



基于集成学习的砂姜黑土含水量高光谱反演研究

王志刚, 黄子琪, 贺成龙, 蔡太义, 冯玉庆, 陆宁静, 窦焕衡

引用本文:

王志刚, 黄子琪, 贺成龙, 蔡太义, 冯玉庆, 陆宁静, 窦焕衡. 基于集成学习的砂姜黑土含水量高光谱反演研究[J]. *农业资源与环境学报*, 2023, 40(6): 1426–1434.

在线阅读 View online: <https://doi.org/10.13254/j.jare.2023.0186>

您可能感兴趣的其他文章

Articles you may be interested in

基于高光谱数据的北疆绿洲农田灰漠土有机质反演

单海斌, 蒋平安, 颜安, 朱磊, 郭星

农业资源与环境学报. 2018, 35(3): 276–282 <https://doi.org/10.13254/j.jare.2017.0256>

基于高光谱的小麦旗叶净光合速率的遥感反演模型的比较研究

吕玮, 李玉环, 毛伟兵, 宫雪, 陈士更

农业资源与环境学报. 2017, 34(6): 582–586 <https://doi.org/10.13254/j.jare.2017.0173>

基于高光谱数据的滴灌甜菜叶绿素含量估算

李宗飞, 苏继霞, 费聪, 李阳阳, 刘宁宁, 樊华, 陈兵

农业资源与环境学报. 2020, 37(5): 761–769 <https://doi.org/10.13254/j.jare.2019.0388>

烤烟叶片镉含量高光谱预测模型的构建

陈楠, 冯慧琳, 杨艳东, 陈萍, 任天宝, 贾方方, 刘国顺

农业资源与环境学报. 2021, 38(4): 570–575 <https://doi.org/10.13254/j.jare.2020.0471>

生物质炭对华北平原4种典型土壤冬小麦生育前期氨挥发的影响

张水清, 张博, 岳克, 孙贇, 张秀玲, 林杉, 黄绍敏

农业资源与环境学报. 2021, 38(1): 127–134 <https://doi.org/10.13254/j.jare.2020.0005>



关注微信公众号, 获得更多资讯信息

王志刚, 黄子琪, 贺成龙, 等. 基于集成学习的砂姜黑土含水量高光谱反演研究[J]. 农业资源与环境学报, 2023, 40(6): 1426-1434.

WANG Z G, HUANG Z Q, HE C L, et al. Hyperspectral inversion study of Vertisol soil moisture content based on ensemble learning[J]. *Journal of Agricultural Resources and Environment*, 2023, 40(6): 1426-1434.

基于集成学习的砂姜黑土含水量高光谱反演研究

王志刚¹, 黄子琪², 贺成龙¹, 蔡太义^{1*}, 冯玉庆¹, 陆宁静¹, 窦焕衡¹

(1. 河南理工大学测绘与国土信息工程学院, 河南 焦作 454000; 2. 哈尔滨师范大学地理科学学院, 哈尔滨 150025)

摘要:为提高砂姜黑土土壤水分的估测精度,本研究以河南省西平县砂姜黑土为研究对象,通过配制不同含水率土壤样本并在室内进行高光谱测量,对土壤样本高光谱数据平滑(SR)、倒对数[LOG(1/R)]、一阶微分(FD)、多元散射校正(MSC)、去包络线(CR)光谱变换后,结合连续投影算法(SPA)识别最佳特征波段,采用偏最小二乘回归(PLSR)、支持向量机回归(SVR)的机器学习方法和堆叠(Stacking)集成学习方法分别构建土壤含水率反演模型。结果表明:经MSC变换后光谱中土壤含水率相关信息增强最多;SPA算法能对砂姜黑土含水率光谱数据进行降维和特征信息提取;经反射光谱MSC变换后由PLSR和SVR集成的Stacking集成模型决定系数最高($R^2=0.963$)、均方根误差最小($RMSE=1.7$)。研究表明,Stacking集成学习模型有效提升了模型的精度和泛化能力,是砂姜黑土含水率最佳反演模型。

关键词:土壤含水率;高光谱;砂姜黑土;堆叠集成;偏最小二乘回归;支持向量机回归

中图分类号:S152.7

文献标志码:A

文章编号:2095-6819(2023)06-1426-09

doi: 10.13254/j.jare.2023.0186

Hyperspectral inversion study of Vertisol soil moisture content based on ensemble learning

WANG Zhigang¹, HUANG Ziqi², HE Chenglong¹, CAI Taiyi^{1*}, FENG Yuqing¹, LU Ningjing¹, DOU Huanheng¹

(1. School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; 2. College of Geographical Science, Harbin Normal University, Harbin 150025, China)

Abstract: To improve the accuracy of soil moisture estimation in Vertisols, this study took the Vertisol in Xiping County, Henan Province, China, as its research object and conducted hyperspectral measurement in the laboratory by configuring soil samples with different moisture contents after implementing smoothing (SR), logarithm of the inverse [LOG(1/R)], first-order differentiation (FD), multiple scattering correction (MSC), and continuum removal (CR) spectral transformation processes on the soil sample hyperspectral data. The best feature bands were identified by combining the successive projection algorithm (SPA) with the machine learning methods of partial least squares regression (PLSR) and support vector machine regression (SVR) and stacking (Stacking) integrated learning methods were used to construct the soil water content inversion model. The results showed that the information related to soil water content was most enhanced in the MSC-transformed spectra. The SPA algorithm was able to downscale and extract feature information from the water content spectral data of the Vertisol. The Stacking integrated model, which integrated PLSR and SVR after MSC transformation based on the reflection spectra, had the highest coefficient of determination ($R^2=0.963$) and the lowest root mean square error ($RMSE=1.7$). This study indicates that the Stacking integrated learning model is the best inversion model for Vertisol moisture content. It effectively improves the accuracy and generalization ability of the model.

