

引用格式:Chai Siyue, Ma Weiling, Liu Gaohuan, *et al.* Accuracy Analysis of GA-PLS based Soil Water Salinity Hyperspectral Characteristics Mining Approach[J]. Remote Sensing Technology and Application, 2015, 30(4): 638-644. [柴思跃, 马维玲, 刘高焕, 等. GA-PLS 方法提取土壤水盐光谱特征的精度分析[J]. 遥感技术与应用, 2015, 30(4): 638-644.]

doi:10.11873/j.issn.1004-0323.2015.4.0638

GA-PLS 方法提取土壤水盐光谱特征的精度分析

柴思跃^{1,2}, 马维玲^{1,2}, 刘高焕¹, 黄 翀¹, 刘庆生¹

(1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院大学, 北京 100049)

摘要:光谱定量遥感已成为土壤盐渍化大尺度调查的有效手段之一,但黄河三角洲地区盐渍化土壤的光谱响应特征尚未明确。以黄河三角洲野外测定土壤体积分含水率、电导率为例,应用遗传偏最小二乘法(GA-PLS)在小样本集条件下提取盐渍土壤的水分—盐分的光谱响应特征,利用蒙特卡罗方法随机模拟结果表明:在不同土壤水盐含量条件下,GA-PLS 方法所提取的光谱特征具有鲁棒性,含水率模型稳定在 23 个波段变量,即响应特征为 365~425、500~515、720~740、755~765 与 955~965 nm;土壤电导率模型的特征集数目为 20 个波段变量,特征为 370~385、405~425、500~535、650~660、755~760 与 1 030~1 050 nm。实验在不同预处理模型下,GA-PLS 算法所建立水盐光谱模型较 PLSR 模型均显示出更高的精度。其中,包络线预处理方法与 GA-PLS 算法相结合效果最优,其水分光谱模型测试集拟合精度(R^2),预测残差平方和(PRSS)与残差预测方差(RPD)分别为 0.88、9.36 与 15.80;土壤光谱模型测试集精度 R^2 、PRSS 与 RPD 分别为 0.71、15.68 与 13.76。

关键词:遗传—偏最小二乘算法(GA-PLS);土壤电导率;高光谱;黄河三角洲

中图分类号:TP 75 **文献标志码:**A **文章编号:**1004-0323(2015)04-0638-07

1 引 言

表层土壤的成土母质、机械组成、含水率与可溶性盐分组成等多种理化特征共同影响土壤可见光—近红外反射光谱,这使得提取土壤盐分光谱响应特征并建立盐分定量光谱模型成为研究难点^[1]。但伴随着机载或星载、手持光谱仪的广泛应用,利用高分辨率光谱数据提取土壤盐分响应信号,并最终定量反演盐分信息已成为可能^[2-4]。

国内外学者在这一领域进行了深入探索,但计量分析方法的优劣直接影响光谱反演土壤盐分特征的准确性。王静等^[5]利用土壤盐分光谱的峰谷作为盐分含量特征;张飞等^[6]、姚远等^[7]应用相关分析法建

立盐分含量(或特定盐分)与光谱波段的线性关系。但受制于样本容量大小对相关系数值的影响,相关系数的不确定性造成以上两篇报道中对土壤盐分光谱区间判定具有差异,甚至得出相反的结论;Csillag 等^[8]与翁永玲等^[9]应用主成分思想用于光谱特征挖掘,研究均给出了土壤反射光谱与土壤盐分的良好回归关系;Goldshleger 等^[10]依据植物对盐分的吸收过程,通过土壤光谱特征建立起土壤盐分与番茄叶片中 Na、Cl 离子的间接关系,并量化植物叶片的含盐量。但其中所涉及方法无论是主成分分析法(PCA)或偏最小二乘法(Partial Least Square, PLS)方法,都是利用全部变量投影到主成分中进行建模,并不能解决选取特征的问题。近年来,人工智能领域中的非线性寻

收稿日期:2014-06-18;修订日期:2015-06-15

基金项目:国家自然科学基金项目“现代黄河三角洲地下水—土壤—大气相互作用模式研究”(41271407),海洋公益性行业科研专项项目(201105020)。

作者简介:柴思跃(1985—),男,北京人,博士研究生,主要从事时空数据挖掘研究。E-mail: chaisy@lreis.ac.cn。

优算法可以成功解决自适应发现 PLS 方法自变量成分特征选择的问题。遗传算法 (Genetic Algorithm, GA) 作为一种优秀的寻优算法,已在光谱分析领域进行广泛应用^[11-13]。

