

基于包络线去除和偏最小二乘的土壤参数光谱反演

彭小婷¹ 高文秀² 王俊杰³

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 深圳大学建筑与城市规划学院,广东 深圳,518060

3 武汉大学资源与环境科学学院,湖北 武汉,430079

摘要:基于江苏省宜兴市 100 个土样的可见光-近红外高光谱反射率(400~2 450 nm)数据,结合包络线去除(continuum removal,CR)与偏最小二乘回归(partial least squares regression,PLSR),构建了土壤重金属砷(As)和有机质(OM)含量的反演模型。结果表明,相比普通 PLSR 模型(模型决定系数 R^2 和预测根均方误差 RMSEP 分别为 0.512,3.090 和 0.621,5.934),CR-PLSR 构建的模型预测能力有明显的改善(R^2 和 RMSEP 分别为 0.763,2.323 和 0.911,4.599)。CR 有效增强了 550、900、1 420、1 900 和 2 200 nm 等波段处的反射光谱特征,根据模型回归系数分析,CR 有效突出的波段正是 As 和 OM 的 CR-PLSR 模型所共用的重要波段。研究表明,CR 能够协助 PLSR 模型重要波段的选择,利用遥感技术结合 CR-PLSR 能够有效提高土壤重金属 As 和 OM 含量的反演精度,从而为土壤质量的遥感监测提供参考。

关键词:高光谱;土壤质量;包络线去除;偏最小二乘回归

中图法分类号:P237

文献标志码:A

土壤质量与土壤中重金属和有机质(OM)含量有着密切联系,有效地监测土壤重金属砷(As)和 OM 含量有利于土壤资源的整治和保护^[1]。近年来,可见光-近红外高光谱技术以其便捷、实时、可靠、非破坏性的优势广泛用于土壤参数测定中^[2-7]。郑光辉、任红艳等通过比较一阶导数、二阶导数、多元散射校正和包络线去除(CR)等不同预处理方法,得出了多元散射校正对研究区 As 含量预测的效果最佳^[3,7]。土壤 OM 含量的反演方法很多,主要有多元逐步回归、主成分回归(PCR)、偏最小二乘回归(PLSR)和 BP 神经网络等^[2,7-8],其中任红艳等利用多种预处理方法与 PCR 相结合,得出 CR-PCR 模型为 OM 预测的最佳模型^[7]。在土壤参数定量反演方面,从数据预处理到模型的研究很多,但是大多数模型精度不高,推广较难。在目前的技术条件下,由于地表状况以及大气和光照影响,很难利用野外光谱和卫星遥感数据精确反演土壤参数,因此,主要集中在相对稳定的实验室条件下获取的室内高光谱数据^[7]。PLSR 是一个全波段建模的标准方法,广泛用于多种土壤参数反演^[2-3]。包络线法最早由 Clark 和 Roush 提出,定义为逐点直线连接随波

长变化的吸收或反射凸出的“峰”值点,并使折线在“峰”值点上的外角大于 180° ^[9],能够突出光谱曲线的吸收和反射特征,增强光谱曲线各波段之间的对比性^[10]。目前,将 CR 与全波段 PLSR 结合预测土壤 As 和 OM 的模型并不多。因此,本文利用室内高光谱遥感数据,结合 CR 和 PLSR(CR-PLSR)进行土壤 As 和 OM 含量预测,以期提高 PLSR 模型的土壤参数的反演精度和稳健性;并与未经 CR 的普通 PLSR 模型进行比较分析,验证该方法的可行性并探究 CR 提高 PLSR 模型精度的实质原因。

1 数据与方法

1.1 土壤样品采集与分析

研究区位于江苏省宜兴市,是典型的农业区,在农用地上设计 30 个采样区(30 m×30 m),每个区内采集 3~4 个表层土壤(0~10 cm)的土样,共 100 个土样,实验室内风干、移除杂物及研磨后,用于土壤参数化学分析和光谱测量。OM 含量测定采用重铬酸钾容量法,As 浓度采用原子荧光分光光度法。

