

DOI: 10.11766/trxb202210040548

李学兰, 李德成, 郑光辉, 曾荣, 蔡凯, 高维常, 潘文杰, 姜超英, 曾隄涛. 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, ZENG Rong, CAI Kai, GAO Weichang, PAN Wenjie, JIANG Chaoying, ZENG Yuntao. Comparative Study on Prediction of Soil Nutrients by Visible-Near Infrared and Mid-infrared Spectroscopy[J]. Acta Pedologica Sinica, 2024, 61(3): 687-698.

可见-近红外与中红外光谱预测土壤养分的比较研究*

李学兰¹, 李德成², 郑光辉¹, 曾荣^{1†}, 蔡凯³, 高维常³, 潘文杰⁴, 姜超英⁴, 曾隄涛⁴

(1. 南京信息工程大学地理科学学院, 南京 210044; 2. 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008; 3. 贵州省烟草科学研究院烟草行业山地烤烟品质与生态重点实验室, 贵阳 550081; 4. 中国烟草总公司贵州省公司, 贵阳 550004)

摘要: 对土壤养分的快速和准确测定有助于适时指导施肥。为进一步研究可见-近红外(350~2 500 nm)与中红外光谱(4 000~650 cm^{-1})对土壤养分的预测能力, 以贵州省 500 个土样为例, 对光谱进行 Savitzky-Golay (SG) 平滑去噪处理, 再用标准正态化 (SNV) 方法进行基线校正, 然后分别应用偏最小二乘回归 (PLSR) 和支持向量机 (SVM) 两种方法进行建模, 探讨了可见-近红外和中红外光谱对土壤全氮 (TN)、全磷 (TP)、全钾 (TK) 和碱解氮 (AN)、有效磷 (AP)、速效钾 (AK) 共六种土壤养分的预测效果。结果表明: (1) 无论基于可见-近红外光谱还是中红外光谱, PLSR 模型的预测精度整体均优于 SVM 模型。(2) 中红外光谱对 TN、TK 和 AN 的预测精度均显著高于可见-近红外光谱, 可见-近红外和中红外光谱均可以可靠地预测 TN 和 TK (性能与四分位间隔距离的比率 (RPIQ) 大于 2.10), 中红外光谱可相对较可靠地预测 AN (RPIQ=1.87); 但两类光谱对 TP、AP 和 AK 的预测效果均较差 (RPIQ<1.34)。(3) 当变量投影重要性得分 (VIP) 大于 1.5 时, PLSR 模型在中红外光谱区域预测 TN 和 TK 的重要波段多于可见-近红外光谱区域, TN 的重要波段主要集中于可见-近红外光谱区域的 1 910 和 2 207 nm 附近, 中红外光谱区域的 1 120、1 000、960、910、770 和 668 cm^{-1} 附近; TK 的重要波段主要集中于可见-近红外光谱区域的 540、2 176、2 225 和 2 268 nm 附近, 中红外光谱区域的 1 040、960、910、776、720 和 668 cm^{-1} 附近。因此, 中红外光谱技术结合 PLSR 模型对土壤养分预测效果较好, 可快速准确预测土壤 TN 和 TK, 可为指导适时施肥提供技术支撑。

关键词: 可见-近红外光谱; 中红外光谱; 土壤养分; 偏最小二乘回归; 支持向量机

中图分类号: S151.9 文献标志码: A

Comparative Study on Prediction of Soil Nutrients by Visible-Near Infrared and Mid-infrared Spectroscopy

LI Xuelan¹, LI Decheng², ZHENG Guanghui¹, ZENG Rong^{1†}, CAI Kai³, GAO Weichang³, PAN Wenjie⁴, JIANG Chaoying⁴, ZENG Yuntao⁴

* 国家自然科学基金项目(42107322)、中国烟草总公司贵州省公司科技项目(201910)、江苏省高校自然科学基金面上项目(20KJB210009)共同资助 Supported by the National Natural Science Foundation of China (No. 42107322), the Science and Technology Project of China National Tobacco Corporation Guizhou Provincial Company (No. 201910), and the Natural Science Foundation of Colleges and Universities of Jiangsu Province, China (No. 20KJB210009)

