

基于可见-近红外光谱的滨海盐土土壤盐分预测方法*

刘 娅^{1,2} 潘贤章^{1†} 王昌昆¹ 李燕丽¹ 周 睿¹ 解宪丽¹ 王 淼³

(1 中国科学院土壤环境与污染修复重点实验室(南京土壤研究所),南京 210008)

(2 中国科学院研究生院,北京 100049)

(3 成都市食品药品检测中心,成都 610045)

PREDICTION OF COASTAL SALINE SOIL SALINITY BASED ON VIS-NIR REFLECTANCE SPECTROSCOPY

Liu Ya^{1,2} Pan Xianzhang^{1†} Wang Changkun¹ Li Yanli¹ Zhou Rui¹ Xie Xianli¹ Wang Miao³

(1 Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

(2 Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

(3 Chengdu Center for Food and Drug Control, Chengdu 610045, China)

关键词 滨海盐土;可见-近红外光谱;预测;电导率;偏最小二乘回归

中图分类号 S151.9 文献标识码 A

我国目前尚存盐渍土 0.35 亿 hm^2 , 潜在盐渍化土壤 0.17 亿 hm^2 , 15 m 等深线以内的浅海和滩涂有 0.14 亿 hm^2 , 黄河河口及长江口以南诸省的滨海盐土都有逐年递增的趋势^[1]。因此,对盐渍化土壤盐渍化程度的连续、实时监测尤为重要。然而,我国在区域盐渍土制图、分区及全国性的监测预报体系建设方面还比较落后^[2],究其原因主要是目前盐渍化土壤信息大都依靠实验室物理分析或者野外快速测定,对区域监测而言效率较低,而一些新型的区域测定电导率的方法,比如磁感大地电导率仪的应用性还不够广泛^[3]。

近年来,遥感尤其是近红外光谱技术的发展为区域盐渍化土壤空间分布及变化监测提供了一种新的途径。早期研究指出,与非盐渍化耕地相比,盐渍土在可见光和近红外波段光谱反射强,且土壤盐渍化程度越高,光谱反射越强^[4]。Farifteh 等^[5]进一步指出,盐分含量的增加会导致大于 1 300 nm (1 400 nm, 1 900 nm 左右)波段的反射率变化。刘焕军等^[6]研究也指出,盐碱土在 400 ~ 2 500 nm 的

范围内主要有五个明显的吸收谷(500、670、1 418、1 915和 2 210 nm 附近),且盐斑土壤由于富含盐分在五个吸收谷范围的吸收特征均显著强于其他土壤。黄明祥^[7]等以浙江省上虞市海涂围垦区为研究区,研究表明处于不同围垦区并具有不同改良程度的土样,其光谱数据具有很好的类可分性。屈永华等^[8]对内蒙古河套灌区的土壤盐分进行了预测研究并建立了回归模型,认为在可见光-近红外波段建立的回归模型优于其他波段。翁永玲等^[9]对青海茶卡-共和盆地土壤盐分进行了图像尺度上的反演,结果表明基于偏最小二乘回归(Partial Least Square Regression, PLSR)方法进行土壤盐分含量的空间预测是一种行之有效的方法。在植被覆盖较少的盐渍化地区,当土壤中的盐分积累到表层(0 ~ 30 cm)时,利用 MODIS 和 ASTER 的反射率差异可以进行盐渍化区域识别^[10-11]。然而,目前土壤盐渍化的遥感监测精度有限,除了因其易受地表覆盖影响,盐渍土光谱特征分析及数据处理方法还不够完善也是重要原因之一。

* 国家自然科学基金项目(41071140)、中国科学院战略性先导科技专项(XDA05050509)、中国科学院创新前沿项目(ISSASIP0717)和国家水体污染控制与治理科技重大专项(2008ZX07101-006-6)资助