表 1 土壤盐分高光谱特征选取的文献简述
Table 1 A short overview of hyper spectral features of soil salinity

文献	光谱仪	量程	处理方法	研究区域	特征波段/nm
王静等 ^[5]	ASD	350~1 050	定性分析	吉林省长岭县	550、720、760、820、940
张飞等 ^[6]	CI700	200~1 100	相关系数	渭一库绿洲	EC、含盐量、TDS 与 405、410、415、820 波段的相关性不明显
姚远等 ^[7]	ASD	350~2 500	相关系数	渭一库三角洲	456、533、686、1 373
Csillag 等 ^[8]	—	495~2 395	主成分分析	California &	550~770、900~1 030、1 270~1 520、1 940~2 150、2 150~
			(MSPCA)	Hungary	2 310、2 330~2 400
翁永玲等 ^[9]	ASD	350~2 500	PLSR	茶卡—共和盆地	83 个波段
Goldshleger 等 ^[10]	ASD	350~2 500	PCA+PLSR	Jezreel Valley	400~560、720、983、1 158、1 345、1 377、1 423、1 538、1 623、
					1 660、1 777、1 804、1 844、1 880、1 950、2 200~2 240、2 350

本文应用土壤电导率作为盐渍化程度指标,旨在应用 GA-PLS 算法分别建立土壤含水率光谱模型与电导率光谱模型,并利用含水率光谱特征结果与现有研究结果的一致性来证明 GA-PLS 算法挖掘土壤电导率光谱特征波段的准确性。实验利用野外采样数据应用蒙特卡洛方法 (Monte Carlo Simulation, MC) 重复模拟验证 GA-PLS 算法在光谱特征水分、盐分建模的鲁棒性。实验在原始反射率、反射率去包络线、吸光度和反射率一阶导数 4 种预处理条件下,分别应用 GA-PLS 与 PLS 算法建立土壤水分、盐分的光谱特征模型,并验证了 GA-PLS 方法在盐渍土水盐模拟预测结果的精度优势。

2 实验方法

2.1 研究区域

黄河三角洲位于渤海湾南岸与莱州湾西岸,背陆面海,地势低平缓和。属暖温带半湿润大陆性季风气候,春季升温快、秋季降温幅度大。年均降雨量 530~630 mm,年蒸发量 1 900~2 400 mm。高蒸降比造成该区域内大部分土壤出现不同程度的盐渍化现象;全年平均风速 3.1~4.6 m/s,风力因素更加剧了这一过程。区域内的土壤发育程度不均,主要包括潮土、盐化潮土和滨海盐土 3 个类型,面积分别为 928、1 344、3 004 km²,分别占总面积的 17%、25%和 55%,其中以滨海盐土分布面积最大。本研究采样区域位于 118°~119°E 和 37°27'~38°6'N 之间,依地势呈东北—西南方向;采样区内分布有农田、荒地以及大量斑块状怪柳、碱蓬等耐盐植被,具备典型盐碱化特征。

2.2 数据采集

现有研究普遍使用室内与室外采样两种光谱采

集方法。部分研究^[2,8]为消除水分对光谱特征的干扰,降低土壤光谱的复杂性,利用烘干土壤样品在室内采集土壤光谱,但这种方法对土样的制备与室内光谱采集人员的经验提出了较高要求;另一部分研究^[4-5]为建立同期星载光谱仪成像对象的一致性,通过光谱仪现场采样方式获取光谱,使得采样光谱具有更广泛的应用价值。

本研究选用现场直接测定土壤光谱的方式。以电导率作为表层土壤含盐量的指标^[14-15],选用 TRI-ME-PICO TDR 便携式土壤水分测量仪测定含水率、电导率。设备在土壤电导率 0~12 dS m⁻¹ 条件下,测定指标重复精度<0.3%。光谱采样选用 SVC-GER 1500 光谱仪,光谱量程为 300~1 100 nm,分辨率为 3.2 nm。利用 2012 年资源 3 号卫星的多光谱高分影像与黄河三角洲 DEM 图为底图选取随机分布点 59 个(图 1)。实验于 2013 年 5 月 15~17 日 10:00~14:00 进行,期间天气晴朗少云,测试人员着深色服装,点位测定选用 1 m 间隔的 Z 字型的采样方法,在同一地点采集 5 个重复光谱与电导率,用采样均值作为区域土壤反射光谱。TDR 所测定体积含水率在 16~55.5%之间,均值为 39.8%;电导率测定值在 2.4~6.34 dS m⁻¹之间,均值为 4.82 dS m⁻¹。

