

南疆枣园盐渍土电导率的近红外高光谱分析

张楠楠^{1,2}, 张 晓^{1*}

(1. 塔里木大学信息工程学院/中国农业科学院农业信息研究所新疆南疆农业信息化研究中心, 新疆 阿拉尔 843300; 2. 中国农业科学院农田灌溉研究所/农业农村部作物需水与调控重点实验室, 河南 新乡 453000)

摘 要:土壤电导率的高低反映土壤中盐分含量的多少。以南疆一枣园盐渍土为研究对象, 运用 Zolix GaiaSorter 近红外成像高光谱仪采集 900~1 700 nm 范围内 142 个土壤样品的光谱数据, 进行 6 种光谱预处理, 分别运用 SPA (连续投影算法) 和 CARS (竞争性自适应重加权法) 提取的特征波段建立 PLS 模型, 与全波段的模型进行对比分析。结果显示, CARS-PLS 的模型优于 SPA-PLS 和 W-PLS, MSC 处理后的 CARS-PLS 模型效果最好。因此利用高光谱技术结合 CARS-PLS 可实现南疆枣园盐渍土的电导率快速准确检测。

关键词:高光谱; 土壤电导率; PLS 建模; 南疆枣园

中图分类号: O433.1; S151.9+3 文献标识码: A

文章编号: 2096-5877(2022)04-0118-04

Analysis of Near Infrared Spectra of Southern Xinjiang Jujube Orchard Conductivity of Saline Soil

ZHANG Nannan^{1,2}, ZHANG Xiao^{1*}

(1. College of Information Engineering, Tarim University/Southern Xinjiang Agricultural Information Research Center, Agricultural Information Institute of CAAS, Alear 843300; 2. Farmland Irrigation Research Institute, CAAS/ Key Laboratory of Crop Water Use and Regulation, Ministry of Agriculture and Rural Affairs, Xinxiang 453000, China)

Abstract: The level of soil conductivity reflects the amount of salt in the soil. In this study, the saline soil of a jujube garden in southern Xinjiang was taken as the research object. The spectral data of 142 soil samples were collected using Zolix GaiaSorter near infrared imaging high spectrometer at the band of 900-1,700 nm. At first six spectral pretreatments were performed, and then SPA and CARS were used for feature band extraction, and model effects of SPA-PLS, CARS-PLS, and full-wavelength PLS soil conductivity at different wavebands were comparatively analysed. The results showed that CARS-PLS model was superior to SPA-PLS and W-PLS, and the MSC-treated CARS-PLS model had the best effect. Therefore, the use of hyperspectral technology combined with CARS-PLS can achieve rapid and accurate conductivity measurement of soil saline soil in Southern Xinjiang jujube garden.

Key words: Hyperspectral; Soil conductivity; PLS modeling; Southern Xinjiang jujube orchard

新疆南疆是我国枣树的重要种植大区, 土壤多为盐渍土, 土壤盐分的高低, 对枣树生长影响较大^[1-3]。土壤电导率可间接反映土壤盐分, 两者存在正相关性。传统获取土壤电导率的方法费时、费力^[4], 现用光谱的方法得到土壤电导率的大

小, 以为南疆盐碱土枣树栽培和红枣产业可持续发展提供参考依据。

近年来, 相当多的专家学者致力于采用不同的高光谱预处理方法和变量的选择问题^[5-8], 使所建立的土壤电导率模型更稳定和精确。如高华光等^[9]运用随机森林算法建立齐长城土壤电导率的高光谱定量模型精度较好; 亚森江·喀哈尔等^[10]用两波段优化算法进行波段优化, 最终建立较好的土壤电导率模型; 李相等^[11]同时对土壤盐分和电导率进行分析, 通过 15 种光谱变换, 得出相比于土壤盐分, 土壤电导率对高光谱信息更加敏感。

目前, 对南疆枣园盐渍土电导率的高光谱研究较少。本文以南疆绿洲区为研究区, 依据近红

收稿日期: 2019-11-20

基金项目: 国家自然科学基金项目(62061041、31960503); 国家自然科学基金青年科学基金项目(32101621); 农业农村部作物需水与调控重点实验室项目(FIRI2018-05-03)