收稿日期:2013-06-03

项目来源:国家自然科学基金资助项目(41171290)。

第一作者:彭小婷,硕士生,主要从事基于遥感技术的土壤和植被参数反演研究。E-mail: pxiaoting91@gmail.com

1.2 土样光谱测量与预处理

采用 ASD FieldSpec®3 便携式光谱仪在实验室内采集土样高光谱反射率数据(350~2 500 nm)。光源为 50 W 卤素灯,探头视场角 25°,光源入射角 15°,土样中心距离光源和探头分别为 40 cm 和 15 cm。每个土样采集 10 次光谱,取其平均值作为原始反射光谱。去除信噪比较低的波段(350~399 nm 和 2 451~2 500 nm),并对光谱数据实施 Savitzky-Golay 平滑。

1.3 建模方法与精度评价

PLSR 是一种全波段建模的标准方法,能够有效地提取对系统解释能力最强的综合变量,使之对因变量有最强的解释能力^[11]。剔除 100 个样本中的 4 个异常样本及 4 个裸地样本,对剩余 92 个样本实行 CR。根据 As 和 OM 含量排序结果,等间隔选取 23 个样本作为预测集,剩余 69 个作为建模集。本文分别用 CR-PLSR 和普通的 PLSR 构建 As 和 OM 反演模型,并对建立的模型和预测精度进行对比。

1) 运用 PLSR 分别对 CR 后和 CR 前的建模集进行全波段(400~2 450 nm)建模,通过“留一法”交叉验证,选择最佳因子数(latent variables,

LVs),确定 PLSR 模型。

2) 利用各自的预测集进行预测。模型精度的评价按照以下参数进行:建模精度通过模型决定系数(coefficients of determination, R^2)、交叉验证根均方误差(root mean square error of cross-validation, RMSECV)、交叉验证模型决定系数(R_{cv}^2)来评价;预测精度根据预测根均方误差(root mean square error of prediction, RMSEP)、预测模型决定系数(R_{pre}^2)评价。 R^2 、 R_{cv}^2 和 R_{pre}^2 越大, RMSECV 和 RMSEP 越小,表明模型的建模、预测精度越高。

2 结果与讨论

2.1 土壤光谱反射率

本文根据 92 个土地样本的原始反射光谱曲线(图 1),土壤反射率在可见光区域(400~780 nm)呈快速上升趋势,只是变化速率不同;红外波段区域(800~2 450 nm)变化较平缓,但在 1 400、1 900 和 2 200 nm 波段处有明显的吸收谷,即水分吸收带,这些波段通常认为是与粘土矿物中所含的 OH^- 有关^[12]。

