(2016CFA027)资助

**作者简介:** 章涛, 1994年生, 华中师范大学城市与环境科学学院硕士研究生 e-mail: zhangt@mails.ccnu.edu.cn

\* 通讯联系人 e-mail: yulei@mail.ccnu.edu.cn

学者们尝试采用连续统去除,一阶微分,高通滤波等预处理方法,消除样品背景及杂散光等因素所引起的噪声和增强特征信号,但难以有效去除白噪声,特别是随机和低频信号<sup>[5-6]</sup>。于雷等<sup>[7]</sup>、廖钦洪等<sup>[8]</sup>基于连续小波变换法将土壤高光谱分解成多尺度小波系数,采用相关性分析法筛选与有机质含量具有极显著相关性的敏感小波系数(sensitive wavelet coefficients, SWC)建立估测模型;蔡亮红等<sup>[9]</sup>采用小波变换对土壤光谱进行分解,通过竞争适应重加权采样法(competitive adaptive reweighted sampling, CARS)算法筛选 SWC,建立土壤水分高光谱估测模型,这些研究表明小波变换方法能够克服白噪声影响,将土壤高光谱分解为不同尺度的小波系数,以更准确地局部描述和分离信号特征<sup>[10-12]</sup>。但是,通过提取 SWC 估测土壤属性,难以充分利用不同尺度小波系数包含的有效光谱信息<sup>[13]</sup>。Pu 等<sup>[14]</sup>通过逐尺度提取全波段小波系数的能量特征(energy features, EF)组成小波能量特征向量(energy features vector, EF<sub>v</sub>)估测森林叶面积指数;宋开山等<sup>[15]</sup>基于小波分析提取大豆叶片光谱的 EF<sub>v</sub>,建立叶绿素 a 含量高光谱估测模型,结果表明基于 EF<sub>v</sub> 的回归模型可提高估测水平,这对构建土壤有机质高光谱估测模

型具有重要的参考价值。

本工作采用小波变换方法将土壤高光谱分解为 10 个尺度的小波系数,利用均方根法计算各尺度的 EF,构建 EF<sub>v</sub> 并建立基于 EF<sub>v</sub> 的土壤有机质含量估测模型。目的是探索土壤高光谱噪声去除和信息提取方法,以提高土壤有机质高光谱估测性能。

## 1 实验部分

### 1.1 土样的采集

试验区位于湖北省潜江市运粮湖管理区[图 1(a)],该区域地势平坦,水源充足,第一产业较为发达,是重要的粮油生产基地。试验区面积约为 375 m×300 m(11.25 ha)的连片田块,土壤类型为水稻土,是潜江市主要土类之一(占全市面积的 55.22%);常年采用中稻-油菜轮作的耕作制度。

2016 年 9 月在试验区内均匀布设 80 个样点,由北向南共 5 行(每行相距 75 m)、由西向东共 16 列(每列相距 25 m)[图 1(b)];土样采集深度为 0~20 cm,用密封袋带回实验室经一系列预处理操作后用于后续分析。