作者简介: 张楠楠(1987-), 男, 副教授, 硕士, 主要从事农业遥感及作物模型方面的研究。

通讯作者: 张 晓, 女, 硕士, 副教授, E-mail: zhangxiaosenu@163.com

外高光谱数据、土壤电导率实测数据,通过多种处理方法对盐渍土光谱进行变换处理消除光谱噪声,运用SPA^[12]和CARS选出特征波长,建立W-PLS、SPA-PLS、CARS-PLS预测模型,并将3种模型进行比较,以期为南疆枣园土壤电导率的预测模型提供一种新的思路和方法。

1 材料与方 法

1.1 研究区概况

本研究以新疆维吾尔自治区南疆绿洲区枣园土壤为研究对象。南疆枣园土壤为盐渍土,每年冬季土壤表层重新聚盐会危害枣树的生长和发育,则需要探明表层土壤电导率。该地区最高气温为35℃,最低气温为-28℃,太阳辐射平均每年133.7~146.3 kcal/cm²。年均日照2 556.3~2 991.8 h,日照率为58.69%。研究区内雨雪稀少,地表蒸发强烈。

1.2 土壤采样

在南疆绿洲区所选的典型样点进行土壤采集分别,分别取表层0~20 cm土壤(正常土壤),地表面层盐分含量较高的土壤(高盐土壤)。将两类盐渍土带回实验室分别进行处理。室内阴干:将两类土样在塑料布上摊开,并做好标记,捏碎较大土块,以利于磨细;捡出杂物,以减小误差。研磨过筛,将阴干后的土样,分别倒入木盘中,用擀面杖或啤酒瓶研磨,通过1 mm筛。将高盐土壤以0 g、0.5 g、1.0 g、1.5 g、……和正常土壤94 g、93.5 g、93.0 g、92.5 g、……混合均匀,每份土壤样品为94 g,共获得142份土壤样品。将每个样品分成两份,用于土壤电导率测量和高光谱数据测定。

1.3 土壤电导率测定

土壤电导率采用的测量仪器电导率仪:DDSJ-308A(上海雷磁),参照《土壤农化分析》^[13]中电导法测定。

1.4 光谱预处理

采用Zolix GaiaSorter近红外成像高光谱仪,光谱测定范围900~1 700 nm(实际测量到1 750 nm),光谱分辨率5 nm,采样间隔4 nm,样本在室内20~25℃环境下进行扫描,获取一维影像和光谱信息,通过自带软件或ENVI软件获取每个枣园盐渍土样品的平均光谱值,每个样本数据测量5次取平均值,导出Excel文件。应用MATLAB2016a软件对采集到的土壤光谱数据进行处理和分析。光谱预处理采用标准化、归一化、移动窗口平滑、卷积平滑和多元散射校正(MSC)5种方法。

1.5 算 法

SPA是通过对向量的投影分析,从光谱矩阵提取有效信息使光谱变量共线性最小的新型变量选择方法^[14]。具体算法步骤见参考文献[15]。CARS算法能够在一定程度上克服变量选择中的组合爆炸问题,筛选出优化的变量子集,提高模型的预测能力^[16]。

1.6 建模方法和模型验证

PLS是多变量统计分析方法,该方法是集主成分分析、普通多元线性回归和典型相关分析于一体的回归分析方法,解决了自变量多重共线性的问题^[17],在光谱分析中得到广泛应用。为有效评价模型精度,本研究选取相关系数(R)、RMSECV、预测均方根误差(RMSEP)、准确率(Precision)、RPD进行模型分析检验。

2 结果与分析

2.1 土壤原始光谱

图1为在900~1 700 nm波长范围内142个土壤样品的感兴趣区域的光谱变化趋势。每条光谱包含256个变量。由图1可知,142个土壤样品光谱曲线趋势类似,无明显差异。142个土壤样品对近红外光的反射和吸收存在差异,在1 300~1 500 nm区域存在较多的吸收和反射峰,无法直接获得与电导率相关的关键波长。因此,必须利用有效算法去除异常样本和提取关键波长。

