

基于高光谱的土壤有机质含量估算研究

刘磊^{1,2}, 沈润平^{1,2*}, 丁国香²

1. 南京信息工程大学, 气象灾害省部共建教育部重点实验室, 江苏 南京 210044

2. 安徽省气象局, 安徽 合肥 230001

摘要 高光谱遥感技术以其光谱分辨率高、波段连续性强、数据丰富的特点, 因而在土壤养分研究中得到广泛应用。通过土壤有机质的高光谱遥感分析, 可以充分了解土壤养分的状况及动态变化, 为指导农业生产及保护农业生态环境提供科学依据。本文基于江西省余江县和泰和县采集的34个红壤土样350~2500 nm波段的光谱曲线, 研究了土壤光谱与土壤有机质含量之间的关系。先对土壤反射率光谱进行两种变换: 一阶微分(R')、倒数的对数 $\log(1/R)$, 然后在提取特征吸收波段的基础上, 运用多元逐步线性回归法和偏最小二乘回归法建立相应的估算模型, 并对模型进行检验。结果表明, 偏最小二乘回归法优于多元逐步线性回归法, 其建立的高光谱估算模型具有快速估算土壤有机质含量的潜力。

关键词 高光谱; 土壤有机质; 多元逐步回归; 偏最小二乘回归

中图分类号: S127; TP79 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2011)03-0762-05

引言

土壤有机质是存在于土壤中所有含碳的有机物质, 其含量的多少, 是土壤肥力的一个重要指标。通过土壤有机质的高光谱遥感分析, 了解土壤的现状, 并据此进行农业生产管理是保证农业高产、优质、高效的重要前提。

对土壤光谱与土壤有机质含量之间关系的研究, Alibas等^[1]发现土壤光谱反射率与有机质含量呈显著负相关。Galvos等^[2]通过室内研究证实了土壤反射光谱在550~700 nm处的吸收峰主要是由有机质引起, 并用AVIRIS高光谱影像资料分析了有机质与土壤反射光谱的关系, 取得了一系列进展。我国土壤光谱与有机质含量关系的研究的开始于20世纪80年代。徐彬彬等^[3]通过对南疆土壤的研究, 建立了用光谱反射系数估算土壤有机质含量的关系式。曾志远^[4]利用资源卫星多光谱图像, 借助统计方法, 将得到的土壤表层有机质含量表示成卫星图像四个波段辐射值的函数, 影响深远。周清等研究结果表明, 不同母质发育的水稻土的有机质含量高光谱模型因粘上矿物不同而选择不同光谱参数模型。谢伯承^[5]、王静^[6]、郑立华^[7]、贺军亮等^[8]在对土壤光谱分析的基础上, 构建出估算有机质含量的最佳模型, 取得了一系列研究成果。

尽管国内外对土壤有机质含量的光谱效应进行了大量的研究工作, 但在认识上还存在一定的分歧, 主要是由于以上研究大多基于多种土壤类型, 不同的土壤, 由于受气候、母质、地形、生物等成土因素不同和成土年龄的差异, 以及人类活动的影响, 其理化特性具有明显差别。土壤组成物质的多样性以及每种组成物质独特光谱特性, 使得各类土壤光谱具有自己的特点。因此, 将研究结果应用到特定土壤时, 都可能产生较大误差。基于此, 本文尝试先对土壤反射率光谱进行数学变换, 在提取特征吸收波段的基础上进行光谱重采样, 去除冗余信息和噪音后, 再运用多元逐步线性回归法和偏最小二乘回归法研究土壤光谱与土壤有机质含量之间的关系, 建立相应的估算模型, 为土壤有机质的定量遥感反演奠定基础。

1 材料与方法

1.1 土壤样本的采集

土壤样本的采集在中国科学院红壤生态实验站和千烟洲生态试验站进行, 属我国中亚热带典型红壤丘陵区。本次研究共采集土壤样品34个, 其中红壤生态实验站采集不同作物条件下的0~20 cm土层土壤样品22个, 千烟洲生态试验站采集土壤样品12个, 覆盖林地、草地、花生地、油菜地、