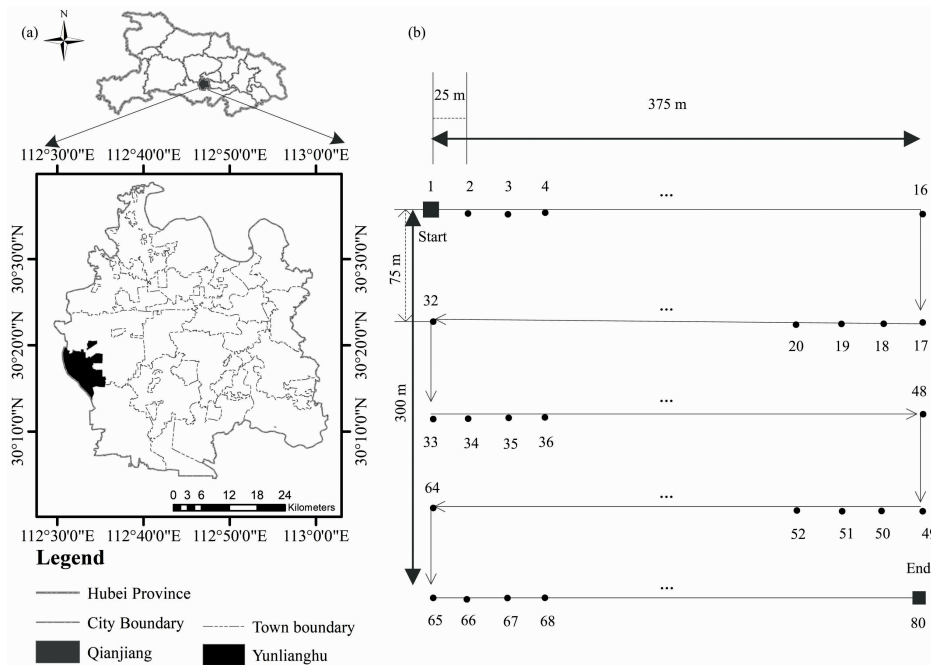


图 1 研究区(a)及采样方案(b)

Fig. 1 The study area (a) and sketch of sampling points (b)

### 1.2 土壤样本集的划分

将 80 个土样作为总体样本集,根据土壤有机质含量从

小到大排序,采用浓度梯度法依次每隔两个样本取一个作为验证集,共计 26 个样本,其余 54 个样本作为建模集(表 1)。

表 1 样本统计特征

Table 1 Descriptive statistics of sample sets

Type of samples	Observations	Min /(g · kg <sup>-1</sup> )	Max /(g · kg <sup>-1</sup> )	Mean /(g · kg <sup>-1</sup> )	Standard deviation /(g · kg <sup>-1</sup> )
Whole set	80	19.92	28.52	23.89	1.82
Calibration set	54	19.92	28.52	24.01	1.95
Validation set	26	20.92	26.40	23.60	1.49

### 1.3 土样光谱数据的测定

采用 ASD FieldSpec3 光谱仪测定土样光谱, 详细仪器参数、采集过程、参数设置及预处理可参见文献[16]。由于两端波长处的信噪比较低, 因此只保留 400~2 400 nm。

### 1.4 小波变换和能量特征向量的计算

采用连续小波变换(CWT)对土壤高光谱数据进行转换, 通过小波分解后, 得到各尺度的小波系数[17]

$$W_f(a, b) \leq f; \Psi_{a, b} \geq \int_{-\infty}^{+\infty} f(\lambda) \Psi_{a, b}(\lambda) d\lambda \quad (1)$$

$$\Psi_{a, b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{\lambda-b}{a}\right) \quad (2)$$

其中,  $\lambda$  为高光谱波段数,  $a$  为尺度参数,  $b$  为位移参数,  $f(\lambda)$  是光谱反射率,  $\Psi_{a, b}$  为基函数,  $W_f(a, b)$  为小波系数。 $W_f(a, b)$  包括  $i$  和  $j$  两个维度表示为  $W_f(a_i, b_j)$ ,  $i$  为分解尺度,  $j$  为波段。

计算第  $i$  个分解尺度的小波系数的均方根, 作为该尺度的  $EF_i$ [式(3)]。

$$EF_i = \sqrt{\frac{1}{j} \sum_{j=1}^j W_f(a_i, b_j)^2} \quad (3)$$

构建  $EF_V$ [式(4)]。

$$EF_V = [EF_1, EF_2, \dots, EF_i] \quad (4)$$

### 1.5 敏感小波系数的确定

为对比  $EF_V$  与 SWC 模型估测效果, 采用相关性分析法筛选每一分解尺度与有机质含量相关性极显著( $p < 0.01$ )的小波系数作为 SWC[式(5)]。其中,  $R^2$  为决定系数,  $X_{sw}$  为样本  $s$  在波长为  $w$  所对应的小波系数,  $Y_s$  为样本  $s$  的土壤有机质含量,  $\bar{X}_w$  为波长为  $w$  的小波系数平均值,  $\bar{Y}$  为土壤有机质含量的平均值。

$$R^2 = \left[ \frac{\sum_{s=1}^n (X_{sw} - \bar{X}_w)(Y_s - \bar{Y})}{\sqrt{\sum_{s=1}^n (X_{sw} - \bar{X}_w)^2} \sqrt{\sum_{s=1}^n (Y_s - \bar{Y})^2}} \right]^2 \quad (5)$$