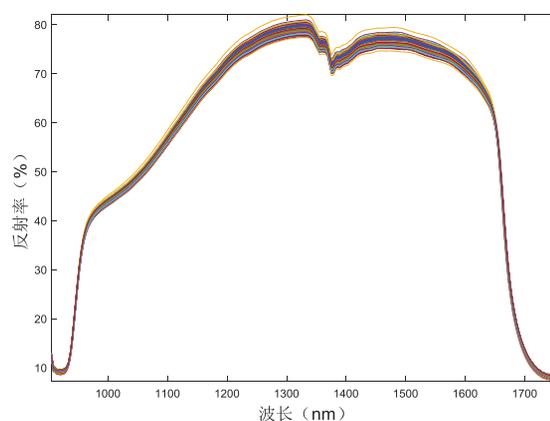


图1 土壤原始光谱

2.2 异常变量样本

通过蒙特卡洛交叉验证(Monte Carlo Cross Validation, MCCV)的方法对异常样本进行筛选。通过有效样本的光谱矩阵与电导率数据矩阵建立PLSR以确定最佳主成分数,然后采用蒙特卡洛采样法随机选取将80%有效样本作为校正集,剩余样本为预测集;通过已确定的主成分数与校正集

样本建立 PLSR 模型, 预测集则用来预测每个样本的预测残差; 计算各个样本预测残差的均值 (MEAN) 和标准差 (SD), 得到样本预测残差的 MEAN-SD 分布。其中 2 个样本为离群样本, 后续试验用去除两个异常样本后的样本进行处理。后续数据处理共 140 个样本, 其中 100 个用于样本建模, 剩余 40 个作为校正集。

2.3 SPA 选择特定波长后的 PLS 模型

采用建模集的 100 个样本 256 条光谱变量建立的 PLS 全谱模型在建模过程中有很大的光谱数据,

并引入干扰变量, 这将会降低模型的预测精度。在全谱范围内不使用光谱预处理, 之后使用 SPXY (Sample set partitioning based on joint x-y distance) 进行校正集样品划分处理, 最后使用 SPA 算法进行光谱变量压缩。当特征波长数为 5 时, SPA-PLS 模型的均方根误差 (RMSE) 最小为 0.311 78。5 个特征波长分别为: 1 443.13、1 501.430 1、1 136.810 1、1 595.13、1 354.98 nm, 其重要性依次减弱。

采用 6 种光谱预处理方法后建立的 SPA-PLS 模型如表 1 所示。

表 1 SPA-PLS 模型的预测结果

模型	预处理方法	交互检验均方根误差	预测均方根误差	准确率	相关系数	相对分析误差
SPA-PLS	原始光谱	0.315 27	0.311 78	0.940 6	0.992 89	8.051 6
	标准化	0.389 34	0.304 55	0.926 9	0.994 77	9.045 9
	归一化	0.390 45	0.280 24	0.923 62	0.995 67	10.266 6
	移动窗口平滑	0.289 02	0.356 91	0.944 47	0.989 67	6.667 5
	卷积平滑	0.188 65	0.307 9	0.948 46	0.992 26	7.848 5
	多元散射校正	0.623 61	0.301 71	0.927 4	0.994 95	9.131 1

综合比较, 经归一化处理后的效果最好, 如图 2 所示, 模型的 RMSECV 和 RMSEP 相对较小, 分别为 0.390 45 和 0.280 24, R 最接近 1 为 0.995 67, Precision 的值为 0.923 62, RPD 最大为 10.266 6。从局部特征波段来看, 说明归一化处理适合土壤电导率的近红外高光谱预处理。