† 通讯作者, E-mail: panxz@issas.ac.cn

作者简介:刘 娅(1986—),女,山东菏泽人,硕士研究生,主要从事土壤遥感应用研究。E-mail: zyliuya1986@sina.com

收稿日期:2011-08-31;收到修改稿日期:2011-12-20

本研究以滨海盐土为研究对象,运用偏最小二乘回归(PLSR)及多元逐步回归方法(SMLR)建立预测模型,并对其结果进行验证,旨在建立适合该区域土壤盐渍化程度的预测模型,为实现土壤盐渍化程度的快速、无损监测提供新的手段,也为遥感区域土壤盐分的定量反演提供理论基础和实验依据。

1 材料与方法

1.1 土壤样品采集与处理

本研究区位于江苏省东台、大丰的一个矩形区域(介于 $E120^{\circ}20'$ ~ $120^{\circ}50'$, $N32^{\circ}50'$ ~ $33^{\circ}00'$),面积为 $50 \text{ km} \times 20 \text{ km}$ 。土壤样点按照 $4 \text{ km} \times 3 \text{ km}$ 的网格布设(图 1),于 2010 年 12 月共采集了 0 ~ 20 cm 表层

样品 102 个。每个样品均按梅花采样法采集,并记录 GPS 位置及相应的环境信息。

由于原始土壤样品的土壤盐分含量分布不够均匀,本研究选取一些已脱盐的土样,采用浓度为 0.5%、1%、1.5%、2%、2.5%、3%、3.5%、4%、4.5%、5% 和 6% 的 NaCl 溶液湿润,以模拟不同盐渍化程度的土壤。原始土样和模拟样品共计 120 个。所有样品在自然状态下风干后,挑出瓦砾及植物残茬等杂质,并磨细,分别过 10 目、60 目及 100 目筛,供光谱及理化分析测试用。

理化分析方法参照文献[12]。鉴于土壤电导率(EC)容易获取,已成为国际上广泛使用的盐分表示方式^[13],本文用电导率代表土壤盐分含量状况。土壤电导率(水土比 5:1)用 DDS-307 电导率仪测定。

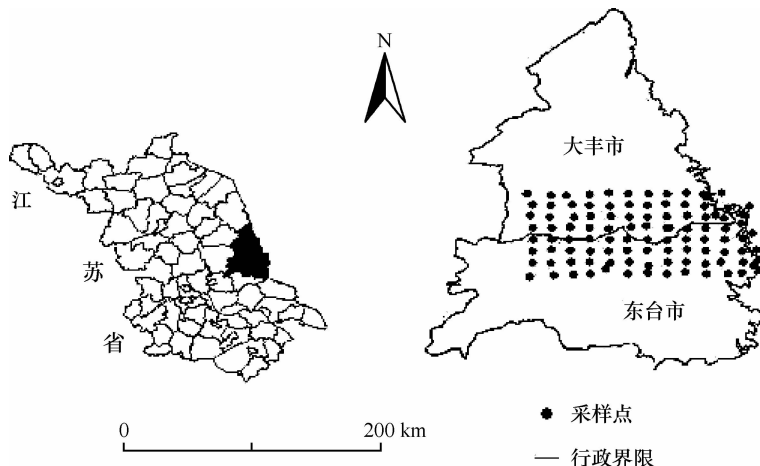


图 1 土壤采样点分布图

1.2 光谱数据采集

光谱采集采用 ASD FieldSpec Pro 地物光谱仪在暗室中进行,该光谱仪光谱范围为 350 ~ 2 500 nm,采样间隔为 1 nm。光谱测量采用 25° 视场角探头,探头到土样表面距离为 13 cm。光源为 50 W 卤素灯,其到样品表面的距离为 40 cm,光源照射方向与垂直方向夹角为 40° 。被测土样装满直径 7 cm 的小盒,表层用直尺轻轻刮平。为了消除阴影影响,每个样品采集 4 个方向光谱,按 90° 旋转样品盒,每个方向采集 5 条光谱曲线,共 20 条曲线。每次测量前均进行白板校正。