† 通讯作者 Corresponding author, E-mail: rzeng@nuist.edu.cn

作者简介: 李学兰 (1997—), 女, 广西桂林人, 硕士研究生, 主要从事土壤遥感方面的研究。E-mail: lixuelan785@163.com

收稿日期: 2022-10-04; 收到修改稿日期: 2023-02-23; 网络首发日期 (www.cnki.net): 2023-03-28

(1. School of Geographical Science, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 3. Upland Flue-cured Tobacco Quality & Ecology Key Laboratory of China Tobacco, Guizhou Academy of Tobacco Science, Guiyang 550081, China; 4. China National Tobacco Corporation Guizhou Provincial Company, Guiyang 550004, China)

Abstract: 【Objective】 Predicting soil nutrients by visible-near infrared (vis-NIR) and mid-infrared (MIR) spectroscopy has the advantages of being fast, cost-effective and environmental friendly. Soil spectra contain abundant information of soil properties, and can be combined with machine learning methods to effectively and accurately predict soil nutrients, which can provide support and guidance for timely fertilization management. The objective of this study was to compare the predictive ability of vis-NIR (350-2 500 nm) and MIR spectroscopy (4 000-650 cm^{-1}) for predicting both the total and available contents of soil nitrogen (N), phosphorus (P) and potassium (K), in order to construct an optimal model for estimation of different nutrient contents. 【Method】 In this study, 500 samples were collected from the surface layers (0-20 cm) of the dryland in Guizhou Province for determination of soil N, P and K contents and spectral analysis. The vis-NIR spectra were measured by Cary 5000 and the MIR spectra by Thermo Scientific Nicolet iS50. Soil spectra were pre-processed by Savitzky-Golay (SG) smoothing for denoising and standard normal variate (SNV) transformation for baseline correction. Partial least squares regression (PLSR) and support vector machine (SVM) were used to predict the contents of total nitrogen (TN), total phosphorus (TP), total potassium (TK), alkali-hydrolyzable nitrogen (AN), available phosphorus (AP) and available potassium (AK). 【Result】 The results showed that: (1) Whether using the vis-NIR spectroscopy or the MIR spectroscopy, the prediction accuracy of PLSR model was better than that of SVM model. (2) The accuracy of MIR spectroscopy for prediction of TN, TK and AN was significantly higher than that of vis-NIR spectroscopy. Vis-NIR and MIR spectroscopy could reliably predict TN and TK (ratio of performance to interquartile distance (RPIQ) > 2.10), while MIR spectroscopy could predict AN with moderate accuracy (RPIQ = 1.87). However, both types of spectra had poor ability to predict TP, AP and AK (RPIQ < 1.34). (3) When the variable in the projection (VIP) score was > 1.5, there were more important bands selected by PLSR models in the MIR region than the vis-NIR region. The important bands selected for estimation of TN were mainly concentrated near 1 910 and 2 207 nm in the vis-NIR region, and centered around 1 120, 1 000, 960, 910, 770, and 668 cm^{-1} in the MIR region. The important bands of TK were mainly distributed around 540, 2 176, 2 225, and 2 268 nm in the vis-NIR region, and around 1 040, 960, 910, 776, 720, and 668 cm^{-1} in the MIR region. 【Conclusion】 Therefore, MIR spectroscopy combined with PLSR model proved to be promising for accurate prediction of soil nutrients, especially for the estimation of TN and TK, and can provide technical support for guiding timely fertilization.

Key words: Visible-near infrared spectroscopy; Mid-infrared spectroscopy; Soil nutrients; Partial least squares regression; Support vector machine

土壤氮 (N)、磷 (P)、钾 (K) 是土壤肥力的重要组成部分^[1]。常规土壤养分测定方法主要为实验室化学测量法, 其周期长、成本高, 存在环境风险。而可见-近红外 (VNIR) 和中红外 (MIR) 光谱预测土壤属性, 具有快速、低成本、低环境风险的优势^[2], 已成为土壤养分预测的研究热点。

利用土壤光谱预测土壤养分的研究自 21 世纪开始逐步增多, 早期的研究主要集中在可见-近红外光谱。如李伟等^[3]利用近红外光谱有效预测了土壤碱解氮 (AN), 但对有效磷 (AP) 和速效钾 (AK) 含量的预测效果较差; 吴茜等^[4]发现土壤可见-近红外光谱可用于对水稻土 N、P、K 速效态预测; Yang

