

内蒙古河套灌区土壤盐分光谱定量分析研究

屈永华¹, 段小亮², 高鸿永², 陈爱萍², 安永清², 宋金玲¹, 周红敏¹, 何涛¹

- 1 北京师范大学遥感科学国家重点实验室(与中科院遥感应用研究所共建), 北京师范大学环境遥感与数字城市北京市重点实验室, 北京师范大学地理学与遥感科学学院, 北京 100875
- 2 内蒙古河套灌区义长灌域管理局, 内蒙古 五原 015100

摘要 研究盐碱化土壤的光谱特性是利用遥感技术实现区域尺度上进行土壤盐碱化监测和评价的工作基础。为了实现对内蒙古河套灌区土壤盐分的定量反演, 于2007年7~8月间采集了河套灌区土壤样品, 进行土壤化学成分及其光谱反射特性的测量。基于统计分析的方法, 建立了土壤盐分与高光谱数据的偏最小二乘回归(PLSR)模型, 对土壤中主要盐分参数进行了反演实验。独立验证结果表明, 对全盐($\%$)、硫酸根离子(SO_4^{2-})、pH值以及钾、钠总含量($\text{K}^+ + \text{Na}^+$)有较好的反演精度, 验证数据的决定系数 R^2 分别是0.728、0.801、0.715和0.734, 预测方差比(RDP)分别是11.79、11.87、11.64、11.63。将以上参数的PLSR模型回归系数聚合在TM的可见光(蓝色、绿色、红色)和近红外波段时, 回归系数在数值上均有显著反应。研究结果为在航空航天遥感尺度上实现土壤盐分的定量反演提供了理论基础和实验依据。

关键词 土壤盐分; 偏最小二乘回归; 高光谱; 遥感

中图分类号: TP7 S15 文献标识码: A DOI: 10.13964/j.issn.100020593(2009)0521362205

引言

对于依靠灌溉补给农业用水的地区, 由于灌溉和排盐方式不当, 土壤盐碱化成为当地土地退化的主要问题。遥感以其宏观、综合、动态、快速等特点, 已成为监测土壤盐碱化的一种新的探测手段^[1]。用遥感技术识别土壤盐碱化程度的方法基础是利用不同盐碱化的土壤在其光谱反应上具有差异性, 因此, 研究土壤盐分的光谱特征, 建立基于光谱分析的土壤盐分定量反演模型是开展基于遥感技术的土壤盐碱化监测的基础工作。

利用地面测量得到的光谱数据, 结合土壤样品的地球化学成分, 对土壤盐分含量进行反演, 是应用光谱遥感技术定量监测盐碱化土壤盐分含量的重要方法^[2]。Mettemicht等从表层土壤结构、土壤颜色、表层植被等多个方面分析了土壤盐碱化遥感监测的影响因素, 指出了用多光谱遥感数据监测土壤盐碱化的可能性和面临的问题^[3]。已有的研究表明, 用高光谱数据分析土壤盐碱化情况, 建立土壤光谱与土壤盐分(或电导率EC)以及其他土壤参数(如pH值)之间的定量模型是实现土壤盐分定量反演的基础^[4-5]。Csillag等研究了表层土壤样品的高光谱数据与土壤中pH值、电导率

(EC)和可交换钠含量(ESC)的回归模型, 基于修正逐步主成分分析的结果, 指出土壤光谱在可见光波段(550~770 nm)、近红外波段(900~1030 nm、1270~1520 nm)以及中红外波段(1940~2150 nm、2150~2310 nm、2330~2400 nm)的反射率对于定量分析盐碱化有较好的指示作用^[6]。

在用高光谱数据估计土壤盐分或其他土壤化学成分的方法研究上, 传统的方法是基于相关分析和统计回归的思想, 通过相关分析, 从高光谱数据中发现对土壤化学成分变化相关性强的波段, 建立这些波段的反射率和土壤成分的回归方程, 并用来进行对未知数据的预测^[2-4, 6, 28]。然而由于高光谱数据的光谱维数较大, 波段之间又具有一定的相关性^[9], 用相关分析的方法从众多波段中选择某几个敏感波段的方法很难保证高光谱的有用信息能被有效利用。因此, 在相关分析的基础之上, 又发展了一种既利用高光谱的细分光谱信息, 又能降低波谱维数的方法, 即主成分分析^[6]。但是, 主成分方法过多地强调了自变量之间矢量变换, 而忽略了自变量对因变量的解释能力, 所以, 最终形成的主成分组合虽然能够减少原始光谱波段之间相关性, 降低波谱维数, 但是, 形成的主成分与因变量的相关性不能保证是最大的。偏最小二乘回归分析综合了主成分分析和相关分析的优点, 近年来在用高光谱数据的农作物理化参数、分析化学等多个方面得到了