2.3 数据预处理

预处理主要包括对光谱采样数据的校正、截取、平滑与滤波。光谱校正分为两步,分别校正参考白板反射率与仪器内参数;截取处理针对光谱仪采样噪声,将光谱范围边缘数据进行剔除,即去除 300~350 nm 与 1 050~1 100 nm 范围内的光谱数据(图 2);在样本数量有限的条件下,平滑处理平衡了土壤采样个数与变量个数(采样数不小于变量数的 20%^[16]),选取比土壤光谱精度略高的窗口(窗口大

小=5 nm)进行平滑处理;本次实验应用 3 种类型滤波模型,即吸光度模型、去包络线模型和一阶导方法对原始信号进行变换。

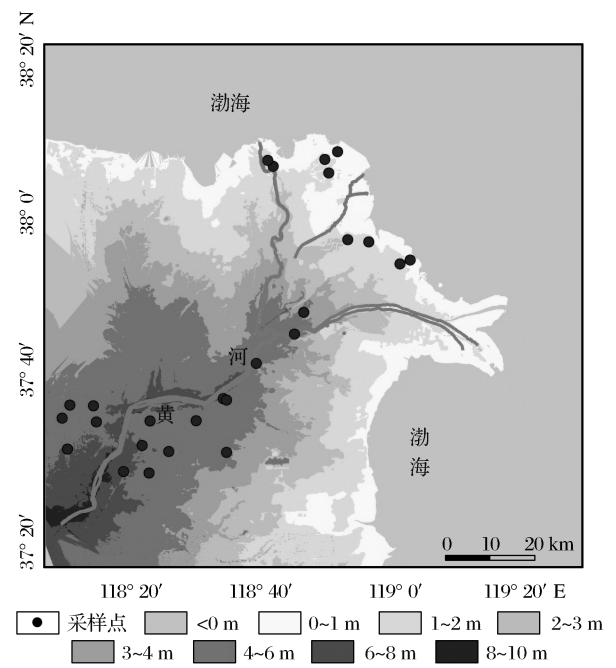


图 1 黄河三角洲土壤样本采样分布图
Fig. 1 Soil sampling distribution map in YDR

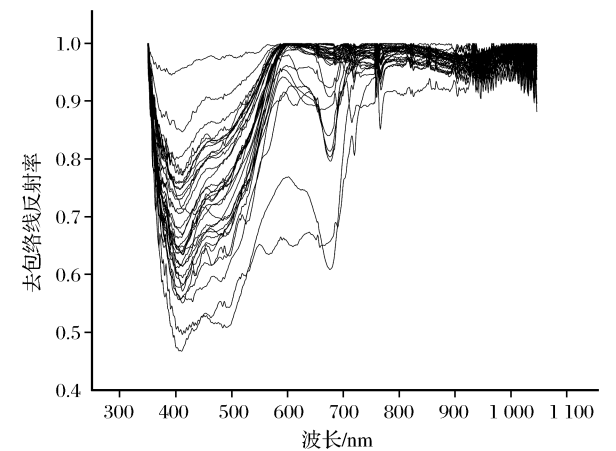


图 2 盐渍化土壤光谱去包络线曲线
Fig. 2 Soil salinity reflection curve with continuum removal

3 结果与评价

3.1 建模方法

PLS 方法的主要思想是将自变量与因变量同时投影到新坐标系内构建一个新的线性模型,使得这个新模型在交叉验证过程中满足交叉验证均方根误差(Root Mean Square Error of Cross Validation, RMSECV)最小^[17]。非线性寻优 GA 算法用于确定此新坐标系。GA 算法引入染色体概念,将变量视为染色

体内的基因。通过随机建立种群,利用适宜度(Fitness)评价种群内个体优劣并繁衍后代,模拟自然界遗传选择规律,以优胜劣汰机制选择更适宜的基因。另外,引入交叉机制模拟种群间的基因交叉,生成新的个体保证了寻优过程的收敛,同时引入变异机制以避免结果终止于局部最优。GA-PLS 具体算法过程可参阅文献[18],流程图详见图 3。其中,留一法交叉验证的评价指标 RMSECV 计算公式如下:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^m (y_i - \bar{y}_i)^2}{m}}$$
 (1)