收稿日期: 2010-03-29, 修订日期: 2010-06-30

基金项目: 国家(973计划)项目(G20000779)和江苏省“青蓝工程”和南京信息工程大学重点基金项目资助

作者简介: 刘磊, 1986年生, 南京信息工程大学硕士研究生 e-mail: paopaogouuse@hotmail.com

* 通讯联系人 e-mail: rpshen@nuist.edu.cn

果园等土地利用类型。土壤样品经风干、碾磨,过 20 目筛。每个样品分成两份,一份用于光谱测量,一份用于化学分析。本研究土壤化学分析工作由中国科学院南京土壤研究所红壤生态试验站完成,土壤有机质含量用重铬酸钾氧化-外加加热法测定^[9]。

1.2 光谱测定

光谱测试仪器是美国 ASD 公司生产的 ASD Field SpecPro FR 型光谱仪,光谱范围为 350~2 500 nm,其中 350~1 000 nm 光谱采样间隔为 1.4 nm,在 1 000~2 500 nm 范围内光谱采样间隔 2 nm,重采样间隔为 1 nm。室内几何测试条件参见文献^[10]。将每个土样测得的 10 条反射光谱曲线经算术平均后则为该土样的实际反射光谱数据。

1.3 光谱数据预处理

1.3.1 光谱数据的数学变换

光谱一阶微分和倒数的对数是两种常用的光谱处理方法^[11],其中,对光谱曲线作一阶微分变换,可以对重叠混合光谱进行分解以便识别,扩大样品之间的光谱特征差异;光谱反射率经倒数的对数变换后,不仅趋向于增强可见光区的光谱差异,而且趋向于减少因光照条件变化引起的乘性因素影响^[12]。土壤样品的一阶微分和倒数的对数的光谱曲线分别见图 1 和图 2。

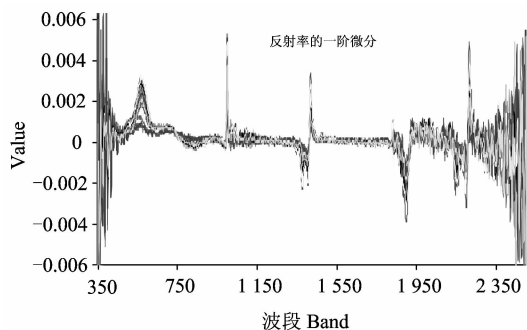


Fig. 1 Spectral curves of first derivative reflectance

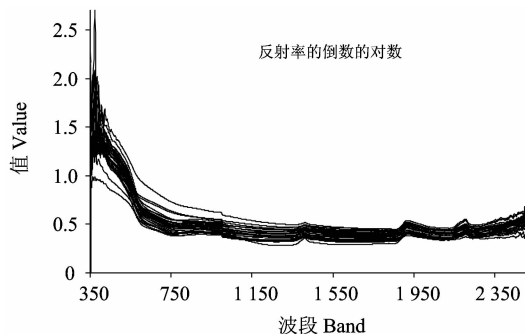


Fig. 2 Spectral curves of inverse-log reflectance

1.3.2 特征吸收带的提取

从前人的相关研究可知,土壤有机质对土壤光谱有吸收作用,表现为降低光谱反射率,因此,本文将基于反射率光谱曲线上的这些特征吸收带进行分析建模,估算土壤中的有机质含量。相较将整个可见光、近红外光谱波段作为研究对象,可能取得更好的效果。本文利用去包络线法^[13]对反射率