等^[5]采用可见-近红外光谱较好地预测了土壤氮含量, 其预测精度决定系数 (R^2) 最高达到 0.70; Devianti 等^[6]同样利用可见-近红外光谱较好地预测了土壤全氮 (TN)、全磷 (TP) 和全钾 (TK) 含量。近年来, 伴随中红外光谱技术的发展与成熟, 应用中红外光谱技术预测土壤养分受到了较多关注。相比于可见-近红外光谱, 中红外光谱的波峰波谷主要源于基频振动引起的能级跃迁, 包含的土壤属性信息更为丰富^[7]。已有研究表明, 与可见-近红外光谱相比, 中红外光谱对土壤养分的预测更为精准, 如蒋璐璐等^[8]利用中红外光谱对 AN 的预测结果与近红外光谱相近, 但对 AP 和 AK 的预测结果更优;

Santos 等^[2]发现中红外光谱对土壤全碳 (TC) 和 TN 含量的预测要优于可见-近红外光谱; Greenberg 等^[9]发现中红外光谱技术可准确地预测土壤 TN 含量, 性能与四分位间隔距离的比率 (ratio of performance to interquartile distance, RPIQ) 高达 2.50; Hati 等^[10]利用中红外光谱较好地预测了有机碳 (SOC) 含量, 而对 AP 和 AK 的含量预测较差。但迄今为止, 结合可见-近红外和中红外光谱预测土壤 N、P、K 全量和有效态的比对研究甚少。此外, 随着多元数据分析技术的发展, 结合使用土壤光谱技术和机器学习方法能更好地实现对土壤属性的预测, 偏最小二乘回归 (PLSR) 和支持向量机 (SVM) 作为常用的机器学习方法, 对土壤属性的预测均取得了较好的结果^[2, 11]。本研究选取贵州全省 500 个土壤样本为研究对象, 研究可见-近红外和中红外波段的光谱特征, 比较两类光谱对 TN、TP、TK 和 AN、AP、AK 的预测能力; 评估 PLSR 和 SVM 两种模型的预测精度, 以期构建预测不同养分含量的最优模型; 并筛选出土壤养分在可见-近红外和中红外光谱区域的重要建模波段, 为快速预测土壤养分含量提供技术支撑。

1 材料与方法

1.1 研究区概况与土样采集和测定

贵州省 (24°37'N~29°13'N, 103°36'E~109°35'E) 位于我国西南地区, 平均海拔 1 100 m 左右, 是典型喀斯特地貌区, 以高原山地为主, 地势西高东低。土壤类型主要有黄壤、黄棕壤、红壤、水稻土等^[12]。

2019 年 11 月至 12 月期间利用代表性采样法在贵州省全省范围内确定了 500 块农田, 主要为烟草和玉米轮作的旱地, 具体分布位置如图 1 所示。在每块农田随机选取了 5~8 个点采集耕作层 (0~20 cm) 土样, 记录采样点的经纬度、海拔高度、成土因素、土地利用类型等重要信息, 将土样充分混合, 采取四分法保留其中的 1.5 kg 土样。将采集的土样在实验室干燥处进行自然风干、去杂、磨碎, 过 60 目 (0.25 mm) 筛, 充分混匀后等分为两份, 分别用于化学组分测定和光谱分析。

