

doi:10.6041/j.issn.1000-1298.2025.06.027

基于混合蛙跳算法的果园土壤全氮含量高光谱预测

冯上奇^{1,2} 袁全春^{2,3} 黄凯^{2,4} 孙元昊^{2,3} 曾锦^{2,4} 吕晓兰^{1,2}

(1. 南京林业大学机械电子工程学院, 南京 210037; 2. 江苏省农业科学院农业设施与装备研究所, 南京 210014;
3. 农业农村部园艺作物农业装备重点实验室, 南京 210014;
4. 农业农村部南方果园(桃、梨)全程机械化科研基地, 南京 210014)

摘要: 土壤全氮含量是土壤重要的养分指标, 基于高光谱数据研究并构建果园土壤全氮含量预测模型, 为准确检测土壤全氮含量提供新方法。以江苏省农业科学院梨园土壤为研究对象, 利用高光谱成像技术获取土壤光谱反射率数据, 引入混合蛙跳算法和竞争性自适应加权采样进行光谱特征提取, 并分别采用全波段和特征波段构建偏最小二乘回归、支持向量机、随机森林和卷积神经网络模型对土壤全氮含量进行估测。结果表明: 原始光谱经过多种预处理方法处理后, 经 SG 卷积平滑联合标准正态变换预处理, 全波段构建的全氮预测模型表现最佳; 基于混合蛙跳算法提取 10 个关键波段, 占总波段数量的 4.08%, 有效降低了数据维度; 基于混合蛙跳算法提取特征波段构建的卷积神经网络模型表现优异, 此模型测试集决定系数为 0.95、均方根误差为 0.21 g/kg、相对分析误差为 3.97。研究表明应用混合蛙跳算法能高效提取特征波段, 降低数据维度, 并且提高了土壤全氮含量估测精度, 为果园土壤全氮含量准确估测提供参考。

关键词: 果园; 土壤全氮; 预测模型; 高光谱成像技术; 混合蛙跳算法; 卷积神经网络

中图分类号: S153.6 **文献标识码:** A **文章编号:** 1000-1298(2025)06-0277-09

OSID:



Hyperspectral Estimation of Total Nitrogen Content in Orchard Soil Based on Shuffled Frog Leaping Algorithm

FENG Shangqi^{1,2} YUAN Quanchun^{2,3} HUANG Kai^{2,4} SUN Yuanhao^{2,3} ZENG Jin^{2,4} LÜ Xiaolan^{1,2}

(1. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

2. Institute of Agricultural Facilities and Equipment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

3. Key Laboratory of Modern Horticultural Equipment, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China

4. Southern Orchard (Peach, Pear) Fully Mechanized Research Base, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China)

Abstract: Soil total nitrogen is an important nutrient index of soil. The soil of pear orchard of Jiangsu Academy of Agricultural Sciences was taken as the research object, the soil spectral reflectance data were obtained by hyperspectral imaging technology, the shuffled frog leaping algorithm and competitive adaptive reweighted sampling in total nitrogen content in orchards was studied and constructed based on hyperspectral data, which provided a method for accurately detecting soil total nitrogen content. The competitive adaptive reweighted sampling were introduced for spectral feature extraction, and the partial least squares regression, support vector regression, random forest and convolutional neural network models were used to estimate the total nitrogen content of the soil by using the full band and the characteristic band, respectively. The results showed that after the original spectrum was processed by a variety of preprocessing methods, it was found that the total nitrogen prediction model constructed by SG convolution smoothing combined with standard normal transform pretreatment had the best performance. Based on the shuffled frog leaping algorithm, totally ten key bands were extracted, accounting for 4.08% of the total number of bands, which effectively reduced the data dimension. The convolutional neural network model constructed based on the shuffled frog leaping algorithm to extract feature bands performed

收稿日期: 2025-03-08 修回日期: 2025-05-15

基金项目: 国家重点研发计划项目(2022YFD2001400)和国家梨产业技术体系项目(CARS-28)

作者简介: 冯上奇(2000—),男,硕士生,主要从事高光谱检测土壤养分装备研究,E-mail: 13186553418@163.com

通信作者: 吕晓兰(1980—),女,研究员,主要从事果园智能农机装备研究,E-mail: lxlan@126.com

well, and the coefficient of determination of the model test set was 0.95, the root mean square error was 0.21 g/kg, and the relative analysis error was 3.97. The results showed that the shuffled frog leaping algorithm can efficiently extract the feature bands, reduce the data dimension, and improve the estimation accuracy of soil total nitrogen content, which provided a reference for the accurate estimation of soil total nitrogen content in orchards.

Key words: orchard; soil total nitrogen; prediction model; hyperspectral imaging technology; shuffled frog leaping algorithm; convolutional neural networks

0 引言

土壤中的全氮含量是衡量土壤肥力的重要指标之一^[1]。对梨果业来说,土壤氮素的过剩和不足均会对梨的品质和产量产生不利影响,并存在污染环境的风险。因此,快速准确预测果园土壤全氮含量对于果园的精准施肥、土壤信息精细化管理、提高氮肥吸收率及减少环境污染具有重要价值。然而,传统的土壤全氮测定方法实验过程繁琐、耗时长、试剂消耗大,难以满足现代农业对土壤养分快速检测的需求。近年来,高光谱技术因其检测速度快、无污染、土壤信息丰富等优点,逐渐成为土壤全氮含量快速检测的重要手段。目前国内外学者利用高光谱技术估测土壤全氮已有较多研究^[2-4]。这些研究表明,高光谱技术在土壤全氮含量检测方面具有较大的应用潜力。然而,尽管高光谱技术在土壤全氮检测中取得了一些进展,大多数研究模型的预测精度仍然不高,决定系数 R^2 多集中在0.61~0.86之间。这主要归因于特征波段提取效果不显著,导致高光谱数据冗余,从而影响模型的预测性能。因此,筛选具有代表性的光谱特征波段是构建高精度高光谱估测模型的关键步骤。目前,特征波段提取方法包括:基于统计分析的特征提取(连续投影算法^[5]、相关系数分析^[6]、主成分分析^[7]、竞争自适应重加采样^[8]);基于机器学习算法的特征提取(极限梯度提升^[9]、遗传算法^[10]);基于深度学习算法的特征提取(自编码器^[11]、生成对抗网络^[12])等。这些方法在提取有效特征波段方面各具优势,但普遍存在特征提取效果不佳、抗噪能力不足的问题,影响了模型的泛化能力。而混合蛙跳算法在高光谱特征提取中,通过整合遗传算法的全局寻优能力和连续投影算法的局部搜索机制,能有效去除数据冗余和噪声波段,筛选出与土壤全氮高相关的光谱特征,为高光谱土壤全氮含量的高效预测提供了新的解决方案。