### 1.6 模型构建与检验

偏最小二乘回归(PLSR)是最常用的高光谱估测模型之一, 其考虑了因变量的作用, 在处理多维数据、解决多重共线性时性能优越[18]。多元线性回归(MLR)是一种较为简单的线性回归方法, 适用于样本数大于自变量数[19]。分别以土壤高光谱(记为: Origin)、SWC 和  $EF_V$  为自变量, 有机质含量为因变量建立 PLSR 和 MLR 模型, 依次表示为 Origin-PLSR, SWC-PLSR,  $EF_V$ -PLSR, SWC-MLR,  $EF_V$ -MLR。通过  $R^2$ , RMSE 和 RPD 来检验估测模型精度[20]。

## 2 结果与讨论

### 2.1 土壤光谱特征分析

不同土壤有机质含量的光谱反射率曲线(图 2)基本特征相似, 可见光部分反射率低于近红外部分; 土样全波段反射率与有机质含量成反比。在可见光部分的土壤反射率曲线与近红外部分差异较大; 在 1 400, 1 900 和 2 200 nm 处, 出现具有不同吸收深度和宽度的吸收谷。

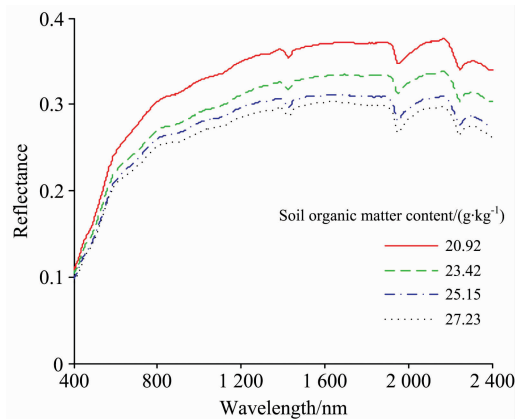


图 2 土壤高光谱反射率

Fig. 2 Spectral reflectance curves of soil

### 2.2 敏感小波系数分析

以 mexh 为小波基函数, 计算土壤高光谱在 10 个尺度上 ( $2^{1-10}$ ) 的小波系数, 分析小波系数与有机质含量的相关性, 得到  $R^2$  分布状况(图 3)。当  $p < 0.01$ , 则  $R^2 > 0.12$ 。由于波长众多, 难以提取敏感波长, 以图中红色部分 ( $R^2 > 0.3$ ) 作为相关性极强的区域。不同分解尺度和波段下的  $R^2$  存在明显差异, 说明经过小波变换, 与土壤有机质相关的细节特征信息被逐层释放, 每一层小波系数信息代表性各异。其中, 红色区域主要分布在 490~500, 500~510, 670~680, 740~760, 1 830~1 840 和 2 300~2 310 nm, 分别对应尺度 5, 4, 3, 3, 7 和 4, 对应最大值波长位置为 494, 508, 672, 752, 1 838 和 2 302 nm, 将上述波长对应的小波系数确定为 SWC。

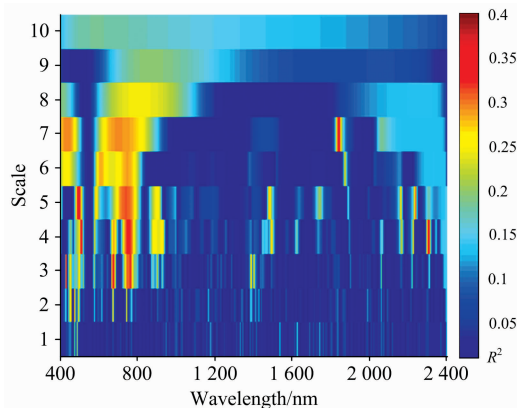


图 3 不同尺度  $R^2$  分布

Fig. 3  $R^2$  distribution at different decomposition scales

### 2.3 小波能量特征向量分析

根据 EF 计算公式, 每个尺度提取一个 EF, 最终每个样本形成 10 维  $EF_V$ 。主成分分析法(PCA)可有效降低数据维度, 以少数几个主成分变量反映原始信息。在本工作中, PCA 用于分析土壤高光谱和  $EF_V$  的第一主成分贡献率及前三个主成分得分的空间离散程度, 判断 EF 转换前后变量的主成分信息变化趋势。 $EF_V$  的第一主成分贡献率为 99.13%, 比原始光谱的第一主成分贡献率高出 3%[图 4(a)], 接近于