Keywords: soil moisture content; hyperspectral; Vertisol; stacking ensemble; partial least squares regression; support vector machine regression

收稿日期: 2023-03-23 录用日期: 2023-08-24

作者简介: 王志刚(1994—)男,河南夏邑人,硕士研究生,研究方向为土地资源管理。E-mail: wzg@home.hpu.edu.cn

*通信作者: 蔡太义 E-mail: caity2008@hpu.edu.cn

基金项目: 河南省高校人文社会科学一般项目(2024ZZJH147); 国家自然科学基金项目(41671225); 河南省重大科技专项(181100110400)

Project supported: General Project of Humanities and Social Sciences in Colleges and Universities of Henan Province(2024ZZJH147); The National Natural Science Foundation of China(41671225); The Key Science and Technology Project of Henan Province, China(181100110400)

土壤含水量(SMC)是农业、水文和生态等领域中的关键指标^[1],实时、快速测定土壤水分对作物增产和粮食安全具有重要意义^[2]。传统土壤水分测定方法检测周期长、范围小,且具有破坏性,这严重限制了对农田水分的精准管理^[3]。

高光谱遥感因具有分辨率高、预测精准,且能快速、实时和大面积监测等优势^[4],受到了国内外学者的广泛关注。Liu等^[5]的研究探明,土壤含水率低于 $0.4\text{ g}\cdot\text{cm}^{-3}$ 时与高光谱反射率呈负相关,高于 $0.4\text{ g}\cdot\text{cm}^{-3}$ 时则呈正相关。吴代晖等^[6]综合评价了土壤含水率和反射率的两种机理模型(物理模型和统计模型),认为后者精度更高。刘伟东等^[7]比较了5种光谱变换方法反演土壤表面湿度的能力,认为一阶微分与差分方法对土壤水分的预测能力最强。姚艳敏等^[8]的研究表明,黑土反射率对数一阶微分建立的含水率预测方程的模拟效果最好,决定系数(R^2)可达0.931。贾学勤等^[9]的研究表明,褐土基于连续投影算法的多元线性回归模型 R^2 可达0.957;张磊^[10]建立了喀斯特地区耕地土壤多元散射校正-主成分分析的神经网络(MSC-PCA-BPNN)模型,显著提高了模型预测能力。张通等^[11]对大豆的相关研究表明,与传统反演算法支持向量机回归(Support Vector Regression, SVR)相比,集成学习(Ensemble Learning)的自适应提升算法(Adaptive Boost, AdaBoost)能提高模型的准确性,可将决定系数 R^2 提高至0.982。Fu等^[12]基于决策树的集成机器学习算法(XGBoost)对模型进行校正后可使 R^2 由0.59显著提升至0.86。Tao等^[13]研究发现,与单个机器学习算法相比,堆叠(Stacking)集成模型可以更

准确和稳定地检测葡萄地土壤含水率。综上,集成学习能够整合多个学习方法的结果进而增强原方法的泛化能力,较单一模型估测精度更高。

砂姜黑土是黄淮海地区典型的中低产土壤(面积约400万 hm^2),恶劣的物理结构是其低产诱因^[14],而监测和调控其土壤含水率则是砂姜黑土结构改良的关键。然而,利用集成学习预测砂姜黑土土壤含水量的研究明显不足。本研究以黄淮海地区砂姜黑土为研究对象进行室内高光谱测量,利用集成学习优化和改进砂姜黑土土壤含水率反演算法,进而提高砂姜黑土土壤水分估算精度,为砂姜黑土区结构精准改良提供理论参考。

1 材料与方 法

1.1 土壤样品采集与处理

河南驻马店西平县(图1)整体地势西高东低,西部为山区,中部为丘岗区,东部为平原区。在对西平县土壤类型分布、地形地貌、空间特征分析后,本研究选择东部平原砂姜黑土区作为研究对象,采样时间为2022年6月,采土样时地表主要作物处于收获期,土样采集数量20个,采集深度0~30 cm。

将不同采集点土壤样品风干后,剔出土壤中植物残体、石砾等侵入体。为探究研究区内砂姜黑土含水率反演模型,消除不同采集点砂姜黑土组成成分之间的差异,将各采样点土壤样品等比例充分混合,将混合后的样品研磨均匀,过2 mm孔径筛,均匀混合后分别装入密封袋备用。设置8个含水率梯度组,每组内有4个重复样,共32份样品,每份样品为30 g,分别放

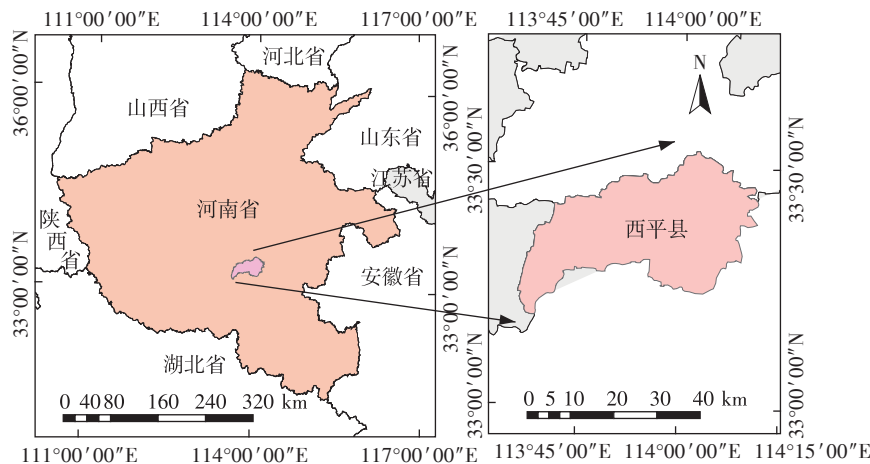


图1 研究区位置示意图

Figure 1 Location diagram of the study area

置在半径 5 cm、深 4.5 cm 的铝盒内,标明组别序号,向 1~8 组分别喷水 1.6、3.2、4.8、6.4、8.0、9.6、11.2 mL 和 12.8 mL。在放置 24 h 土壤充分浸润、吸收表面自由水后,对土壤样本进行光谱测量,得到不同含水率条件下的光谱曲线。测量完成后,立刻采用烘干法测定样本的土壤含水率。