光谱曲线上的特征吸收带进行提取。图 3 为土壤样本反射率光谱曲线的包络线。以包络线为背景,去掉包络线后便为光谱波段深度曲线(图 4)。即将每一样点土壤反射率归一到对应的光谱背景上,有利于光谱曲线之间特征波段的比较,并且可以有效地突出光谱曲线的吸收和反射特征。光谱波段深度曲线计算公式是

$$R'(\lambda) = 1 - R(\lambda)/R_c(\lambda)$$

式中, R' , R , R_c 分别是光谱波段深度、原始光谱和光谱包络线, λ 为波长。

从图 4 可以发现,原土壤反射率光谱曲线去包络线后,其特征吸收带大致为 480~580, 820~950, 1 010~1 060, 1 360~1 500, 1 880~2 020, 2 160~2 240 nm。

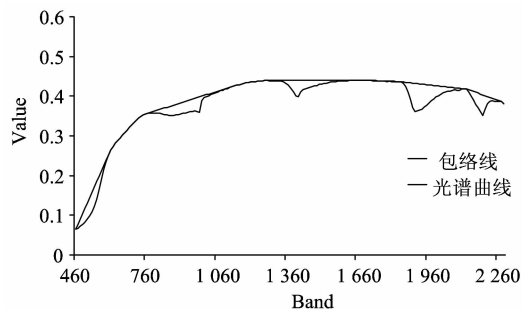


Fig. 3 Spectral curve and hull curve of some soil sample

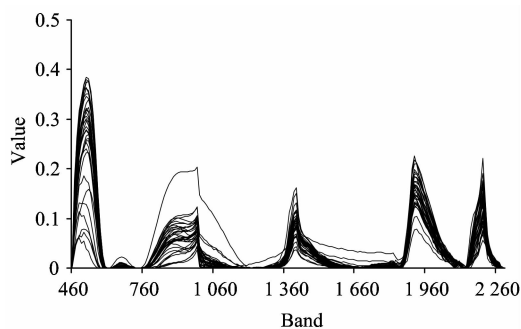


Fig. 4 Band-depth of soil samples

1.3.3 光谱重采样

由于光谱仪在数据输出时对 350~2 500 nm 的光谱数据进行了 1 nm 为间隔的重采样,总共 2 151 个波段,这使得原始光谱曲线中相邻波段之间存在信息重合,整个光谱数据冗余。同时,由于光谱仪在不同波段对能量响应的差异,原始光谱曲线存在许多“毛刺”噪声,光谱曲线不太平滑。此外,无论是经典的多元逐步回归法,还是偏最小二乘回归法,都存在变量多样本少的问题,这必然给分析、处理带来一定的困难。因此,在维持光谱原有基本特征的前提下。本研究对特征吸收带的光谱数据以 10 nm 为间隔进行算术平均运算,得到 64 个新的反射率波段,并对 64 个波段进行一阶微分和倒数的对数变换。

2 结果与分析

2.1 有机质的多元逐步回归模型

本文将 34 个土壤样本数据分为两组,第一组 25 个样本

进行多元逐步回归分析^[14], 建立回归模型, 第二组 9 个样本用于模型的精度检验。分别将反射率、一阶微分、倒数的对数这三个光谱指标的 64 个波段作为自变量, 土壤有机质含量实测值作为因变量, 进行多元逐步回归分析。设定变量方差贡献显著水平 0.05 作为选入和剔除变量的标准, 建立有机质含量的高光谱反演模型(Y 为有机质含量的反演值)。

(1) 反射率模型表达式: $Y = 22.34 - 692 * b_{820} + 673 * b_{880}$ (b_{820} , b_{880} 分别为 820~829 和 880~889 nm 波段范围的反射率平均值)。

(2) 一阶微分模型表达式: $Y = 33.71 - 10.988 * b_{570} - 84.491 * b_{1470} - 12.658 * b_{1900}$ (b_{570} , b_{1470} , b_{1900} 分别为 570~579, 1470~1479, 1900~1909 nm 波段范围的一阶微分的平均值)。

(3) 倒数的对数模型表达式: $Y = 9.349 + 111 * b_{820} - 45 * b_{490}$ (b_{820} , b_{490} 分别为 820~829 和 490~499 nm 波段范围的反射率倒数的对数的平均值)。