本文以江苏省农业科学院梨园土壤为研究对象,采用SG卷积平滑、标准正态变换、多元散射校正、一阶导数、二阶导数5种方法进行预处理,运用混合蛙跳算法和竞争性自适应加权采样两种方法筛选特征波段进行对比,结合偏最小二乘回归、随机森

林、支持向量机、卷积神经网络分别建立果园土壤全氮含量估测模型,以期为果园土壤养分指标测定提供方法参考与技术支撑。

1 实验

1.1 研究区概况

研究区位于江苏省农业科学院,地处于江苏省南京市玄武区(东经118°46',北纬32°03')。研究区占地面积约60 hm²、地势平坦、属亚热带湿润季风气候区。年均日照时长2 000 h,年均气温15.4℃,年均降水量1 100 mm。土壤类型以黄棕壤为主,适宜多种农作物种植与实验。

1.2 土壤样本采集与处理

本研究在江苏省农业科学院的典型梨园区开展土壤采样工作。野外采样时,将每个采样点地表植物及石块清除,在每个单元按照五点取样法的原则采集耕层(0~20 cm)的土壤。保留1 kg土样,分成两部分,一部分用于土壤高光谱数据采集,另一部分用于化学法测定土壤全氮含量。

1.3 土壤全氮测定

土壤全氮含量测定采用凯氏定氮法,表1为土壤样本全氮含量(质量比)统计结果。由表1可知,样本中土壤全氮含量具有较大的离散性,属于中等变异水平。

表1 土壤全氮含量统计

Tab.1 Statistics of soil total nitrogen content

样本类型	样本数	最大值/ (g·kg ⁻¹)	最小值/ (g·kg ⁻¹)	均值/ (g·kg ⁻¹)	标准差/ (g·kg ⁻¹)	变异系数/%
总样本	90	3.91	0.86	1.68	0.93	55.36
训练集	60	3.91	0.86	1.69	0.94	55.62
测试集	30	3.87	0.88	1.65	0.93	56.36

为保证训练集和测试集在全氮含量分布上的均匀性,采用分层抽样法对90个土壤样本进行随机划分,以比例2:1划分训练集和测试集。训练集与测试集在最大值、最小值、平均值、标准差等统计指标上与总样本接近,变异系数分别为55.62%和56.36%,表明训练集与测试集在全氮含量分布上具有一致性。其中,测试集的全氮含量范围(0.88~

3.87 g/kg)完全落在总样本范围(0.86 ~ 3.91 g/kg)之内,说明样本划分合理,具有良好的代表性和分布梯度,有助于后续建模回归分析的稳定性与可靠性。

1.4 高光谱成像系统及数据获取

高光谱成像系统主要由计算机、光谱仪、可移动平台和光源等组成。实验设备有无锡双利合谱科技有限公司生产的 Gaia Sorter-Dual 型双相机全波段高光谱分选仪(全波段光谱范围 350 ~ 2 500 nm)。

采集图像时,将处理好的土壤样品平放在培养皿中并压平压实,放置于可移动平台上,通过操作计算机中的 Spec View 图像采集软件,使移动平台以速度 1 cm/s 带动土壤样本进入暗箱内,在暗箱内采集土壤样品原始高光谱图像。采集完毕后,对采集的原始高光谱图像进行黑白板校正,并计算每个样本区域内所有像素点光谱反射率的平均值,以生成一个平均光谱数据作为此样本的高光谱反射率数据。由于光谱曲线头部和尾部噪声较大,这种噪声会降低后续模型的稳定性与精度,进而影响模型对全氮含量预测的敏感性和泛化能力^[13]。故每个土壤样本只保留 400 ~ 1 000 nm 的光谱数据作为有效光谱数据,共计 245 个波段。图 1 为高光谱样本图像,图 2 为所有样本原始平均光谱曲线。

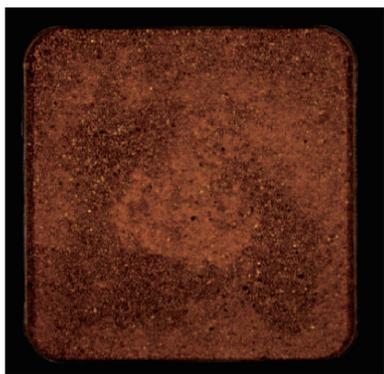


图 1 样本图像

Fig.1 Sample image

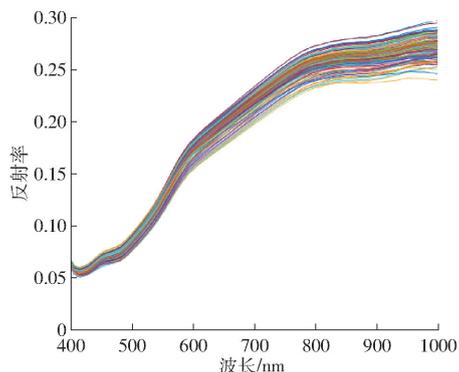


图 2 原始平均光谱曲线

Fig.2 Original average spectral curves

1.5 预处理及特征波长提取

在高光谱数据处理中,原始光谱数据常受到环

境条件、仪器性能、样品状态等多种因素影响,易引入噪声、基线漂移和散射干扰^[14]。这些干扰不仅掩盖了与土壤全氮含量相关的有效光谱特征,还可能导致后续模型不稳定、预测精度下降。因此,在建模分析前对原始光谱数据进行预处理是十分必要的。采用 SG(Savitzky Golay)卷积平滑算法与标准正态变换(Standard normal variate, SNV)、多元散射校正(Multiplicative scatter correction, MSC)、一阶导数(First derivative, FD)、二阶导数(Second derivative, SD)4种预处理方法组合对原始光谱进行预处理,表示为 SG + SNV、SG + MSC、SG + FD、SG + SD。