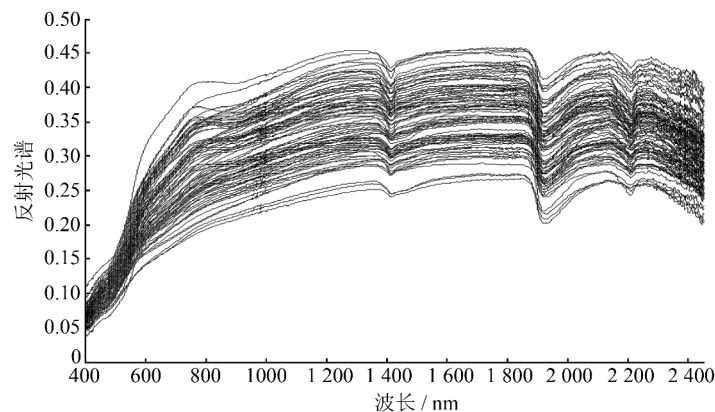


图 1 土壤原始反射光谱图

Fig.1 Original Soil Reflectance Spectrum

2.2 土壤光谱数据包络线去除分析

经过 CR 后的光谱吸收带更加明显,光谱曲线都归一化到 0~1 之间(图 2(b))。相比 CR 前的反射光谱曲线(主要突出 1 420、1 900、2 200 nm 波段处的吸收特征),从 CR 后的光谱曲线上可以明显看出 550、900、1 420、1 900、2 200 和 2 400 nm 波段有明显的吸收谷。可见光范围内的 550 nm 为 OM 的响应区。可见光-近红外的 600~930 nm 波段是铁的氧化物矿物光谱的主要吸收区,与 OM 含量密切相关。然而 CR 前的土壤光谱上并没有显示这两个吸收带,CR 后的光谱

也使得 2 100~2 300 nm 之间的吸收谷更加明显,主要是由残余在土壤中少量水分以及空气中吸收水分引起的。任红艳等^[7]利用实验室内光谱测定了土壤 As 污染浓度及 OM 含量,得出预测 As 和 OM 含量使用的共同重要波段是 450、1 000、1 400、1 900、2 050、2 200、2 250、2 400 和 2 470 nm。这些重要波段与 CR 有效突出的波段大致保持一致,因而 CR 后的反射光谱数据与 As 和 OM 含量之间的关系更加明显。

2.3 建模结果与讨论

本文分别用 CR-PLSR 和普通 PLSR 对 As

和 OM 进行建模和预测,结果见表 1。CR-PLSR 校正模型和预测模型的 RMSE 均小于 PLSR 模型, R^2 都大于 PLSR 模型,尤其 OM 的 R^2 从 0.621 显著增加到 0.911, RMSEP 减少 22.5%, R_{cv}^2 和 R_{pre}^2 分别达到 0.581 和 0.683,显著改善了 PLSR 模型的能力,能够定量预测 As 和 OM 含量。从图 3 可以看出,As 和 OM 的 CR-PLSR 与 PLSR 模型的回归系数 (regression coefficient, RC) 大致趋势保持一致,但是相比 PLSR 模型,CR-PLSR 模型 RC 变化幅度显著增大,各自分布范围为 $-20 \sim 20$ 和 $-110 \sim 110$ (相应的 PLSR 模型 RC 介于 $-10 \sim 5$ 以及 $-10 \sim 10$ 之间)。对应波段尤其是重要波段处的 RC 明显增加,不同波段处 RC 对比更加明显。相对于 As,OM 的 CR-PLSR 模型 RC 范围几乎是 PLSR 模型的 10 倍而且波段之间的 RC 波动最大,这可能是因为 OM 有显著的光谱特征,而 As 没有光谱特征。这也能解释为何 CR 对 OM 的 PLSR 建模和预测能力的改善程度大于对 As 建模和预测能力的改善程度。两者 CR-PLSR 模型和 PLSR 模型的 RC 在一些特定波段 (如 550、700、900、1 420、1 900 和 2 200 nm) 处均存在明显的波峰和波谷,而这些波

段正是 As 和 OM 预测重要的波段,与前人提出的 As 和 OM 预测的重要波段保持一致^[5,7]。由 § 2.2 可知这些重要波段大都是 CR 后显著突出的波段 (550、900、1 420、1 900、2 200 和 2 400 nm)。由此可知,CR 后的光谱曲线突出的 6 大吸收特征波段几乎全部参与 As 和 OM 建模中,由相应波段的 RC 可知,这些波段对模型贡献最大。为了探究为何 CR-PLSR 模型 RC 的较大波动性,分别计算 As 和 OM 含量与 CR 前后的光谱之间的相关系数 (correlation coefficient, CC) (图 3 (a) 与 3 (f))。相比未 CR 的光谱与 As 和 OM 含量之间的相关系数 (分别为 $-0.449 9 \sim 0.343$ 和 $-0.607 4 \sim 0.1023$), CR 后的相关系数 (分别为 $-0.515 4 \sim 0.456 4$ 和 $-0.508 5 \sim 0.665 0$) 明显增强且对比强烈,其中 OM 相关系数变化更加明显,导致 CR-PLSR 模型的 RC 增强因而波动也较大。这也可能是与 CR 的作用有关,CR 是一种归一化处理方法,增强了光谱曲线各波段之间的对比性并将其归一化到一个一致的光谱背景上,因而也对增强 PLSR 模型的 RC 和改善模型的精度起到重要作用。