土壤 N、P、K 含量的测定参考《土壤调查实验室分析方法》^[13], TN 采用重铬酸钾、硫酸消化—蒸馏法测定; TP 采用酸溶—钼锑钨比色法测定; TK

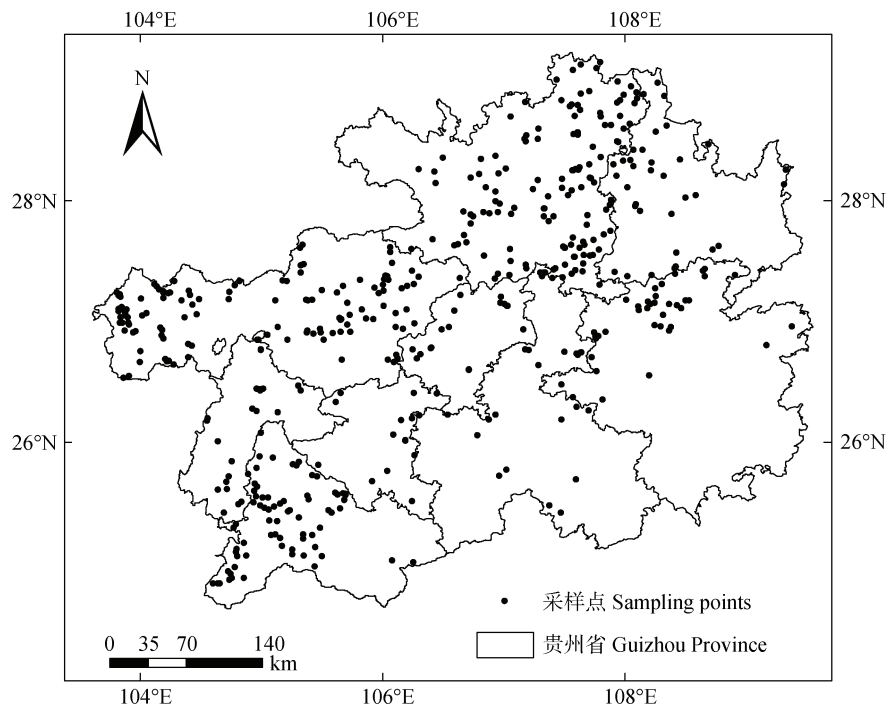


图 1 研究区采样点分布图

Fig. 1 Spatial distribution of sampling sites

采用酸溶—火焰光度法测定；AN 采用碱解扩散法测定；AP 采用浸提—钼锑钒比色法测定；AK 采用乙酸铵浸提—火焰光度法测定。

1.2 光谱数据的采集和预处理

测定光谱前，首先将土样于 45℃ 烘箱烘烤 24 h 置于干燥器中待测。可见-近红外光谱的测定仪器为紫外-可见-近红外 (UV-Vis-NIR) 分光光度计 (Agilent Cary 5000, 马来西亚), 波长范围为 350~2 500 nm, 其中 350~700 nm 的光谱分辨率小于等于 0.048 nm, 700~2 500 nm 的分辨率小于等于 0.20 nm, 光谱重采样至 1 nm, 每次选取 1~2 g 样品置于样品仓; 中红外光谱测定仪器为傅立叶变换红外光谱 (FT-IR, Thermo Scientific Nicolet iS50, 美国), 测定范围为 4 000~400 cm^{-1} , 光谱分辨率小于等于 0.09 cm^{-1} , 将粉末状土样置于光滑圆盘上, 旋回顶针压实样品, 测定土样的光谱透过率, 同一样品重复压实测量三次取平均值, 最后通过 OMNIC 软件转化为吸光度。

在构建光谱模型之前, 首先进行光谱预处理。在比较不同预处理方法对模型预测精度的影响后, 优选了如下的预处理方法: 对可见-近红外和中红外光谱进行 Savitzky-Golay (SG) 卷积平滑处理, 以期去除光谱噪声的影响; 再应用标准正态化 (SNV) 方法对平滑后的中红外光谱进行基线校正, 以消除基线漂移^[14]。

1.3 预测模型的建立与检验

本研究采用偏最小二乘回归 (PLSR) 和支持向量机 (SVM) 两种方法来构建光谱预测模型。PLSR 是应用最为广泛的线性建模方法, 结合了主成分回归 (PCR) 与线性回归的优势, 该方法建模时可允许样本量少于变量, 对特征向量有很好的辨识能力, 能够通过数据降维将光谱转化为有限的潜在变量 (latent variable, LV) 来预测土壤性质; SVM 通过结构风险最小化原理来提高泛化能力, 其优点是能够较好地解决非线性、高维数和局部极小等实际问题^[15]。

为避免随机划分建模集和验证集对模型精度的影响, 本文采用留一法交叉验证 (Leave-one-out cross validation, LOOCV)。该方法是指假设有 N 个样本, 每次仅留下一个样本作为预测集, 其余 $N-1$ 个样本做建模集^[16]。PLSR 和 SVM 均采用留一法交叉验证来构建土壤养分的预测模型。

预测模型的精度通过实测值与预测值的决定系数 (R^2)、均方根误差 (RMSE) 和性能与四分位间隔距离的比率 (Ratio of performance to interquartile distance, RPIQ) 来进行评价。RPIQ 综合考虑了预测误差和测量值本身的变异, 且不对测量值的正态性作出任何假设。相对于相对分析误差 (RPD), 它对模型的精度评定更为客观^[17]。当模型的 R^2 越大、RMSE 越小、RPIQ 越大, 说明模型预测精度越高, 反之越低。

