

DOI: 10.11766/trxb201902190024

曹肖奕, 丁建丽, 葛翔宇, 王敬哲. 基于光谱指数与机器学习算法的土壤电导率估算研究[J]. 土壤学报, 2020, 57 (4): 867–877.

CAO Xiaoyi, DING Jianli, GE Xiangyu, WANG Jingzhe. Estimation of Soil Electrical Conductivity Based on Spectral Index and Machine Learning Algorithm[J]. Acta Pedologica Sinica, 2020, 57 (4): 867–877.

基于光谱指数与机器学习算法的土壤电导率估算研究

曹肖奕^{1, 2, 3}, 丁建丽^{1, 2, 3†}, 葛翔宇^{1, 2, 3}, 王敬哲^{1, 2, 3}

(1. 新疆大学资源与环境科学学院, 乌鲁木齐 830046; 2. 新疆大学绿洲生态教育部重点实验室, 乌鲁木齐 830046; 3. 新疆大学智慧城市与环境建模自治区普通高校重点实验室, 乌鲁木齐 830046)

摘要: 土壤盐分是干旱区土壤盐渍化评价的重要指标。以新疆维吾尔自治区渭干河-库车河三角洲绿洲为例, 基于土壤电导率 (Electrical conductivity, EC) 及可见光-近红外 (Visible and near infrared, VIS-NIR) 光谱数据, 通过蒙特卡罗交叉验证 (Monte Carlo cross validation, MCCV) 确定 364 个有效样本。采用原始光谱 (Raw reflectance, R) 及其经过微分、吸光度 (Absorbance, Abs)、连续统去除 (Continuum removal, CR) 等 6 种预处理后的数据构建光谱指数。基于筛选出的 21 个最优指数, 采用 BP 神经网络 (Back propagation neural network, BPNN)、支持向量机 (Support vector machine, SVM)、极限学习机 (Extreme learning machine, ELM) 三种算法对 EC 进行估算, 并引入偏最小二乘回归 (Partial least squares regression, PLSR) 进行比较。结果表明: 在基于 R 与 6 种光谱预处理数据构建的 21 个最优光谱指数之中, R_FD_RSI (R₁₉₁₃, R₂₁₄₂) 表现最佳 ($r = 0.649$); 与 PLSR 相比, 机器学习算法能够显著提高模型的估算精度, R^2 提高了 34.55%。三种机器学习算法模型中, ELM 表现最优 ($R^2 = 0.884$, RMSE = 3.071 mS·cm⁻¹, RPIQ = 2.535)。本研究中所构建的光谱指数在兼顾遥感机理的同时能深度挖掘更多的隐含信息, 并且基于机器学习算法的土壤 EC 估算模型精度显著提高, 为干旱区土壤盐分定量估算提供了科学参考。

关键词: 光谱; 土壤电导率; 光谱预处理; 光谱指数; 机器学习

中图分类号: O433.4 文献标志码: A

Estimation of Soil Electrical Conductivity Based on Spectral Index and Machine Learning Algorithm

CAO Xiaoyi^{1, 2, 3}, DING Jianli^{1, 2, 3†}, GE Xiangyu^{1, 2, 3}, WANG Jingzhe^{1, 2, 3}

(1. College of Resources and Environmental Science, Xinjiang University, Urumqi 830046, China; 2. Key Laboratory of Oasis Ecology under Ministry of Education, Xinjiang University, Urumqi 830046, China; 3. Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, Xinjiang University, Urumqi 800046, China)

Abstract: **【Objective】** Soil salinity is an important indicator for evaluation of soil salinization in arid regions. It is of great practical significance to grasp real-time information about salinization to disaster prevention, ecology stabilization and harmonization of human-land relationship in this area. **【Method】** A total of 400 soil samples were collected from the Weigan

* 国家自然科学基金项目 (41771470) 资助 Supported by the National Natural Science Foundation of China (No. 41771470)

† 通讯作者 Corresponding author, E-mail: watarid@xju.edu.cn

作者简介: 曹肖奕 (1994—), 男, 新疆哈密人, 硕士研究生, 主要从事干旱区遥感应用方面的研究。E-mail: yi_0645@163.com

收稿日期: 2019-02-19; 收到修改稿日期: 2019-05-10; 优先数字出版日期 (www.cnki.net): 2019-06-05

River-Kucha River Delta Oasis in the Xinjiang Uygur Autonomous Region of China in October 2017, and prepared, separately, with distilled water into suspensions, 5 : 1 in ratio, from which soil extracts were obtained for analysis of electrical conductivity (EC) and VIS-NIR (visible-near infrared) spectral reflectances in the laboratory. Based on the obtained data and the Monte Carlo cross validation (MCCV), 364 samples were determined to be valid. After the raw spectrum reflectances (R) were pre-processed with differential, absorbance (Abs), continuum removal (CR) and three others, 21 spectral indices were selected and established. Then based on the 21 optimal spectral indices, EC was assessed using the back propagation neural network (BPNN), support vector machine (SVM), and extreme learning machine (ELM), compared to the partial least squares regression (PLSR), and validated with its root mean square error (RMSE), determination coefficient (R^2) and ratio of performance to interquartile range (RPIQ). 【Result】 Results show: the 21 optimal spectral indices based on R and its pretreated data are R_NDSI (R_{2006} , R_{2257}), R_DSI (R_{1882} , R_{2010}), R_RSI (R_{2006} , R_{2257}), R_FD_NDSI (R_{1376} , R_{2142}), R_FD_DSI (R_{1376} , R_{2124}), R_FD_RSI (R_{1913} , R_{2142}), R_SD_NDSI (R_{416} , R_{2470}), R_SD_DSI (R_{894} , R_{1373}), R_SD_RSI (R_{689} , R_{2355}), ABS_NDSI (R_{2005} , R_{2168}), ABS_DSI (R_{2006} , R_{2257}), ABS_RSI (R_{2006} , R_{2168}), ABS_FD_NDSI (R_{876} , R_{2490}), ABS_FD_DSI (R_{1376} , R_{2123}), ABS_FD_RSI (R_{1913} , R_{2142}), ABS_SD_NDSI (R_{1081} , R_{1725}), ABS_SD_DSI (R_{858} , R_{1374}), ABS_SD_RSI (R_{709} , R_{2355}), CR_NDSI (R_{2119} , R_{2261}), CR_DSI (R_{2119} , R_{2261}), and CR_RSI (R_{2119} , R_{2261}), among which R_FD-RSI (R_{1913} , R_{2142}) is the optimal ($r = 0.649$) one. Compared with the PLSR, the machine learning algorithm (MLA) could significantly improve accuracy of the model, with the R^2 increased by 34.55%. Among the three models using the machine learning algorithm, ELM was the best ($R^2 = 0.884$, RMSE = 3.071 mS·cm⁻¹, RPIQ = 2.535). 【Conclusion】 In this study, different spectral pretreatment methods were used to obtain 21 optimal spectral indices. In constructing the spectral indices in this study, besides considering the remote sensing mechanism, it is advisable to explore in depth more implicit information. Compared with the traditional linear model, the MLA-based soil EC estimation model is obviously higher in accuracy. All the findings in this study may serve as a scientific reference for quantitative estimation of soil salinity in arid regions.

