论文

中国主要土壤高光谱反射特性分类与有机质光谱预测模型

史舟^{①⑤*}, 王乾龙^①, 彭杰^②, 纪文君^①, 刘焕军^③, 李曦^①,
Raphael A VISCARRA ROSSEL^④

① 浙江大学环境与资源学院, 杭州 310058;

② 塔里木大学植物科学学院, 阿拉尔 843300;

③ 东北农业大学资源与环境学院, 哈尔滨 150030;

④ CSIRO Land & Water, Bruce E. Butler Laboratory, Canberra ACT 2601, Australia;

⑤ 浙江大学唐仲英传感材料及应用研究中心, 杭州 310058

* E-mail: shizhou@zju.edu.cn

收稿日期: 2013-05-09; 接受日期: 2013-07-18; 网络版发表日期: 2014-04-16

国家自然科学基金项目(编号: 41271234, 40871100)、教育部新世纪优秀人才支持计划项目(编号: NCET-10-0694)和浙江省杰出青年科学基金项目(编号: R5100140)资助

摘要 土壤可见-近红外漫反射光谱是当前对地遥感观察和土壤近地传感器研究的重要方向, 同时也被认为是土壤数字制图、精确农业和土壤资源调查等方面最重要的数据获取技术. 从中国西藏、新疆、黑龙江、海南等地采集 16 种土类的 1581 个土壤样本, 经干燥过筛后统一采用 ASD 光谱仪测量了其室内可见-近红外反射光谱(350~2500 nm). 对所有的土壤光谱数据采用 Savitzky-Golay 平滑加一阶微分进行转换, 来减少大样本数据受到实验室光学测试环境条件差异的影响, 然后对数据进行主成分变换降维处理. 引入模糊 *k*-means 方法进行大样本光谱数据的最佳分类数目计算, 并将中国土壤光谱数据分成五类, 各自代表了不同的土壤矿物和有机组分, 主要类型与国际同行类似成果有可比性. 最后提出了采用土壤光谱分类方法结合偏最小二乘回归法(PLSR)方法建立土壤有机质的光谱分类-局部预测模型, 结果比未分类直接采用 PLSR 方法的一阶微分-全局预测模型的精度有了显著提高, 其预测模型的 R^2 和 RPD 两个指数分别从 0.697 和 1.817 提高到 0.899 和 3.158.

关键词漫反射光谱
可见-近红外光谱
土壤有机质
光谱数据库
中国

地物光谱特性是卫星遥感解译和近地传感器探测的理论基础, 是对地遥感观测和地面智能感知技术研究的重要组成部分. 土壤作为地球表面五大典型地物之一, 其可见光-近红外反射光谱研究对当前

精准农业、土壤数字制图、土壤资源遥感调查等工作有着重要作用, 备受众多科研人员的关注(Gerbbers 和 Adamchuk, 2010; Sanchez 等, 2009). 从 20 世纪 70 年代开始, 大量研究已经充分肯定了利用土壤可见-近

中文引用格式: 史舟, 王乾龙, 彭杰, 等. 2014. 中国主要土壤高光谱反射特性分类与有机质光谱预测模型. 中国科学: 地球科学, 44: 978-988

英文引用格式: Shi Z, Wang Q L, Peng J, et al. 2014. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. Science China: Earth Sciences, doi: 10.1007/s11430-013-4808-x

红外反射光谱来区分不同土壤类型(Viscarra Rossel 和 Webster, 2011), 以及预测土壤有机质、黏土矿物、质地、水分、重金属含量等特性的能力(Stenberg 等, 2010; Kuang 等, 2012; Song 等, 2012). 最近的研究还将光谱范围进一步拓展到中红外部分(Reeves, 2010). 但由于不同区域土壤类型多样, 成土因素复杂, 造成不同类型土壤的高光谱特性和预测模型的差异很大, 需要我们在国家尺度乃至全球尺度对不同类型土壤进行大样品的测试分析, 建立相应的光谱数据库, 来全面评价土壤光谱特性和预测能力.