高光谱数据具有高维度、高冗余的特点,直接利用全波段信息建模不仅会导致计算复杂度上升,还会降低模型泛化能力。因此,选择高效、准确的特征提取方法是建模过程中的关键步骤,对于提升土壤全氮含量预测模型精度具有重要价值^[15]。采用混合蛙跳算法(Shuffled frog leaping algorithm, SFLA)和竞争性自适应加权采样(Competitive adaptive reweighted sampling, CARS)两种算法对光谱数据进行特征波段提取。SFLA是一种基于群体智能的启发式优化算法,结合了局部搜索和全局信息交换的策略。在应用于光谱特征波段选择时,其优点在于具备较强的全局搜索能力和鲁棒性,适用于处理高维、非线性的光谱特征优化问题。CARS其优势在于能有效识别出对模型贡献最大的变量,剔除冗余或共线性强的波段,提升模型泛化能力,尤其适用于多重共线性严重的高光谱数据^[16-17]。

1.6 回归分析

利用偏最小二乘回归(Partial least squares regression, PLSR)、随机森林(Random forest, RF)、支持向量机(Support vector machine, SVM)、卷积神经网络(Convolutional neural network, CNN)4种方法构建土壤全氮含量估测模型。使用高光谱数据构建模型时,分别以原始光谱、经4种组合方法预处理后的光谱全波段和筛选的特征波段为自变量,以土壤全氮含量为因变量,对模型参数进行选择,通过交叉验证,确定最佳模型。

1.7 模型评价指标

采用决定系数 R^2 、均方根误差(RMSE)、相对分析误差(RPD)对预测模型进行评价。决定系数 R^2 越接近 1, RMSE 越接近 0, 说明预测模型的精度越高。RPD 用于评价模型的整体性能,当 RPD 小于 1.4 时,说明模型预测能力较弱,当 RPD 在 1.4 ~ 2.0 之间,说明模型预测能力一般,当 RPD 大于等于 2.0 时,说明模型具备良好预测能力^[19]。模型评价指标具体计算公式参考文献[18]。

2 结果分析

2.1 不同全氮含量土壤高光谱反射率特征

图3为不同全氮含量土壤原始光谱曲线。

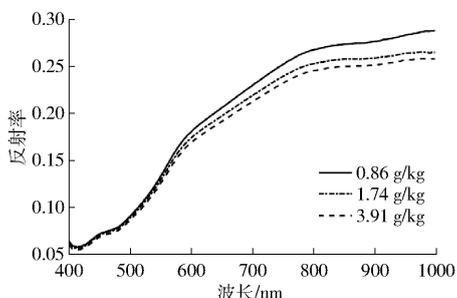


图3 不同全氮含量土壤原始光谱曲线

Fig.3 Original spectral curves of soils with different total nitrogen contents

如图3所示,不同全氮含量土壤平均反射率曲线存在明显差异,在400~600 nm范围内光谱反射率迅速上升,600~1000 nm范围内光谱反射率曲线呈平缓波动变化。此外,400~1000 nm范围内全氮含量越高,平均反射率整体下降得越明显,这种差异为后续模型建立提供了基础。

2.2 基于全波段的全氮含量估测模型建立与验证

原始光谱经4种方法预处理后,利用PLSR、RF、SVM和CNN对土壤全氮含量进行建模预测,训练集和测试集效果见表2。从表2可知,使用不同预处理方法和建模方法建立的土壤全氮含量预测模型精度差异明显。原始光谱经SG+SNV预处理后使用RF建模方法建立的土壤全氮含量估测模型效果明显优于其他模型,模型(SG+SNV-RF)训练集和测试集的 R^2 分别为0.95、0.91、RMSE分别为0.16、0.26 g/kg,RPD分别为4.75、3.49,此模型的估测精度最高且误差最小。整体来看,原始光谱经SG+SNV方法预处理后,使用4种建模方法建立的全氮含量模型精度较原始光谱建模均有提高,故在后续运用特征波段建模分析中选择SG+SNV预处理方法对原始光谱进行预处理。

2.3 土壤全氮含量与原始光谱及SG+SNV预处理后光谱相关性

为探究土壤全氮含量与光谱特征之间的关系,分析了原始反射率光谱与经过SG+SNV预处理后的光谱数据与土壤全氮含量之间的相关性。标准正态变换是一种常用的光谱预处理方法^[20]。主要功能在于校正散射效应,消除样品粒径不均、表面粗糙度等因素对光谱数据带来的干扰,从而增强光谱对目标变量的响应能力。从图4可以看出,原始反射率光谱曲线与土壤全氮含量在整个波段内存在一定的相关性,在400~500 nm之间呈现出较强的负相

表2 土壤全氮含量预测模型建模与验证

Tab.2 Establishment and validation of predictive models for soil total nitrogen content

建模方法	预处理方法	训练集			测试集		
		R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	RPD	R^2	RMSE/ ($\text{g}\cdot\text{kg}^{-1}$)	RPD
PLSR	无	0.85	0.27	2.58	0.68	0.49	1.81
	SG+MSC	0.87	0.24	2.82	0.73	0.45	1.95
	SG+SNV	0.87	0.23	2.86	0.71	0.46	1.90
	SG+FD	0.90	0.29	3.21	0.68	0.49	1.80
	SG+SD	0.92	0.26	3.58	0.42	0.48	1.34
RF	无	0.91	0.20	4.50	0.86	0.38	2.94
	SG+MSC	0.92	0.18	4.23	0.85	0.35	2.68
	SG+SNV	0.95	0.16	4.75	0.91	0.26	3.49
	SG+FD	0.94	0.18	4.73	0.78	0.31	2.98
	SG+SD	0.92	0.25	3.59	0.79	0.31	2.14
SVM	无	0.93	0.19	4.23	0.66	0.45	1.74
	SG+MSC	0.95	0.22	4.71	0.72	0.45	1.96
	SG+SNV	0.94	0.24	4.64	0.72	0.48	1.88
	SG+FD	0.98	0.19	4.74	0.51	0.48	2.31
	SG+SD	0.95	0.20	4.62	0.66	0.46	1.92
CNN	无	0.96	0.22	3.89	0.74	0.49	2.01
	SG+MSC	0.98	0.24	3.26	0.74	0.45	2.22
	SG+SNV	0.97	0.25	3.96	0.76	0.42	2.87
	SG+FD	0.98	0.21	3.81	0.75	0.38	2.67
	SG+SD	0.97	0.23	3.55	0.72	0.36	2.26