收稿日期: 200802212 修订日期: 2008205216

基金项目: 国家自然科学基金项目(40601059), 国家重点基础研究发展计划项目(2007CB714407)和长江学者和创新团队发展计划项目资助

作者简介: 屈永华, 1972年生, 北京师范大学地理学院讲师, Email: qyh@bnu.edu.cn

©1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

广泛应用。研究表明,用偏最小二乘(PLS)回归方法对高光谱与地物特征参数之间进行回归分析,建立地物参数与高光谱数据之间的回归模型,是分析地物高光谱特征、揭示光谱特征波段与地表参数定量关系的一个很好的尝试^[9,21]。

本文利用内蒙河套灌区土壤地面测量光谱与同步测量的土壤样品化学分析数据,运用统计学方法对灌区的盐碱化信息与土壤光谱信息进行统计回归,建立灌区盐碱化信息与土壤光谱信息之间的偏最小二乘回归模型,为在遥感尺度上对土壤盐碱化进行定量评价提供了理论基础。

1 数据获取

1.1 研究区概况

内蒙古河套灌区位于内蒙古西部的巴彦淖尔市南部,西与乌兰布和沙漠相接,东至包头市郊区,南临黄河,北抵阴山山脉。在北纬 40°20′~41°18′,东经 106°20′~109°19′之间。东西长 250 km,南北宽达 50 km,总土地面积 11 897.13 km² (1 784.6 万亩)。由于该地区年均降水量只有 139~222 mm,而年均蒸发量达 1 999~2 346 mm,降水量小而蒸发量大,是典型的/无灌溉则无农业/地区。但是,由于各种因素的影响,当地土地盐碱化问题相当突出。经过多年的改土治碱,耕地得到了很大改善,但灌区土壤盐碱化仍然困扰着灌区的可持续发展。据灌区土壤调查资料统计,灌区耕地中仍有 160 万亩中、重度盐化耕地。通过遥感技术实现土壤盐碱化动态监测,能够为灌区的改土、治盐以及防止次生盐碱化提供大尺度、多时相的基础数据。

1.2 野外实验

为了获取土壤化学成分与土壤光谱的配套数据,从 2007 年 7 月 30 日至 8 月 2 日,在河套灌区进行了野外实验,获取了土壤样品和光谱数据。采样点在河套灌区的中游位置,分别位于中游南部(ZN),中游中部(ZZ)和中游北部(ZB),如图 1 所示。其中在 ZN 样区属于轻度盐碱化地区,ZZ 样区属于中度盐碱化地区,ZB 样区的盐碱化程度则相对要严重一些。在采样期间,这 3 个样区的地表覆盖类型主要为农作物、林地、杂草、水体和裸土。在每个样区,我们布置了大约 10~20 个样方,在每个样方中心点附近用取土环获取土样,采样深度为 0~5 cm。

1.3 光谱测量

在获取土壤样品之后,一部分被用来做土壤化学成分测试,另一部分被留作土壤光谱测量。对采集的土样经过通风干燥、过筛(5 mm 筛)处理,使得土壤颗粒均匀,去除杂质和水分的影响,然后进行土壤光谱测试,测试时将土壤样品均匀平铺在一块面积约为 0.5 m @ 0.15 m 黑色背景布上,选择光照条件好的晴天进行测量,测量时间在上午十点到下午两点之间。光谱测量设备为 ASD 便携式光谱仪,光谱波长范围是 0.35~1.1 μm,视场角为 25°,重采样之后的光谱分辨率为 1 nm,测量时光谱仪探头距离目标物垂直距离为 110 cm,保证地面视场在一个半径约 0.22 m 的范围内,样品光谱测量前后进行标准参考板校正,转换为地表反射率。