国际上关于不同类型大样本土壤反射光谱特性的系统研究最早见于 20 世纪 70 年代 Condit (1972) 的报道, 他采集和测试了美国 36 个州 160 个土样的反射光谱(320~1000 nm), 并根据谱线形状特性将土壤反射光谱分为三种类型; 而最为经典的当数 Stoner 和 Baumgardner (1981) 在 20 世纪 80 年代初测定了美国和巴西 485 个土壤的反射光谱(520~2320 nm), 提出了五类土壤反射光谱曲线: 有机质控制型、最小改变型、铁影响型、有机质影响型和铁控制型. 然而他们获得的土壤信息仍很有限, 不能很好的支持土壤光谱响应机理、光谱分类和预测模型等方面的系统研究. 从 2006 年开始, 国际上分别由美国国家土壤调查中心的 Brown 等和澳大利亚 CSIRO 的 Raphael 等牵头开展了两次全球土壤光谱库的收集和研究工作(Brown 等, 2006). 其中 2008 年开始 Raphael 等牵头的全球土壤数据建设和应用研究工作, 目前已收集了 90 个国家 16000 个土样的光谱数据和相应的理化性质数据, 其中包括中国的数据, 试图通过全球大样本光谱数据的比较分析, 建立适用于不同类型的土壤光谱分类方法或理化特性光谱预测模型. 本课题组参加了此项全球性工作. 围绕着全球、国家尺度的大样本土壤光谱库的应用, 最近国际上先后提出了不同的思路, 核心思想是如何处理大样本光谱数据的分异性, 设法建立相对均一的子样本集来进行预测建模. 如 Goge 等(2012)利用全法国 2200 个土壤样品光谱库, 结合 Local 算法进行地方土壤有机碳、CEC、黏土等属性的快速预测. Brown (2007)利用全球 4184 个土壤样品光谱库, 结合 BRT (Boosted regression trees)数据挖掘技术对乌干达某流域内的土壤有机碳、黏土等进行预测中得到了很好的应用. Wetterlind 和 Stenberg (2010)利用瑞典全国光谱数据

库信息结合 Spiking 方法建立田间尺度的土壤有机碳和机械组成等关键属性的预测模型. 但是在大量光谱数据分割、类似样本局部搜索、本地对照样本嵌入等规则研究上还存在很多的问题(Guerrero 等, 2010; Ramirez-Lopez 等, 2013).

中国虽然陆续开展了部分区域土壤反射光谱特性的研究, 但主要是针对土壤分类的光谱研究工作. 如早期的戴昌达(1981)、徐彬彬等(1991)、吴豪翔和王人潮(1991), 以及近期的刘焕军等(2008)根据土壤反射特性曲线的形态与斜率等特征, 进行了土壤光谱分类工作. 同时在区域或小样本土壤属性光谱预测方面开展了很多的工作. 但是与国际同行工作类似, 到目前为止还未系统的收集中国不同区域不同土壤类型以及大样本的数据来解释土壤光谱预测建模的潜力.

本研究从中国西藏、新疆、黑龙江、海南等地采集了 16 种土类的 1581 个土壤样本, 获取其室内可见-近红外漫反射光谱(350~2500 nm). 研究如何利用反射光谱强度与曲线特征来进行土壤光谱分类, 并解决最佳分类数目的计算问题; 研究利用土壤反射光谱特性建立土壤有机质含量的光谱预测模型, 并通过先光谱分类, 后偏最小二乘回归(PLSR)预测建模的方法来提高光谱预测的精度, 探求可应用于中国不同类型土壤的有机质光谱预测方法.

1 数据采集与测试

1.1 土壤采集

采集土壤样品均为 0~20 cm 的表层土壤, 由各合作单位野外采集, 从中国十多个省市区域共收集到 16 个土类 1581 个土壤样品, 初步建立了中国范围的土壤反射光谱样本数据库. 土壤样本经风干、研磨并通过 2 mm 孔筛, 采用四分法分成两份, 分别供化学分析和光谱测试用. 采用重铬酸钾容量法-外加热法分析有机质含量(中国土壤学会, 2000). 表 1 为供试土壤样品的基本情况, 16 种土类按照平均有机质含量从高到低排列, 可以看出土壤样品有机质含量变幅较大, 从 1.66 到 74.84 g/kg, 这为全面解析土壤有机质反射光谱特性研究提供了较完整的样本数据. 有机质含量较高的主要是东北的草甸土、黑土、黑钙土以及南方的水稻土, 有机质相对低的主要是发育程度低的风沙土、初骨土以及滨海盐土.

表 1 不同类型土壤有机质含量的统计特征^{a)}

土壤类型(土类)	采样区域	样本数	有机质(g/kg)			
			最小值	最大值	均值	均方差
草甸土	吉林	52	16.9	47.0	30.46	6.51
水稻土	浙江/福建/湖北/湖南/四川/新疆	552	3.2	74.84	29.66	12.08
黑土	黑龙江/吉林	104	12.0	58.6	28.67	7.50
黑钙土	吉林	138	11.0	43.6	25.39	5.63
冲积土	吉林	11	11.3	32.0	23.96	6.85
砖红壤	海南	19	11.3	34.84	20.60	6.58
紫色土	四川/浙江/湖南/江西	107	1.66	43.4	20.23	9.40
褐土	山西	12	5.62	32.2	18.48	8.50
冷棕钙土	西藏	38	1.67	38.5	18.46	7.02
红壤	浙江/江西/湖北/湖南	172	3.39	59.3	18.40	10.44
石灰土	云南/浙江	57	4.35	36.3	17.01	9.82
赤红壤	福建	39	4.44	61.5	16.21	12.10
风沙土	吉林	18	8.2	32.8	15.28	6.95
粗骨土	浙江	21	6.1	35.9	15.08	8.15
潮土	河南/浙江/湖北/湖南	127	5.32	33.77	12.82	4.10
滨海盐土	浙江	114	3.71	31.5	12.46	4.53
合计	/	1581	1.66	74.84	23.14	11.56