关性,相关系数最低,达到约-0.85,随后在可见-近红外波段内逐渐趋于正相关,但整体波动较小,相关性较弱。而在经过SG+SNV联合预处理后光谱数据的相关性变化更加明显。SG+SNV预处理显著提高了某些关键波段与土壤全氮的相关性,在420~460 nm和820~880 nm两个波段内均出现了显著的正相关峰值,最大相关系数接近0.85。所以相比原始光谱,SG+SNV处理不仅增强了信号与土壤氮含量之间的联系,还有效抑制了噪声干扰,提高了光谱曲线的一致性和稳定性。

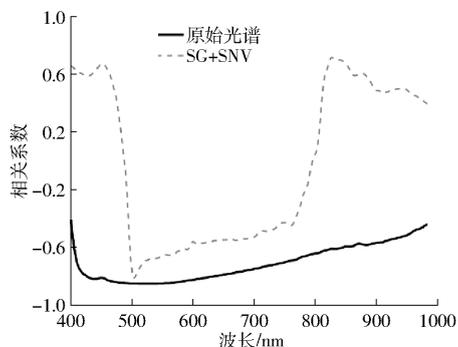


图4 土壤全氮含量与光谱数据相关分析

Fig.4 Correlation analysis of soil total nitrogen content and spectral data

2.4 基于SFLA和CARS提取特征波段

2.4.1 基于混合蛙跳算法筛选特征波段

混合蛙跳算法(SFLA)是一种基于群体协作和

局部搜索相结合的启发式优化算法^[21],能够在保证建模性能的基础上,显著减少冗余信息。图5展示了SFLA在迭代过程中对245个波段的选择概率分布情况,可以观察到部分波段在多次迭代中被频繁选择,表明这些波段与土壤全氮含量具有较强的敏感性。研究中设定了选择概率阈值(如图5中红线所示,下同),只有概率高于该阈值的波段被最终保留。图6则展示了所选特征波段在光谱曲线中的位置分布情况。可以看出,最终筛选出10个关键波长,分别位于:561.5、599.6、770.95、773.45、775.95、893.17、895.75、898.35、945.22、974.15 nm, 占全波段数据的4.08%。这些波段主要集中在红边区(680~750 nm)及近红外区(750~1000 nm),对应若干含氮有机化合物及水分的典型吸收特征,体现了其在土壤全氮光谱反演中的代表性^[22]。具体分析如下:561.5 nm与599.6 nm位于可见光区,受土壤颜色、有机质含量影响,间接反映土壤中腐殖酸、氨基酸类氮源的存在;770.95~775.95 nm位于红边区,该区域与土壤胶体中结合氮(如蛋白氮、有机胺类物质)的吸收有关;893.17、895.75、898.35 nm属于近红外区,接近N—H和C—H基团的弱吸收带,常与蛋白质、肽链、腐殖质中的含氮化合物有关,是土壤中全氮有机组分的重要光谱响应区域;945.22 nm和974.15 nm分别对应O—H伸缩振动的第一与第二吸收带,受铵态氮 NH_4^+ 、可溶性有机氮的共同影响,由于土壤全氮中常含有水解性氮,这两个波段可用于间接表征全氮含量^[23]。这些波段不仅具有明确的化学键基础(如N—H、O—H和C—H吸收特征),还能反映与土壤全氮密切相关的有机氮和无机氮组分的光谱响应特性,充分说明了SFLA在特征波段筛选中的有效性和合理性,为提高全氮预测精度提供了数据基础^[24]。

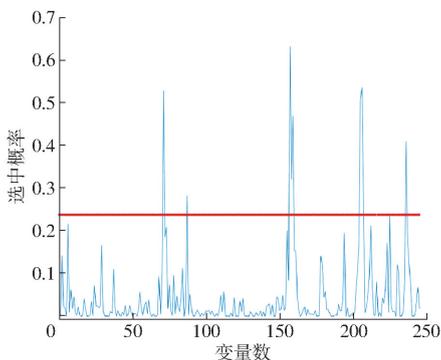


图5 SFLA选择波段概率

Fig. 5 SFLA band selection probability

2.4.2 基于竞争性自适应加权采样筛选特征波段 竞争性自适应加权采样(CARS)的逐步筛选机

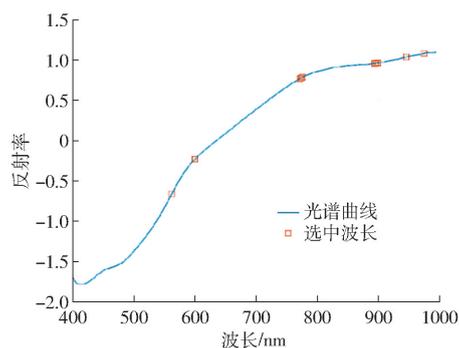


图6 SFLA最终选中波长分布情况

Fig. 6 Distribution of SFLA final selected bands

制有效避免了传统特征选择方法中存在的过拟合问题,显著提升模型预测性能。因此,CARS已被广泛应用于土壤养分监测、作物生长分析以及环境监测等高光谱数据处理领域^[25]。CARS基于蒙特卡洛采样生成多个训练子集;每个子集使用偏最小二乘回归,得到回归系数;根据回归系数绝对值来评估波段的重要性;在多轮迭代中逐步“淘汰”不重要的波段,保留对模型效果贡献大的;最终输出一个或多个表现最优的波段组合。与此同时,CARS在特征选择过程中引入了一种竞争机制,通过对每一轮迭代的特征权重进行重新评估,结合遗传算法中的“竞争”思想,优胜劣汰,逐步优化特征子集的表现。这种自适应加权策略能够动态调整每个波段权重,使得在特征筛选的同时,模型泛化能力得到提升^[26]。