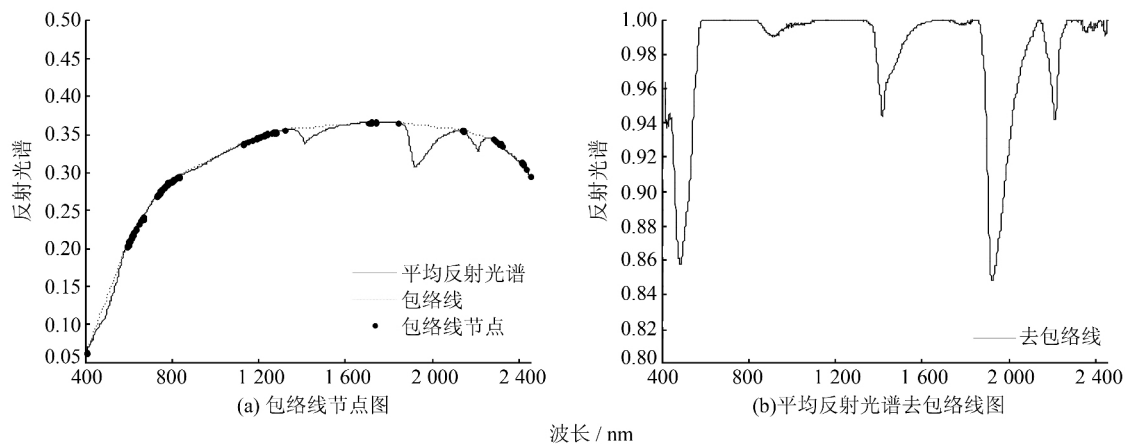


图 2 土样的平均反射光谱

Fig. 2 Curve of Average Reflectance Spectrum

表 1 As 和 OM 的 CR-PLSR、PLSR 模型的建模及预测结果

Tab. 1 Modelling and Prediction Results of CR-PLSR and PLSR Models for As and OM

土壤参数	CR-PLSR						PLSR					
	R^2	LVs	RMSECV	RMSEP	R_{cv}^2	R_{pre}^2	R^2	LVs	RMSECV	RMSEP	R_{cv}^2	R_{pre}^2
As	0.763	8	1.736	2.323	0.531	0.522	0.512	7	1.760	3.090	0.500	0.240
OM	0.911	11	5.943	4.599	0.581	0.683	0.621	6	6.005	5.934	0.564	0.474

3 结 语

本研究表明,CR-PLSR 模型能够快速、精确

的估算土壤中 As 和 OM 含量。CR 能够显著提高 PLSR 建模和预测能力,以改善 OM 建模和预测能力最佳。通过模型的 RC 分析,CR 有效突出的波段是 As 和 OM 的 CR-PLSR 模型所共用的