a) 表中土壤类型参照中国土壤发生分类系统(中国土壤普查办公室, 1996)

1.2 光谱测试

采用美国 ASD 公司的 FieldSpec Pro FR 型光谱仪进行土壤测试, 其波长范围是 350~2500 nm, 采样间隔为 1.4 nm (350~1000 nm 区间)和 2 nm (1000~2500 nm 区间), 重采样间隔为 1 nm, 输出波段数为 2150. 光谱测量在一个能控制光照条件的暗室内进行, 光源为卤素灯, 距土壤样品表面 70 cm, 天顶角 30°. 土壤样本分别放置在直径 10 cm、深 1.5 cm 的盛样器皿内, 土样表面刮平. 传感器探头位于土壤样本表面垂直上方 15 cm 处. 采用 25°视场角探头, 探头接受土壤光谱的区域为直径 6.7 cm 的圆, 远小于盛样器皿面积, 保证探头接收的均为土壤的反射光谱. 测试之前先进行白板校正. 每个土样采集 10 条光谱曲线, 算术平均后得到该土样的实际反射光谱数据.

2 方法

2.1 光谱数据预处理

对光谱数据的预处理主要包括两方面内容: (1) 去噪和数据变换处理. 每个土样的光谱曲线去除噪声较大的边缘波段(350~399 和 2401~2500 nm), 保留 400~2400 nm 波段的反射光谱数据用于后续研究.

为减少实验室光学环境场差异和磨样过筛的影响, 对所有的光谱数据我们尝试了采用 Savitzky-Golay 平滑加一阶微分或二阶微分变换、多元散射校正 (Multiplicative Scatter Correction, MSC)、标准正态变量变换(Standard Normal Variate, SNV) 等处理方法. 通过对后期预测效果的比较, 我们选用 Savitzky-Golay 平滑加一阶微分数据转换方法, Savitzky-Golay 平滑采用二次多项式和左右各 10 个样点平滑. 通过微分处理计算光谱梯度, 还可以确定光谱弯曲点, 以及最大值与最小值波长位置; (2) 数据主成分处理. 为压缩土壤高光谱数据, 采用主成分分析方法将原始 2000 个波段(400~2400 nm)的反射光谱数据进行压缩, 然后对转换后的前几个主成分得分数据进一步分析.

2.2 光谱数据分类方法

光谱分类是将具有相似光谱反射率和曲线特征的土壤分为一类, 常用的方法包括系统聚类、线性判别分析、*k* 均值聚类等方法. 但在进行中国大样本土壤光谱分类时, 关键问题是如何给出最佳的分类数目, 目前很多聚类分析方法都需要人为设置阈值或人为分析判别来确定, 不能给出最佳类别数目的定量判别指标. 本研究引入模糊 *k* 均值聚类法, 基本思

想是将一个数据集分成 k 个类别, 寻找目标函数的迭代最小化, 其优势是能给出最佳分类数的分析指标. 模糊 k 均值方法的目标函数(Bezdek, 1981)为

$$J_{\phi}(M, C) = \sum_{i=1}^n \sum_{j=1}^k \mu_{ij}^{\phi} d_{ij}^2, \quad (1)$$

其中, n 是属性数据的个数, 此处为研究所选取的主成分个数; k 是类别个数; μ_{ij} ($1 \leq i \leq n$, $1 \leq j \leq k$) 表示数据矩阵 X 中第 i 个土壤样本 x_i 属于聚类中心矩阵 C 中第 j 个聚类中心 c_j 的隶属度; M 为隶属度矩阵; $(d_{ij})^2$ 等于 x_i 与 c_j 在特征向量上距离的平方; ϕ 为模糊加权指数 ($\phi \geq 1$), 它控制了不同类别间共用数据的数目. 当 $\phi=1$ 时, 就会发生硬聚类, 即没有数据共用现象, 此时为一般的 k 均值聚类法, 即用最小二乘法找出一个恰当的 k 组分类矩阵 M 和一个由恰当的聚类中心向量 c_j 组成的矩阵 C , 使得目标函数 $J_{\phi}(M, C)$ 达到最小; 当 $\phi > 1$ 时用加权最小二乘法找出一个恰当的 M 和 C 使得目标函数达到最小. ϕ 越大, 共用数据数目增加, 最终分类变得较不明显.