针对 PLSR 的最佳预测模型, 采用变量投影重要性 (Variable in the projection, VIP)^[18] 的得分来筛选预测目标属性的重要光谱波段, 具体计算公式如下:

$$\text{VIP}_k(a) = K \sum_a \omega_{ak}^2 \left(\frac{\text{SSY}_a}{\text{SST}_t} \right) \quad (1)$$

式中, $\text{VIP}_k(a)$ 是 LV 为 a 的模型中第 k 个预测变量的重要性得分, ω_{ak} 是第 a 个 LV 对应的第 k 个变量的载荷权重, SSY_a 是 a 个 LV 的 PLSR 模型中解释预测变量 Y 的总平方和, SST_t 为 Y 的总离差平方和, K 为预测变量的个数。VIP 分数是对每个解释变量的重要性的衡量。一般认为 VIP 得分大于 1 的有影响波段是重要波段, VIP 得分越大说明该土壤属性对应的波长在 PLSR 模型的重要性越强^[19]。

2 结果与讨论

2.1 土壤氮磷钾含量的统计特征

本研究根据贵州省土壤养管理情况, 并参考全国第二次土壤普查养分分级标准, 将土壤养分分为 5 个等级 (仅 TP 为 4 个等级), 并结合 N、P、K 含量统计情况 (表 1), 计算各属性在每个分级标准下含量所占比重 (表 2)。其中, TN 含量适中, 大多数含量低于 1.73 $\text{g}\cdot\text{kg}^{-1}$, 范围为 0.55~3.51 $\text{g}\cdot\text{kg}^{-1}$, 变异系数低 (26.85%); TP 含量偏低, 平均值仅为 0.91 $\text{g}\cdot\text{kg}^{-1}$, 范围为 0.11~10.73 $\text{g}\cdot\text{kg}^{-1}$, 变异系数达到 62.29%; TK 含量在分级范围内占比较均匀, 处于 2.17~41.37 $\text{g}\cdot\text{kg}^{-1}$ 之间, 值域较宽, 变异系数为 50.44%。AN、AP 和 AK 含量平均值分别为 150.8、32.30 和 349.0 $\text{mg}\cdot\text{kg}^{-1}$, AN 含量适中,

变化范围为 36.75~475.9 mg·kg⁻¹, 变异系数为 29%, 别处于 0.26~251.8 mg·kg⁻¹ 和 40.00~148.0 mg·kg⁻¹ 与 TN 变异系数相近; AP 和 AK 含量比较富足, 分 之间, 变异系数分别为 79.41%和 54.37%。

表 1 土壤氮磷钾含量描述性统计特征

Table 1 Descriptive statistical characteristics of soil N, P and K contents

土壤属性 Soil properties	全氮 TN	全磷 TP	全钾 TK	碱解氮 AN	有效磷 AP	速效钾 AK
最大值 Max	3.51	10.73	41.37	475.9	251.8	1 480
最小值 Min	0.55	0.11	2.17	36.75	0.26	40.00
平均值 Mean	1.73	0.91	14.79	150.8	32.30	349.0
标准差 Std	0.46	0.56	7.46	43.73	25.65	189.7
变异系数 CV/%	26.85	62.29	50.44	29.00	79.41	54.37

注: TN、TP 和 TK 的单位为 g·kg⁻¹, AN、AP 和 AK 的单位为 mg·kg⁻¹。Note: The units of TN, TP and TK are g·kg⁻¹, and the units of AN, AP and AK are mg·kg⁻¹.