Key words: Spectral; Soil electrical conductivity; Spectral pretreatment; Spectral index; Machine learning

土壤盐渍化是干旱半干旱地区土地退化的主要形式之一,是绿洲农业发展的重要限制因子,它不仅影响人类正常的生产生活,同时威胁绿洲的生态安全与稳定^[1-2]。新疆维吾尔自治区位于我国西北干旱区,是重要的耕地储备库与农垦基地,也是国家“一带一路”重大倡议实施的核心区,但该区域长期受到土壤盐渍化灾害的影响,导致土地退化速率加快、农业产量和环境质量不断下降,影响“一带一路”倡议建设布局;加之人口数量激增,现有土地难以满足当地居民的用地需求,盐渍地的复垦势必会成为解决人地矛盾的主要途径。因此,及时掌握盐渍化信息对该区域的灾害防治、生态稳定以及协调人地关系具有极为重要的现实意义。此外,实现干旱区土壤盐分的定量估算和精准监测也已经成为近年来环境遥感领域的研究热点并受到了广大研究学者的关注^[3]。

可见光-近红外(Visible and near infrared, VIS-NIR)光谱数据能够准确反映物质的光谱信息,将 VIS-NIR 光谱数据进行预处理之后可减少背景噪

声的影响,具有环境因子干扰小、光谱特征表现性强等优点,从而在国内外土壤盐分的估算上被广泛应用,成为其定量研究的重要手段^[4-5]。张贤龙等^[6]研究发现一阶微分和一阶对数倒数预处理下,土壤盐分的反演精度较高, R^2 分别为 0.64 和 0.67;朱赞等^[7]根据 6 种光谱预处理发现平滑二阶微分预处理构建的偏最小二乘回归(Partial least squares regression, PLSR)土壤盐分模型效果最好($r=0.825$, RMSE=1.256 g·kg⁻¹);Brunner^[8]通过归一化植被指数对新疆博斯腾湖地区的土壤盐渍化状况进行评价;姚远等^[9]指出基于一阶微分预处理后的光谱数据构建的 SI2 与土壤电导率(Electrical conductivity, EC)所建立的土壤盐渍化监测模型效果最佳($R^2=0.831$);Nawar 等^[10]利用土壤反射光谱与土壤 EC 对干旱区土壤盐分进行了预测,结果表明基于多元自适应回归样条(Multivariate adaptive regression splines, MARS)构建的预测模型效果最好(剩余预测偏差 Residual predictive deviation, RPD > 2.00)。纵观国内外学者研究,利用 VIS-NIR 光谱预测土壤

盐分主要是通过不同光谱预处理或构建光谱指数等方式来实现,但这些方法仅考虑了土壤盐分与不同预处理之间的相互作用,光谱指数的构建也仅基于原始光谱数据,未能深度挖掘土壤盐分与多种预处理方案下不同指数间的有效信息,同时由于干旱区土壤属性多呈偏态分布,植被、水分和土壤系统的相互作用使得土壤盐分与土壤光谱之间存在复杂的非线性问题,一般的线性回归估算模型难以提供综合解决方案^[11]。而不同预处理构建的二维光谱指数能有效地削弱噪声干扰,综合考量多波段间的协同作用,在一定程度上减少单一波段造成的误差;况且机器学习算法可以有效地解决非线性函数逼近问题,进而提升土壤属性 VIS-NIR 模型的定量估算能力。

基于此,本研究以新疆维吾尔自治区渭干河-库车河绿洲为研究靶区,基于室内获取的 VIS-NIR 数据,探索不同光谱预处理下二维光谱指数与土壤 EC 间的关系,并在此基础上利用 BP 神经网络(Back propagation neural network, BPNN)、支持向量机(Support vector machine, SVM)和极限学习机(Extreme learning machine, ELM)三种机器学习算法构建土壤 EC 估算模型,以期提高研究区土壤 EC 的估算精度,为干旱区土壤盐度监测及盐渍化灾害防治提供科学依据与新思路。

1 材料与方 法

1.1 研究区概况

渭干河-库车河三角洲绿洲(简称渭-库绿洲, 82°10'~83°40'E、41°06'~41°40'N)位于新疆维吾尔自治区塔里木盆地中北部,行政范围主要由库车、沙雅与新和 3 个县构成。渭-库绿洲属温带大陆性干旱气候,年平均气温为 11.6℃,年均降水量仅为 52 mm,年均蒸发量约为 1 991~2 864 mm,土壤质地类型以轻壤与砂壤为主,土壤构成物颗粒细、透水性较差。绿洲自然植被以怪柳(*Tamarix*)、盐节木(*Halocnemum strabliaceum*)、盐爪爪(*Kalidium foliatum*)、花花柴(*Karelinia caspica*)等盐生植被为主,人工植被主要有棉花、玉米以及冬小麦^[12]。由于渭-库绿洲降雨量与蒸发量的差异悬殊,使得该地区普遍存在土壤盐渍化现象,植被覆盖不断减少,

生态环境恶化,致使绿洲发展受到严重威胁。

1.2 样品采集与分析

野外样品采集时间为 2017 年 10 月,根据研究区景观特征,选取绿洲、荒漠以及交错带等景观布设 100 个样点进行土壤样品采集(图 1),每个样点采集 4 层(0~10 cm、10~20 cm、20~40 cm、40~60 cm)土样,共 400 个样品用样品袋封装带回实验室。经过自然风干并剔除碎石及植被等其他杂质后,研磨通过 2 mm 孔筛,再将每个样品分为 2 份,分别用来测定土壤 EC 和 VIS-NIR 光谱数据。过筛后的土壤样品按照水土质量比 5:1 的比例提取土壤浸出液,使用 Cond 7310 型土壤测试仪(Wissenschaftlich Technische Werkstätten 公司,德国)测定土壤 EC。