研究引入模糊性能指数 FPI 、改进分类熵 MPE 和聚类独立性指数 S 来确定合适的聚类数, 作为最佳光谱分类数(Guo 等, 2013). 模糊性能指数的值在 0 到 1 之间变动, 若该值接近 0 表示聚类时共用数据较少, 类的划分明显. 改进分类熵表示模拟数据矩阵 X 的模糊 K 分区的分解量, 是度量各类间模糊程度的一个综合指标. 聚类独立性指标表示类与类之间相对分明性, 值越小, 聚类效果越好. 这三类指标都是值越小, 分类效果越好. 所以本研究通过对不同分类数目进行聚类分析, 然后采用这三个评价指标的比较来确定最佳的分类数目.

2.3 光谱预测建模与验证方法

由于土壤反射光谱数据的复杂性, 越来越多的多元数据分析技术被用于建立土壤属性与反射光谱之间的关系模型. 最常用的方法包括多元逐步线性回归、偏最小二乘回归、人工神经网络、支持向量机、回归树、随机森林等(纪文君等, 2012).

其中, 偏最小二乘回归(Partial least squares regression, PLSR)是一种基于因子分析的非参数回归方法, 在光谱分析中最为经典和常用. 该方法集成了主成分分析、典型相关分析和线性回归分析的优点, 实现了数据结构的简化, 解决了自变量之间多重相关的问题, 适合用于光谱分析这种自变量较多的情况.

本研究为了比较不同土壤类型光谱建模的潜力分析, 以及与国内外同行的相关研究进行横向比较, 选择了最常用的 PLSR 方法来预测建模.

预测模型采用 Leave-one-out 方法进行检验, 并利用决定系数(R^2)、均方根误差(RMSE)以及测定值标准偏差与标准预测误差的比值 RPD 为精度评价指标. 验证决定系数(R^2)及 RPD 越大, 均方根误差(RMSE)越小, 说明预测效果越好. 另外, 当 $1.5 < \text{RPD} < 2$ 时表明模型只能对样品高含量与低含量进行粗略估测, 当 $2.0 < \text{RPD} < 2.5$ 时表明模型具有较好的定量预测能力, 当 $2.5 < \text{RPD} < 3.0$ 时模型具有很好的预测能力, 当 $\text{RPD} > 3.0$ 时模型具有极好的预测能力.

研究用到的 Fuzzy k -means 分类、主成分分析、微分转换、PLSR 建模和验证等处理分别在 FuzzyME, Unscrambler X10.1 软件和 R 软件中完成.

3 结果与讨论

3.1 不同类型土壤光谱特性与分类

全部的土壤样本按照土纲和土类进行划分, 同一土类的光谱求取平均后得到其反射光谱曲线. 我们选取了初育土纲和铁铝土纲两类, 从图 1 可见不同类型土壤反射光谱曲线存在差异, 但又有一定的规律性. 如铁铝土纲的红壤、赤红壤、砖红壤三个土壤, 其光谱反射强度、曲线变化的斜率和特征吸收谷的强度等基本类似, 光谱曲线在可见光部分呈陡坎型, 在 900 nm 附近出现强烈的 Fe^{3+} 的吸收谷. 而初育土纲的五个土类, 按照光谱曲线形状可以分为陡坎型(石灰土、初育土、紫色土)和缓斜型(风沙土和冲积土)两种不同类型, 其中缓斜型的两个土类都采自东北吉林省, 而陡坎型的三类采自南方的浙江、湖南、江西、四川和云南省. 如徐彬彬等(1991)就认为在中国, 黄淮以南的土壤, 一般铁对光谱的贡献要大于有机质; 而山海关以外的东北地区, 则相反, 是有机质对光谱的影响占主导位置. 显然, 土壤反射光谱特性不但与土壤有机质、氧化铁、黏土矿物等特性直接相关, 还存在明显的地带性或区域分异规律.