1.3 光谱数据预处理

对每个样品测得的 20 条光谱曲线进行拼接校正,然后取其算术平均值,作为该样品的原始光谱特征曲线。采用 Savitzky-Golay 多项式(二次多项

式,9 点)对原始光谱进行了平滑处理。由于 350 ~ 380 nm 和 2 400 ~ 2 500 nm 噪声比较大,信噪比较低,研究中去除了上述波段。

1.4 模型建立与验证

为了比较不同样本集对预测结果的影响,本文采用四种建模集和验证集的选择模式:将所有 120 个样本按 EC 值大小降序排列后,分别按隔一取一作为验证样本(即 60 个建模样 + 60 个验证样,模式 1),隔二取一作为验证样本(80 个建模样 + 40 个验证样,模式 2),隔三取一作为验证样本(90 个建模样 + 30 个验证样,模式 3)以及全部样本建模并全部验证(120 个建模样 + 120 个验证样,模式 4)。

1.4.1 建模方法 本文采用 PLSR 建模方法进行土壤电导率的预测。PLSR 吸收了主成分分析中提取主成分的思想,实现了数据结构的简化^[14],解

决了自变量之间多重相关的问题^[15],同时又克服了主成分分析对自变量有较强解释能力,但是对因变量解释能力不够的缺点^[16],适合用于光谱分析这种自变量较多的情况。为了比较不同建模方法对四种模式建模结果的影响,本文还采用多元逐步回归分析(Stepwise Multiple Linear Regression, SMLR)对所有样品 380~2 400 nm 经 10 nm 重采样后的反射率和电导率对数值进行了四种模式下的建模和预测。

建模过程中验证方法采用交叉验证(Full Cross Validation),即给定一个 i 个样品的矩阵 X ,将其中 1 个样品暂时剔除在 PLSR 建模计算之外,由保留下来的 $i-1$ 的数据计算模型参数向量,然后,以这 $i-1$ 个计算模型预测被剔除的样品的成分含量 y ,以同样的方法重复对每一个样品进行预测,直到每一个样品都被剔除过一次。

为了减少异常样本对预测和验证精度的影响,本文剔除残差大于 3 倍样本集标准差,以及杠杆值(leverage)大于 3 倍建模样本集杠杆值均值的异常样本^[17]。建模在 Unscrambler 9.7 软件中完成,模型因子数直接采用软件建议的个数。

1.4.2 模型验证 模型的验证主要选用建模集交互验证决定系数(Coefficient of determination in cross validation, R_{cv}^2)、交叉验证均方根误差(Root mean square error of cross-validation, $RMSE_{cv}$)、验证

集决定系数(Coefficient of determination in Prediction, R_p^2)、验证集标准预测误差(Standard error of Prediction, SEP)以及测定值标准偏差与标准预测误差的比值(Ratio of standard deviation to standard error of prediction, RPD)^[18]为模型评价参数以及模型精度验证评价参数。 R_{cv}^2 、 R_p^2 及 RPD 越大, $RMSE_{cv}$ 和 SEP 越小,说明预测效果越好。已有研究认为,当 $R_p^2 > 0.91$, $RPD > 2.5$ 时,该模型能精确预测盐分含量;当 $0.82 < R_p^2 < 0.9$, 并且 $RPD > 2$ 的模型能很好的预测盐分含量; $0.66 < R_p^2 < 0.81$, 且 $RPD > 1.5$ 时,该模型能近似预测盐分含量;当 $0.5 < R_p^2 < 0.65$ 时,该模型预测能力较差^[19]。

模型验证在软件 Unscrambler 9.7 中完成,数据统计及图表的制作在 Excel 2007 和 SPSS 18.0 中完成。

2 结果与分析

2.1 土壤样本电导率描述性统计

由表 1 可知土壤样本电导率实测值偏度大于 0,呈偏正态分布,而经对数转换过的值偏度接近 0,因此本研究在建模及预测时均采用电导率的对数值。预测完成之后,再将电导率的对数值转换成实际值,进行后续数据处理。

表 1 土壤样本电导率统计结果(以模式 2 为例)

	样本数 (个)	电导率实测值			电导率对数值
		平均值($mS\ cm^{-1}$)	标准差($mS\ cm^{-1}$)	偏度	偏度
全部样本	120	1.98	2.58	1.21	0.41
建模样本	80	1.94	2.54	1.21	0.42
验证样本	40	2.06	2.69	1.17	0.39