其中:m 为训练集个数。

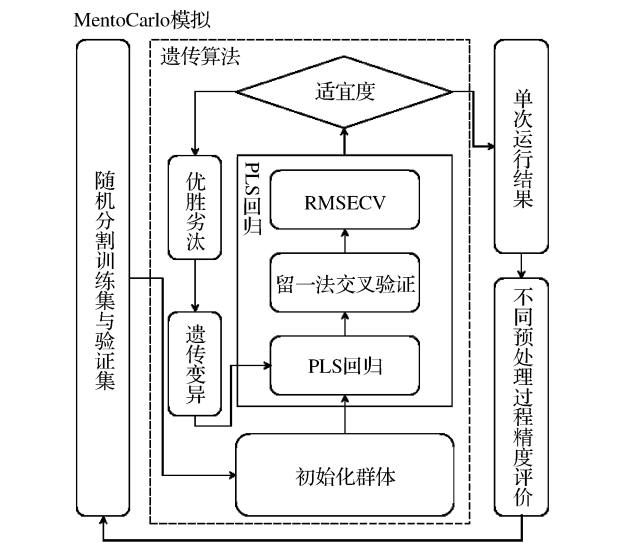
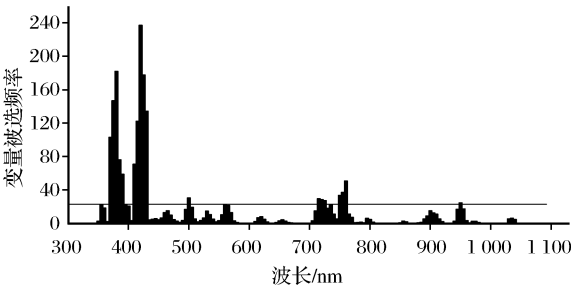


图 3 GA-PLS 土壤水盐光谱特征建模及验证流程图
Fig. 3 Flow chart of soil spectral feature selection and validation using GA-PLS

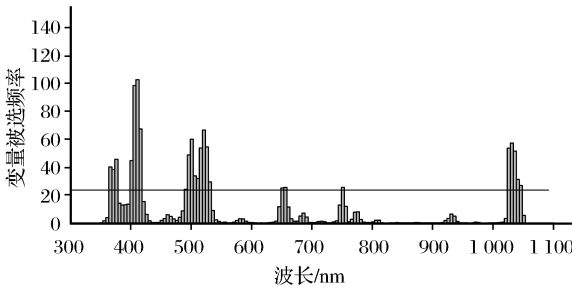
在 Matlab7.0 平台下,对 140 个波段变量进行 GA-PLS 运算。设定种群大小为 30,遗传概率为 0.6、变异概率为 0.01、遗传算法迭代次数为 200 次。每次迭代过程中,优势基因(波段特征变量)在所设定的竞争模式模式下保留。在 PLS 方法交叉验证计算过程中,依变量负载值大小对变量排序,通过逐一计算误差值 RMSECV,选取最小 RMSECV 所对应的特征变量数目即是最优拟合特征数。以土壤电导率建模为例(图 4),当选取变量数为 20 时,误差平方和达到最小值 0.216,选取参数数量大于 20 时反而 RMSECV 升高。这表明随着新变量的加入,模型精度呈 U 型变化,当变量个数大于临界值后,模型精度受到冗余信息影响不断恶化,使得土壤电导率模拟精度逐步降低。

统计建模结果表明(图 5):土壤含水率对应变

量集数目为 23 个,分别是 365~425、500~515、720~740、755~765 与 955~965 nm 波段。特别在 400~800 nm 波段内,实验所得土壤含水率的光谱特征接近 Pope 等^[19]所报导纯水可见光光谱范围内的弱反射峰值一侧(图 6),为光谱一阶导绝对值较大区域。同样方法所得土壤电导率特征光谱集数目为 20 个,分别对应 370~385、405~425、500~535、650~660、755~760 与 1 030~1 050 nm 波段。土壤电导率测定值受到含水率和盐含量共同作用的影响,其光谱反射特征主要分布在蓝光波段与近红外波段。以上分析结果表明,GA-PLS 算法所提取的土壤反射光谱细节特征具有准确性。