1.3 光谱测定与预处理

土壤光谱反射率的测定使用 ASD FieldSpec 3 型(Analytical Spectral Devices 公司,美国)光谱仪(波段 350~2 500 nm),实验在可控光照条件的暗室内进行。光谱的采样间隔:350~1 000 nm 为 1.4 nm,1 000~2 500 nm 为 2 nm,重采样间隔为 1 nm。将制备好的 400 个土壤样品分别装入直径为 12 cm、深 3 cm 盛样皿中,将表面刮平。测量时使用 50 W 的卤素灯为光源,光源距土壤样品表面 50 cm,天顶角为 15°,探头至待测样品表面距离为 10 cm,每次光谱测定之前均进行白板标定以消除暗电流的影响^[11]。本实验各土壤样品于 8 个方向采集共 24 条光谱曲线,取 24 条光谱曲线的算术平均值作为该土样的原始光谱数据。

1.4 数据处理

将土壤原始光谱数据进行 Savitzky-Golay(S-G)滤波平滑(2 阶多项式+9 点移动平均)处理,选取平滑处理后的数据进行后续研究。为了提高样品代表性以及建模有效性,利用蒙特卡洛交叉验证(Monte Carlo cross validation, MCCV)对样品光谱和土壤 EC 进行验证并剔除异常样本,降低异常值对建模结果的影响^[13],最终从 400 个样品中确定 364 个有效样品。

为了减少背景噪声的干扰,突显光谱曲线的混合特征信息,增强光谱灵敏度,对土壤原始光谱反射率进行光谱预处理,共得到光谱原始反射率(Raw reflectance, R)、反射率的一阶微分(First derivative,

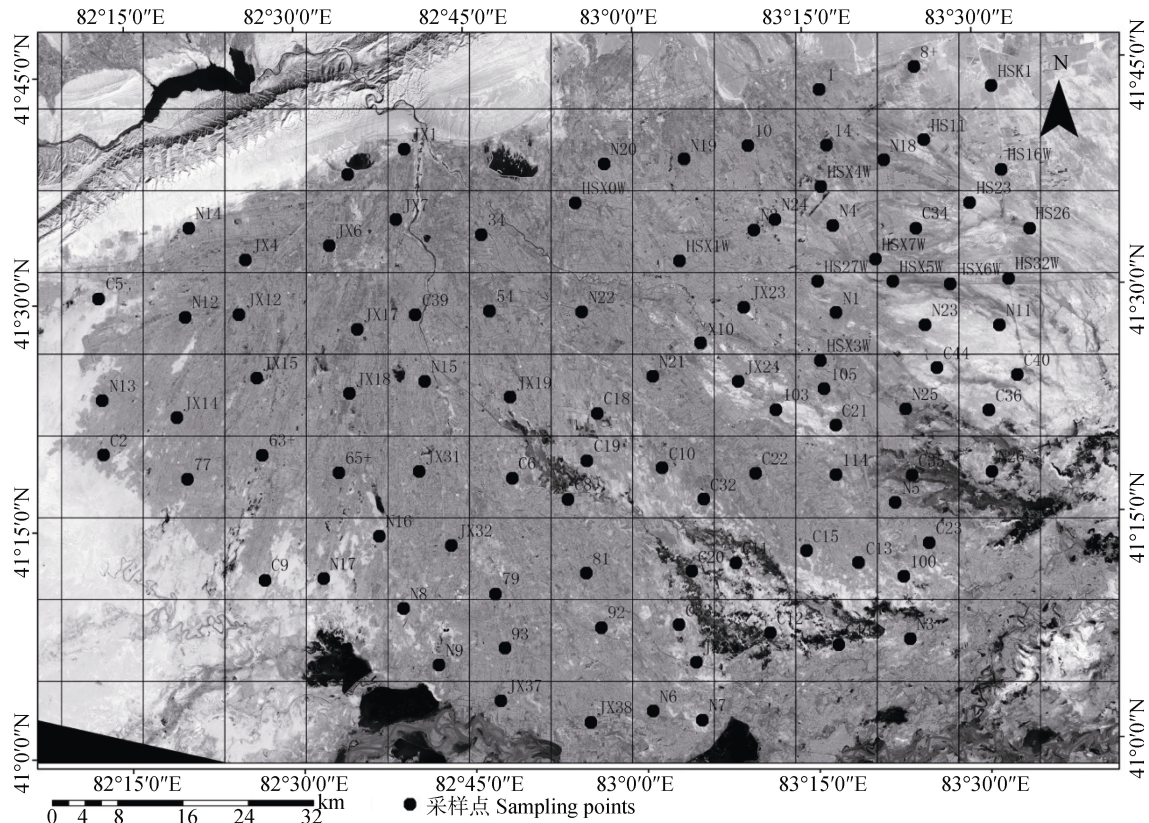


图1 研究区位置及采样点分布图

Fig. 1 Location of the study area and distribution of the sampling points

R_FD) 及二阶微分 (Second derivative, R_SD)、吸光度 (Absorbance, Abs)、吸光度的一阶微分 (Abs_FD) 及二阶微分 (Abs_SD)、连续统去除 (Continuum removal, CR) 7 种光谱形式作为构建光谱指数的光谱数据集。

1.5 光谱指数的构建

为提高光谱信息的利用率与模型监测的准确性, 减少冗余信息, 参照土壤盐分指数^[14-16], 利用土壤光谱反射率构建归一化光谱指数 (Normalized difference spectral index, NDSI)、差值光谱指数 (Difference spectral index, DSI) 及比值光谱指数 (Ratio spectral index, RSI), 具体计算公式如下:

$$\text{NDSI}_{(\lambda_1, \lambda_2)} = \frac{R_{\lambda_1} - R_{\lambda_2}}{R_{\lambda_1} + R_{\lambda_2}} \quad (1)$$

$$\text{DSI}_{(\lambda_1, \lambda_2)} = R_{\lambda_1} - R_{\lambda_2} \quad (2)$$

$$\text{RSI}_{(\lambda_1, \lambda_2)} = \frac{R_{\lambda_1}}{R_{\lambda_2}} \quad (3)$$

式中, R_{λ_1} 、 R_{λ_2} 分别为波长在 λ_1 和 λ_2 处的反射率。

1.6 建模方法与精度检验

本研究拟采用 BPNN^[17]、SVM^[18-20]和 ELM^[21]机器学习算法构建土壤 EC 估算模型, 对比分析三种模型在干旱区对土壤 EC 的估算效果, 其中 SVM 选用鲁棒性高的 RBF 核函数, 惩罚参数设定为 2, ELM 将隐藏层节点数设定为 30。上述 3 种方法均在 Matlab R2014b 软件中完成。