1.2 土壤反射光谱测量及光谱预处理

采用美国 ASD Filed Spec 4 Wide-Res 便携式光谱仪(光谱范围为 350~2 500 nm)进行室内土壤反射光谱测量。将土壤样品表面刮平,采用 Hi-Brite 探头补充光源及光谱测量,探头放置于样品上方 5 cm 处。利用白板进行标准化校正,将土壤样品 90°转动 3 次,在 4 个方向重复采集 5 次,共得到 20 条光谱曲线,取平均值作为原始光谱反射率,以降低土壤样品光谱受各向异性因素的影响。所得土壤光谱反射率在波段 350 nm 和 2 500 nm 处受噪音影响波动明显异常,因此在经过光谱仪配套数据处理软件 ViewSpec Pro 预处理后,选取 400~2 400 nm 范围内的光谱用于分析^[15]。

在 Matlab 软件中对光谱数据进行高斯加权移动平均法平滑,窗口宽度设置为 9,以去除光谱中细微噪声,从而较大程度上保留光谱变化的原始特征。光谱平滑后很大程度上保留着原始光谱曲线的信息,因此本研究中以平滑后的光谱反射率数据作为原始数据进行计算。

1.3 光谱变换与特征波段选择

光谱变换通过对原始光谱进行数学变换来削弱噪声和无用高光谱信息并增强特征信息。本研究中除了研究平滑后的反射率光谱(Smoothed reflectance, SR)外,还采用了 4 种光谱数学变化方法:反射率倒数的对数变换[LOG(1/R)]^[16]、反射率的一阶微分变换(First-order differentiation, FD)^[10]、多元散射校正变换(Multivariate scatter correction, MSC)^[17]和去包络线变换(Continuum removal, CR)^[18]。

在特征波段选择过程中,土壤样品反射率曲线首先经过不同形式数学变换获得 5 种光谱曲线,再对 5 种光谱曲线每个波长的反射率与土壤含水率进行皮尔逊相关性分析以及显著性检验,通过 $P < 0.01$ 水平显著相关的波长为初步筛选的特征波段。通过连续投影算法(Successive projections algorithm, SPA)^[19]将初步筛选的特征波段进一步筛选为最佳特征波段。以上计算在 Matlab 软件中进行。

1.4 模型构建与精度检验

首先构建偏最小二乘法回归(Partial least

squares regression, PLSR)^[20]模型和支持向量机回归(Support vector machine regression, SVR)^[21]模型,并以 PLSR 和 SVR 作为基模型建立了堆叠集成(Stacking ensemble)学习模型^[22],通过交叉验证^[23]计算出上述构建模型的均方根误差(RMSE)、决定系数(R^2)和相对分析误差(RPD)^[24](表 1),对模型精度以及泛化能力进行评估。RMSE 越小,模型泛化能力越好; R^2 越接近 1,模型越稳定。RPD 小于 1.4 时模型不具有预测能力,介于 1.4~2.0 时具有一定的预测能力,而大于 2.0 时模型具有极好的预测能力。

具体操作如下:选取 32 份土壤样品并随机分为 4 组,每次以其中 3 组作为训练集构建基于交叉验证的 PLSR 模型和基于交叉验证的 SVR 模型;并以 PLSR 和 SVR 作为基模型,以线性回归模型为元模型将基模型预测结果作为新特征值,训练 Stacking 集成模型。然后以剩余的一组数据对模型进行验证并评价。通过测试发现,本实验样本建模过程中增加到 40 次交叉验证时的评价结果趋于稳定,因此本实验中以 50 次交叉验证得出的 RMSE、 R^2 、RPD 均值作为模型最终精度。以上计算在 Matlab 软件中进行,设置交叉验证折数为 4,PLSR 函数主成分数设置为 5,SVR 模型使用线性核函数,设定核尺度为“auto”。

表 1 精度校验方法公式

Table 1 Precision calibration method formula

公式 Formula	定义 Definition
$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y_{ipre})^2}{\sum_{i=1}^n (Y_i - Y_{iavg})^2}$	决定系数(R^2)可用来评价构建的回归方程的拟合程度及优劣程度, R^2 随变量个数的增加而增大
$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - Y_{ipre})^2}{n}}$	均方根误差(RMSE)用来判断观测值与真实值的偏差大小
$RPD = \sqrt{\frac{\sum_{i=1}^n (Y_i - Y_{iavg})^2}{\sum_{i=1}^n (Y_i - Y_{ipre})^2}}$	相对分析误差(RPD)是验证集标准差与均方根误差的比值

注: Y_i 、 Y_{ipre} 和 Y_{iavg} 分别为样本 i 的真实值、预测值和平均值; n 为样本数。

Note: Y_i , Y_{ipre} and Y_{iavg} are the measured, predicted and mean values of sample i , respectively; n is the number of samples.

2 结果与分析

2.1 土壤含水率与光谱曲线

由表 2 可知,砂姜黑土样品含水率介于 11.14%~44.31% 之间,取同一梯度组的重复样含水率平均值作为该组含水率,并绘制不同含水率下砂姜黑土光谱

反射率曲线(图2)。

由图2可知,不同梯度样品光谱曲线具有相近的形态特征,差异在于主要吸收峰的位置偏移和深度改变,整体上表现为随含水率增加反射率下降,11.14%与16.20%含水率之间反射率曲线间隔较大。在400~500 nm范围内各梯度反射率几乎重叠而无法分辨,500~1 000 nm范围内各梯度间反射率曲线十分接近而难以辨认。这是由于该部分土壤光谱反射率主要受土壤样品质地与组成等因素影响,而在土壤充分混合后样品之间土壤组成差异被排除,不同样品间光谱

曲线趋同,说明含水率变化对该范围内光谱反射率变化贡献不明显。在1 000~2 400 nm近、中红外光波段内,可以明显观察到光谱反射率都随土壤含水率的增加而降低。

2.2 光谱变换

LOG(1/R)光谱曲线拥有与原光谱曲线几乎相反的变化规律,整体的特征信息并未得到明显增强(图3a)。FD光谱曲线经数学变换形态发生极大改变(图3b),反射率在0.005~-0.007 5范围内围绕零值上下

表2 不同梯度组土壤样本平均含水率

Table 2 Average moisture content of soil samples group

组 Group	加水量 Water addition/mL	平均土壤含水率 Average soil moisture/%
1	1.6	11.14
2	3.2	16.20
3	4.8	20.41
4	6.4	25.53
5	8.0	30.36
6	9.6	35.59
7	11.2	40.18
8	12.8	44.31