(a) 土壤水分光谱响应特征



(b) 电导率光谱响应特征

图 5 水分和电导率光谱响应特征结果
Fig. 5 Water and EC spectral characters

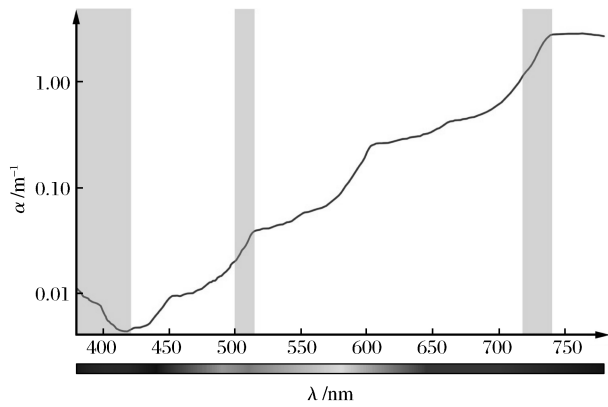


图 6 可见光范围内纯水的弱反射峰值(由 POPE^[19]报道修改)
Fig. 6 Reflectance of water in visible light

3.2 鲁棒性评价——以电导率为例

GA-PLS 算法在寻优路径上的随机性造成特征选择数目的不确定性,即每次运行结果间具有差异,甚至存在陷入局部最优的概率。为避免这种特殊情况所产生伪解的问题,本文应用 MC 模拟方法验证 GA-PLS 算法所得光谱特征的鲁棒性。方法依照 10% 比例将 59 个土壤样本划分为 53 个样本的训练集与 6 个样本的测试集采用随机划分训练集与验证集并进行 GA-PLS 重复计算,在人为创造土壤水

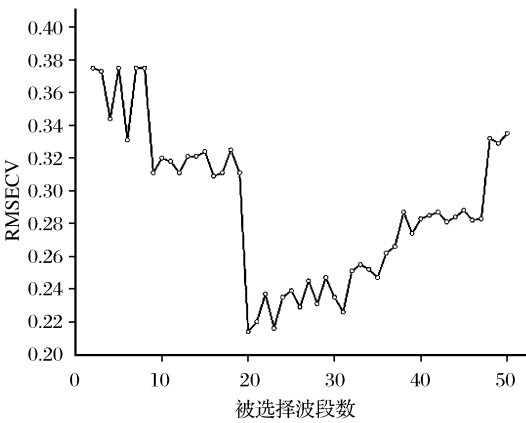


图 4 土壤电导率光谱特征交叉验证结果
Fig. 4 Cross validation result of soil EC spectral model

分、盐分区间不同条件下,验证土壤电导率特征光谱的一致性。MC 实验通过统计光谱特征区间,将光谱波段作为 x 轴,将 MC 模拟次数作为 y 轴,将在遗传算法计算过程中选择的优势基因被遗传的次数作为 z 轴,通过 500 次重复计算所得到的可视化结果表明(图 7):沿 x 轴方向,土壤电导率信息在可见光—近红外波段主要包含 405~425、755~760 与 1 030~1 050 nm 被选频率远远高于其他波段,这表明土壤电导率同时受到土壤中水含量、盐含量的影

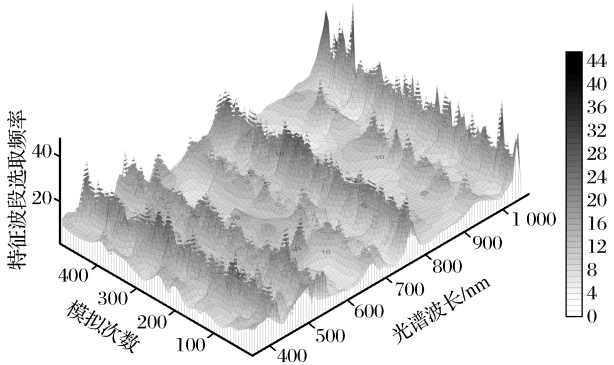


图 7 MC 模拟土壤电导率光谱响应特征可视化图
Fig. 7 Visualization of soil water spectral characters in Monte Carlo simulation