可以看出, 入选波段大多集中在 400~1000 nm 的范围, 土壤光谱的可见光区域主要为有机质的响应区。表 1 为

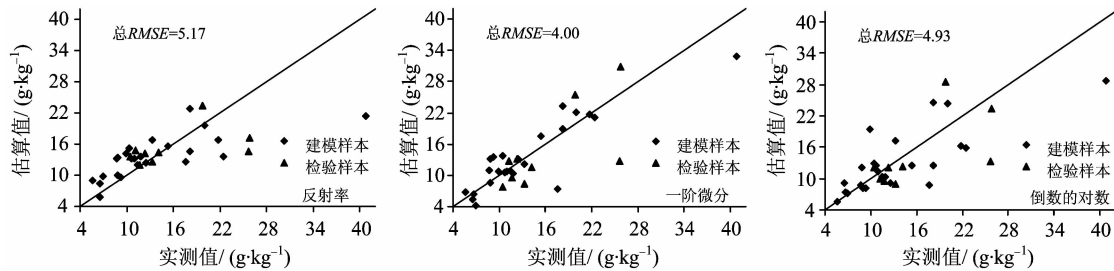


Fig. 5 Comparison of the measured SOM and the values estimated by SMLR model

2.2 有机质的偏最小二乘回归模型

与构建多元逐步回归模型方法相同, 进行偏最小二乘回归分析^[15]。模型的检验结果见表 2。

Table 2 Test results of PLSR model for soil organic matter (SOM)

模型	建模		模型检验		均方根误差
	决定系数 R^2	相关系数	平均相对误差/%		
反射率	0.97	0.84	16		1.91
一阶微分	0.99	0.58	22		3.09
倒数的对数	0.96	0.91	17		2.12

从表 2 可以看出, 反射率、一阶微分、倒数的对数这 3 个模型的建模精度都很高, 决定系数都在 0.96 以上。反射

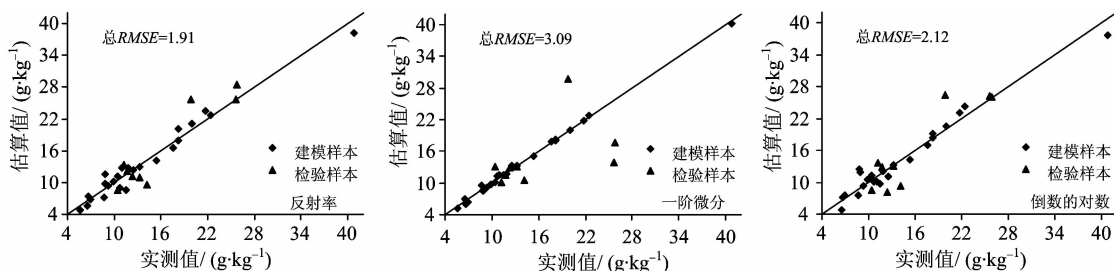


Fig. 6 Comparison of the measured SOM and the values estimated by PLSR model

土壤有机质多元逐步回归模型的检验结果。观察发现, 反射率、倒数的对数这 2 个模型的建模精度和检验精度均不高, 而一阶微分模型的效果相对较好, 这可能是光谱进行了一阶微分后扩大了样本光谱特征之间的差异所致。

Table 1 Test results of SMLR model for soil organic matter (SOM)

模型	建模		模型检验		均方根误差
	决定系数 R^2	相关系数	平均相对误差/%		
反射率	0.52	0.49	20		5.17
一阶微分	0.78	0.71	24		4.00
倒数的对数	0.58	0.64	22		4.93

图 5 为有机质含量实测值和估算值的比较, 横坐标是土壤样品有机质含量的实测值, 纵坐标是回归模型的反演值。样点越接近 1:1 线, 说明模型反演值越接近实测值, 模型的反演效果越好。因此, 一阶微分在多元逐步线性回归模型中表现较好, 反演能力较好。