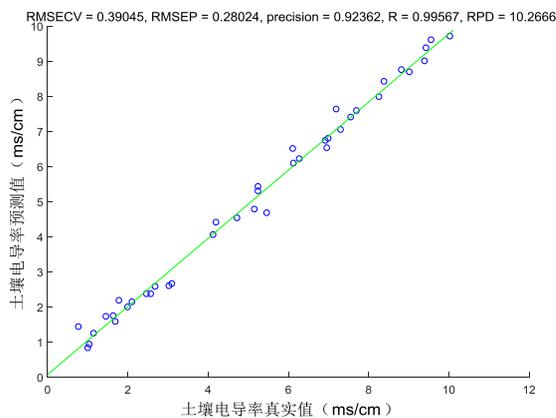


图 2 归一化的 SPA-PLS 模型参数图

2.4 基于 CARS 选择特定波长后的 PLS 模型

CARS 基本原理是基于指数衰减函数和自适应重加权采样技术, 优选出对应回归模型中决定系数较大的波长点集合。图 3 为在算法运行期间, 随采样次数的增加, 样本选取的波长点数、内部交叉验证均方差 RMSECV 和各个波长点回归系数的变化。图 3(a) 是特征变量提取的过程, 在前 3 次采样中波长点减少速度较快, 随后减慢, 说明

算法在特征波长点选取中具有“粗选”和“精选”两个阶段。图 3(b) 中随着 MC 采样次数的增加, 回归模型的 RMSECV 先减小后增大, 在第 25 次采样时降至最低点, 提取到的特征波长点数为 24, 对应的回归系数如图 3(c) 所示。

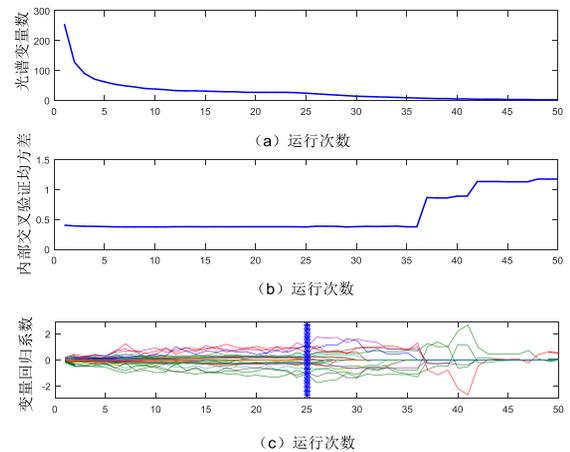


图 3 CARS 提取最优变量

本研究 CARS 算法设置的蒙特卡罗 (MC) 采样次数为 100, 采用 6 种光谱预处理后的模型如表 2 所示。经 MSC 预处理的光谱建立的模型效果最好, 如图 4 所示, precision 为 0.931 12, RMSEP 为 0.245 66, RMSECV 为 0.0383 48, RPD 为 11.662 5。CARS 算法提取的 25 个最优波长点建立的 PLS 模型, R 为 0.996 57, 其 RMSECV 最小值所对应的波长点集合即为 CARS 提取的最优特征波段组合。

表2 CARS-PLS模型的预测结果

模型	预处理方法	交互检验均方根误差	预测均方根误差	准确率	相关系数	相对分析误差
CARS-PLS	原始光谱	0.336 15	0.313 64	0.943 83	0.992 64	8.003 6
	标准化	0.419 99	0.315 64	0.914 16	0.993 71	8.728 1
	归一化	0.108 58	0.272 55	0.921 27	0.995 58	10.556 3
	移动窗口平滑	0.251 48	0.386 21	0.935 15	0.986 89	6.161 6
	卷积平滑	0.427 89	0.327 99	0.937 28	0.991 15	7.367 7
	多元散射校正	0.038 348	0.245 66	0.931 22	0.996 57	11.625 5