100%，表明  $EF_V$  第一主成分所包含的信息量较土壤高光谱有所增加，且具有极强的信息代表性。前三个主成分得分所建立的三维散点图[图 4(b)]更为形象地表明  $EF_V$  的降维效果，原始光谱各样本主成分得分散点分布广泛，样本间差异显著，主成分所含信息分散； $EF_V$  各主成分得分散点十分聚集，表明噪声消除后样本间差异降低。通过 PCA 分析， $EF_V$  在降低光谱数据维度、消减光谱噪声的同时提高了样本变量信息贡献率。

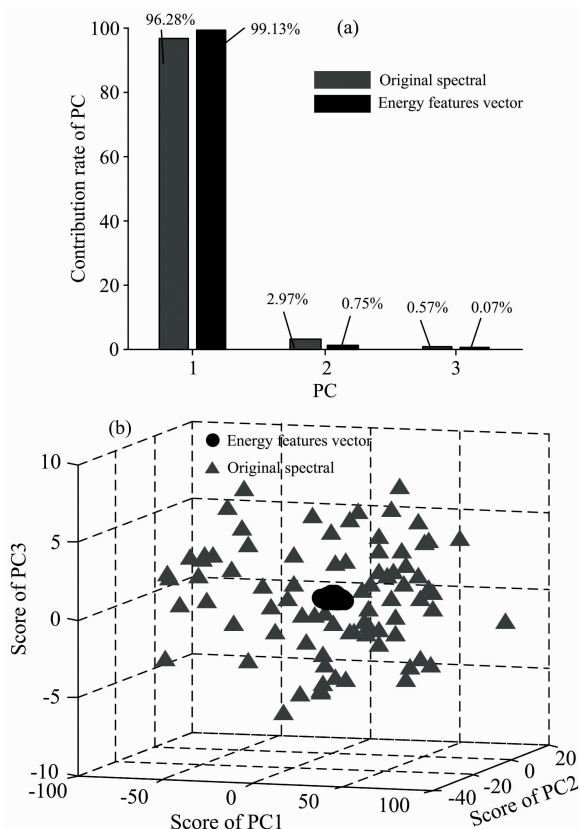


图 4 土壤高光谱、能量特征向量前三个主成分贡献率(a)及得分散点(b)

Fig. 4 The first three principal component contribution rate (a) and the score scatter point (b) of soil original spectral and the root mean square energy feartures vector

## 2.4 估测模型比较

由表 2 可以看出，同一自变量集的 MLR 模型估测效果要优于 PLSR 模型，表明当变量数小于样本数时(在此 SWC 和  $EF_V$  自变量数分别为 6 和 10 个)，MLR 模型具有更好的

适用性，可更为准确地表达土壤有机质同高光谱间的关系，简化模型结构。由此表明，在选取估测模型时，应充分考虑变量及样本特征。

基于不同自变量集的最优模型分别为 Origin-PLSR, SWC-MLR,  $EF_V$ -MLR 模型，对应  $R^2$  依次为 0.70, 0.73 和 0.77，表明 SWC-MLR,  $EF_V$ -MLR 模型估测效果要优于 Origin-PLSR 模型。在所有估测模型中  $EF_V$ -MLR 模型表现最优，RPD 达到 1.82，可较为准确地估测土壤有机质含量。利用小波分析可以降低噪声干扰，放大土壤有机质敏感信息。

通过提取 SWC 和  $EF_V$  均可降低数据维度，减少模型复杂度，提升模型估测性能。基于  $EF_V$  的模型要优于 SWC 模型，表明 EF 在小波系数的基础上，在每个尺度上计算小波分解系数的能量，形成特征向量，相比于代表部分敏感信息的 SWC,  $EF_V$  更为综合地表征了光谱信息。同时，EF 无需事先获取待测属性且数据压缩维度可控，相比于基于相关性分析提取 SWC, 该方法更具推广性和可重复性，可尝试用于多领域高光谱数据提取。