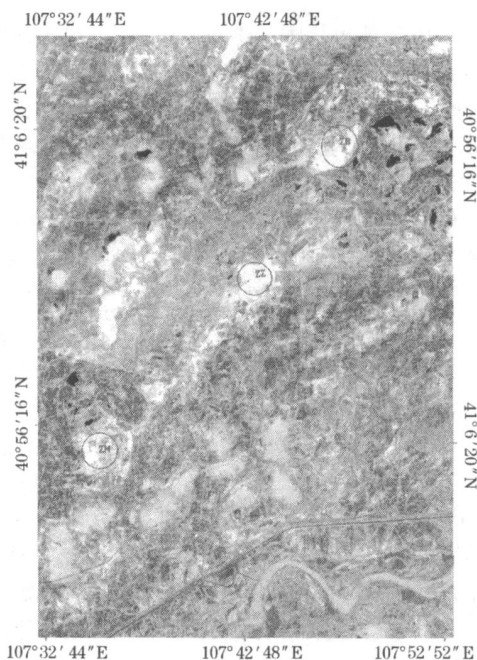


Fig 1 Sampling p bts in Hetao irrigation district
(TM imagery with standard pseudo colors)

1.4 土壤化学成分化验

土样取回登记后,放在通风干燥的房间里,让其自然风干,然后用无 Cl⁻及无 Ca²⁺的蒸馏水按照土壤和水的比例为 1B 5 来配置土壤浸提液,进而测定土壤液体中的 pH 值、CO₃²⁻, HCO₃⁻, SO₄²⁻, Cl⁻, Ca²⁺, Mg²⁺, K⁺ + Na⁺ 的含量,在后期数据处理中,根据土壤中各个盐分的含量计算土壤中盐度,即全盐含量(%)。

2 定量反演模型

由于高光谱数据含有波段数目比较多,一般有几百到上千波段。虽然高光谱为揭示地物的反射特征提供了很好的数据,但是也带来了一系列的问题。首先是各个波段之间具有很大的相关性,传统的最小二乘回归方法对具有相关性的数据进行建模时,由于自变量之间相关性影响,线性方程的解不稳定,有时甚至无法求得解析解。其次是波段的个数远远大于观测样本的个数,常规的最小二乘回归方程要求观测样本应该大于等于自变量个数的 2 倍。因此用常规的最小二乘法很难对高光谱数据进行回归建模。偏最小二乘回归是在传统的最小二乘回归的基础之上,引入了主成分分析以及方差分析的思想而发展起来的一种多变量回归分析方法^[1,32]。基于偏最小二乘方法对土壤盐分反演的主要思想是从高光谱数据中提取能够最大解释盐分含量变化的主成分,这些主成分是原始光谱数据的线性组合,各个主成分之间是正交的,即保证提取的各个成分之间的相关性最小,同时要求这些主成分能够对因变量的相关性最大,即能最好地解释因变量的变化。所以,通过这种方法,能够在大大减少光谱维数的同时,保证选择的变量能够最大限度地反映出土壤盐分含量的

变异特征。

我们利用 ParLes 软件^[16]建立了土壤盐分和高光谱数据之间的偏最小二乘回归模型。在建模过程中,从 54 个样品中选出 42 个用来建立偏最小二乘回归模型,其余的 12 个用来对回归模型进行独立验证,通过比较验证的精度来考查模型的适用性。

模型的精度用决定系数 R^2 , 相对均方根误差 RMSE%, 预测方差比 (RPD) 这 3 个参数进行判断。 R^2 是一个判断模型预测值和实测值之间相关性大小的指标。 R^2 取值在 0 (最差) 到 1 (最好) 之间。RMSE (root mean square error) 是一个表示模型预测误差大小的指标,但是,由于 RMSE 的大小和预测值的量纲有关,故通常使用相对均方根误差 RMSE% 来表示。RPD (rate of prediction to deviation) 是预测值与预测误差这两个方差的比值,也是一个表达预测值和实测值统计相关性大小的一个指标。从统计上来说,一个好的反演模型,应该是 R^2 和 RPD 尽量大,而 RMSE% 尽量要小。一般来说,如果一个模型的 $R^2 > 0.17$, $RPD > 1.15$, 则模型的预测结果的可信度较高,反之,则说明模型的预测结果可信度低,模型不可靠^[5]。

(%)、硫酸根离子 (SO_4^{2+})、pH 值以及钾、钠总含量 ($\text{K}^+ + \text{Na}^+$) 的 RPD 和 R^2 较大,表明估算的结果能够和实测数值有很好的相关性。其中对 pH 值的反演精度最高,其相对均方根误差为 71.4%,其他 3 个参数的预测误差在 20% ~ 30% 之间。在另外的 5 个参数中(氯离子、镁离子、钙离子、碳酸根和碳酸氢根),模型对它们的反演结果具有很低的 R^2 (从 0.147 ~ 0.1523) 和很低的 RPD (从 1.11 ~ 1.164),表明用当前的数据对氯离子、镁离子、钙离子、碳酸根和碳酸氢根的反演精度较低。