通常样本集划分方法包括随机抽样 RS (Random sampling)、K-S (Kennard-Stone) 和基于联合 x-y 距离 SPXY (Simple set portioning based on joint x-y distance), SPXY 用于考虑基于 K-S 方法的特征空间因子, 能有效覆盖多维向量空间, 进而改善建模精度, 提升模型预测能力。SPXY 详细步骤参考文献[22]。

本研究样本依据 SPXY 进行划分, 选取 273 个样本作为建模集, 91 个样本作为验证集。为了评估基于 BPNN、SVM 和 ELM 的土壤 EC 实测值和预测值建模效果和精度, 本研究选用决定系数

(Coefficient of determination, R^2)、均方根误差(Root mean squared error, RMSE)和四分位数间距(Ratio of performance to interquartile range, RPIQ)三种参数来进行评估。其中 R^2 值与模型精度成正比, 越接近于 1, 代表模型拟合精度越高; RMSE 值与模型精度成反比, 越接近于 0, 代表模型实测值与预测值偏差越小, 预测能力越强; RPIQ 是四分位数间距和 RMSE 的比值, 四分位数间距即样本的 75% 和 25% 出现值之间的差值^[23], 一般认为, $RPIQ < 1.7$ 则模型可信度低, $1.7 \leq RPIQ < 2.2$ 表

明模型有较均衡预测能力, $RPIQ \geq 2.2$ 则视为模型效果极佳^[24]。

2 结 果

2.1 样本土壤电导率统计

表 1 为土壤 EC 的描述性统计特征。本研究全样本的 EC 均值为 $6.73 \text{ mS}\cdot\text{cm}^{-1}$, 建模集与验证集所对应的均值分别为 $6.84 \text{ mS}\cdot\text{cm}^{-1}$ 和 $6.38 \text{ mS}\cdot\text{cm}^{-1}$, 全样本的均值介于建模集与验证集之间。

表 1 土壤样品 EC 统计特征

Table 1 Statistical characteristics of soil electrical conductivity of the soil samples/ ($\text{mS}\cdot\text{cm}^{-1}$)

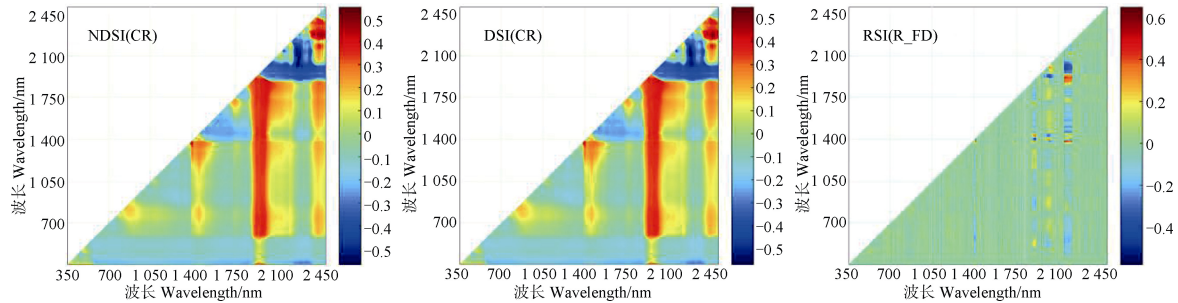
样本类型	样本数	最大值	最小值	均值	标准差
Type of samples	Observations	Maximum	Minimum	Mean	Standard deviation
全集 Whole sets	364	71.60	0.11	6.73	8.90
建模集 Calibration sets	273	71.60	0.11	6.84	8.90
验证集 Validation sets	91	59.10	0.17	6.38	8.94

2.2 最优光谱指数的构建

在 Matlab R2014b 软件中, 计算 7 种光谱形式与土壤 EC 的相关系数 r , 最终得到 7 种光谱形式下的 NDSI、DSI 和 RSI 共计 21 组 VIS-NIR 二维相关系数, 并优选出最佳波段组合。在光谱原始反射率 R 下 NDSI、DSI 和 RSI 的最大相关系数分别为 -0.483 、 0.470 、 -0.482 , 最佳响应波段组合分别为 (R_{2006} , R_{2257})、(R_{1882} , R_{2010})、(R_{2006} , R_{2257}); 在 R_{FD} 下 NDSI、DSI 和 RSI 的最大相关系数分别为 0.516 、 -0.539 、 0.649 , 最佳响应波段组合分别为 (R_{1376} , R_{2142})、(R_{1376} , R_{2124})、(R_{1913} , R_{2142}); 在 R_{SD} 下 NDSI、DSI 和 RSI 的最大相关系数分别为 -0.524 、 0.430 、 -0.428 , 最佳响应波段组合分别为 (R_{416} , R_{2470})、(R_{894} , R_{1373})、(R_{689} , R_{2355}); 在 Abs 下 NDSI、DSI 和 RSI 的最大相关系数分别为 0.502 、 0.488 、 0.504 , 最佳响应波段组合分别为 (R_{2005} , R_{2168})、(R_{2006} , R_{2257})、(R_{2006} , R_{2168}); 在 Abs_{FD} 下 NDSI、DSI 和 RSI 的最大相关系数分别为 -0.539 、 0.488 、 0.642 , 最佳响应波段组合分别为 (R_{876} , R_{2490})、(R_{1376} , R_{2123})、(R_{1913} , R_{2142}); 在 Abs_{SD} 下 NDSI、DSI 和 RSI 的最大相关系数分别为 -0.539 、 -0.412 、 0.431 ,

最佳响应波段组合分别为 (R_{1081} , R_{1725})、(R_{858} , R_{1374})、(R_{709} , R_{2355}); 在 CR 下 NDSI、DSI 和 RSI 的最大相关系数分别为 -0.567 、 -0.568 、 -0.567 , 最佳响应波段组合分别为 (R_{2119} , R_{2261})、(R_{2119} , R_{2261})、(R_{2119} , R_{2261})。由于篇幅限制, 图 2 仅列出 3 种代表性指数。