2.2 土壤样本光谱特征分析

按照王遵亲等^[20]对滨海盐土的分级标准以及蔡阿兴等^[21]提出的苏北地区电导率和全盐含量的换算公式,以在非盐化、轻度盐化、中度盐化、强度盐化和盐土五个级别内的 5 个土壤样本为例,对其经 10 nm 重采样的原始反射光谱曲线进行说明(图 2)。从图上可以看出滨海盐土的反射光谱曲线有三个明显的吸收谷,即 1 400 nm、1 900 nm 和 2 200 nm,这正是水分的三个吸收谷^[22]。且土壤样品的光谱曲线形态相似,在可见光部分,差异并不明显,但是在近红外区域差异明

显,且有规律性,总体趋势是电导率越高反射率越低。这是由于滨海盐土富含氯化物,而氯化物容易吸收空气中的水汽,导致土壤湿度增加,反射率降低^[23-24]。

对原始反射光谱数据和电导率进行相关性分析发现,相关性较好的波段为 500、1 570、1 960、2 070、2 330、2 170、1 910、1 840 和 2 240 nm,这其中的部分波段与已有学者的结论类似^[6,25]。

2.3 PLSR 不同建模样本数量建模结果比较

对于样本的四种处理方式分别进行建模及模型验证,其结果如表 2 所示。

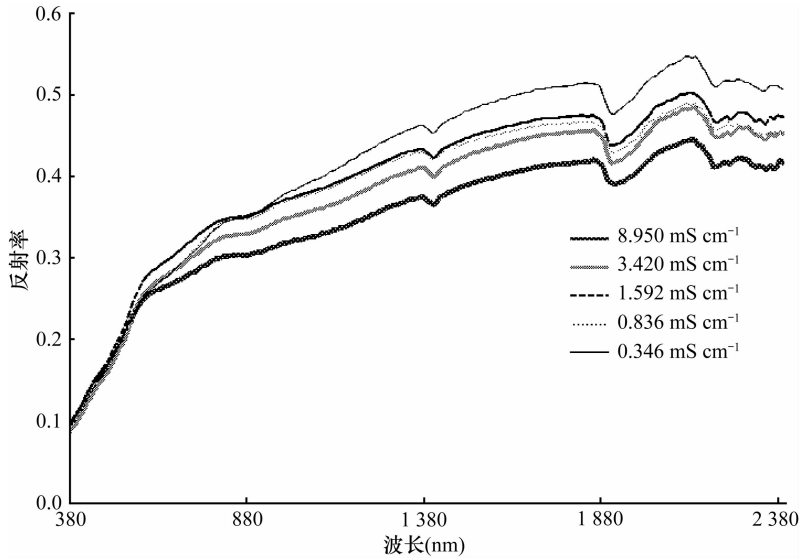


图2 不同盐渍化程度土壤样品原始光谱特征图

表2 PLSR 不同建模及验证集的建模和验证结果

	建模集交叉验证					验证集验证				
	实际样本数 (个)	选择样本数 (个)	因子数 (个)	均方根误差 RMSE _{cv} (mS cm ⁻¹)	决定系数 R _{cv} ²	实际样本数 (个)	选择样本数 (个)	决定系数 R _p ²	标准预测 误差 SEP (mS cm ⁻¹)	RPD
模式1	59	60	6	0.31	0.77	56	60	0.90	0.76	3.08
模式2	78	80	6	0.28	0.83	38	40	0.86	0.93	2.52
模式3	82	90	6	0.27	0.83	30	30	0.53	2.51	1.08
模式4	109	120	10	0.22	0.89	117	120	0.84	0.99	2.45

注:RPD 为测定值标准偏差与标准预测误差的比值

由表2可知,模式1、模式2、模式3引入的因子数均为6个,模式3交叉验证的决定系数 R_{cv}^2 较高,但是SEP值最高,RPD值最低;模式1的 R_p^2 及RPD值最高;模式2的RPD值及决定系数 R_p^2 较模型1稍低;模式4建模和验证样本数量在四种模式中最多,交叉验证的决定系数 R_{cv}^2 最高,RMSE_{cv}最低,引入的因子数最多为10个,模型验证的决定系数 R_p^2 及