但是, 由于土壤反射光谱特性和土壤分类系统一样类型众多而复杂, 而使得建立土壤光谱分类与土壤类型之间的相互关系存在很大难度. 图 2 为发育于不同母质的海南岛砖红壤的反射光谱曲线, 从图中可见发育于玄武岩的砖红壤反射光谱与花岗岩和

图1 不同类型土壤的反射率均值曲线比较

(a) 初育土土纲; (b) 铁铝土土纲

图2 不同母质发育的砖红壤反射光谱曲线比较

砂页岩发育的砖红壤反射光谱特征差异非常大,特别是图2中玄武岩发育的砖红壤反射光谱曲线与图1(b)中所有砖红壤土样的平均反射光谱曲线完全不同,说明同一类型土壤还与发育的母质有很大的关系,我们必须对土壤样品进一步细分到亚类,甚至到土种来比较其反射光谱的差异。但是按照中国第二次土壤普查所采用的土壤发生分类系统,中国有12个土纲,61个土类,235个亚类,约2500个土种(中国土壤普查办公室,1996),这为我们所进行的土壤光谱研究是一个极大的挑战。这也从另外角度说明,完全依靠土壤分类体系来分别建立不同土壤特性光谱预测模型的方法不太现实,只有在将来全部采集完来自不同类型土壤的众多样本后,才有某种可行性。因

此,我们提出依靠土壤反射光谱特性进行土壤光谱分类后建模的思路。

由于土壤高光谱测试数据量大,各波段之间又存在多重性相关的问题,因此先对全部1581个土样的一阶微分处理后数据采用主成分分析方法进行数据降维处理,得到第一组分(PCA1, 38%)、第二组分(PCA2, 12%)和第三组分(PCA3, 6%),合计贡献率达56%。研究引入模糊聚类法进行土壤光谱合理分类和最佳分类数目的计算。先将前三个主分数据归一化处理后作为输入变量用于模糊 k 均值聚类分析中。取最大迭代次数为300,收敛阈值为0.001,模糊加权指数为1.5(Guo等,2013)。为了找出最佳分类数目,分别产生2,3,4,5,6,7,8,9,10个类别,计算并比较每一类的模糊性能指数 FPI 、改进分类熵 MPE 和聚类独立性指数 S 的值,三者均是越小越好。图3是将所有土壤按照反射光谱特性划分为2,3,4,5,6,7,8,9,10类时得到的 FPI , MPE 和 S 值。可以看出,当分类数目为5时, FPI , MPE 和 S 值均为最小,因此本研究建立的中国光谱数据库,最佳的分类数目为5。

对所有1581个土壤样品根据其光谱特征分成五类,各类别的聚类中心和前两个主分值的分布见图4所示。从图4可见,土壤光谱分类主要依据PCA1的值按大小顺序进行划分,也就是按照反射光谱强度的大小进行分类。再是根据谱线的形状特征,这体现在五个聚类中心的PCA2值存在差异,类型2,3和4的PCA2值(-0.0019, -0.0026和-0.0018)基本在0左

图3 模糊聚类分成不同类别时的 *FPI*, *MPE* 和 *S* 值

右, 谱线的形状总体差异不大; 而类型 1 和类型 5 的 *PCA2* 值分别为 0.0073 和 0.006, 即谱线的形状相对特殊(见图 5)。从图 5 可以看出, 随着土壤有机质含量的升高土壤光谱反射率总体上呈下降趋势, 其中类型 5 由 335 个土壤样品组成, 主要来自东北地区的黑龙江和吉林的黑土(103 个)、黑钙土(138 个)、草甸土(52 个)、冲积土(11 个)、风沙土(18 个)五个土类共 322 个, 占类型 5 所有样本数的 96%, 而且土壤有机质含量最高, 平均达到 26.83 g/kg。土壤颜色较深, 中细结构, 属于有机质控制型光谱曲线, 在可见光波段, 光谱曲线是平缓上升的。类型 1 由 44 个土壤样品组成, 主要来至南方福建和海南的赤红壤(18 个)、砖红壤(7 个)、红壤(9 个)三个土类共 34 个, 占类型 1 所有

样本数的 77%, 土壤有机质含量很低, 仅为 10.84 g/kg, 氧化铁含量中等, 在 900 nm 处存在明显的氧化铁吸收谷, 属于铁影响型光谱曲线。中间类型 2, 3, 4 反射率曲线形状基本一致, 属于土壤有机质和氧化铁相互共同影响的过渡型土壤。类型 4 中有 11 个玄武岩发育的砖红壤和赤红壤, 将其从类型 4 中单独划出来计算其平均反射率曲线(见图 5 中的类型 4-1 线), 虽然两者的平均有机质含量相当(分别为 25.28 和 25.47 g/kg), 但类型 4-1 的反射率线比类型 4 明显要低, 这主要是受到土壤氧化铁的影响, 是一种铁控制型曲线。一般来说, 土壤中氧化铁含量增加, 会导致土壤反射率在可见光波段, 特别是在 700~1100 nm 处明显下降。同时在红外波段范围吸收强烈, 在 1.45 和 1.95 μm 处水的吸收特征不显著。但本研究中由于类型 4-1 的样本太少, 因此在后面的有机质预测建模中没有单独划分为一类。总之, 从光谱分类结果与土壤类型之间相互比较来看, 土壤地带性和理化特性对土壤光谱分类有着主要的控制作用, 但也不是完全对应, 存在混分现象, 这其中的关系需要更多土壤样本的测试和更详细的分类研究来解释。