二维同步相关光谱是一种以光谱强度为两个独立光学变量的函数技术, 将光谱信号扩展至第二维上, 对角线上的强度峰, 相当于光谱强度变化在观察周期内自相关函数值, 其值大小代表光谱强度动态涨落的总强度; 该技术具有提高光谱分辨率、简化包含重叠峰复杂光谱的特点, 通过选择相关的光谱信号研究土壤属性间的相互作用^[25]。本研究为验证光谱指数及光谱组合波段的合理性, 构建了光谱二维同步相关图。如图 3 所示, 光谱在 $350\sim 880 \text{ nm}$ 附近、 $1\ 500\sim 2\ 100 \text{ nm}$ 附近以及 $2\ 200\sim 2\ 450 \text{ nm}$ 附近具有较高的相关性, 而光谱指数所优选的光谱组合波段大都位于上述波段范围内, 丁建丽等^[26]选取 $1\ 942$ 、 $2\ 010 \text{ nm}$ 构建最佳盐度指数, 姚远等^[9]通过光谱变换选择 456 、 533 、 686 、 $1\ 373 \text{ nm}$ 为盐分估算敏感波段, 说明光谱指数的构建以及相对应的光谱组合优选具有合理性。



注：色柱表示土壤 EC 与光谱指数的相关系数 r ，x、y 轴为 350~2500 nm 的波段。深红色与深蓝色代表土壤 EC 与光谱指数高的相关系数。Note: The colorbar illustrates the value of the correlation coefficient (r) between soil EC and spectral indices, and the x-axes and y-axes indicate the wavebands of 350~2500 nm. Dark red and blue portrays a high r between soil EC and the spectral indices.

图 2 土壤 EC 与不同指数的 VIS-NIR 二维矩阵系数图

Fig. 2 VIS-NIR two-dimensional matrix coefficient diagram of soil EC relative to spectral index

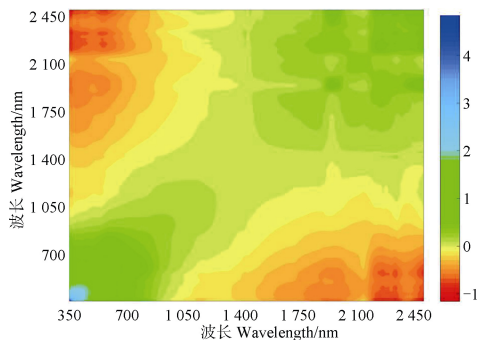


图 3 土壤 EC 光谱二维同步相关图

Fig. 3 Soil EC spectral two-dimensional synchronous correlation diagram

为进一步验证光谱指数对土壤 EC 定量估算的有效性，本研究分析了 21 个光谱指数与土壤 EC 的相关性并尝试构建线性回归模型，结果如表 2 所示。对比分析发现 7 种光谱形式所构建的 21 个光谱指数与土壤 EC 的 $|r|$ 均大于 0.4，且均在 0.01 水平上显著（显著性检验阈值 $P^{**} = \pm 0.128$ ），其中基于 R_FD-RSI (R_{1913} , R_{2142}) 构建的光谱指数与土壤 EC 相关性最强， r 为 0.649。通过分析 21 个线性回归模型的决定系数 R^2 可知，在 R 构建光谱指数的模型中，基于 R-DSI (R_{1882} , R_{2010}) 的线性回归模型预测精度最优， R^2 为 0.314；在不同光谱预处理构建光谱指数模型中，基于 CR-DSI (R_{2119} , R_{2261}) 的线性回归模型预测精度最优， R^2 为 0.423。在同种预处理不同光谱指数构建模型中，基于 DSI 模型的预测精度均最高，仅在 Abs 预处理中 RSI 模型预测效果最佳， R^2

为 0.372。综合分析可知，21 个光谱指数虽能反映与土壤 EC 的相关性，但单一指数对于定量估算土壤 EC 的精度较低，难以解释二者之间的复杂关系，因此考虑采用多指数联合建模来估算土壤 EC。

2.3 土壤 EC 估算模型

将 21 个光谱指数作为土壤 EC 预测模型所需的独立变量，经 MCCV 剔除异常值后的土壤 EC 数据作为响应变量，分别构建土壤 EC 的 BPNN、SVM 和 ELM 三种机器学习算法估算模型，并引入传统线性回归模型 PLSR 进行对比，模型结果、评价指标如表 3 所示。

由表 3 可知，在四种建模方法所构建的土壤 EC 估算模型中，PLSR 模型验证集 $R^2=0.657$ 、 $RMSE=5.487 \text{ mS}\cdot\text{cm}^{-1}$ 、 $RPIQ=1.455$ ，相比于三种机器学习模型估算效果较差，估算能力较弱。在三种机器学习模型中，ELM 模型 RPIQ 达到 2.535 ($RPIQ > 2.2$)，属于极好预测，而 BPNN 模型和 SVM 模型仅能较好地估算土壤 EC ($1.7 < RPIQ < 2.2$)，且 SVM 模型的 RPIQ 大于 BPNN 模型。对比分析三种模型的建模集及预测集的 R^2 、 $RMSE$ 可知，ELM 预测精度最高，SVM 预测精度次之，BPNN 预测精度最低。通过综合评估四种模型的评价参数可知，对土壤 EC 估算的建模效果依次为 $ELM > SVM > BPNN > PLSR$ ，说明 ELM 模型可以精准地定量估算土壤 EC。图 4 为 ELM 模型验证集验证效果。

表 2 不同光谱指数的土壤 EC 回归模型

Table 2 Regression model of soil electrical conductivity relative to spectral index