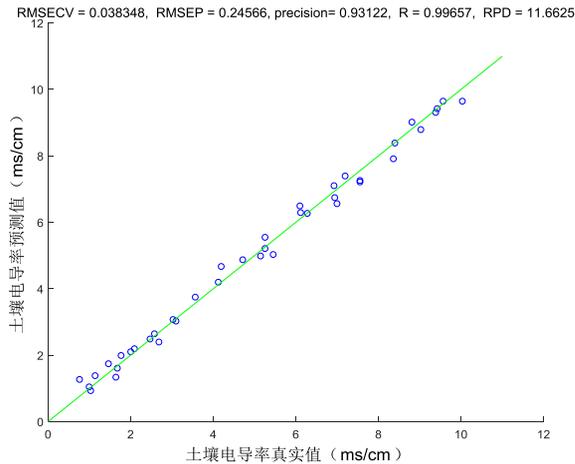


图4 MSC预处理后的CARS-PLS模型

2.5 与全谱PLS模型比较

剔除异常样本建立的全谱PLS模型见表3。综合比较,归一化处理的效果最好,结果如图5所示。

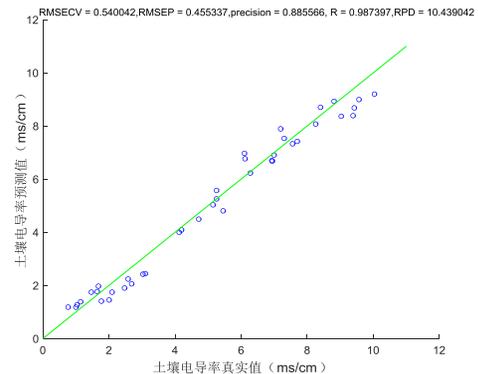


图5 归一化后的全谱PLS模型

表3 PLS模型的预测结果

模型	预处理方法	交互检验均方根误差	预测均方根误差	准确率	相关系数	相对分析误差
PLS	原始光谱	0.058 413	0.357 564	0.923 598	0.989 803	8.015 941
	标准化	0.114 317	0.327	0.920 786	0.992 931	7.709 952
	归一化	0.540 042	0.455 337	0.885 566	0.987 397	10.43 9042
	移动窗口平滑	0.819 020	0.585 429	0.898 234	0.969 267	4.907 872
	卷积平滑	0.832 450	0.583 159	0.900 121	0.970 446	4.938 939
	多元散射校正	0.114 388	0.326 988	0.920 787	0.992 931	7.710 921

3 讨论与结论

本研究以南疆枣园盐渍土为研究对象,利用近红外高光谱技术联合SPA和CARS算法提取特征波长,比较SPA-PLS、CARS-PLS、W-PLS共3类模型,发现经6种处理后,SPA-PLS模型较W-PLS模型大部分指标有所改善,但从RPD来看,前者略逊于后者,是由于SPA筛选特征变量时把有用信息也筛选掉。CARS-PLS模型较W-PLS模型好,是由于CARS筛选的变量较SPA多,保留有用信息较多,从而提高预测精度。总而言之,CARS-PLS模型评价指标均较优,且经MSC预处理的光谱建立的模型效果最好, Precision为0.931 22, RMSEP为0.245 66, RMSECV为0.038 348, RPD为11.662 5。

比较全局波段和局部特征波段的模型,经SPA和CARS算法波长筛选后波长数减少98%和90%以上,在较大程度上简化了校准模型,降低了计算复杂度,并具有很好的稳定性。

根据模型评价指标,经6种预处理后的模型都达到状态稳定、可用的标准(RPD均大于3),结果较好是由于整个实验在实验室环境下进行,并且盐渍土样品均经过筛选处理,尽量去除外界环境的干扰。该方法可在南疆枣园土壤中使用,因不同地区土壤理化参数不同,是否能在其他地区使用需要进一步验证。

参考文献:

- [1] 王东明,贾媛,崔继哲.盐胁迫对植物的影响及植物盐适应性研究进展[J].中国农学通报,2009,25(4):124-128.

(下转第154页)