Table 1 Evaluation indications of PLSR model

因变量	主成分个数	评价指标			
		R^2	RMSE	RMSE%	RPD
S%	6	0.1728	0.1347	241.6	1.179
SO_4^{2-}	6	0.1801	1.1642	291.9	1.187
pH	5	0.1715	0.1705	71.4	1.164
$\text{K}^+ + \text{Na}^+$	5	0.1734	6.016	281.9	1.163
HCO_3^-	5	0.161	0.1824	401.5	1.164
CO_3^{2-}	4	0.1436	2.1525	851.0	1.136
Cl^-	4	0.1523	3.1393	301.9	1.146
Ca^{2+}	4	0.1498	0.1144	601.0	1.127
Mg^{2+}	5	0.147	0.1889	181.4	1.11

3 结果分析

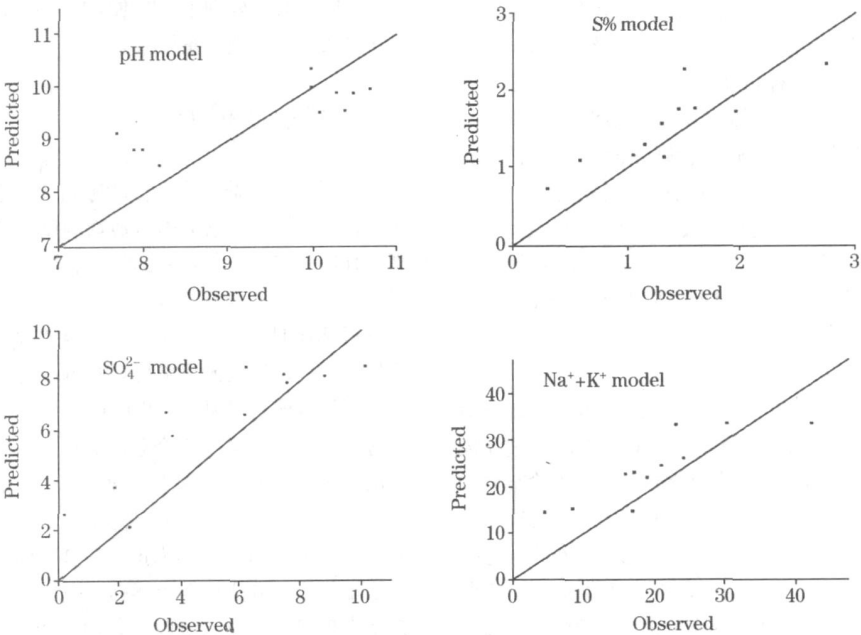
针对每一种土壤化学成分,建立偏最小二乘回归模型,所用到的因子个数和模型的精度评价指标如表 1 所示。

从表 1 可以看出,在高光谱数据的大约 700 个波段的反射率数据中,根据偏最小二乘方法,需要主成分个数为 4 ~ 6 个。在用定量估计模型用验证数据进行反演的时候,对全盐

用验证数据对模型进行检验的结果和实测数据的比较如图 2 所示。偏最小二乘回归模型的表达式的形式如下

$$Y = B @ X + B_0 \tag{1}$$

其中 B 为回归系数, B_0 是截距, Y 为盐分数据, X 为高光谱中的各个波段反射率值。



Figl2 Comparison between the observed value and the predicted values

B 是一个 n 维向量, n 为高光谱的波段数。为简单起见,将系数向量 B 按照 LandSat 卫星的 TM (Thematic mapper) 在可见光和近红外的波段设置分别进行平均处理,得到了与 TM 传感器对应的 4 个波段的回归系数。分别表示蓝色波段 (450 ~ 515 nm), 绿色波段 (525 ~ 605 nm), 红色波段 (630 ~ 690 nm) 和近红外波段 (775 ~ 900 nm) 对土壤盐分参数的贡