图7为CARS不同波段的被选择概率分布情况以及图8被选中的波长分布。得到了10个波长:411.42、466.02、768.45、775.95、780.97、823.95、864.83、895.75、913.9、974.15 nm,同样也占全波段的4.08%。

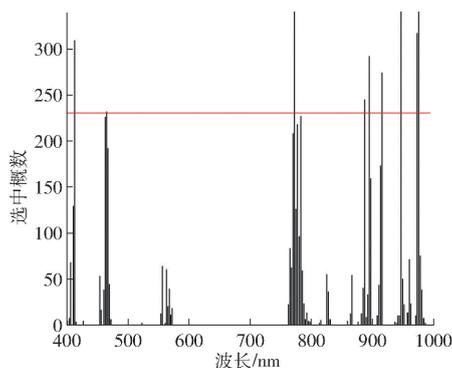


图7 CARS各波长累积被选概率图

Fig. 7 Cumulative selected frequencies for each wavelength of CARS

同样,在基于CARS的特征波段优化过程中,这些位于红边区和近红外区的特征波段往往被优先保留,原因在于它们与土壤全氮含量的光谱响应更为显著,能够更好地反映土壤中的有机质和氮素分布。CARS精简冗余波段、突出重要特征。因此,CARS

筛选出的这些特征波段能够更精确地捕捉土壤全氮含量的变化趋势,大幅提升模型预测精度。更重要的是,通过对高维数据的高效降维,CARS避免了传统方法中信息冗余和过拟合的问题,增强了模型泛化能力。

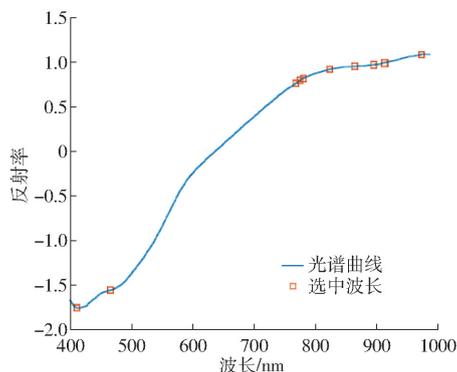


图8 CARS最终选中波长分布情况

Fig. 8 Final distribution of selected wavelength by CARS

2.5 基于特征波段的全氮含量预测模型构建

图9为2种特征提取方法下及全波段全氮含量预测模型测试集散点图。在全波段光谱下,RF建立的预测模型精度最高,效果最好(测试集 $R^2 = 0.86$ 、RMSE为0.38 g/kg、RPD为2.94)。在使用SG+SNV预处理原始数据后,综合比较不同建模方法在验证集上的RPD,模型精度排序从小到大依次为:CNN、RF、PLSR、SVM。其中,SFLA-CNN预测模型预测性能最优(测试集 $R^2 = 0.95$ 、RMSE为0.21 g/kg、RPD为3.97)。由图9可知,在SFLA-CNN模型下,测试集预测值和真实值分布在1:1线附近的分布最均匀;而且使用SFLA提取特征波段构建的预测模型比经过CARS特征提取的模型更靠近1:1线,说明在SFLA下模型预测性能更好。

3 讨论

3.1 不同预处理方法对预测模型影响

已有研究表明,不同的光谱预处理方法可在不同程度上消除噪声,增强特征信号,从而提高光谱与土壤全氮等养分之间的相关性。文献[27]应用基线校正加归一化处理来预处理原始光谱,而后建立的土壤全氮含量预测模型,预测精度最高。文献[28]对原始光谱经29种预处理变换,确定了基于SG卷积平滑和对数变换预处理后构建的模型是最适用于土壤速效氮含量的预测模型,可实现对土壤速效氮含量的定量估测。本研究采用SG、FD、SD、MSC和SNV进行组合。进一步证实SG加SNV效果最佳,更能映射出全氮含量的特征信号。从原理上看,SG通过在局部滑动窗口内进行多项式拟合,既能去除光谱曲线中的高频噪声,又能保持光谱

曲线的整体趋势,避免了信息的过度丢失;而SNV主要用于校正散射效应,通过对每条光谱进行中心化和归一化处理,使光谱之间的特征相关性增强。两者结合后,SG先对光谱进行平滑降噪,提供更干净的输入数据,再由SNV进一步消除散射引起的基线偏移和振幅波动,从而最大限度地保留与全氮相关的光谱信息,提升模型识别全氮光谱特征的能力。相比SG和FD、SD或MSC组合等方法,SG+SNV组合在降噪与标准化之间实现了良好平衡,更有助于后续建模的稳定性和准确性。

3.2 不同特征提取方法比较

特征提取作为高光谱建模中的关键环节,直接关系到模型的稳定性与泛化能力。如文献[29]采用随机森林、差分进化算法等提取土壤养分特征波段,发现基于随机森林-差分进化-层次分析法提取特征波段的BP预测模型预测精度最高,说明特征提取对机器学习建模具有有效性和可行性。文献[30]利用相关分析和离散小波变换提取特征参数,结论表明离散小波变换特征提取后的模型要优于全波段和以敏感波段作为输入构建的模型。采用混合蛙跳算法(SFLA)和竞争性自适应加权采样(CARS)两种方法筛选特征波段进行对比,结果表现为基于SFLA提取特征波段构建的预测模型比CARS构建的模型预测性能更好。

3.3 不同建模方法比较

在高光谱数据建模中,选择合适的回归算法对于提高预测精度至关重要。常见的建模方法包括偏最小二乘回归(PLSR)、支持向量机(SVM)和随机森林回归(RF)等。文献[31]采用PLSR、RF和SVM分别建立土壤全氮含量估测模型,结果表明基于倒数的一阶微分建立的SVM模型预测精度最高,是预测滨海湿地土壤全氮含量的最优模型。文献[32]采用PLSR、SVM、RF和反向传播神经网络(BPNN)4种建模方法分别构建土壤全氮含量反演模型。研究发现基于主成分分析降维的BPNN模型预测能力最好,为南疆农田土壤全氮含量快速估测提供一定的参考价值。文献[33]采用全波段原始光谱结合PLSR、BP神经网络(BPNN)和SVM,分别构建不同的山地红壤全氮含量高光谱估算模型。研究结论表明,基于SVM构建的土壤全氮含量估算模型精度最高,可以较精确地估测庐山山地红壤全氮含量。为了进一步评估不同建模方法对土壤总氮含量预测精度的影响,本研究分别使用PLSR、SVM、RF和CNN模型,并结合最优预处理(SG+SNV)及最优特征提取方法(SFLA)进行建模比较。结果发现SG+SNV-SFLA-CNN模型表现最佳,测试集