表 2 不同估测模型的效果比较

Table 2 Comparison of different estimation model results

Model	Calibration set		Validation set		
	$R^2$	RMSE/ ( $g \cdot kg^{-1}$ )	$R^2$	RMSE/ ( $g \cdot kg^{-1}$ )	RPD
Origin-PLSR	0.78	0.79	0.70	1.03	1.45
SWC-PLSR	0.68	1.08	0.64	1.15	1.30
$EF_V$ -PLSR	0.76	0.90	0.68	1.08	1.38
SWC-MLR	0.77	0.82	0.73	0.99	1.51
$EF_V$ -MLR	0.81	0.74	0.77	0.82	1.82

## 3 结论

基于小波变换方法提取 EF, 有效减少了土壤光谱噪声影响，提升了土壤有机质高光谱估测精度。研究表明，小波变换可放大光谱细节特征，降低噪声干扰，SWC 可代表部分与土壤有机质相关性较强的光谱信息；EF 充分利用了各尺度小波系数包含的光谱信息，提高了光谱数据信噪比，具有极佳的信息代表性；在所有模型中， $EF_V$ -MLR 模型表现最优，可有效降低数据维度与模型复杂度，对于快速估测田间土壤属性，推动农业精准化管理具有一定的参考价值。但 EF 代表的是不同尺度小波系数整体信息，如何更加准确地结合不同尺度的局部特征，值得深入研究。

## References

- [1] Dotto A C, Dalmolin R S D, Caten A T, et al. Geoderma, 2018, 314: 262.
- [2] ZHOU Qian-qian, DING Jian-li(周倩倩, 丁建丽). Acta Pedologica Sinica(土壤学报), 2018, 55(2): 313.
- [3] Rossel R A V, Behrens T, Ben-Dor E, et al. Earth-Science Reviews, 2016, 155: 198.
- [4] Nouri M, Gomez C, Roger J M, et al. Geoderma, 2017, 298: 54.
- [5] ZHANG Rui, LI Zhao-fu, PAN Jian-jun(张锐, 李兆富, 潘剑君). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(1): 175.

- [ 6 ] CHEN Hong-yan, ZHAO Geng-xing, LI Xi-can, et al(陈红艳, 赵庚星, 李希灿, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2012, 45(7): 1425.
- [ 7 ] YU Lei, HONG Yong-sheng, ZHOU Yong, et al(于雷, 洪永胜, 周勇, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2016, 36(5): 1428.
- [ 8 ] LIAO Qin-hong, GU Xiao-he, LI Cun-jun, et al(廖钦洪, 顾晓鹤, 李存军, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2012, 28(23): 132.
- [ 9 ] CAI Liang-hong, DING Jian-li(蔡亮红, 丁建丽). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2017, 33(16): 144.
- [10] Wang H F, Huo Z G, Zhou G S, et al. *Plant Physiology & Biochemistry*, 2016, 98: 39.
- [11] CHEN Hong-yan, ZHAO Geng-xing, LI Xi-can, et al(陈红艳, 赵庚星, 李希灿, 等). *Chinese Journal of Applied Ecology(应用生态学报)*, 2011, 22(11): 2935.
- [12] Lark R M. *Geoderma*, 2016, 266: 84.
- [13] Yun Y H, Wang W T, Tan M L, et al. *Analytica Chimica Acta*, 2014, 807(1): 36.
- [14] Pu R, Gong P. *Remote Sensing of Environment*, 2004, 91(2): 212.
- [15] SONG Kai-shan, ZHANG Bai, WANG Zong-ming, et al(宋开山, 张柏, 王宗明, 等). *Chinese Journal of Plant Ecology(植物生态学报)*, 2008, 32(1): 152.
- [16] ZHU Ya-xing, YU Lei, HONG Yong-sheng, et al(朱亚星, 于雷, 洪永胜, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2017, 50(22): 4325.
- [17] Bruce L M, Li J. *Geoscience & Remote Sensing IEEE Transactions on*, 2001, 39(7): 1540.
- [18] Yu X, Liu Q, Wang Y B, et al. *Catena*, 2016, 137: 340.
- [19] LI Pei-zhe(李培哲). *Statistics and Decision(统计与决策)*, 2012, (24): 89.
- [20] Xu S X, Shi X Z, Wang M Y, et al. *Geoderma*, 2016, 267: 174.