光谱形式	光谱指数	相关系数	土壤电导率线性模型	决定系数
Spectral forms	Spectral indices	Correlation coefficient	Fitting model of soil electrical conductivity	R^2
R	NDSI (R_{2006} , R_{2257})	-0.483	$y = -0.000\ 2x + 0.011\ 4$	0.285
	DSI (R_{1882} , R_{2010})	0.470	$y = 0.000\ 2x - 0.000\ 5$	0.314
	RSI (R_{2006} , R_{2257})	-0.482	$y = -0.000\ 4x + 1.023$	0.284
R_FD	NDSI (R_{1376} , R_{2257})	0.516	$y = 0.023\ 8x - 0.575\ 4$	0.222
	DSI (R_{1376} , R_{2124})	-0.539	$y = -0.000\ 009x - 0.000\ 2$	0.377
	RSI (R_{1913} , R_{2142})	0.649	$y = 0.191\ 3x - 0.006\ 5$	0.279
R_SD	NDSI (R_{416} , R_{2470})	-0.524	$y = 0.102\ 1x - 2.239\ 3$	0.019
	DSI (R_{894} , R_{1373})	0.430	$y = 0.000\ 001x + 0.000\ 01$	0.138
	RSI (R_{689} , R_{2355})	-0.428	$y = -0.011x + 0.336\ 4$	0.090
Abs	NDSI (R_{2005} , R_{2168})	0.502	$y = 0.000\ 6x + 0.015\ 4$	0.371
	DSI (R_{2006} , R_{2257})	0.488	$y = 0.000\ 2x - 0.009\ 9$	0.282
	RSI (R_{2006} , R_{2168})	0.504	$y = 0.00\ 11x + 1.030\ 3$	0.372
Abs_FD	NDSI (R_{876} , R_{2490})	-0.539	$y = -0.519\ 6x + 1.759\ 3$	0.016
	DSI (R_{1376} , R_{1373})	0.488	$y = 0.000\ 005x + 0.000\ 2$	0.344
	RSI (R_{1913} , R_{2142})	0.642	$y = 0.209\ 9x + 0.001\ 3$	0.272
Abs_SD	NDSI (R_{1081} , R_{1725})	-0.539	$y = 0.033\ 3x + 0.493\ 4$	0.020
	DSI (R_{858} , R_{1374})	-0.412	$y = -0.000\ 000\ 9x + 0.000\ 001$	0.175
	RSI (R_{709} , R_{2355})	0.431	$y = 0.152\ 3x - 0.939\ 7$	0.044
CR	NDSI (R_{2119} , R_{2261})	-0.567	$y = -0.000\ 1x + 0.015\ 5$	0.422
	DSI (R_{2119} , R_{2261})	-0.568	$y = -0.000\ 2x + 0.030\ 3$	0.423
	RSI (R_{2119} , R_{2261})	-0.567	$y = -0.000\ 3x + 1.031\ 4$	0.421

表 3 土壤 EC 估算结果

Table 3 Estimation for soil electrical conductivity

模型	变量个数	建模集		验证集			
		Number of variables	Calibration sets		Validation sets		
			RMSE/ ($\text{mS}\cdot\text{cm}^{-1}$)	R^2	RMSE/ ($\text{mS}\cdot\text{cm}^{-1}$)	R^2	RPIQ
PLSR	21	8.830	0.667	5.487	0.657	1.455	
BPNN	21	5.486	0.637	4.280	0.817	1.793	
SVM	21	5.282	0.650	3.526	0.841	1.948	
ELM	21	5.179	0.666	3.071	0.884	2.535	

3 讨 论

光谱仪测定的光谱数据通常包含样本的有效信息以及光谱叠加所产生的无效信息，而光谱预处理

能有效压缩噪声对目标光谱的影响，同时能将非线性问题转变为线性问题，是对光谱数据进行挖掘及后续定量建模的重要前提^[11]。微分处理可以降低噪声影响的敏感性，分辨重叠峰，提高分辨率及灵敏

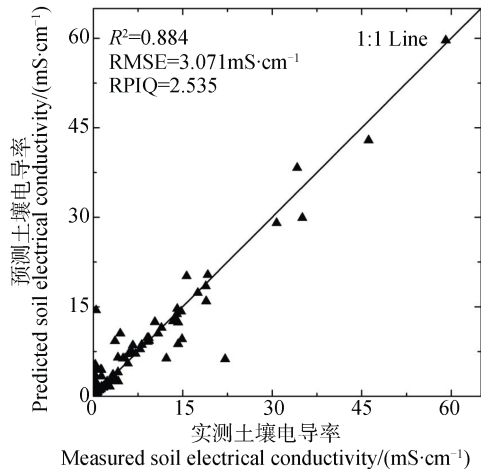


图4 ELM模型土壤EC预测值与实测值

Fig. 4 Relationship between measured and ELM model predicted soil electrical conductivities

度,降低基线效应^[27]。FD、SD有利于限制低频噪声对目标光谱的影响,CR与Abs能够去除土壤中特定物质由于电子跃迁或分子振动引起的特征吸收带之后的吸收特性,突出目标物的光谱特征^[28]。光谱预处理虽能有效降低噪声干扰,突显研究特征,但是一定程度上造成数据的二次冗余,导致模型不能达到最佳精度。光谱指数通过波段的数学运算即可放大波段间的微弱关联、简化模型、去除冗余信息^[29],从而有效提升土壤盐分估算模型的精度。前人研究中^[6-8, 30-31],光谱指数的构建多基于光谱原始反射率或单一预处理方式,张贤龙等^[6]认为基于原始光谱反射率构建DSI、RSI、NDSI三种光谱指数模型效果理想;而Zhang等^[30]仅用分数阶微分处理后的光谱指数构建盐分估算模型,精度得到显著提升。上述研究学者均考虑光谱指数作为建模因子对土壤盐分进行估算,然而,未能考虑到多种预处理方式对光谱指数的作用,模型的构建具有局限性。本研究对土壤光谱进行多种预处理,也通过构建多波段组合的光谱指数作为建模因子,但将多种光谱预处理后的21个光谱指数引入估算模型中,增加了分析建模的可选择性,提高了模型的预测性能。本文中,21个光谱指数与土壤EC均具有良好的相关性, $|r|$ 区间为(0.412~0.649),其中R_FD和CR预处理的效果最优,其NDSI、DSI、RSI的 $|r|$ 均大于0.5,属于最优光谱预处理;而21个光谱指数中,R_FD-RSI(R_{1913} , R_{2142})与土壤EC($r=0.649$)为最优光谱指数。

土壤盐分主要是由碳酸盐、卤化物、硫酸盐和硼酸盐四种主要化学基团组成,它们的共同作用影响着土壤盐分的光谱特性,其中在400、430、450 nm处 Fe^{2+} 、 Fe^{3+} 有较强的光谱响应^[32],在1400 nm处 OH^{-} 起主导作用,在2200、2300 nm主要受Si-OH键、碳酸盐矿物中的 CO_3^{2-} 离子、土壤硅酸盐矿物及其阳离子-OH键的影响^[32-33]。Csillag等^[34]确定了VIS(0.55~0.77 μm)、NIR(0.90~1.30 μm)和MIR(1.94~2.15、2.15~2.30、2.33~2.4 μm)五个光谱范围用于表征不同盐化和碱化的土壤,发现NIR及MIR区域的敏感波段具有识别土壤盐度状况的能力,同时还发现,在干旱条件下,盐在VIS蓝光区域附近具有较高的反射率值。本研究的二维同步相关光谱在350~880 nm附近、1500~2100 nm附近以及2200~2450 nm附近具有强烈的自相关性,说明土壤盐分在这些波段范围内协同性很强,而优选的光谱组合波段大都位于上述范围内,这与Csillag等^[34]、Nawar等^[10]和Bannari等^[35]研究结果相似,进一步佐证了光谱指数的有效性与科学性。本研究不仅从统计学的角度证明了光谱指数法的可行性,并且验证了其内涵的物理意义,结果表明光谱指数能够在兼顾遥感机理的同时,最大程度地削弱噪声的影响,动态提取土壤EC光谱信息。但同时发现本研究中光谱响应范围与上述学者研究发现不完全一致,这可能是由于土壤属性差异和空间异质性导致的。