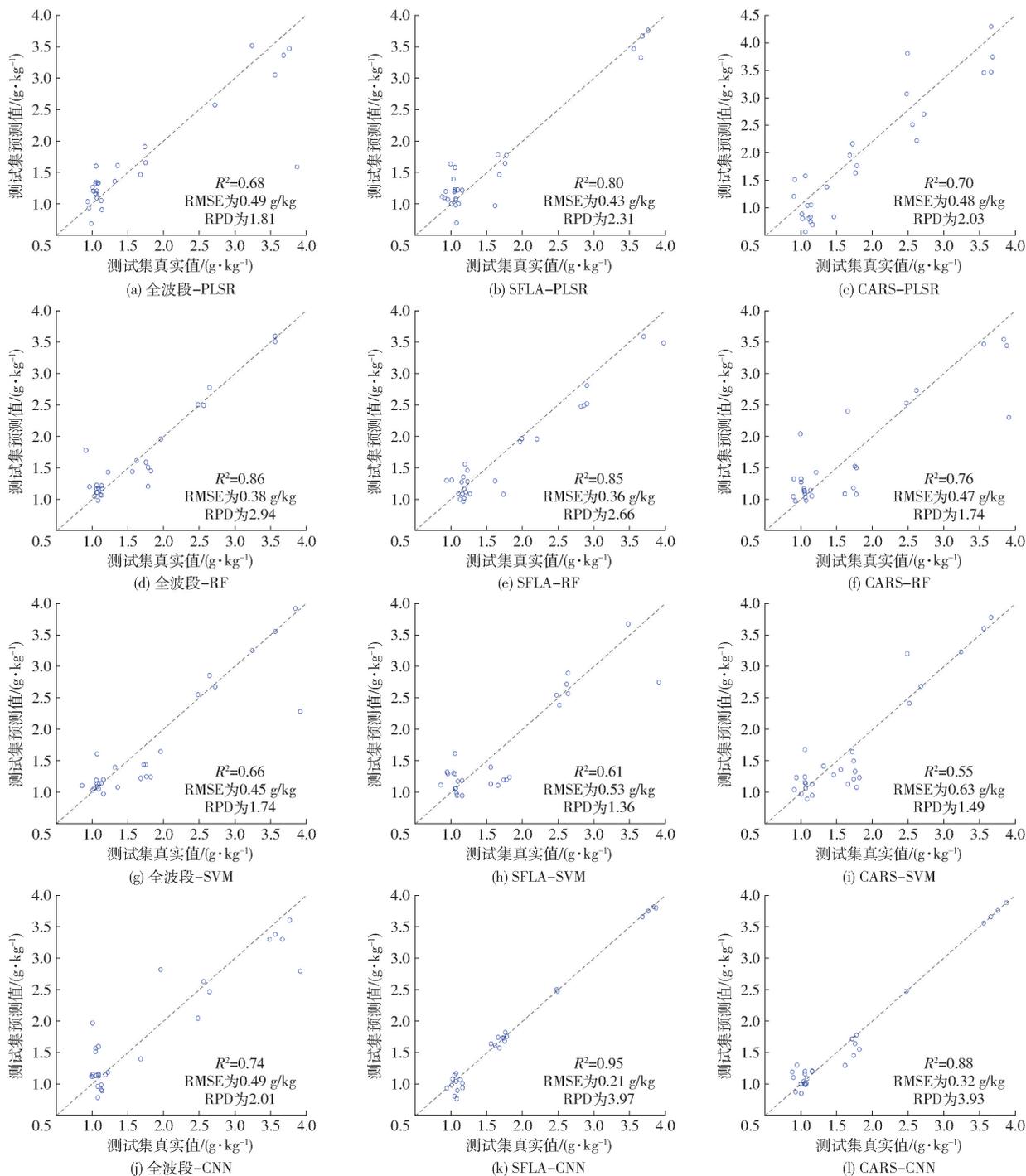


图9 2种特征提取方法下全氮含量预测模型的测试集散点图

Fig. 9 Test distribution plots of total nitrogen content prediction models by two feature extraction methods

R^2 、RMSE、RPD 分别为 0.95、0.21 g/kg、3.97;与文献[31-33]相比精度更高。

4 结论

(1)在 400~1 000 nm 范围内,不同全氮含量的土壤光谱反射率存在明显差异,且同一波段下光谱反射率随全氮含量的增加而下降。

(2)原始光谱经 SG + SNV 预处理后,建立的全氮含量预测模型精度显著提高。基于 SG + SNV 预处理方法,使用特征波段建模时,相同建模方法下用

SFLA 构建的全氮含量估测模型精度均高于使用 CARS 建立的模型。

(3)基于 SG + SNV - SFLA 建立的 CNN 模型 R^2 为 0.95, RMSE 为 0.21 g/kg, RPD 为 3.97,是预测果园土壤全氮含量的最优模型。模型的 R^2 从 0.86 显著提升至 0.95,说明该模型对土壤全氮含量的拟合能力和预测精度得到了明显改善。同时, RMSE 的降低体现了预测误差显著减少,而 RPD 的增加则表明模型泛化能力更强。运用高光谱成像技术能够实现果园土壤全氮含量的精确估测。