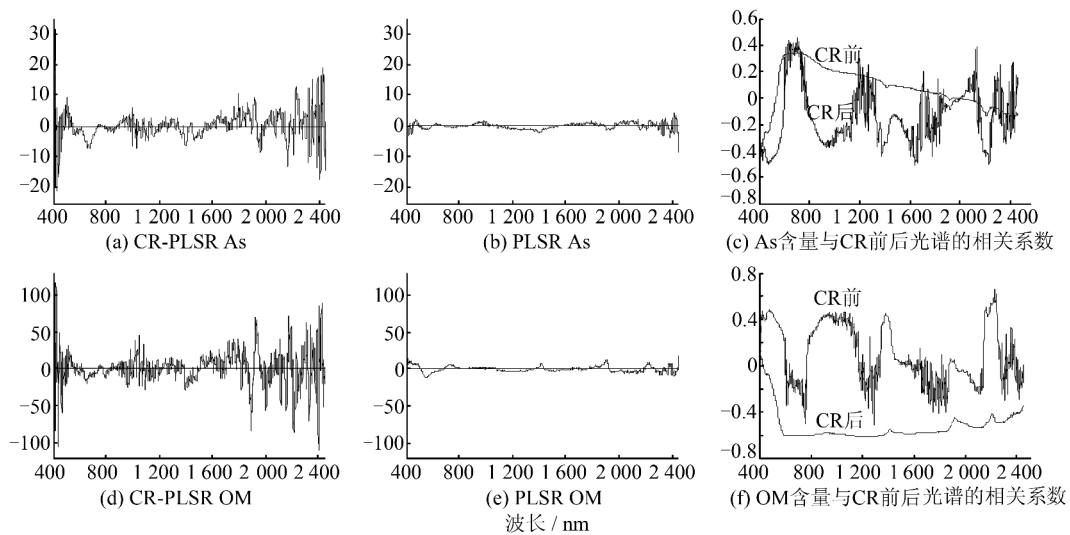


图 3 As 和 OM 的 CR-PLSR 模型和 PLSR 模型的 RC 以及与 CR 前后的光谱数据相关系数图

Fig. 3 Regression Coefficients of CR-PLSR and PLSR Models of As and OM Content and Pearson's Correlation Coefficient with Spectrum Before CR and After CR

重要波段: 550、700、900、1 420、1 900 和 2 200 nm, 更加支持了利用具有显著光谱特性的土壤属性(OM)与重金属之间的相关性, 运用光谱数据能够快速测定土壤中重金属含量的理论假设。今后的研究可尝试将 CR-PLSR 方法用于其他研究区域以及野外光谱甚至卫星遥感数据中用以验证和推广。

参 考 文 献

- [1] Li Fuyan, Li Xuming, Wu Pengfei, et al. Correlation Between Heavy Metal Pollution and Basic Properties of Agricultural Soils in Hainan Province[J]. *Soils*, 2009, 41(1): 49-53 (李福燕, 李许明, 吴鹏飞, 等. 海南省农用地土壤重金属含量与土壤有机质及 pH 的相关性[J]. *土壤*, 2009, 41(1): 49-53)
- [2] Zhu Dengsheng, Wu Di, Song Haiyan, et al. Determination of Organic Matter Contents and pH Values of Soil Using Near Infrared Spectroscopy [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2008, 24(6): 196-199 (朱登胜, 吴迪, 宋海燕, 等. 应用近红外光谱法测定土壤的有机质和 pH 值[J]. *农业工程学报*, 2008, 24(6): 196-199)
- [3] Zheng Guanghui, Zhou Shenglu, Wu Shaohua. Prediction of As in Soil with Reflectance Spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2011, 31(1): 173-176 (郑光辉, 周生路, 吴绍华. 土壤砷含量高光谱估算模型研究[J]. *光谱学与光谱分析*, 2011, 31(1): 173-176)
- [4] Ren H Y, Zhuang D F, Singh A N, et al. Estimation of As and Cu Contamination in Agricultural Soils Around a Mining Area by Reflectance Spectroscopy: A Case Study [J]. *Pedosphere*, 2009, 19(6): 719-726
- [5] Gomez C, Lagacherie P, Coulouma G. Continuum Removal Versus PLSR Method for Clay and Calcium Carbonate Content Estimation from Laboratory and Airborne Hyperspectral Measurements [J]. *Geoderma*, 2008, 148(2): 141-148
- [6] Xie Bocheng, Xue Xuzhang, Liu Weidong, et al. Hull-curve-method Based Extraction and Analysis of Soil Spectral Characteristics [J]. *Acta Pedologica Sinica*, 2005, 42(1): 171-175 (谢伯承, 薛绪掌, 刘伟东, 等. 基于包络线法对土壤光谱特征的提取及其分析[J]. *土壤学报*, 2005, 42(1): 171-175)
- [7] Ren Hongyan, Zhuang Dafang, Qiu Dongsheng, et al. Analysis of Visible and Near-infrared Spectra of As-Contaminated Soil in Crop-lands Beside Mines [J]. *Spectroscopy and Spectral Analysis*, 2009, 29(1): 114-118 (任红艳, 庄大方, 邱冬生, 等. 矿区农田土壤砷污染的可见-近红外反射光谱分析研究[J]. *光谱学与光谱分析*, 2009, 29(1): 114-118)
- [8] Zheng Lihua, Li Minzan, Pan Luan, et al. Estimation of Soil Organic Matter and Soil Total Nitrogen Based on NIR Spectroscopy and BP Neural Network [J]. *Spectroscopy and Spectral Analysis*, 2008, 28(5): 1 160-1 164 (郑立华, 李民赞, 潘鸾, 等. 基于近红外光谱技术的土壤参数 BP 神经网络预测[J]. *光谱学与光谱分析*, 2008, 28(5): 1 160-1 164)
- [9] Clark R N, Roush T L. Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications [J]. *J Geophys Res*, 1984, 89(B7): 6 329-6 340