机器学习具有自主学习的能力,可以很好地解决土壤盐分监测中复杂的非线性函数逼近问题。Khosravi等^[20]利用机器学习算法估算了土壤的重金属含量,结果表明ELM模型具有较PLSR模型更好的效果;葛翔宇等^[36]利用机器学习算法对土壤含水量进行了定量估算,指出ELM模型的精度最高。本研究尝试分别建立21个光谱指数与土壤EC的线性回归模型和PLSR模型,发现线性回归模型估算精度均不理想($R^2=0.657$);而以ELM模型为代表的机器学习算法表现最为突出, R^2 提高了34.55%,较Nawar等^[10]建立的PLSR模型 R^2 也提高了14.81%,因此本研究构建的ELM估算模型效果最佳。作者同上述研究者均得到了机器学习预测能力优于线性模型、ELM为最佳机器学习算法的结果,并认为其在土壤属性定量研究中将拥有广阔的应用前景。

光谱的一阶、二阶微分预处理虽然提高了模型

的光谱差异和精度，但与原始光谱之间仍存在一定差异，可能会忽略信息并限制模型精度的提升，若将分数阶微分引入则会增加更多的光谱差异，提取更多有效信息，从而进一步提升模型精度^[11]。本研究的土壤 EC 估算模型出现验证集 R^2 大于建模集的现象，这可能是由于采样点分布类型多样化（荒漠、沙地、湿地、农田、荒漠-绿洲交错带），土壤的理化性质受自然和人为干扰因素大，土壤盐分分布的随机性高，导致土壤 EC 呈偏态分布等原因造成的。因此，在后续研究工作中一方面将探讨分数阶微分的预处理方案，以提升光谱利用率，提高光谱灵敏度及建模准确性，提升模型泛化能力；另一方面，今后将进一步讨论土地类型、土壤质地等因素对估算效果的影响。同时，还将在扩大样本数量的基础上建立区域土壤光谱库，以期提升 VIS-NIR 在土壤属性估算中的应用潜力并将 VIS-NIR 光谱与星载遥感观测系统相结合，实现干旱区半干旱区土壤盐渍化的协同监测^[36]。

4 结 论

本研究通过实测土壤 EC 及 VIS-NIR 数据，利用 R 和 6 种光谱预处理数据构建了 21 个最优光谱指数，其中 $R_{FD-RSI}(R_{1913}, R_{2142})$ 为最优指数。机器学习算法在土壤 EC 估算模型建模过程中，模型精度和预测能力得到了大幅提高，验证集决定系数 R^2 较线性回归模型提高了 34.55%，实现了对土壤 EC 的精准估算。在 BPNN、SVM 和 ELM 三种机器学习算法估算模型中，根据建模预测精度及预测能力可以得知三种模型的建模效果：ELM > SVM > BPNN，其中 ELM 为土壤 EC 最优估算模型。

参考文献 (References)

- [1] Ding J L, Wang F. Environmental modeling of large-scale soil salinity information in an arid region: A case study of the low and middle altitude alluvial plain north and south of the Tianshan Mountains, Xinjiang[J]. Acta Geographica Sinica, 2017, 72 (1): 64—78. [丁建丽, 王飞. 干旱区大尺度土壤盐度信息环境建模——以新疆天山南北中低海拔冲积平原为例[J]. 地理学报, 2017, 72 (1): 64—78.]
- [2] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote Sensing of Environment, 2003, 85 (1): 1—20.
- [3] Wang F, Ding J L, Wei Y. Analysis of drought characteristics over countries and regions of “the belt and road initiatives” in recent one hundred years[J]. Journal of Geo-Information Science, 2017, 19(11): 1442—1455. [王飞, 丁建丽, 魏阳. “一带一路”国家和地区百年尺度干旱化特征分析[J]. 地球信息科学学报, 2017, 19(11): 1442—1455.]
- [4] Zovko M, Romić D, Colombo C, et al. A geostatistical Vis-NIR spectroscopy index to assess the incipient soil salinization in the Neretva River valley, Croatia[J]. Geoderma, 2018, 332: 60—72.
- [5] Peng J, Ji W J, Ma Z Q, et al. Predicting total dissolved salts and soluble ion concentrations in agricultural soils using portable visible near-infrared and mid-infrared spectrometers[J]. Biosystems Engineering, 2016, 152: 94—103.
- [6] Zhang X L, Zhang F, Zhang H W, et al. Optimization of soil salt inversion model based on spectral transformation from hyperspectral index[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34 (1): 110—117. [张贤龙, 张飞, 张海威, 等. 基于光谱变换的高光谱指数土壤盐分反演模型优选[J]. 农业工程学报, 2018, 34 (1): 110—117.]
- [7] Zhu Y, Shen G R, Xiang Q Q, et al. Spectral characteristics of soil salinity based on different pre-processing methods[J]. Chinese Journal of Soil Science, 2017, 48 (3): 560—568. [朱贇, 申广荣, 项巧巧, 等. 基于不同光谱变换的土壤盐含量光谱特征分析[J]. 土壤通报, 2017, 48 (3): 560—568.]
- [8] Brunner P, Li H T, Kinzelbach W, et al. Generating soil electrical conductivity maps at regional level by integrating measurements on the ground and remote sensing data[J]. International Journal of Remote Sensing, 2007, 28 (15): 3341—3361.
- [9] Yao Y, Ding J L, Zhang F, et al. Research on model of soil salinization monitoring based on hyperspectral index and EM38[J]. Spectroscopy and Spectral Analysis, 2013, 33 (6): 1658—1664. [姚远, 丁建丽, 张芳, 等. 基于高光谱指数和电磁感应技术的区域土壤盐渍化监测模型[J]. 光谱学与光谱分析, 2013, 33 (6): 1658—1664.]
- [10] Nawar S, Buddenbaum H, Hill J. Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: A case study from Egypt[J]. Arabian Journal of Geosciences, 2015, 8 (7): 5127—5140.
- [11] Wang J Z, Ding J L, Abulimiti A, et al. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (VIS-NIR) spectroscopy, Ebinur Lake Wetland, Northwest China[J]. PeerJ, 2018, 6: e4703. <https://doi.org/10.7717/peerj.4703>.
- [12] He B Z, Ding J L, Wang F, et al. Research on data mining