参 考 文 献

- [1] 安柏耸,王雪梅,黄晓宇,等. 基于随机蛙跳波段选择算法的土壤铅含量高光谱估测[J]. 光谱学与光谱分析, 2023, 43(10): 3302-3309.
AN Baizhuo, WANG Xuemei, HUANG Xiaoyu, et al. Hyperspectral estimation of soil lead content based on random frog jump band selection algorithm[J]. Spectroscopy and Spectral Analysis, 2023, 43(10): 3302-3309. (in Chinese)
- [2] 李学兰,李德成,郑光辉,等. 可见-近红外与中红外光谱预测土壤养分的比较研究[J]. 土壤学报, 2024, 61(3): 687-698.
LI Xuelan, LI Decheng, ZHENG Guanghui, et al. Comparative study on prediction of soil nutrients by visible-near-infrared and mid-infrared spectroscopy[J]. Acta Pedologica Sinica, 2024, 61(3): 687-698. (in Chinese)
- [3] KIM M, LEE J, BACK I, et al. Estimation of total nitrogen content in topsoil based on machine and deep learning using hyperspectral imaging[J]. Agriculture, 2023, 13(10): 1975.
- [4] 张东辉,赵英俊,秦凯,等. 光谱变换方法对黑土养分含量高光谱遥感反演精度的影响[J]. 农业工程学报, 2018, 34(20): 141-147.
ZHANG Donghui, ZHAO Yingjun, QING Kai, et al. Effect of spectral transformation method on accuracy of hyperspectral remote sensing inversion of nutrient content in black soil[J]. Transactions of the CSAE, 2018, 34(20): 141-147. (in Chinese)
- [5] 张恒,梁太波,冯文强,等. 基于连续投影算法的土壤全氮和碱解氮含量高光谱估测[J]. 中国烟草科学, 2023, 44(5): 103-113.
ZHANG Heng, LIANG Taibo, FENG Wenqiang, et al. Hyperspectral estimation of soil total nitrogen and alkali hydrolyzable nitrogen content based on continuous projection algorithm[J]. Chinese Tobacco Science, 2023, 44(5): 103-113. (in Chinese)
- [6] 刘雯,韩玲,刘明,等. 基于高分五号高光谱波段选择的矿区周边土壤 Cd 含量反演[J]. 激光与光电子学进展, 2023, 60(17): 362-369.
LIU Wen, HAN Ling, LIU Ming, et al. Inversion of Cd content in soil around mining area based on GF-5 hyperspectral band selection[J]. Advances in Laser and Optoelectronics, 2023, 60(17): 362-369. (in Chinese)
- [7] 江振蓝,陈付勋,罗双飞,等. 基于多光谱变换和主成分分析的土壤全铁含量随机森林模型反演[J]. 浙江农业学报, <https://link.cnki.net/urlid/33.1151.S.20250317.1236.008>.
- [8] 路皓翔,张静,李灵巧,等. 最小角回归结合竞争性自适应重加权采样的近红外光谱波长选择[J]. 光谱学与光谱分析, 2021, 41(6): 1782-1788.
LU Haoxiang, ZHANG Jing, LI Lingqiao, et al. Wavelength selection of near-infrared spectroscopy based on minimum angle regression combined with competitive adaptive reweighted sampling[J]. Spectroscopy and Spectral Analysis, 2021, 41(6): 1782-1788. (in Chinese)
- [9] 李祥,刘帅,陈发明,等. 基于 SSA-XGBoost 算法的水质叶绿素 a 遥感高光谱反演[J]. 环境科学与技术, 2024, 47(12): 169-176.
LI Xiang, LIU Shuai, CHEN Faming, et al. Hyperspectral inversion of water quality chlorophyll a based on SSA-XGBoost algorithm[J]. Environmental Science and Technology, 2024, 47(12): 169-176. (in Chinese)
- [10] 牛芳鹏,李新国,白云岗,等. 遗传算法和连续投影算法结合的土壤有机碳含量高光谱估算模型[J]. 光谱学与光谱分析, 2023, 43(7): 2232-2237.
NIU Fangpeng, LI Xinguo, BAI Yungang, et al. Hyperspectral estimation model of soil organic carbon content based on genetic algorithm and continuous projection algorithm[J]. Spectroscopy and Spectral Analysis, 2023, 43(7): 2232-2237. (in Chinese)
- [11] 刘芳华,宋文杰. 基于改进反馈卷积自编码器的高光谱图像降维[J]. 现代电子技术, 2024, 47(19): 94-99.
LIU Fanghua, SONG Wenjie. Hyperspectral image dimensionality reduction based on improved feedback convolutional autoencoders[J]. Modern Electronic Technology, 2024, 47(19): 94-99. (in Chinese)
- [12] 丁繁昌,郭宝峰,贾响响,等. 基于生成对抗网络的高光谱图像特征提取[J]. 光电子·激光, 2021, 32(8): 852-861.
DING Fanchang, GUO Baofeng, JIA Xiangxiang, et al. Hyperspectral image feature extraction based on generative adversarial network[J]. Optoelectronics & Laser, 2021, 32(8): 852-861. (in Chinese)
- [13] WANG Hongyu, WANG Juan, MA Rongrong, et al. Soil nutrients inversion in open-pit coal mine reclamation area of loess plateau, China: a study based on ZhuHai-1 hyperspectral remote sensing[J]. Land Degradation & Development, 2024, 35(17): 5210-5223.
- [14] 张紫玥,刘晓,杜丽丽,等. 基于随机蛙跳降维和 PSO-BPNN 的高光谱水体总氮遥感反演[J]. 激光与光电子学进展, 2025, 62(19): 1930002.
ZHANG Ziyue, LIU Xiao, DU Lili, et al. Remote sensing inversion of total nitrogen in hyperspectral water body based on random frog jump dimensionality reduction and PSO-BPNN[J]. Progress in Laser and Optoelectronics, 2025, 62(19): 1930002. (in Chinese)
- [15] LI Hao, JU Weiliang, SONG Yamei, et al. Soil organic matter content prediction based on two-branch convolutional neural network combining image and spectral features[J]. Computers and Electronics in Agriculture, 2024, 217: 108561.
- [16] SIMONE P, STEFANO P, RAFFAELE C, et al. Hyperspectral remote sensing of agriculture and vegetation[J]. Remote Sensing, 2020, 12: 3665.
- [17] MOHAMED F G, COSIMO D, PAOLO S, et al. Deep learning techniques for hyperspectral image analysis in agriculture: a