- [10] Tong Qingxi, Zhang Bing, Zheng Fenlan. Hyperspectral Remote Sensing[M]. Beijing: Higher Education Press, 2006:145-148 (童庆禧,张兵,郑芬兰. 高光谱遥感[M]. 北京: 高等教育出版社, 2006: 145-148)
- [11] Wang Huiwen. Partial Least Squares Regression Method and Application[M]. Beijing: National Defence Industry Press, 2006:119-125 (王惠文. 偏最小二乘回归方法及应用[M]. 北京:国防工业出版社, 2006: 119-125)
- [12] Xu Binbin. The Reflection Spectrum Research of Soil Profile[J]. *Soils*, 2000, 32(6):281-287 (徐彬彬. 土壤剖面的反射光谱研究[J]. 土壤, 2000, 32(6): 281-287)

Inversion of Soil Parameters from Hyperspectra Based on Continuum Removal and Partial Least Squares Regression

PENG Xiaoting¹ GAO Wenxiu² WANG Junjie³

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China

3 School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China

Abstract: In the study, we combined continuum removal (CR) with a partial least squares regression (PLSR) method to build inversion models of soil As concentration and organic matter (OM) content based on hyperspectral data between 400-2 500 nm of 100 soil samples which were collected in Yixing region of Jiangsu Province. The results show that, compared with the common PLSR models (model's determination coefficients (R^2) and root mean square error of prediction (RMSEP) are 0.512, 3.090 and 0.621, 5.934, respectively), CR-PLSR models present an improvement of As and OM inversion models (R^2 and RMSEP are 0.763, 2.323 and 0.911, 4.599, respectively). In addition, CR effectively strengthens the spectral characteristics of 550, 900, 1420, 1900 and 2 200 nm wavelengths. The analysis of regression coefficients of CR-PLSR and PLSR models of As and OM demonstrate several important wavelengths enhanced by CR and are simultaneously used for CR-PLSR models of As and OM. Thus CR is able to help PLSR conduct wavelength selection and enhances the inversion precision of soil As and OM content effectively. This finding can be used as a reference for remote sensing monitoring of soil quality.

Key words: hyperspectra; soil quality; continuum removal (CR); partial least squares regression (PLSR)

First author: PENG Xiaoting, postgraduate, specializes in inversion research of soil and vegetation parameters based on remote sensing technology. E-mail: pxiaoting91@gmail.com

Foundation support: The National Natural Science Foundation of China, No. 41171290.