- of salinization information based on phenological characters <https://doi.org/10.7717/peerj.4703>. *Acta Ecologica Sinica*, 2017, 37 (9): 3133—3148. [何宝忠, 丁建丽, 王飞, 等. 基于物候特征的盐渍化信息数据挖掘研究 <https://doi.org/10.7717/peerj.4703>. *生态学报*, 2017, 37 (9): 3133—3148.]
- [13] Vohland M, Ludwig M, Thiele-Bruhn S, et al. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection[J]. *Geoderma*, 2014, 223/225 (1): 88—96.
- [14] Allbed A, Kumar L, Sinha P. Soil salinity and vegetation cover change detection from multi-temporal remotely sensed imagery in Al Hassa Oasis in Saudi Arabia[J]. *Geocarto International*, 2018, 33 (8): 830—846.
- [15] Li Z, Zhang F, Feng H K, et al. Research on the estimation of salt ions of vegetation leaves based on band combination[J]. *Acta Optica Sinica*, 2017, 37 (11): 325—339. [李哲, 张飞, 冯海宽, 等. 基于波段组合的植被叶片盐离子估算研究[J]. *光学学报*, 2017, 37 (11): 325—339.]
- [16] Peng J, Chi C M, Xiang H Y, et al. Inversion of soil salt content based on continuum-removal method[J]. *Acta Pedologica Sinica*, 2014, 51 (3): 459—469. [彭杰, 迟春明, 向红英, 等. 基于连续统去除法的土壤盐分含量反演研究[J]. *土壤学报*, 2014, 51 (3): 459—469.]
- [17] Feng L, Zhu S S, Lin F C, et al. Detection of oil chestnuts infected by blue mold using near-infrared hyperspectral imaging combined with artificial neural networks[J]. *Sensors*, 2018, 18 (6): 1944.
- [18] Faris H, Hassonah M A, Al-Zoubi A M, et al. A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture[J]. *Neural Computing and Applications*, 2018, 30 (8): 2355—2369.
- [19] Hong Y S, Chen Y Y, Yu L, et al. Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS-NIR spectroscopy[J]. *Remote Sensing*, 2018, 10 (3): 479.
- [20] Khosravi V, Doulati Ardejani F, Yousefi S, et al. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods[J]. *Geoderma*, 2018, 318: 29—41.
- [21] Ge X, Wang J, Ding J, et al. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring[J]. *PeerJ*, 2019, 7: e6926.
- [22] Galvão R K, Araujo M C, José G E, et al. A method for calibration and validation subset partitioning[J]. *Talanta*, 2005, 67 (4): 736—740.
- [23] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy[J]. *Trends in Analytical Chemistry*, 2010, 29 (9): 1073—1081.
- [24] Chabrilat S, Foerster S, Steinberg A, et al. Prediction of common surface soil properties using airborne and simulated EnMAP hyperspectral images: Impact of soil algorithm and sensor characteristic[C]// *IEEE Geoscience and Remote Sensing Symposium*. IEEE, 2014: 2914—2917.
- [25] Guo Y L. Two-dimensional infrared and near infrared correlation spectroscopic study of the hydrophilic polymers[D]. Shanghai: Fudan University, 2008. [郭宜鲁. 二维红外及近红外相关光谱对亲水性高分子的研究[D]. 上海: 复旦大学, 2008.]
- [26] Ding J L, Wu M C, Liu H X, et al. Study on the soil salinization monitoring based on synthetical hyperspectral index[J]. *Spectroscopy and Spectral Analysis*, 2012, 32 (7): 1918—1922. [丁建丽, 伍漫春, 刘海霞, 等. 基于综合高光谱指数的区域土壤盐渍化监测研究[J]. *光谱学与光谱分析*, 2012, 32 (7): 1918—1922.]
- [27] Qiao X X, Wang C, Feng M C, et al. Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques[J]. *Spectroscopy Letters*, 2017, 50 (3): 156—163.
- [28] Aldabaa A A A, Weindorf D C, Chakraborty S, et al. Combination of proximal and remote sensing methods for rapid soil salinity quantification[J]. *Geoderma*, 2015, 239/240: 34—46.
- [29] Wang J Z, Ding J L, Ma X K, et al. Detection of soil moisture content based on UAV-derived hyperspectral imagery and spectral index in oasis cropland[J]. *Transactions of the Chinese Society of Agricultural Machinery*, 2018, 49 (11): 164—172. [王敬哲, 丁建丽, 马轩凯, 等. 基于光谱指数的绿洲农田土壤含水率无人机高光谱检测[J]. *农业机械学报*, 2018, 49 (11): 164—172.]
- [30] Zhang F, Wang X P, Kung H T, et al. Estimating soil salt content using fractional derivatives and optional spectral indices in the Ebinur Lake Oasis, Northwestern China[J]. *Data-Enabled Discovery and Applications*, 2018, 2 (1): 6.
- [31] Hong Y S, Yu L, Chen Y Y, et al. Prediction of soil organic matter by VIS-NIR spectroscopy using normalized soil moisture index as a proxy of soil moisture[J]. *Remote Sensing*, 2017, 10 (2): 28.
- [32] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. *Remote Sensing of Environment*, 2003, 85 (1): 1—20.
- [33] Peng J, Biswas A, Jiang Q S, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. *Geoderma*, 2019, 337: 1309—1319.

- [34] Csillag F, Pásztor 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.
- [35] Bannari A, El-Battay A, Bannari R, et al. Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape[J]. *Remote Sensing*, 2018, 10 (6): 855.
- [36] Ge X Y, Ding J L, Wang J Z, et al. Estimation of soil moisture content based on competitive adaptive reweighted sampling algorithm coupled with machine learning[J]. *Acta Optica Sinica*, 2018, 38(10): 393—400. [葛翔宇, 丁建丽, 王敬哲, 等. 基于竞争适应重加权采样算法耦合机器学习的土壤含水量估算[J]. *光学学报*, 2018, 38 (10): 393—400.]

(责任编辑：檀满枝)