率、倒数的对数的检验精度较高, 其对检验样本估算的平均相对误差分别为 16% 和 17%, 估算值和实测值的相关系数分别达到 0.84 和 0.91。而一阶微分模型的精度相对较差。

图 6 给出了有机质含量实测值和估算值的比较, 可以看出, 反射率、倒数的对数 2 个模型的估算精度较高, 建模样点和检验样点基本上都分布在 1:1 线附近, 而一阶微分模型效果相对略差, 特别是对检验样本的估算中, 有 3 个样点已经偏离了 1:1 线。将图 6 与图 5 进行对比可以发现, 偏最小二乘回归模型的精度明显优于多元逐步回归模型, 其估值更逼近实测值。可能的原因是, 土壤有机质成分复杂, 功能团多样, 其光谱特性可能很难用几个波段解释, 即可能不存在“敏感”波段, 多元逐步回归模型只利用了几个波段的光谱信息, 而偏最小二乘回归则从整各波段的光谱信息中提取

能够最大解释土壤有机质含量不同的主成分,同时要求这些主成分能够对有机质含量的相关性最大,大大的减少了光谱维数。可以说,用偏最小二乘回归方法对高光谱与土壤有机质含量之间进行回归分析,是分析土壤高光谱特征、揭示光谱特征波段与土壤有机质含量关系的一个很好的尝试。

3 小 结

本文将光谱数据以 10 nm 为间隔进行算术平均运算,在保留尽量多的光谱信息并维持光谱原有基本特征的前提下,对于减小数据量,尤其是去除冗余信息起到了很好的作用。基于反射率光谱曲线上的特征吸收带进行分析建模,估算土壤中的有机质含量,在偏最小二乘回归中取得了较好的效果。运用反射率、一阶微分(R')、倒数的对数 $\log(1/R)$ 这三种光谱数据参与多元逐步回归和偏最小二乘回归分析,发现一阶微分在多元逐步线性回归中表现较好,而在偏最小二乘

回归中,反射率和倒数的对数表现较好。但模型建立的好坏要根据相关系数和检验点两者结论的综合进行评定。偏最小二乘回归在处理样本容量小、自变量多、变量间存在严重多重相关性方面具有独特的优势,能减少或消除噪声因素的干扰,比多元线性逐步回归能更有效地提取和利用光谱中的重要信息。其验证精度比起多元逐步回归分析总体来说要高一些,但优势并不很明显。从模型的结果来看,偏最小二乘回归法优于多元逐步线性回归法。其中,反射率、倒数的对数的偏最小二乘回归模型达到较高精度,其对检验样本估算的平均相对误差分别为 16% 和 17%,估算值和实测值的相关系数分别达到 0.84 和 0.91,因而具有快速估算土壤有机质含量的潜力。

致谢:中国科学院遥感应用研究所路鹏博士后、研究生董晶晶,北京师范大学研究生孙知文和南京信息工程大学遥感学院 02 级杨安民、朱万强同学参与了野外土壤样品采集工作,特此致谢!