表 2 预测精度评价
Table 2 Training set prediction accuracy

预处理	建模方法	含水率			电导率		
		R^2	PRSS	RPD	R^2	PRSS	RPD
反射率 R	PLSR	0.64	37.45	9.11	0.61	45.97	11.29
	GA-PLS	0.66	32.49	10.83	0.65	36.72	12.31
吸光度 $\text{Log}(1/R)$	PLSR	0.56	71.57	5.70	0.56	88.74	6.85
	GA-PLS	0.59	67.57	7.77	0.55	82.08	7.15
去包络线 CR	PLSR	0.81	18.66	12.43	0.67	18.15	13.76
	GA-PLS	0.88	9.36	15.80	0.71	15.68	14.87
一阶差分	PLSR	0.69	31.81	11.37	0.63	38.94	8.60
$dR/d\lambda$	GA-PLS	0.70	27.25	9.49	0.67	32.49	9.44

响,蓝光波段与近红外波包含较多的信息;另外,以 515、655 与 910 nm 为中心的 3 个区间包含部分土壤电导率信息,具有丘陵状峰值。

沿 y 轴方向,500 次随机模拟结果所选取的盐分响应特征波段数目与被选择次数具有差异。这是遗传算法随机给定特征变量初值所造成的结果,表现为两种不同现象:①给定值建模适宜度较差,寻优过程收敛速度慢,各波段“均匀”地被选取,被选择变量的数目高,但最终寻优结果为全局变量。这表现为水分光谱响应特征在图中可见光蓝光波段与近红外波段区域峰值多,但本应出现的主峰值变为弱峰值;②给定值建模适宜度较好,寻优过程快速收敛,导致其他特征波段没有机会参与运算,结果收敛于局部最优区间,表现为被选择变量数目少而频率高,例如图 7 中 910、1 035 nm 附近的孤峰现象。重复模拟统计结果表明(图 8),被选变量数目呈正态分布态势,峰值出现在 17 附近,表明不同水盐范围内条件下,GA-PLS算法所提取的土壤电导率光谱特

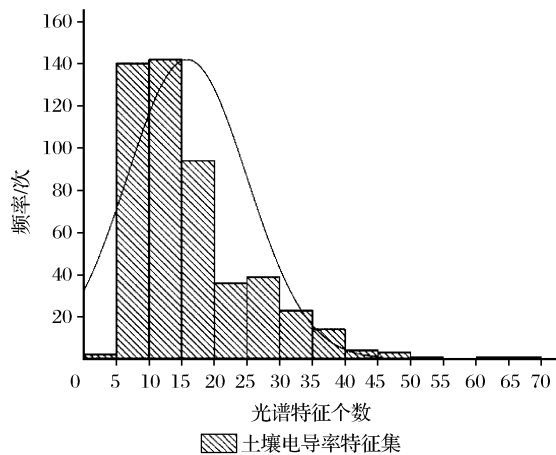


图 8 MC 模拟的 GA-PLS 算法所选特征变量统计图
Fig. 8 The histogram of features selected by GA-PLS algorithm under MC simulation

征波段个数稳定,受到水盐范围影响和随机因素影响较小,算法具有鲁棒性。

3.3 精度评价

验证集精度评价选用拟合精度 R^2 ,残差预测平方和 PRSS(Predicted Residual Sum of Squares)与预测残差方差 RPD(Residual Predictive Deviation)作为指标。模型预测结果 PRSS 值越小,RPD 越大且大于 2.5,表明模型预测的能力越强,PRSS 与 RPD 计算公式如下:

$$\text{PRSS} = \sum_{j=1}^t (Y_j - \bar{Y}_j)^2 \tag{2}$$

其中:t 为测试集个数。

$$\text{RPD} = \text{SD}_{\text{test}} / \text{SEP}_{\text{test}} \tag{3}$$

其中:test 为测试集。

精度分析结果表明,在去包络线处理下最优精度最优,其中 GA_PLS 含水率光谱模型 R^2 ,PRSS 与 RPD 达到 0.88,9.36 与 15.80,较 PLSR 模型分别提高 8.64%、49.83%和 27.11%;同理,GA_PLS 电导率光谱模型 R^2 ,PRSS 与 RPD 达到 0.71、15.68与 13.76,较 PLSR 模型分别提高 5.97%、13.61%和 8.01%。

4 结 语

一般而言,土壤在特定时刻的电磁特性具有确定性。因此,由电磁波时域反射原理所得的表观电导率信号与光谱仪接收的光谱反射率信号二者之间存在天然的联系。但反射率高维信号中的大量冗余信息淹没了这种关联模式。