3.2 不同类型土壤有机质光谱预测

土壤有机质在可见光波段存在较宽的吸收波段, 主要受到了土壤发色团和有机质本身黑色的影响, 因此在视觉上表现为暗黑色的土壤比亮色的土壤有机质含量更高; 而在近红外波段, 主要受到 NH , CH

图4 五类土壤光谱类型的前两个主分值和类中心分布

图5 五类土壤漫反射光谱曲线和平均有机质(OM)含量

和 CO 等基团的分子振动的倍频与合频吸收影响 (Stenberg 等, 2010). 目前国内多数的土壤有机质光谱特性研究多是针对某一地区或某一类型的土壤(韩瑞珍等, 2010; 卢艳丽等, 2011), 本研究通过中国大样本量的分析, 找到不同地区不同类型土壤共同的有机质光谱响应波段, 为大范围土壤有机质的光谱预测提供可行性分析. 图 6 是中国 1581 个土壤样本在全波段范围(400~2400 nm)有机质含量与一阶微分转换后光谱反射率的相关系数曲线, 相关系数较高的峰值出现在 580, 820, 1400 和 2200 nm 四个波段. 图 6 中可见光和三级倍频区存在两个相关波段为 580 和 820 nm, 其中 820 nm 波段附近存在 C—H 吸收带, 这与土壤有机质直接相关. 而在 580 nm 波段相关系数超过 0.4, 而且存在 560~620 nm 高相关性的宽波段, 这与彭杰等通过室内土壤有机质人工处理条件下获得的土壤有机质反射光谱特征波段 570~630 nm 基本重叠(彭杰等, 2013), 同时与其他研究类似(Krishnan 等, 1980; Bartholomeus 等, 2008). 由于可见光波段能量辐射密度大, 在这一波段提取有机质响应波段进行测量时, 仪器响应数相对较高, 有利于提高测量精度, 特别是可见光的宽波段更有利于卫星传感器探测波段的设计. 但在可见光波段, 有机质和氧化铁一起作为土壤最重要的赋色成分, 两者含量的相对高低会对土壤有机质可见光反射光谱产生影响. 图 6 中近红外一级倍频区, 1400 nm 波段是主要的水分吸收带, 会受到土壤表面吸附水、黏土矿物 O—H 晶格结构水等吸收影响. 在合频区的 2200 nm 附近主要存

在 Al—OH 黏土矿物的吸收带. 而有机质在水分吸收波段达到显著水平, 主要是因为有机质的吸湿作用, 有机质含量越高土壤的吸附水含量越高的缘故. 另外, 在 2300 nm 附近存在一个有机质相关的 C—H 特征峰(Viscarra Rossel 和 Behrens, 2010).

按照土壤类型分组, 然后利用 PLSR 方法建立土壤有机质光谱预测模型, 其预测结果的精度评价见表 2. 从表 2 可见, PLSR 光谱建模精度与建模的样本数呈负相关, 从十几个样本的冲积土、褐土建模预测结果 R_{adj}^2 接近 1, 到 552 个样本的水稻土建模预测结果 R_{adj}^2 只有 0.601. 这主要原因是水稻土来至浙江、四川、新疆、湖南、湖北、福建等不同区域及不同发育母质, 其反射光谱的分异性较大, 而冲积土、褐土等小样本土样来至同一区域, 土壤本身的理化特性分异较小. 因此, 目前很多研究报道的土壤有机质光谱预测精度很高, 但其应用的普适性还必须通过大样本来验证. 这里也再次说明, 即使每一种土壤类型都建立了各自的土壤有机质光谱预测模型, 但是当来自不同地区不同发育母质的同一类型土壤进行预测时, 其预测精度同样不能保证, 其预测模型的稳定性存在问题.

3.3 基于中国土壤光谱库的有机质光谱预测

从前述结果可见, 按照传统土壤分类分别建立各自土壤有机质光谱模型存在很大难度. 因此, 本研究先按照反射光谱信息将中国土样分成五类, 然后采用 PLSR 方法建立各自类别的土壤有机质光谱预

图 6 土壤有机质含量与一阶微分转换后光谱反射率的相关系数曲线

表 2 不同土类土壤有机质光谱 PLSR 预测结果

土类	样品数	RMSE	R_{adj}^2
水稻土	552	7.946	0.601
红壤	172	7.194	0.574
潮土	127	2.615	0.691
黑钙土	138	2.37	0.82
滨海盐土	114	2.053	0.788
紫色土	107	4.748	0.754
黑土	104	3.316	0.808
石灰土	57	2.961	0.901
草甸土	52	2.109	0.902
赤红壤	39	4.385	0.863
冷棕钙土	38	1.686	0.940
粗骨土	21	3.198	0.813
砖红壤	19	0.965	0.976
风沙土	18	0.858	0.984
褐土	12	0.002	1
冲积土	11	0.004	1