RPD值较模式1、模式2低。结果表明,模式3验证的样本划分模式的建模效果最差,预测精度最低,而模式2和模式1建模结果较好,预测精度较好。

2.4 SMLR 模型及其与PLSR比较

对于多元逐步回归分析,设定变量进入模型的 F 检验概率为0.05,模型剔除变量的 F 检验概率为0.1,建模结果如表3所示。

表3 SMLR 不同建模及验证集的建模和验证结果

	建模集交叉验证					验证集验证				
	实际样本数 (个)	选择样本数 (个)	因子数 (个)	均方根误差 RMSE _{cv} (mS cm ⁻¹)	决定系数 R _{cv} ²	实际样本数 (个)	选择样本数 (个)	决定系数 R _p ²	标准预测 误差 SEP (mS cm ⁻¹)	RPD
模式1	60	60	6	0.28	0.82	—	—	—	—	—
模式2	80	80	8	0.21	0.90	38	40	0.81	1.10	2.14
模式3	90	90	10	0.20	0.91	30	30	0.56	2.71	1.00
模式4	109	120	9	0.22	0.89	—	—	—	—	—

注:RPD 为测定值标准偏差与标准预测误差的比值

从预测结果上看,模式 1 和模式 4 做异常值检验时,会剔除掉过多的样本,导致很多正常样本被误认为异常样本删掉,虽然最终的决定系数 R_p^2 和 RPD 值很高,但是会损失掉样品集的很多信息,因此本文不认为该模型具有代表性,略去对该模型建模结果的说明。相比而言,模式 2、模式 3 能比较均匀地涵盖样本集信息,模式 2 的 SEP 值为 1.10 mS cm^{-1} , RPD 为 2.14,模式 3 的 SEP 值为 2.71 mS cm^{-1} , RPD 为 1.00。结果表明,基于

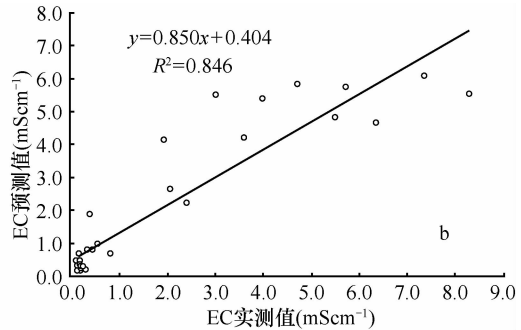
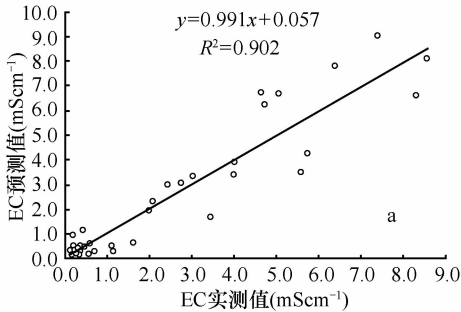


图3 PLSR方法模式1(a)和模式2(b)电导率实测值和预测值关系图

3 讨论

在建立可见光-近红外光谱 PLSR 模型时,建模样品选择非常重要。通常的目标是,要求所建立模型在最佳主成分的预测值与实测值的相关系数 R^2 最大且交叉验证预测均方差最小,否则预测能力和精度不高;所选样品数量适当,样品数不能太少,否则模型缺乏代表性,同时样品数不能太多,否则模型建立时间很长^[26]。本文尝试了隔一选一、隔二选一、隔三选一及全部选入这四种建模方式,模式 4 (全部样本都选入)的预测值与实测值的相关系数 R^2 最大,且交叉验证预测均方差最小,该模型显然不是最优建模方式;模式 3,即隔三个选一个作为验证样本的模型,经验证该模型预测精度较差,预测值与实测值回归方程的决定系数 R^2 仅 0.53, RPD 值仅 1.08;因此作者认为最佳的建模样品选择方式为隔一选一(60 个建模样 + 60 个验证样)和隔二选一的方式(80 个建模样 + 40 个验证样),这两种选择方式既避免了数据量过多造成的重复、冗余,又能保证样品分布均匀、有代表性,因此能取得较好的建模效果。