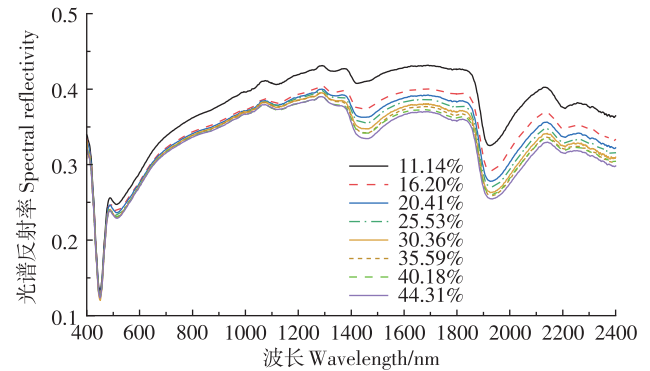


图2 不同含水率条件下的光谱曲线(SR)

Figure 2 Smoothed spectral curves under different moisture content(SR)

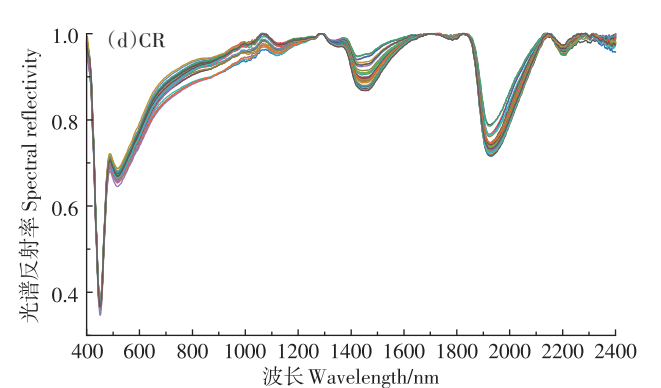
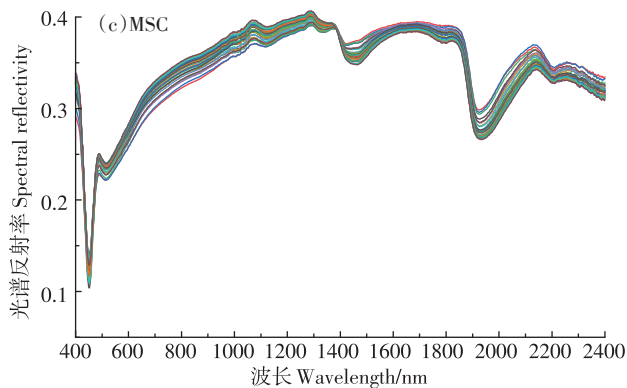
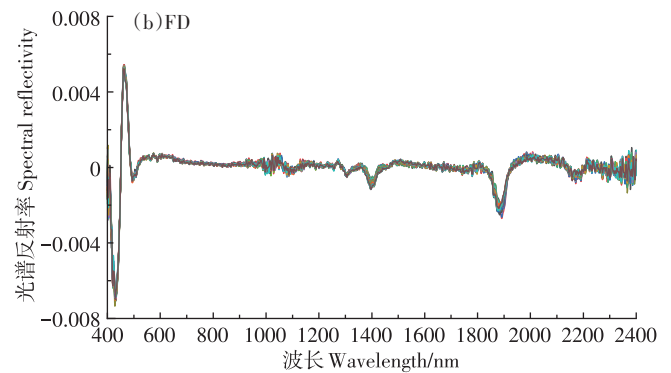
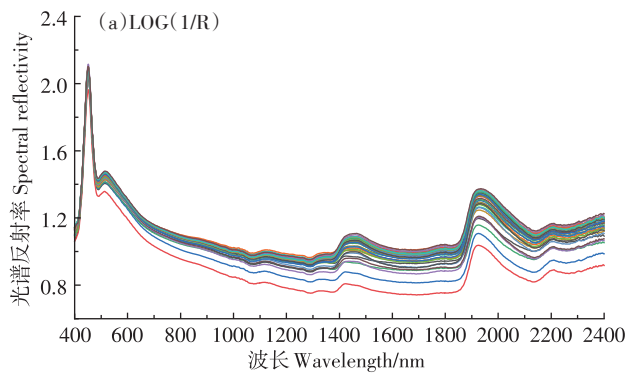


图3 砂姜黑土4种光谱变换曲线

Figure 3 Four spectral transformation curves of Vertisol

波动,其中400~500 nm范围内出现明显起伏峰谷,而1 400、1 900 nm和2 200 nm处的波谷则是常见的水分吸收峰。MSC变换后的光谱曲线与原光谱曲线相比,放大了含水率变化对500~2 400 nm范围内光谱反射率变化的贡献,同时削弱了三个土壤水分吸收峰的差异(图3c)。CR处理的数据较原始数据在400~2 400 nm的波段范围内局部光谱反射率变化得到了增强,主要体现在1 400、1 900、2 200 nm三个水分吸收峰处的差异性(图3d)。

2.3 最佳特征波段选取

2.3.1 相关性分析

SR、LOG(1/R)、FD、MSC和CR五种光谱数据分别与土壤含水率进行皮尔逊相关性分析与显著性分析,获得显著波段信息(表3)及光谱反射率相关系数曲线(图4)。通过相关系数曲线图可以看出SR、LOG(1/R)的

表3 五种光谱反射率与土壤含水率相关系数最大值及对应波长
Table 3 Maximum correlation coefficient between reflectivity and soil moisture content and corresponding bands

光谱变换 Spectral transformation	显著相关波段总数 The total number of significantly correlated bands	最显著波长 The most significant wavelength/nm	相关系数 Correlation coefficient
SR	2 001	423	-0.892
LOG(1/R)	2 001	422	0.894
FD	771	1 381	-0.930
MSC	1 996	1 406	-0.978
CR	1 366	1 385	-0.975

注:表中显著波段指反射率与含水率显著相关且通过 $P<0.01$ 显著性检验的波段。

Note: Significant bands in the table refer to the bands that have a significant correlation between reflectivity and moisture content and pass the $P<0.01$ significance test.