测模型, 我们称为光谱分类-局部建模, 然后与未分类的两种全局建模(原始数据和一阶微分处理后数据)进行预测结果比较(见表 3)。图 7 和 8 分别为全局建模和光谱分类-局部建模各自得到的土壤有机质光谱预测结果。

从表 3 和图 7 可见, 原始数据-全局建模预测结果的 R^2 和 RPD 分别为 0.653 和 1.697, 经过一阶微分处理后的全局建模预测得到了部分提高, R^2 和 RPD 分别为 0.697 和 1.817。但是这两个全局预测建模方法的 RPD 值都小于 2, 表明模型只能粗略的估测土壤

表 3 光谱分类建模与全局建模的土壤有机质光谱预测精度比较

土类	样品数	RMSE	R^2	RPD	
原始数据-全局建模	1581	6.811	0.653	1.697	
一阶微分-全局建模	1581	6.357	0.697	1.817	
光谱分类- 局部建模	类型 1	44	0.268	0.998	20.044
	类型 2	144	2.205	0.893	3.053
	类型 3	378	3.251	0.876	2.838
	类型 4	680	4.612	0.881	2.895
	类型 5	335	2.450	0.903	3.167
	合计	1581	3.660	0.899	3.158

有机质含量的高低。因此, 我们采用一阶微分转换后的光谱数据进行分类建模。从表 3 和图 8 可见, 光谱分类后 PLSR 局部建模的预测精度得到极显著提高, 总体预测的 R^2 和 RPD 分别提高到 0.899 和 3.158。而且从表 3 可见, 五类光谱分类的 PLSR 建模预测结果的 R^2 都显著得到了提高, RPD 值都超过了 2.5, 说明预测模型具有很好的预测能力。因此, 将来对新采集土样的有机质光谱预测, 可以先进行光谱分类判别, 然后选择该类型的光谱预测模型, 这样操作流程简单, 而且预测精度的稳定性较好。

4 讨论和小结

本研究通过对来自中国西藏、新疆、黑龙江、海南等地 16 种土类的 1581 个土壤样本的室内可见-近

图7 土壤有机质光谱全局建模预测

(a) 原始数据-全局建模; (b) 一阶微分-全局建模

图8 一阶微分处理后的土壤有机质光谱分类-局部建模预测结果

红外反射光谱数据的分析表明, 不同类型土壤的反射光谱曲线存在一定的规律性. 但是土壤可见-近红外反射光谱只能反映土壤部分理化特性的差异, 还不能很好的揭示不同类型土壤的类间差异, 比如同一种土类由于发育母质的不同其反射光谱差异悬殊

就说明了这点. 因此, 目前而言土壤可见-近红外反射光谱只能作为土壤分类的一个间接指标.

研究中引入模糊 k -means 方法提出了面向大样本光谱数据求取最佳土壤光谱分类数目的方法, 将中国主要类型土壤按照反射光谱特性, 定量的划分

为五种典型光谱类型, 解决了土壤光谱分类依靠人为判别或是传统土壤发生分类为划分依据的弊端. 研究所得到的有机质控制型、氧化铁影响型土壤反射光谱类型完全印证了国际同行的研究成果(Stoner 和 Baumgardner, 1981), 而其他三种过渡类型的可区分性解释还需要更多的土壤样本来印证.

同时提出了采用土壤光谱分类方法代替传统土壤分类方法进行光谱预测分类建模的新思路. 研究结合土壤光谱分类和 PLSR 方法建立的土壤有机质光谱分类-局部建模比一阶微分-全局建模的预测精度有了显著提高, 其预测模型的 R^2 和 RPD 两个指数分别从 0.697 和 1.817 提高到 0.899 和 3.158. 本研究

结果回答了当前土壤可见-近红外反射光谱研究领域一个核心问题, 即如何解决不同类型土壤反射光谱的差异性对土壤属性光谱预测模型的可靠性和普适性的影响. 目前而言, 中国不同类型土壤只建立一个预测模型, 或是中国近 2500 个土种各自建立预测模型, 都还不可行. 而本研究解决这个问题的关键是提出了大样本土壤样品的光谱分类. 因此, 后续工作将进一步扩大不同类型土壤光谱数据的样本量. 从大样本量数据挖掘方法着手, 来提高土壤光谱分类的准确性和可重复性, 从而提高大尺度区域光谱异质性土壤的理化属性的反演精度, 并为精确农业中土壤有机质信息快速获取设备的研制提供方法支持.