References

- [1] Al-Abbas A H, Swain P H, Baumgarder M F. *Soil Sci.*, 1972, 114(6): 477.
- [2] Galvao L S, Vitorello I. *Remote Sens. Environ.*, 1998, 63: 166.
- [3] XU Bin-bin(徐彬彬). *Soils(土壤)*, 2000, 32(6): 281.
- [4] ZENG Zhi-yuan(曾志远). *Soils(土壤)*, 1987, 19(2): 72.
- [5] XIE Bo-cheng, XUE Xu-zhang, LIU Wei-dong, et al(谢伯承, 薛绪掌, 刘伟东, 等). *Acta Pedologica Sinica(土壤学报)*, 2005, 42: 171.
- [6] WANG Jing, HE Ting, LI Yu-huan, et al(王 静, 何 挺, 李玉环, 等). *Journal of Remote Sensing(遥感学报)*, 2005, 4: 438.
- [7] ZHENG Li-hua, LI Min-zan, PAN Luan, et al(郑立华, 李民赞, 潘 雯, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2008, 28(5): 1160.
- [8] HE Jun-liang, JIANG Jian-jun, ZHOU Sheng-lu, et al(贺军亮, 蒋建军, 周生路, 等). *Journal of Agricultural Mechanization Research(农机化研究)*, 2009, (3): 30.
- [9] DU Sen, GAO Xiang-zhao, et al(杜 森, 高祥照, 等). *Technical Specification of Soil Analysis(土壤分析技术规范)*. Beijing: China Agriculture Press(北京: 中国农业出版社), 2006. 36.
- [10] SHEN Run-ping, DING Guo-xiang, WEI Guo-shuan, et al(沈润平, 丁国香, 魏国栓, 等). *Acta Pedologica Sinica(土壤学报)*, 2009, 46(3): 391.
- [11] LU Yan-li, BAI You-lu, YANG Li-ping, et al(卢艳丽, 白由路, 杨俐苹, 等). *Scientia Agricultura Sinica(中国农业科学)*, 2007, 40(9): 1989.
- [12] HE Ting, WANG Jing, LIN Zong-jian, et al(何 挺, 王 静, 林宗坚, 等). *Geomatics and Information Science of Wuhan University(武汉大学学报·信息科学版)*, 2006, 31(1): 975.
- [13] XU Yong-ming(徐永明). *Acta Pedologica Sinica(土壤学报)*, 2006, 9(5): 709.
- [14] Grossman Y L, Ustin S L, Jaquemoud S, et al. *Remote Sensing of Environment*, 1996, 56: 1.
- [15] JIANG Hong-wei, XIA Jie-lai(蒋红卫, 夏结来). *Journal of the Fourth Military Medical University(第四军医大学学报)*, 2003, 24(3): 280.

Studies on the Estimation of Soil Organic Matter Content Based on Hyper-Spectrum

LIU Lei^{1, 2}, SHEN Run-ping^{1, 2*}, DING Guo-xiang²

1. Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science & Technology, Nanjing 210044, China

2. Anhui Meteorological Bureau, Hefei 230001, China

Abstract Hyperspectral remote sensing technology can be extensively applied in soil nutrient research due to its three special advantages, high spectral resolution, strong waveband continuity as well as a great deal of spectral information. Based on analyzing the soil organic matter content using hyper-spectral remote sensing technology, soil nutrients status and its dynamic changes can be fully understood, thus providing the scientific basis for guidance of the agricultural production and protection of agricultural ecological environment. The present paper studies the relationship between soil spectrum and soil organic fraction based on spectrum curves (ranging from 350 to 2 500 nm) of 34 soil samples, which were collected in Yujiang and Taihe County, Jiangxi Province. First, soil reflection spectrum was mathematically manipulated into first derivative reflectance spectra (FDR) and inverse-log spectra ($\log(1/R)$); second, the relationship between soil spectrum and soil organic fraction was investigated by stepwise multiple linear regression (SMLR) and partial least square regression (PLSR) on the ground of characteristic absorption; third, corresponding estimation model was built and examined. The result conveys that spectral data are compressed by carrying out arithmetic average operation by 10nm for intervals. The first derivative of the reflectivity is an effective spectrum indicator, in the stepwise multiple linear regression analysis of soil organic matter, for the first derivative transformation, the regression models' precision of establishment and verification increased. The model built by PLSR method based on the characteristic absorption bands precedes that of SMLR. In the PLSR model of soil reflection spectrum and the inverse-log spectra, the test samples' average of relative error is 16% and 17% respectively, the correlation coefficient between retrieval value and measured value is 0.84 and 0.91 respectively, for it's faster to estimate the soil organic fraction.

Keywords Hyper-spectrum; Soil organic matter; Stepwise multiple linear regression; Partial least square regression

(Received Mar. 29, 2010; accepted Jun. 30, 2010)

* Corresponding author