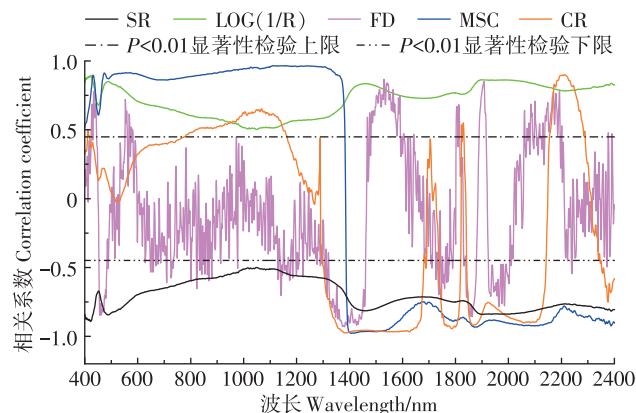


图4 变换光谱反射率相关系数显著性曲线

Figure 4 Correlation coefficient significance curve of transformation spectral reflectance

相关性曲线均较为平滑且平稳,整体相关性都比较高,相关性曲线中最大值分别为-0.892和0.894。砂姜黑土SR光谱反射率相关系数曲线整体呈现负相关,但经MSC变换后在400~1 400 nm呈正相关,1 400~2 400 nm呈负相关,相关系数曲线不再呈单一相关性,经MSC变换后的砂姜黑土光谱反射率与含水率的相关性在1 406 nm波段得到最大值(-0.978),提高了整体相关性。经FD变换后相关系数曲线上起伏频率变快,与土壤含水率整体相关性下降但仍有部分波段相关性较SR有所增强,显著相关波段数量最少,即实现了数据降维。砂姜黑土的CR相关系数曲线频繁上下波动,变化趋势与MSC相关性曲线大致相似。

2.3.2 SPA算法选取最佳特征波段

综合上述五种光谱数据相关系数曲线与显著性检验结果,初步筛选出在 $P<0.01$ 显著水平上显著相关的特征波段集,在Matlab软件中使用SPA算法分别对SR、LOG(1/R)、FD、MSC和CR的特征波段集进行二次筛选,得到最佳特征波段集(表4)。SR与LOG(1/R)曲线形态近似对称,故有近乎一致的最佳特征波段,分布在400、650、1 000~1 100、1 400 nm范围附近,属于可见光及近红外部分;FD曲线呈现频繁波动,在SPA算法筛选后保留了8个波段,但几乎全部集中在480、800~900 nm范围附近;MSC曲线的特征波段分布在400、1 385、1 694 nm和2 193 nm附近,可以看出经过MSC变换后红外光谱部分的信息被放大提取;而CR曲线筛选后的波段集中在800、878、1 270 nm附近,全位于近红外范围内。

此外,在SR光谱曲线中选择400、900、1 400、1 900

表4 砂姜黑土最佳特征波段

Table 4 The best spectral characteristic band for the moisture content of Vertisol

数学变换 Mathematic manipulation	波段数 Number of wavelength	最佳特征波段 Best feature band/nm
SR_SPA	5	400, 408, 650, 1 122, 1 422
NO_SPA	5	400, 900, 1 400, 1 900, 2 400
LOG(1/R)_SPA	5	403, 408, 650, 1 028, 1 405
FD_SPA	8	498, 780, 862, 865, 890, 897, 900, 913
MSC_SPA	5	400, 1 382, 1 387, 1 694, 2 193
CR_SPA	5	796, 802, 878, 1 267, 1 275

注:NO_SPA是SR光谱反射率未经SPA算法而挑选出的波长,用以对比SPA算法对模型精度的提升。下同。

Note:NO_SPA is the wavelength selected by SR spectral reflectance without SPA algorithm screening. It is used to compare the improvement of the model accuracy by the SPA algorithm. The same below.

nm 和 2 400 nm 波长的反射率作为没有经过 SPA 算法筛选的对照波段组(NO_SPA),用来对比验证 SPA 特征选择算法对模型构建的优化。

2.4 土壤含水率反演模型遴选

在每次 4 折交叉验证中,将 32 个样品数据随机分成 4 份,依次以其中 1 份为验证集,其余 3 份为建模集进行 4 次训练,取每次训练的 *RMSE*、*R*²、*RPD* 的均值作为一次交叉验证模型的精度。分别计算三种建模方法在 50 次交叉实验中的 *RMSE*、*R*²、*RPD* 的均值作为该模型的最终评价结果(表 5)。

从表 5 中可以看到没有经过 SPA 算法筛选的波段集 NO_SPA 建模精度最低,*R*² 最高的模型为 SVR 模型(0.846),而经过 SPA 筛选的 SR_SPA 建模精度基模型 *R*² 最高达 0.893,集成模型 *R*² 为 0.917,具有明显提升;从光谱变换的效果分析,除 FD 光谱外,LOG(1/R)、MSC 及 CR 光谱建模效果都很好,单模型和集成模型 *R*² 均在 0.9 以上,均比 SR 模型精度高,而以 FD 为基础的集成模型 *R*² 也在 0.9 以上。以上所有模型的 *RPD* 值均大于 2,说明所有模型都具有较好的预测性,其中以 MSC 变换为基础进行建模效果最好,单模型与集成模型的 *R*² 介于 0.933~0.963 之间,*RPD* 值最高达 6.54。

从不同的模型构建方法来看,PLSR 模型和 SVR 模型效果相近,相同光谱的两种单模型精度各有优劣但差异不大。而除了 NO_SPA 波段集外,以 PLSR 模型和 SVR 模型为基模型进行集成的 Stacking 集成模型的精度均高于两个基模型,说明集成法可有效提升模型精度。

将 5 种光谱最佳特征波段集和 1 个对照波段集,分别输入到对应的 6 个集成模型中得到砂姜黑土含水率预测值,结合含水率实测值绘制预测拟合散点图(图 5)。从图 5 中可以明显看出 FD 集成模型(图 5c)的预测值与实测值的散点更分散,说明该模型稳定性

最低;而 NO_SPA 集成模型(图 5f)的散点拟合直线斜率最小,最偏离直线 $y=x$,说明该模型预测性最差;MSC 集成模型(图 5d)的预测值散点整体上更集中于直线 $y=x$,说明其具有优越的预测能力。