多元逐步回归是在多元线性回归的基础上发展起来的,可以选择较少的变量建立自变量与因变

SMLR 方法的建模,模式 2 验证的样本划分模式下效果最好。

将 PLSR 的两个最好模式结果与 SMLR 最好模式预测结果比较发现,PLSR 模式 1 和模式 2 的 SEP 值均较 SMLR 的模式 2 小,而它们的预测模型决定系数 R_p^2 及预测方差比(RPD)均较大,因此本文认为 PLSR 所建模型的预测精度优于 SMLR 方法。

采用 PLSR 方法的两种样本集划分方式的预测结果如图 3 所示。

量的关系。何挺等^[27]运用了逐步回归分析等方法建立了土壤有机质与土壤反射光谱之间的关系。结果表明土壤有机质含量与反射率对数的一阶微分之间的关系最为敏感。王静等^[28]建立了 550、720、760、820 和 940 nm 波段与盐含量的多元线性回归方程($R^2 = 0.81$, RMSE = 0.25)。张飞等^[29]利用后向多元线性回归分析,建立了部分盐分因子的预测模型。本文运用 SMLR 方法,采用 80 个建模样 + 40 个验证样的组合方式建立模型,所建模型的 SEP 值为 1.10 mS cm^{-1} , RPD 为 2.14,验证模型的决定系数 R_p^2 为 0.81,证明了 SMLR 方法也可较准确的预测土壤电导率,但是其对建模样本数要求较高,只有模式 2 能取得较好的建模效果,模式 1、模式 3、模式 4 建模效果均较差。

就 PLSR 与 SMLR 预测结果比较而言,PLSR 建模精度优于 SMLR 模型。这可能是由于逐步回归建模仅有少量波段的光谱反射率能进入模型,而且这些波段之间仍然存在一定的共线性,模型对于原始光谱信息利用较少,而 PLSR 针对自变量较多,变量相互之间共线性严重,且样本数少于变量数的情况进行回归计算,参与建模的因子之间不存在共线性,而且其代表了几乎所有波段,包含了更多的信息,因此,本文认为,PLSR 可以更好地构建光谱与土壤盐分含量之间的关系模型。

4 结 论

本文采用 PLSR 和 SMLR 方法对不同建模和预测样本集划分模式下的光谱和电导率关系进行了分析, 结果表明 PLSR 方法建模和预测效果优于 SMLR。对 PLSR 而言, 模式 1 以及模式 2 均能取得最好的预测精度, 预测方差比 (RPD) 最高可达 3.08, 表明高光谱方法能比较准确地预测土壤电导率, PLSR 建模方法也具有一定的推广性, 可以应用于相似区域的土壤盐分含量光谱方法建模。该方法同时也为利用遥感卫星数据监测滨海区域的盐碱化状况提供了新的途径。