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

献^[17]。

下面我们给出有关 $\text{S}\%$, SO_4^{2-} , pH , $\text{Na}^+ + \text{K}^+$ 四个变量

转换后的回归系数, 如图 3 所示。

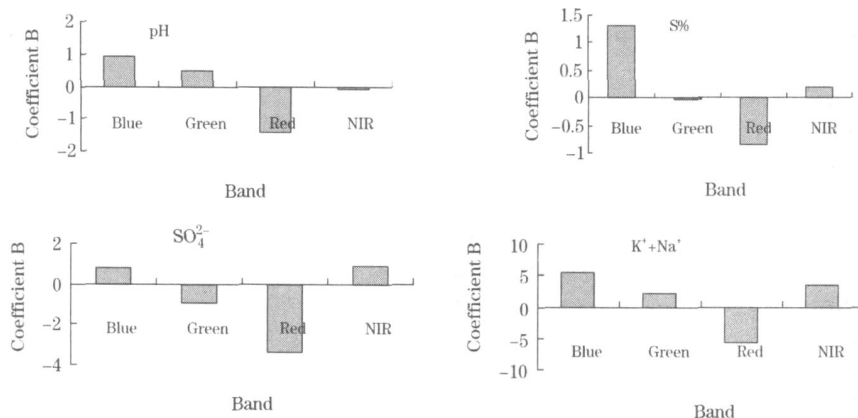


Fig 3 Regression coefficient on the 4 bands of Landsat TM sensor

由回归系数分布图可知, 在可见光波长范围内对上述 4 个土壤参数的贡献较大的为红色和蓝色波段, 而绿色波段的贡献较小。同时也可看出, 红色波段的反射率对上述 4 个土壤参数呈负相关关系。在近红外波段, 光谱反射率对 $\text{S}\%$, SO_4^{2-} 和 $\text{K}^+ + \text{Na}^+$ 这 3 个参数有明显的贡献, 而对 pH 值的贡献较小 (见图 3)。

建立地面模型的目的是为了指导在遥感卫星尺度上对该地区的盐碱化情况进行大范围的监测。我们用地面测量的高光谱数据和土壤的主要盐分数据建立的土壤盐分的定量估算模型, 通过对定量模型的分析可以知道, 在可见光和近红外波段能够找到反演盐度、 pH 值以及硫酸根和钾钠含量这 4 个参数的敏感波谱响应, 用这些波段的反射率建立偏最小二乘回归模型, 可以对上述 4 个土壤盐碱化的指标进行比较精确的估计。这些敏感波段基本上和目前常用的遥感资源卫星的传感器在可见光和近红外波段的设置一致^[17], 这就为我们在进一步的工作中从地面光谱反演到卫星遥感反演提供了很好的根据。

4 结论与讨论

在对内蒙古河套灌区土壤盐碱化定量模型的研究中, 通过对测试得到的全盐含量以及 8 个化学成分用高光谱数据进行定量分析, 建立了它们的定量估算模型。在定量模型中, 利用偏最小二乘方法比较精确地估计了全盐 ($\text{S}\%$)、 pH 值、硫酸根 (SO_4^{2-})、钾钠含量 ($\text{K}^+ + \text{Na}^+$) 这些表示盐碱化程度

的参数。对于其余参数, 如碳酸根 (CO_3^{2-}), 碳酸氢根 (HCO_3^-), 氯离子 (Cl^-), 钙离子 (Ca^{2+}), 镁离子 (Mg^{2+}) 等, 根据目前的数据分析, 很难从遥感数据中精确估计。对这些变量的估计, 一方面需要考虑将遥感波谱从可见光近红外扩展到短波红外, 另一方面, 需要进一步发展新的土壤盐分定量估计模型。目前我们用到的偏最小二乘模型虽然能够大大降低高光谱数据的维数, 充分利用波段组合来描述盐碱化参数的变化, 但是, 这种方法也具有一定的局限性。因为偏最小二乘回归模型毕竟是一种线性模型, 它对自变量 (光谱数据) 和因变量 (土壤盐分) 之间如果存在线性关系的时候能够很好地描述, 当它们之间没有明显的线性关系的时候, 则这种方法难以做到很好的估计。因此, 在后续的工作中, 一是要在更高的波段范围内获取数据, 另一方面是探索基于非线性方法的土壤盐分定量估算模型。

本研究, 主要是对地面获取的高光谱数据进行了分析并建立了定量模型, 目的是为了指导下一步基于遥感影像进行土壤盐碱化的定量估计。因此, 在后续工作中还需要进一步收集遥感影像, 对本文的地面定量模型作进一步修正, 最终建立基于遥感影像的土壤盐碱化定量估计模型, 实现区域尺度上的土壤盐碱化监测与评价。

致谢: 本研究得到欧盟 CEOP2AEG IS (FP72ENV229072L, 资助号 (21291)) 部分资助。光谱测试得到了北京市农林科学研究院的刘良云研究员和北京师范大学地理学与遥感科学学院的任华忠等的支持, 在此表示感谢。