- review[J]. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 2024, 12: 100062.
- [18] 崔玉露, 杨玮, 王炜超, 等. 基于光谱学原理的便携式土壤有机质检测仪设计与实验[J]. *农业机械学报*, 2021, 52(增刊): 323–328, 350.
CUI Yulu, YANG Wei, WANG Weichao, et al. Design and experiment of portable soil organic matter detector based on spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 323–328, 350. (in Chinese)
- [19] 王炜超, 杨玮, 崔玉露, 等. 基于 CatBoost 算法与图谱特征融合的土壤全氮含量预测[J]. *农业机械学报*, 2021, 52(增刊): 316–322.
WANG Weichao, YANG Wei, CUI Yulu, et al. Prediction of soil total nitrogen content based on CatBoost algorithm and spectrum feature fusion[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(Supp.): 316–322. (in Chinese)
- [20] WAN Mengxue, FAN Yanan, JIAO Wentao, et al. Feasibility of using Vis–NIR spectroscopy and PXRF spectrometry to estimate regional soil cadmium concentration[J]. *Journal of Environmental Sciences*, 2024, 145: 88–96.
- [21] 邹宇博, 马振宇, 焦庆斌, 等. 水体参数高光谱反演模型对比研究[J]. *光谱学与光谱分析*, 2023, 43(3): 949–954.
ZOU Yubo, MA Zhenyu, JIAO Qingbin, et al. Comparative study on hyperspectral inversion models of water parameters[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(3): 949–954. (in Chinese)
- [22] WANG Wenqian, WANG Hongjia, LI Ruiyan et al. Selection of spectral parameters and optimization of estimation models for soil total nitrogen content during fertilization period in apple orchards[J]. *Horticulturae*, 2024, 10(4): 358.
- [23] 崔炜楠, 聂志刚, 李广, 等. 基于改进的混合蛙跳算法对旱地小麦籽粒生长模型参数的优化[J]. *中国农业科学*, 2023, 56(12): 2274–2287.
CUI Weinan, NIE Zhigang, LI Guang, et al. Optimization of grain growth model parameters of dryland wheat based on improved hybrid frog jumping algorithm[J]. *Scientia Agricultura Sinica*, 2023, 56(12): 2274–2287. (in Chinese)
- [24] 程介虹, 陈争光. 改进的联合区间随机蛙跳算法的近红外光谱波长选择[J]. *光谱学与光谱分析*, 2020, 40(11): 3451–3456.
CHENG Jiehong, CHEN Zhengguang. Wavelength selection of near-infrared spectroscopy based on improved joint interval stochastic frog jumping algorithm[J]. *Spectroscopy and Spectral Analysis*, 2020, 40(11): 3451–3456. (in Chinese)
- [25] 樊泳灼, 李新国. 湖滨绿洲棕漠土有机碳含量高光谱估算[J]. *江苏农业学报*, 2023, 39(6): 1341–1348.
FAN Yongzhuo, LI Xinguo. Hyperspectral estimation of organic carbon content in brown desert soil in lakeside oasis[J]. *Jiangsu Journal of Agricultural Sciences*, 2023, 39(6): 1341–1348. (in Chinese)
- [26] AHMED L, PAUL V, KELTOUM K, et al. Ensemble band selection for quantification of soil total nitrogen levels from hyperspectral imagery[J]. *Remote Sensing*, 2024, 16(14): 2549.
- [27] 董桂梅, 杨仁杰, 杨延荣, 等. 近红外光谱定量分析土壤全氮含量预处理方法探讨[J]. *光谱学与光谱分析*, 2018, 38(10): 77–78.
DONG Guimei, YANG Renjie, YANG Yanrong, et al. Discussion on pretreatment method for quantitative analysis of soil total nitrogen content by near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(10): 77–78. (in Chinese)
- [28] 方向, 金秀, 朱娟娟, 等. 基于可见-近红外光谱预处理建模的土壤速效氮含量预测[J]. *浙江农业学报*, 2019, 31(9): 1523–1530.
FANG Xiang, JIN Xiu, ZHU Juanjuan, et al. Prediction of soil available nitrogen content based on visible–near-infrared spectroscopy pretreatment modeling[J]. *Journal of Zhejiang Agricultural Sciences*, 2019, 31(9): 1523–1530. (in Chinese)
- [29] 刘凯, 王玉峰, 彭志青, 等. 基于机器学习的高光谱土壤养分特征波段提取方法[J]. *光学学报*, 2025, 45(6): 230–239.
LIU Kai, WANG Yufeng, PENG Zhiqing, et al. Hyperspectral soil nutrient characteristic band extraction method based on machine learning[J]. *Acta Optica Sinica*, 2025, 45(6): 230–239. (in Chinese)
- [30] 张娟娟, 牛圳, 马新明, 等. 基于离散小波的土壤全氮高光谱特征提取与反演[J]. *光谱学与光谱分析*, 2023, 43(10): 3223–3229.
ZHANG Juanjuan, NIU Zhen, MA Xinming, et al. Extraction and inversion of soil total nitrogen hyperspectral features based on discrete wavelets[J]. *Spectroscopy and Spectral Analysis*, 2023, 43(10): 3223–3229. (in Chinese)
- [31] 张清文, 吴风华, 宋敬茹, 等. 基于光谱变换的滨海湿地土壤全氮含量建模预测[J]. *土壤*, 2023, 55(4): 880–886.
ZHANG Qingwen, WU Fenghua, SONG Jingru, et al. Modeling and prediction of soil total nitrogen content in coastal wetland based on spectral transformation[J]. *Soils*, 2023, 55(4): 880–886. (in Chinese)
- [32] 白子金, 彭杰, 罗德芳, 等. 新疆农田土壤全氮含量的中红外光谱反演模型[J]. *光谱学与光谱分析*, 2022, 42(9): 2768–2773.
BAI Zijin, PENG Jie, LUO Fang, et al. Mid-infrared spectroscopy inversion model of soil total nitrogen content in farmland in southern Xinjiang[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(9): 2768–2773. (in Chinese)
- [33] 冀荣华, 李常昊, 郑立华, 等. 基于稀疏自注意力和可见-近红外光谱的土壤氮含量预测[J]. *农业机械学报*, 2024, 55(10): 392–398, 409.
JI Ronghua, LI Chaogao, ZHENG Lihua, et al. Prediction of soil nitrogen content based on sparse self-attention and visible–near-infrared spectroscopy[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(10): 392–398, 409. (in Chinese)