## Determination of Soil Organic Matter Content Based on Hyperspectral Wavelet Energy Features

ZHANG Tao<sup>1,2,3</sup>, YU Lei<sup>1,2,3\*</sup>, YI Jun<sup>1,2,3</sup>, NIE Yan<sup>1,2,3</sup>, ZHOU Yong<sup>1,2,3</sup>

1. Hubei Provincial Key Laboratory for the Analysis and Simulation of Geographical Process, Central China Normal University, Wuhan 430079, China
2. College of Urban and Environmental Science, Central China Normal University, Wuhan 430079, China
3. Research Institute for Sustainable Development of Central China Normal University, Wuhan 430079, China

**Abstract** There is no silver-bullet solution of eliminating noise during the acquisition process of soil hyperspectral. As the noise interference, the observations of soil spectra are in low signal-to-noise ratio, which affects the estimation accuracy of soil organic matter content. This paper attempts to adopt the wavelet energy features method to reduce the noise in soil hyperspectral and improve the estimation accuracy of soil organic matter content. The Yunlianghu Farm of Qianjiang City, Hubei Province, located in the hinterland of Jiangnan Plain, was selected as the experimental area, and 80 samples of paddy soil with a depth of 0~20 cm were collected in September 2016. After pretreatment (air drying, milling, sieving), soil sample spectral reflectance and determine soil organic matter content were collected in the laboratory. The concentration gradient method was employed to divide the whole sample set (80 samples) into a calibration set (54 samples) and a validation set (26 samples). Continuous wavelet transformation was performed using mexh as a wavelet basis function, transforming the soil hyperspectral into sensitive wavelet coefficients of 10 decomposition scales. Then the root mean square of the wavelet coefficients was calculated scale by scale to define wavelet energy features, and the wavelet energy features vector was determined by the wavelet energy features. The correlation coefficients between the wavelet coefficients and the organic matter content were calculated scale by scale and wavelength by wavelength, and the wavelet coefficient which reaches the extremely significant level ( $p < 0.01$ ) was defined as the sensitive wavelet coefficients. Principal component analysis was conducted to calculate the principal component loads of soil hyperspectral and wavelet energy features vector, respectively. The trend of principal component information of modeled independent variables before and after wavelet energy features transformation would be judged from the difference between the first principal component contribution rate and the spatial dispersion of the first three principal component scores degree. Moreover, regression models were established based on wavelet energy features vector and sensitive wavelet coefficients, respectively, to estimate soil or-

ganic matter content. The results showed that with the increase of soil organic matter content, the full-band reflectance decreased, but the spectral reflectance curves of different soil samples were similar, and the reflectance in the near-infrared bands (780~2 400 nm) was higher than that in the visible bands (400~780 nm). The sensitive wavelet coefficients corresponded to wavelengths of 494, 508, 672, 752, 1 838, and 2 302 nm. The first principal component contribution rates of soil hyperspectral and wavelet energy features vector were 96.28% and 99.13%, respectively. The first three principal component scatter points of wavelet energy features vector were more spatially aggregated than those of soil hyperspectral, which demonstrated that the wavelet energy features method effectively reduces the influence of noise. Comparing the estimation models of soil organic matter content, the multivariate linear regression model adopting wavelet energy features vector as the independent variable had the highest estimation accuracy, whose determination coefficients ( $R^2$ ), relative estimate deviation (RPD), and the root mean squared error (RMSE) of validation set were 0.77, 1.82, and 0.82, respectively. Therefore, the wavelet energy features method which is proved to raise the signal-to-noise ratio of the data without adding to the complexity, could improve the estimation accuracy of soil organic matter and realize the dimensional reduction of soil hyperspectral data. This method can be applied to studies like on-the-go soil properties measurement and soil quality monitoring.

**Keywords** Soil hyperspectral; Wavelet coefficients; Wavelet energy features; Soil organic matter; Paddy soil

(Received Aug. 27, 2018; accepted Dec. 21, 2018)

\* Corresponding author