参 考 文 献

- [1] WENG Yongling, GONG Peng (翁永玲, 宫 鹏). Geographical Sciences (地理科学), 2006, 26(3): 369.
- [2] KANG Qing, ZHANG Zengxiang, ZHAO Xiaoli, et al. (亢 庆, 张增祥, 赵晓丽, 等). Journal of Arid Land Resources and Environment (干旱区资源与环境), 2006, 20(3): 144.
- [3] Mettemicht G I, Zinck J A. Remote Sensing of Environment, 2003, 85(1): 1.
- [4] Ben D R, Parkin A, Banin A, et al. International Journal of Remote Sensing, 2002, 23(6): 1043.
- [5] Farifteh J, Meer F V D, Atzberger C, et al. Remote Sensing of Environment, 2007, 110(1): 59.

- [6] Csillag F, Pasztor L, Bihl L L. Remote Sensing of Environment. 1993, 43(3): 231.
- [7] LI Haizhao, Brunner P, LI Wenpeng et al(李海涛, Brunner P, 李文鹏, 等). Hydrogeology and Engineering Geology(水文地质工程地质), 2006, 23(5): 75.
- [8] FU Qinghua, NI Shaoxiang, WANG Shixing et al(扶卿华, 倪绍祥, 王世新, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2007, 27(1): 48.
- [9] Nguyen H T, Lee B W. European Journal of Agronomy. 2006, 24(4): 349.
- [10] Zhou Y P, Jiang J H, Lin W Q et al. Talanta. 2007, 71(2): 848.
- [11] Wu J A, Luko R E, Corke H. Journal of Cereal Science. 2006, 44(2): 117.
- [12] Durand A, Devos Q, Ruckebusch C, et al. Analytica Chimica Acta. 2007, 595(122): 72.
- [13] WANG Huiven(王惠文). Partial Least Square Regression Method and Its Application(偏最小二乘回归方法及其应用). Beijing: National Defence Industry Press(北京: 国防工业出版社), 1999. 45.
- [14] WANG Jizhuo, HUANG Wenjiang, LAO Caizhen et al(王纪华, 黄文江, 劳彩莲, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(7): 1319.
- [15] JI Haizhan, WANG Pengxin, YAN Taizhi(吉海彦, 王鹏新, 严泰来). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(3): 514.
- [16] Ross I V R A. Chemometrics and Intelligent Laboratory Systems. 2008, 90(1): 72.
- [17] MEI Anxin, PENG Wangliu, QIN Qinding et al(梅安新, 彭望, 秦其明, 等). Introduction to Remote Sensing(遥感导论). Beijing: Higher Education Press(北京: 高等教育出版社), 2001. 156.

Quantitative Retrieval of Soil Salinity Using Hyperspectral Data in the Region of Inner Mongolia Hetao Irrigation District

QU Yonghua¹, DUAN Xiaoliang², GAO Hongyong², CHEN Aoping², AN Yongqing², SONG Jinting¹,
ZHOU Hongmin¹, HE Tao¹

1. Research Center for Remote Sensing and GIS, Dept. Geography, Beijing Normal University, China State Key Laboratory of Remote Sensing Science, Beijing 100875, China
2. Administration of Inner Mongolia Hetao Irrigation District, Wuyuan 015100, China

Abstract In the present paper, to investigate the spectral property of salinized soil and the relationship between the soil salinity and the hyperspectral data, the field soil samples were collected in the region of Hetao irrigation, Neimeng in the northwest China from the end of July to the beginning of August. The partial least squares regression (PLSR) model was established based on the statistical analysis of the soil ions and the reflectance of hyperspectra. The independent validation using data which are not included in the calibration model reveals that the proposed model can predicate the main soil components such as the content of total ions ($S\%$), SO_4^{2+} , PH and $K^+ + Na^+$ with higher determination coefficients (R^2) of 0.728, 0.1801, 0.1715 and 0.1734 respectively. And the ratio of prediction to deviation (RPD) of the above predicted value is larger than 11.6, which indicates that the calibrated PLSR model can be used as a tool to retrieve soil salinity with accurate results. When the PLSR model's regression coefficients were aggregated according to the wavelength of visual (blue, green and red) and near infrared bands of Landsat Thematic Mapper (TM) sensor, some significant response values were observed, which indicates that the proposed method in this paper can be used to analyse the remotely sensed data from the space-based platform.

Keywords Soil salinity; Partial least squares regression; Hyperspectrum; Remote sensing

(Received Feb. 12, 2008; accepted May 16, 2008)