本文应用 GA-PLS 方法挖掘出黄河三角洲盐渍土的土壤水分—盐分光谱响应模型,并验证了该模型的可靠性。研究结果证明:

(1)GA-PLS 算法在自变量多、变量间高度相关且样本集远小于变量集的条件下,所建模型能够高

精度提取土壤水分、盐分弱反射特征波段,且具有鲁棒性。

(2)建立适用于含水率在 16%~55.5%且表观电导率在 2.4~6.34 dS m⁻¹间的黄河三角洲盐渍化土壤水分与电导率光谱模型。其中,土壤水分光谱响应波段为 365~425、500~515、720~740、755~765 与 955~965 nm 波段;电导率光谱响应波段为 370~385、405~425、500~535、650~660、755~760 与 1 030~1 050 nm 波段。

然而,GA-PLS 作为机器学习的一种算法,不可避免地受到土壤样本数目限制。若在今后研究中采用大样本数据,所建立土壤电导率光谱反射模型的表现将更为精确与实用。

参考文献(References):

[1] Metternicht G I,Zinck J A. Remote Sensing of Soil Salinity: Potentials and Constraints[J]. Remote Sensing of Environment,2003,85(1):1-20.

[2] Hamzeh S,Naseri A A,AlaviPanah S K,*et al.* Estimating Salinity Stress in Sugarcane Fields with Spaceborne Hyperspectral Vegetation Indices[J]. International Journal of Applied Earth Observation and Geoinformation,2013,21:282-290.

[3] Zhang Xiaoguang,Huang Biao,Ji Junfeng,*et al.* Quantitative Prediction of Soil Salinity Content with Visible-near Infrared Hyper-spectral in Northeast China[J]. Spectroscopy and Spectral Analysis,2012,(8):2075-2079. [张晓光,黄标,季峻峰,等. 基于可见近红外高光谱的东北盐渍土盐分定量模型研究[J]. 光谱学与光谱分析,2012,(8):2075-2079.]

[4] Sidike A,Zhao S,Wen Y. Estimating Soil Salinity in Pingluo County of China Using QuickBird Data and Soil Reflectance Spectra[J]. International Journal of Applied Earth Observation and Geoinformation,2014,26:156-175.

[5] Wang Jing,Liu Xiangnan,Huang Fang,*et al.* Salinity Forecasting of Saline Soil based on ANN and Hyperspectral Remote Sensing[J]. Transactions of the Chinese Society of Agricultural Engineering,2009,(12):161-166. [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[J]. 农业工程学报,2009,(12):161-166.]

[6] Zhang Fei,Tashpolat Tiyp,Ding Jianli,*et al.* Correspondence Analysis of Relationship between Characteristics and Spectral of Soil Salinization[J]. Acta Pedologica Sinica,2009,46(3):513-519. [张飞,塔西甫拉提·特依拜,丁建丽,等. 基于对应分析的土壤盐渍化现状特征及其与光谱关系研究[J]. 土壤学报,2009,46(3):513-519.]

[7] Yao Yuan,Ding Jianli,Ardak·Kelimu,*et al.* Research on Remote Sensing Monitoring of Soil Salinization based on Measured Hyperspectral and EM38 Data[J]. Spectroscopy and Spectral Analysis,2013,(7):1917-1921. [姚远,丁建丽,阿尔达克·克里木,等. 基于实测高光谱和电磁感应数据的区域土

壤盐渍化遥感监测研究[J]. 光谱学与光谱分析,2013,(7):1917-1921.]

[8] Csilag F,Psillag L,Biehl L L. Spectral Band Selection for the Characterization of Salinity Status of Soils[J]. Remote Sensing of Environment,1993,43(3):231-242.

[9] Weng Yongling,Qi Haoping,Fang Hongbin,*et al.* PLSR-based Hyperspectral Remote Sensing Retrieval of Soil Salinity of Chaka-Gonghe Basin in Qinghai Province[J]. Acta Pedologica Sinica,2010,47(6):1255-1263. [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[J]. 土壤学报,2010,47(6):1255-1263.]

[10] Goldshleger N,Chudnovsky A,Ben-Binyamin R. Predicting Salinity in Tomato Using Soil Reflectance Spectra[J]. International Journal of Remote Sensing,2013,34(17):6079-6093.