致谢 感谢东北林业大学、东北农业大学、四川农业大学、湖南农业大学、华中农业大学、云南农业大学、山西农业大学、河南农业大学、江西农业大学、中国热带农业科学院、西藏大学、塔里木大学、浙江大学等相关研究人员提供测试分析土样和部分光谱数据.

参考文献

- 戴昌达. 1981. 中国主要土壤光谱反射特性分类与数据处理的初步研究. 见: 国家遥感中心, 编. 遥感文选. 北京: 科学出版社. 315–323
- 韩瑞珍, 宋韬, 何勇. 2010. 基于可见/近红外光谱的土壤有机质含量预测. 中国科学 F 辑: 信息科学, 40(增刊): 111–116
- 纪文君, 李曦, 李成学, 等. 2012. 基于全谱数据挖掘技术的土壤有机质高光谱预测建模研究. 光谱学与光谱分析, 32: 2393–2398
- 刘焕军, 张柏, 张渊智, 等. 2008. 基于反射光谱特性的土壤分类研究. 光谱学与光谱分析, 28: 624–628
- 卢艳丽, 白由路, 杨俐苹, 等. 2011. 东北平原不同类型土壤有机质含量高光谱反演模型同质性研究. 植物营养与肥料学报, 17: 456–463
- 彭杰, 周清, 张杨珠, 等. 2013. 有机质对土壤光谱特性的影响研究. 土壤学报, 50: 76–83
- 吴豪翔, 王人潮. 1991. 土壤光谱特征及其定量分析在土壤分类上的应用研究. 土壤学报, 28: 177–185
- 徐彬彬, 季耿善, 朱永豪. 1991. 中国陆地背景和土壤光谱反射特性的地理分区的初步研究. 环境遥感, 6: 142–151
- 中国土壤普查办公室. 1996. 中国土种志. 北京: 中国农业出版社
- 中国土壤学会. 2000. 土壤农业化学分析方法. 北京: 中国农业科技出版社
- Bartholomeus H M, Schaepman M E, Kooistra L, et al. 2008. Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, 145: 28–36
- Bezdek J C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press
- Brown D J, Shepherd K D, Walsh M G, et al. 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, 132: 273–290
- Brown D J. 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma*, 140: 444–453
- Condit H R. 1972. Application of characteristic vector analysis to the spectral energy distribution of daylight and the spectral reflectance of American soils. *Appl Optics*, 11: 74–86
- Gerbbers R, Adamchuk V I. 2010. Precision agriculture and food security. *Science*, 327: 828–831
- Goge F, Joffe R, Jolivet C, et al. 2012. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. *Chemometr Intell Lab*, 110: 168–176
- Guerrero C, Zornoza R, Gomez I, et al. 2010. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. *Geoderma*, 158: 66–77
- Guo Y, Shi Z, Li H Y, et al. 2013. Application of digital soil mapping methods to identify salinity management classes in coastal lands of central China. *Soil Use Manage*, 29: 445–456
- Krishnan P, Alexander J D, Butler B J, et al. 1980. Reflectance technique for predicting soil organic matter. *Soil Sci Soc Am J*, 44: 1282–1285

- Kuang B Y, Mahmood H S, Quraishi M Z, et al. 2012. Sensing soil properties in the laboratory, in situ, and on-line: A review. *Adv Agron*, 114: 155–223
- Ramirez-Lopez L, Behrens T, Schmidt K, et al. 2013. Distance and similarity-search metrics for use with soil vis-NIR spectra. *Geoderma*, 199: 43–53
- Reeves III J B. 2010. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, 158: 3–14
- Sanchez P A, Ahamed S, Carre F, et al. 2009. Digital soil map of the world. *Science*, 325: 680–681
- Song Y X, Li F L, Yang Z F, et al. 2012. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. *Appl Clay Sci*, 64: 75–83
- Stenberg B, Viscarra Rossel R A, Mouazen A M, et al. 2010. Visible and near infrared spectroscopy in soil science. *Adv Agron*, 107: 163–215
- Stoner E R, Baumgardner M F. 1981. Characteristic variations in reflectance of surface soils. *Soil Sci Soc Am J*, 45: 1161–1165
- Viscarra Rossel R A, Behrens T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158: 46–54
- Viscarra Rossel R A, Webster R. 2011. Discrimination of Australian soil horizons and classes from their visible-near infrared spectra. *Eur J Soil Sci*, 62: 637–647
- Wetterlind J, Stenberg B. 2010. Near-infrared spectroscopy for within-field soil characterization: Small local calibrations compared with national libraries spiked with local samples. *Eur J of Soil Sci*, 61: 823–843