3 讨论

3.1 土壤光谱反射率影响因素剔除

土壤样品光谱反射率受土壤结构、质地、养分等多因素影响。姚艳敏^[8]在研究东北黑土土壤水分高光谱反演时认为,由于只在室内进行测定,土壤属性空间分布的非均质性可以不予考虑。本研究中土壤样品与姚艳敏研究样品类似,同样基于剔除不必要影响因素,凸显土壤含水率对反射率的影响,故将来自不同地块的同类土壤均匀混合^[25-26],以获得相对均质的土壤样品,减少模型的不确定性。本研究模型 *R*² 介于 0.818~0.963 之间,与韩陈等^[25]的研究模型(*R*²=0.86~0.99)接近,均比较稳定。

3.2 土壤含水率模型预测精度

精度是评价模型是否构建成功的关键指标。本研究中集成学习下的模型 *R*² 为 0.963,较集成前 *R*² 提高了 0.022~0.030,相较于姜传礼等^[27]使用的 BPNN 模型 *R*²(0.941)提高了 0.022,也高于张颖帝等^[2]研究砂姜黑土含水率反演的多元线性回归模型 *R*²(0.89)。原因可能有两个方面:一是在制备样品时,姜传礼和张颖帝未考虑不同区域土壤样品组分差异;二是两个研究使用的多元线性回归模型和多元逐步回归模型都是线性或单一模型,无法很好地处理数据的共线性或非线性问题^[28]。本研究中 SPA 算法可有效处理自变量之间的多重共线性^[29],从而提高模型的稳定性和可靠性;SVR 模型可处理非线性关系^[30],故预测性能提高。

3.3 土壤含水率模型泛化能力

泛化能力是指模型对新鲜样本的适应能力^[31]。

表 5 土壤含水率反演模型的精度评价

Table 5 Accuracy evaluation of soil moisture content inversion model

数据处理 Data processing	PLSR			SVR			Stacking 集成模型 Stacking ensemble model		
	<i>R</i> ²	<i>RMSE</i>	<i>RPD</i>	<i>R</i> ²	<i>RMSE</i>	<i>RPD</i>	<i>R</i> ²	<i>RMSE</i>	<i>RPD</i>
SR_SPA	0.893	2.8	3.97	0.885	2.9	3.87	0.917	2.5	4.22
NO_SPA	0.830	3.6	2.99	0.846	3.5	3.15	0.818	3.4	3.04
LOG(1/R)_SPA	0.932	2.2	5.20	0.934	2.2	5.06	0.945	1.9	5.77
FD_SPA	0.843	3.5	3.25	0.834	3.6	3.02	0.901	2.6	4.06
MSC_SPA	0.941	2.0	5.66	0.933	2.1	5.24	0.963	1.7	6.54
CR_SPA	0.914	2.5	4.22	0.904	2.7	4.16	0.928	2.2	4.62