- 机使用与维修,2018(6):82.
- [5] 施 虎,梅雪松.现代液压传动技术的若干新特点及其发展趋势[J].机床与液压,2017,45(23):158-166.
- [6] 金 梅,吴崇友,韩树钦.液压传动与控制技术在农机上的应用与发展趋势[J].机床与液压,2017,45(23):172-176.
- [7] 付 威,刘玉冬,坎 杂,等.果园修剪机械的发展现状与趋势[J].农机化研究,2017,39(10):7-11.
- [8] 王 哲,王丽红,付 威,等.酿酒葡萄修剪装置的设计[J].农机化研究,2017,39(3):105-110.
- [9] 董 祥,张 铁,严荷荣,等.PJZ-1型酿酒葡萄剪枝机设计与试验[J].农业工程,2018,8(5):95-100.
- [10] 王志强,刘凤之,王孝娣,等.葡萄修剪机的研制与试验[J].果树学报,2018,35(9):1141-1146.
- [11] 阿尔孜古丽·吾买尔.立架种植葡萄修剪机的研制与展望[J].时代农机,2016,43(3):48-49,51.
- [12] 龙 魁.往复式葡萄修剪机的设计与试验研究[D].乌鲁木齐:新疆农业大学,2014.
- [13] 王 赞.浅谈工程机械机电液一体化技术的发展趋势[J].城市建设理论研究(电子版),2018(9):147.
- [14] 赵大勇,李向军,刘宏梅.液压技术在现代农业机械中的应用现状与发展趋势[J].时代农机,2017,44(7):45.
- [15] 凌 杰.液压技术在现代农业机械中的应用现状与趋势[J].南方农机,2017,48(6):53-54.
- [16] 王志军.液压技术的发展现状与趋势[J].电子制作,2016(20):85.
- [17] 王 博,李海燕.液压传动技术应用现状及发展趋势[J].中州煤炭,2016(9):95-97,111.
- [18] 王 燕,刘书伟.槟榔花研究进展[J].东北农业科学,2017,42(5):51-55.

(责任编辑:王 昱)

(上接第121页)

- [2] 位 杰,蒋 媛,张 琦,等.骏枣和灰枣对NaCl胁迫适应性的比较研究[J].北方园艺,2013(24):21-24.
- [3] 王玉刚,肖笃宁,李 彦.三工河流域中上游绿洲土壤盐渍化的时空动态[J].中国沙漠,2008,28(3):478-484.
- [4] Meeterlicht G I, Zinck J A. Remote sensing of soil salinity: potentials and constraints[J]. Remote Sensing of Environment, 2003, 85(1): 1-20.
- [5] 郑立华,李民赞,安晓飞,等.基于近红外光谱和支持向量机的土壤参数预测[J].农业工程学报,2010,26(14):81-87.
- [6] 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,2009,25(12):161-165.
- [7] 石吉勇,邹小波,赵杰文,等.近红外光谱技术快速无损诊断黄瓜植株氮、镁元素亏缺[J].农业工程学报,2011,27(8):283-287.
- [8] 袁越明,熊 伟,方勇华,等.差分偏振FTIR光谱法探测水面溢油污染的模型分析[J].红外与激光工程,2011,40(3):408-412.
- [9] 高华光,于瑞阳.齐长城土壤电导率的高光谱定量估测模型[J].测绘与空间地理信息,2016,42(5):195-198.
- [10] 亚森江·喀哈尔,杨胜天,尼格拉·塔什甫拉提,等.基于分数阶微分优化光谱指数的土壤电导率高光谱估算[J].生态学报,2019,39(19):1-12.
- [11] 李 相,丁建丽,侯艳军,等.干旱半干旱区土壤含盐量和电导率高光谱估算[J].冰川冻土,2015,37(4):1050-1058.
- [12] Mário César Ugulino Araújo, Teresa Cristina Bezerra Saldanha, Roberto Kawakami Harrop Galvão, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.
- [13] 鲍士旦.土壤农化分析[M].北京:中国农业出版社,2000:178-188.
- [14] 陈 斌,孟祥龙,王 豪.连续投影算法在近红外光谱校正模型优化中的应用[J].分析测试学报,2007,26(1):66-69.
- [15] 洪 涯,洪添胜,代 芬,等.连续投影算法在砂糖橘总酸无损检测中的应用[J].农业工程学报,2010,26(S2):380-384.
- [16] 张华秀,李晓宁,范 伟.近红外光谱结合CARS变量筛选方法用于液态奶中蛋白质与脂肪含量的测定[J].分析测试学报,2010,29(5):430-434.
- [17] Tao J S, Yang Y F, Li Y H. Comparison of paper tensile strength prediction models based on PLS and SVM methods [J]. Journal of South China University of Technology (Natural Science Edition), 2014, 42(7): 132-137.

(责任编辑:王 昱)