[11] Atzberger C,Guzber M,Baret,*et al.* Comparative Analysis of Three Chemometric Techniques for the Spectroradiometric Assessment of Canopy Chlorophyll Content in Winter Wheat[J]. Computers and Electronics in Agriculture,2010,73(2):165-173.

[12] Song K,Li L,Li S,*et al.* Hyperspectral Retrieval of Phycocyanin in Potable Water Sources Using Genetic Algorithm of Canopy Chlorophyll Content in Winter[J]. International Journal of Applied Earth Observation and Geoinformation,2012,18:368-385.

[13] Prieto N,Oliveri P,Leardi R,*et al.* Application of a GAal of Applied Earth Observation and Geoinformation Genetic Algorithm[J]. Sensors and Actuators B: Chemical,2013,183:52-57.

[14] Wu Yueru,Wang Weizhen,Wang Haibing,*et al.* Analysis of Variation of Soil Salt with New Electric Conductivity Index[J]. Acta Pedologica Sinica,2011,48(4):869-873. [吴月茹,王维真,王海兵,等. 采用新电导率指标分析土壤盐分变化规律[J]. 土壤学报,2011,48(4):869-873.]

[15] Liu Qingsheng,Liu Gaohuan. The Simple Analysis on water-Salt and Field Spectral Characteristics of Calcaric Fluvisols in the Yellow River Delta[J]. Chinese Agricultural Science Bulletin,2008,24(3):253-257. [刘庆生,刘高焕. 现代黄河三角洲潮土水盐与野外光谱特征浅析[J]. 中国农学通报,2008,24(3):253-257.]

[16] Leardi R,Nørgaard L. Sequential Application of Backward Interval Partial Least Squares and Genetic Algorithms for the Selection of Relevant Spectral Regions[J]. Journal of Chemometrics,2004,18(11):486-497.

[17] Kruger U,Xie L,“Partial Least Squares”in Statistical Monitoring of Complex Multivariate Processes[M]. Singapore: John Wiley & Sons,Ltd,2012:375-409.

[18] Leardi R,Lupidi A. Genetic Algorithms Applied to Feature Selection in PLS Regression: How and When to Use Them[J]. Chemometrics and Intelligent Laboratory Systems,1998,41(2):195-207.

[19] Pope R M,Fry E S. Absorption Spectrum(380~700 nm)of Pure Water II; Integrating Cavity Measurements[J]. Applied Optics,1997,36:8710-8723.

Accuracy Analysis of GA-PLS based Soil Water Salinity
Hyperspectral Characteristics Mining Approach

Chai Siyue^{1,2}, Ma Weiling^{1,2}, Liu Gaohuan¹, Huang Chong¹, Liu Qingsheng¹

(1. *Institute of Geographic Sciences and Natural Resources Research ,
Chinese Academy of Sciences ,Beijing 100101,China ;*
2. *University of Chinese Academy of Sciences ,Beijing 100049,China*)

Abstract: Hyperspectral remote sensing data is one of effective ways which can be used to retrieve salinity quantitatively in soil monitoring. But the quantitative structure-property relationship between soil salinity and soil spectral reflection characters has not been found in yellow river delta region. Genetic Algorithm with Partial Least Square kernel(GA-PLS)method is applied to mine spectral features of volumetric moisture content(V%)and Electrical Conductivity(EC)using the in-stu salinity soil sampling in Yellow River Delta region. MC simulation result shows GA-PLS method mines stable characters numbers and fitness under different of water-salt level, which prove the robustness of the algorithm. Therefore, the spectral features of V% exist in 365~425, 500~515, 720~740, 755~765 and 955~965 nm bands, compared with the spectral features of EC_b appear in 370~385, 405~425, 500~535, 650~660, 755~760 and 1 030~1 050 nm bands. According to the experiment result, through 4 different preprocessing approaches, water content model and electric conductivity model of both PLS and GA-PLS are all evaluated by R^2 , Predicted Residual Sum of Squares(PRSS)and Residual Predictive Deviation(RPD), GA-PLS models got the better point in prediction accuracy rather than PLS regression. The continuum removal approach leads to the highest prediction accuracy among all other preprocessing methods, with R^2 , PRSS and RPD equal 0. 88, 9. 36 and 15. 80 in soil water content model and 0. 71, 15. 68 and 13. 76 in EC model.

Key words: Genetic algorithm-partial least square(GA-PLS); Soil electrical conductivity; Hyperspectral; Yellow river delta