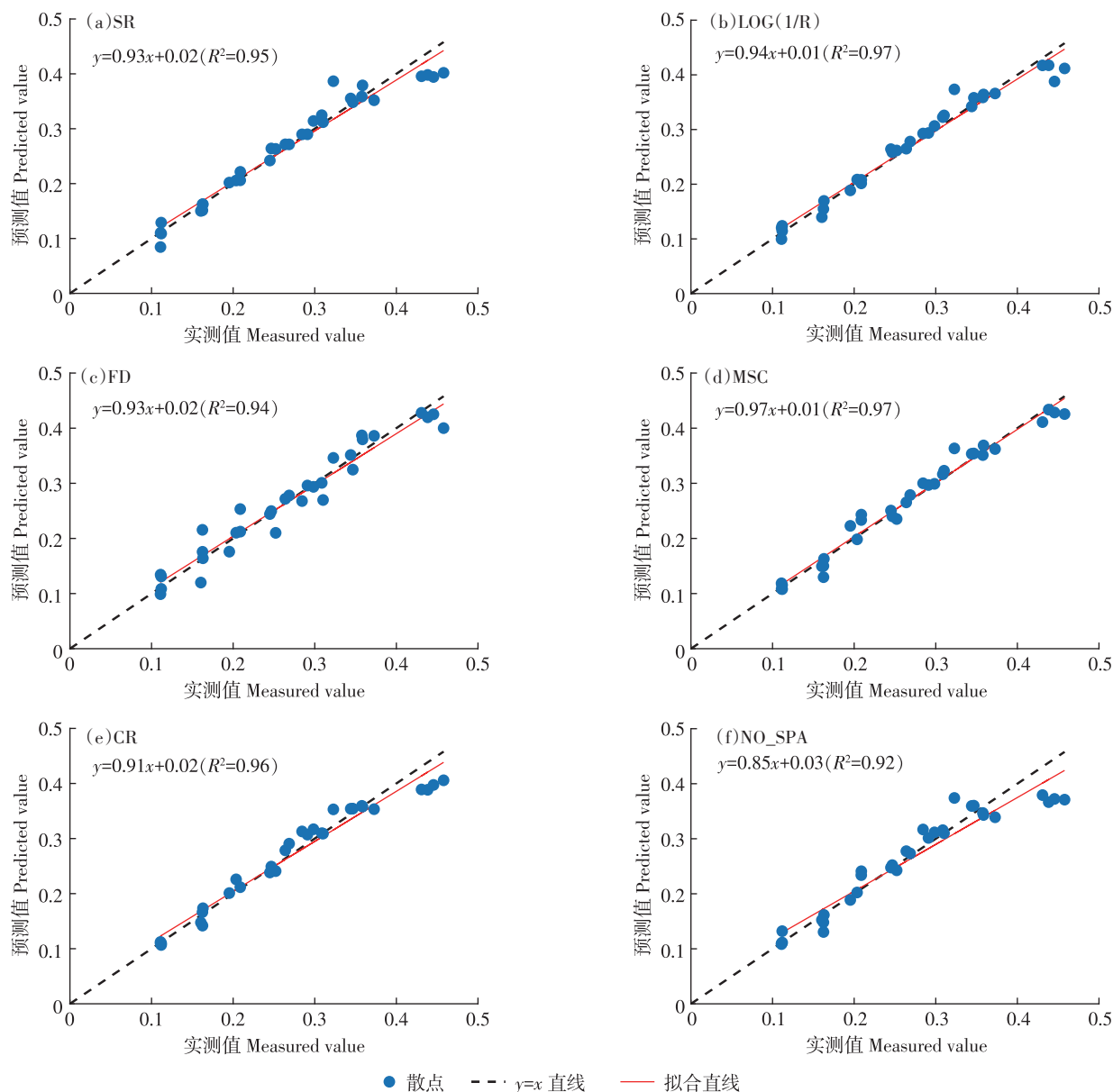


图5 砂姜黑土含水率预测值散点图

Figure 5 The scatter diagram of predicted value of moisture content in Vertisol

本研究发现, Stacking集成模型的RMSE较PLSR和SVR模型降低了0.1~0.9, 表明泛化能力明显增强。增强原因可能是Stacking综合利用了PLSR模型的特征提取能力和SVR模型的非线性建模能力, 深层次原因还需进一步探索研究。

4 结论

(1)多元散射校正(MSC)是增强砂姜黑土土壤含水率信息最多的光谱变换方式。

(2)连续投影算法(SPA)可有效提取砂姜黑土高光谱中含水率特征信息, 并能够实现高维数据降维。

(3)砂姜黑土含水率最佳反演模型是经反射光谱MSC变换后由偏最小二乘回归(PLSR)和支持向量机回归(SVR)集成的Stacking集成模型($R^2=0.963$)。与传统单一算法模型相比, Stacking集成模型提高了模型的精度及泛化能力。

参考文献:

- [1] 葛翔宇, 丁建丽, 王敬哲, 等. 一种基于无人机高光谱影像的土壤墒情检测新方法[J]. 光谱学与光谱分析, 2020, 40(2): 602-609. GE X Y, DING J L, WANG J Z, et al. A new method for predicting soil moisture based on UAV hyperspectral image[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(2): 602-609.

- [2] 张颖帝, 张佳宝, 李晓鹏. 基于高光谱的砂姜黑土含水量反演研究[J]. 土壤, 2017, 49(3):630-634. ZHANG Y D, ZHANG J B, LI X P. Inversion of soil moisture of Shajiang black soil by hyper-spectra[J]. *Soil*, 2017, 49(3):630-634.
- [3] 胡建东, 李林泽, ABDULRAHEEM M I, 等. 介电特性土壤水分测定方法研究进展[J]. 河南农业大学学报, 2021, 55(4):603-611, 638. HU J D, LI L Z, ABDULRAHEEM M I, et al. Research progress of soil water content measurements using dielectric properties[J]. *Journal of Henan Agricultural University*, 2021, 55(4):603-611, 638.
- [4] 费帅鹏, 禹小龙, 兰铭, 等. 基于高光谱遥感和集成学习方法的冬小麦产量估测研究[J]. 中国农业科学, 2021, 54(16):3417-3427. FEI S P, YU X L, LAN M, et al. Research on winter wheat yield estimation based on hyperspectral remote sensing and ensemble learning method[J]. *Scientia Agricultura Sinica*, 2021, 54(16):3417-3427.
- [5] LIU W D, BARET F, GU X F, et al. Relating soil surface moisture to reflectance[J]. *Remote Sensing of Environment*, 2002, 81(2/3):238-246.
- [6] 吴代晖, 范闻捷, 崔要奎, 等. 高光谱遥感监测土壤含水量研究进展[J]. 光谱学与光谱分析, 2010, 30(11):3067-3071. WU D H, FAN W J, CUI Y K, et al. Review of monitoring soil water content using hyperspectral remote sensing[J]. *Spectroscopy and Spectral Analysis*, 2010, 30(11):3067-3071.
- [7] 刘伟东, BARET F, 张兵, 等. 高光谱遥感土壤湿度信息提取研究[J]. 土壤学报, 2004, 41(5):700-706. LIU W D, BARET F, ZHANG B, et al. Extraction of soil moisture information by hyperspectral remote sensing[J]. *Acta Pedologica Sinica*, 2004, 41(5):700-706.
- [8] 姚艳敏, 魏娜, 唐鹏钦, 等. 黑土土壤水分高光谱特征及反演模型[J]. 农业工程学报, 2011, 27(8):95-100. YAO Y M, WEI N, TANG P Q, et al. Hyper-spectral characteristics and modeling of black soil moisture content[J]. *Transactions of the CSAE*, 2011, 27(8):95-100.
- [9] 贾学勤, 冯美臣, 杨武德, 等. 基于SPA-MLR方法的土壤含水量光谱预测模型研究[J]. 干旱地区农业研究, 2018, 36(3):266-269, 291. JIA X Q, FENG M C, YANG W D, et al. Study on the spectral prediction model of soil moisture content based on SPA-MLR method[J]. *Agricultural Research in Arid Areas*, 2018, 36(3):266-269, 291.
- [10] 张磊. 基于高光谱技术的喀斯特地区耕地土壤含水量反演[D]. 贵阳: 贵州大学, 2022:31-39. ZHANG L. Inversion of soil moisture content of cultivated land in karst areas based on hyperspectral technology[D]. Guiyang: Guizhou University, 2022:31-39.
- [11] 张通, 金秀, 饶元, 等. 基于无人机多光谱的大豆旗叶光合作用量子产量反演方法[J]. 农业工程学报, 2022, 38(13):150-157. ZHANG T, JIN X, RAO Y, et al. Inversing photosynthesis quantum yield of the soybean flag leaf using a UAV-carrying multispectral camera[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(13):150-157.
- [12] FU D S, GUEYMARD C A, YANG D Z, et al. Improving aerosol optical depth retrievals from Himawari-8 with ensemble learning enhancement: validation over Asia[J]. *Atmospheric Research*, 2023, 284:106624.
- [13] TAO S Y, ZHANG X, FENG R, et al. Retrieving soil moisture from grape growing areas using multi-feature and stacking-based ensemble learning modeling[J]. *Computers and Electronics in Agriculture*, 2023, 204:107537.
- [14] RUAN R J, ZHANG Z B, WANG Y K, et al. Long-term straw rather than manure additions improved least limiting water range in a Vertisol[J]. *Agricultural Water Management*, 2022, 261:107356.
- [15] 张秋霞, 张合兵, 刘文锴, 等. 高标准基本农田建设区域土壤重金属含量的高光谱反演[J]. 农业工程学报, 2017, 33(12):230-239. ZHANG Q X, ZHANG H B, LIU W K, et al. Inversion of heavy metals content with hyperspectral reflectance in soil of well-facilitated capital farmland construction areas[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(12):230-239.
- [16] 张秀全, 李志伟, 郑德聪, 等. 基于近红外-可见光高光谱的堆叠泛化模型褐土有机质预测[J]. 光谱学与光谱分析, 2023, 43(3):903-910. ZHANG X Q, LI Z W, ZHENG D C, et al. VIS-NIR hyperspectral prediction of soil organic matter based on stacking generalization model[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(3):903-910.
- [17] 付艳华, 刘晶, 毛亚纯, 等. 径向神经网络的苏打盐碱地重金属定量反演[J]. 光谱学与光谱分析, 2022, 42(5):1595-1600. FU Y H, LIU J, MAO Y C, et al. Experimental study on quantitative inversion model of heavy metals in soda saline alkali soil based on RBF neural network[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(5):1595-1600.
- [18] 韩雨, 刘焕军, 张新乐, 等. 基于冠层光谱的水稻穗颈瘟病害程度预测模型[J]. 光谱学与光谱分析, 2021, 41(4):1220-1226. HAN Y, LIU H J, ZHANG X L, et al. Prediction model of rice panicles blast disease degree based on canopy hyperspectral reflectance[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(4):1220-1226.
- [19] 蔡太义, 王志刚, 杨刘帅, 等. 西平县高标准农田Zn含量高光谱反演研究[J]. 农业环境科学学报, 2022, 41(10):2223-2231. CAI T Y, WANG Z G, YANG L S, et al. Hyperspectral inversion model of Zn in high standard farmland soil in Xiping County[J]. *Journal of Agro-Environment Science*, 2022, 41(10):2223-2231.
- [20] 王启元, 赵艳玲, 房铄东, 等. 基于多光谱遥感的裸土土壤含水量反演研究[J]. 矿业科学学报, 2020, 5(6):608-615. WANG Q Y, ZHAO Y L, FANG S D, et al. Inversion of soil moisture in bare soil based on multi-spectral remote sensing[J]. *Journal of Mining Science and Technology*, 2020, 5(6):608-615.
- [21] 许素安, 王家祥, 刘勇. 近红外光谱结合改进鲸鱼算法优化模型BAS-WOA-SVR检测藤椒油掺伪[J]. 光谱学与光谱分析, 2023, 43(2):569-576. XU S A, WANG J X, LIU Y. Detection of adulteration of vine pepper oil by near-infrared spectroscopy combined with improved whale optimization algorithm model BAS-WOA-SVR[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(2):569-576.
- [22] 袁培森, 杨承林, 宋玉红, 等. 基于Stacking集成学习的水稻表型组学实体分类研究[J]. 农业机械学报, 2019, 50(11):144-152. YUAN P S, YANG C L, SONG Y H, et al. Classification of rice phenomics entities based on stacking ensemble learning[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(11):144-152.
- [23] 钟亮, 钱家炜, 储学远, 等. 利用高光谱遥感技术监测小麦土壤重金属污染[J]. 农业工程学报, 2023, 39(5):265-270. ZHONG L, QIAN J W, CHU X Y, et al. Monitoring heavy metal contamination of