参 考 文 献

- [1] 俞仁培, 陈德明. 我国盐渍土资源及其开发利用. 土壤通报, 1999, 30(4): 158—159
- [2] 张士功, 邱建军, 张华. 我国盐渍土资源及其综合治理. 中国农业资源与区划, 2000, 21(1): 52—56
- [3] 杨劲松, 姚荣江. 基于电磁感应仪的表聚型土壤盐渍剖面特征解译研究. 水文地质工程地质, 2007(5): 67—72
- [4] Rao B R M, Sankar T R, Dwivedi R S, et al. Spectral behavior of salt-affected soils. *International Journal of Remote Sensing*, 1995, 16(12): 2125—2136
- [5] Farifteh J, Meer F, Meijde M, et al. Spectral characteristics of salt-affected soils: A laboratory experiment. *Geoderma*, 2008, 145: 196—206
- [6] 刘焕军, 张柏, 王宗明, 等. 基于反射光谱特征的土壤盐碱化评价. 红外与毫米波学报, 2008, 27(2): 138—142
- [7] 黄明祥, 史舟, 李艳, 等. 基于地面光谱的海涂土壤开发程度评价研究——以浙江省上虞市海涂围垦区为例. 遥感学报, 2004, 8(4): 378—384
- [8] 屈永华, 段小亮, 高鸿永, 等. 内蒙古河套灌区土壤盐分光谱定量分析研究. 光谱学与光谱析, 2009, 29(5): 1362—1366
- [9] 翁永玲, 戚浩平, 方洪宾, 等. 基于 PLSR 方法的青海茶卡-共和盆地土壤盐分高光谱遥感反演. 土壤学报, 2010, 47(6): 1255—1263
- [10] 陈康, 金晓媚, 万力, 等. 应用 MODIS 数据进行土壤盐渍化评价方法研究. 科技导报, 2009, 27(14): 59—63
- [11] 李海涛, Brunner P, 李文鹏, 等. ASTER 遥感影像数据在土壤盐渍化评价中的应用. 水文地质工程地质, 2006, 5: 75—79
- [12] 鲁如坤. 土壤农业化学分析方法. 北京: 中国农业科技出版社, 2000: 106—109
- [13] 刘广明, 杨劲松, 姜艳. 江苏典型滩涂区地下水及土壤的盐分特征研究. 土壤, 2005, 37(2): 163—168
- [14] 和燕. 主成分回归与偏最小二乘回归方法比较. 成都电子科技大学学报, 2003, 25(4): 34—37
- [15] 郑光辉. 江苏省部分地区土壤属性高光谱定量估算研究. 南京: 南京大学, 2010
- [16] 金伟, 赵红霞, 高芝. 基于 PLS 建模在近红外光谱分析中的应用展望. 现代农业科学, 2008, 15(11): 10—11
- [17] 闵顺耕, 李宁, 张明祥. 近红外光谱分析中异常值的判别与定量模型优化. 光谱学与光谱分析, 2004, 24(10): 1205—1209
- [18] Chang C W, Laird D A, Mausbach M J, et al. Near-infrared reflectance spectroscopy principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 2001, 65(2): 480—490
- [19] Farifteh J, Meer F, Atzberger C, et al. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, 2007, 110(1): 59—78
- [20] 王遵亲, 祝寿泉, 俞仁培, 等. 中国盐渍土. 北京: 科学出版社, 1993
- [21] 蔡阿兴, 陈章英, 蒋正琦, 等. 我国不同盐渍地区盐分含量与电导率的关系. 土壤, 1997, 27(1): 54—57
- [22] Dor B. Quantitative remote sensing of soil properties. *Advances in Agronomy*, 2002, 75: 173—243
- [23] Karavanova E I, Shrestha D P, Orlov D S. Application of remote sensing techniques for the study of soil salinity in semi-arid Uzbekistan//Bridges E M, Hannam I D, Oldeman L R, et al. Response to land degradation. USA: Oxford & IBH Publishing, 2001: 261—273
- [24] 戚浩平, 翁永玲, 赵福岳, 等. 茶卡-共和盆地土壤盐分与光谱特征研究. 国土资源遥感, 2010, 21(增刊): 4—8
- [25] Weng YL, Gong P, Zhu ZL. A spectral index for estimating soil salinity in the Yellow River Delta Region of China using EO-1 hyperion data. *Pedosphere*, 2010, 20(3): 378—388
- [26] 祝诗平. 基于 PCA 与 GA 的近红外光谱建模样品选择方法. 农业工程学报, 2008, 24(9): 126—130
- [27] 何挺, 王静, 林宗监, 等. 土壤有机质光谱特征研究. 武汉大学学报: 信息版, 2006, 31(11): 975—979
- [28] 王静, 刘湘南, 黄方, 等. 基于 ANN 技术和高光谱遥感的盐渍土盐分预测. 农业工程学报, 2009, 25(12): 161—166
- [29] 张飞, 塔西甫拉提·特依拜, 丁建丽, 等. 渭干河-库车河三角洲绿洲土壤盐渍化现状特征及其与光谱的关系. 环境科学研究, 2009, 22(2): 227—235