- wheat soil using hyperspectral remotesensing technology[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(5): 265-270.
- [24] 郭飞, 许镇, 马宏宏, 等. 基于PCA的土壤Cd含量高光谱反演模型对比研究[J]. *光谱学与光谱分析*, 2021, 41(5): 1625-1630. GUO F, XU Z, MA H H, et al. A comparative study of the hyperspectral inversion models based on the PCA for retrieving the Cd content in the soil[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(5): 1625-1630.
- [25] 韩陈, 唐强, 韦杰. 紫色土和黄壤含水率的室内光谱反演[J]. *水土保持通报*, 2021, 41(5): 174-180, 190. HAN C, TANG Q, WEI J. Estimating soil moisture content of purple soil and yellow soil using laboratory spectral conversion models[J]. *Bulletin of Soil and Water Conservation*, 2021, 41(5): 174-180, 190.
- [26] 杨曦光, 于颖. 基于试验反射光谱数据的土壤含水率遥感反演[J]. *农业工程学报*, 2017, 33(22): 195-199. YANG X G, YU Y. Remote sensing retrieval of soil moisture content based on experimental reflectance spectrum data[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(22): 195-199.
- [27] 姜传礼, 赵健赞, 丁圆圆, 等. SPA算法与机器学习的黄河源土壤水分反演[J]. *光谱学与光谱分析*, 2023, 43(6): 1961-1967. JIANG C L, ZHAO J Y, DING Y Y, et al. Study on soil water retrieval technology of Yellow River source based on SPA algorithm and machine learning[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(6): 1961-1967.
- [28] 刘红卫. 线性回归模型中多重共线性问题的应对策略及其几点改进[D]. 成都: 西南交通大学, 2006: 1-24. LIU H W. Coping strategies for multicollinearity problems in linear regression models and some improvements[D]. Chengdu: Southwest Jiaotong University, 2006: 1-24.
- [29] 黄平捷, 李宇涵, 俞巧君, 等. 基于SPA和多分类SVM的紫外-可见光光谱饮用水有机污染物判别方法研究[J]. *光谱学与光谱分析*, 2020, 40(7): 2267-2272. HUANG P J, LI Y H, YU Q J, et al. Classification of organic contaminants in water distribution systems developed by SPA and multi-classification SVM using UV-Vis spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(7): 2267-2272.
- [30] 承姿辛, 刘玉敏, 邱光琦. 基于可靠性优化的空心半轴轻量化设计[J]. *机械设计*, 2023, 40(1): 34-39. CHENG Z X, LIU Y M, QIU G Q. Lightweight design of hollow half shaft based on reliability optimization[J]. *Journal of Machine Design*, 2023, 40(1): 34-39.
- [31] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 24-36. ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Publishing House, 2016: 24-36.