表 2 土壤氮磷钾分级标准及各等级所占比重

Table 2 Grading standards of soil N, P and K contents and the proportion of each grade

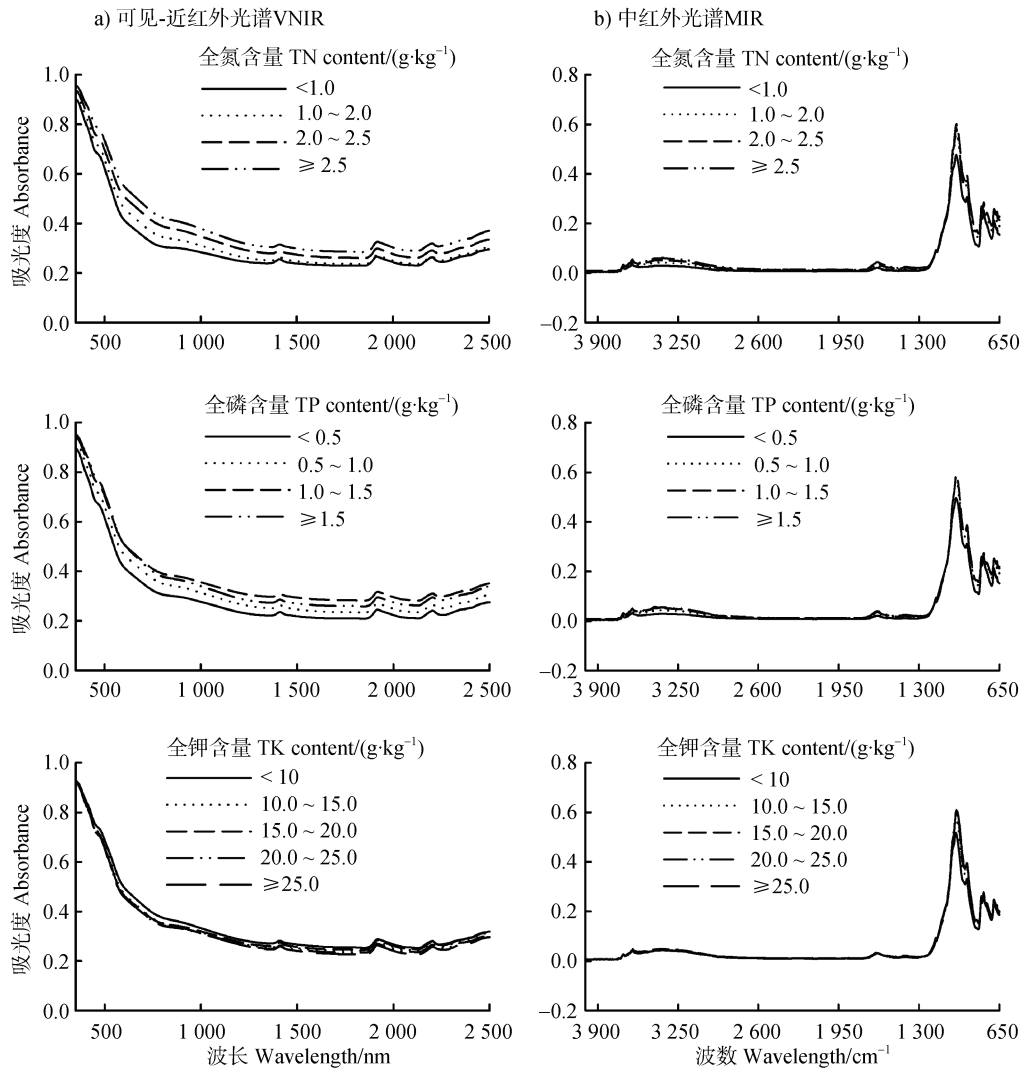
分级指标 Grading index	全氮 TN		全磷 TP		全钾 TK	
	分级 Grading/ (g·kg ⁻¹)	所占比重 proportion/%	分级 Grading/ (g·kg ⁻¹)	所占比重 proportion/%	分级 Grading/ (g·kg ⁻¹)	所占比重 proportion/%
极低 Very low	<0.5	0	<0.5	3.97	<10	14.10
低 Low	0.5~1.0	2.73	0.5~1.0	55.20	10~15	21.44
中 Middle	1.0~2.0	61.22	1.0~1.5	27.48	15~20	23.37
高 High	2.0~2.5	27.92	≥1.5	13.36	20~25	19.25
极高 Very high	≥2.5	8.13			≥25	21.84

分级指标 Grading index	碱解氮 AN		有效磷 AP		速效钾 AK	
	分级 Grading/ (mg·kg ⁻¹)	所占比重 proportion/%	分级 Grading/ (mg·kg ⁻¹)	所占比重 proportion/%	分级 Grading/ (mg·kg ⁻¹)	所占比重 proportion/%
极低 Very low	<65	0.26	<10	2.18	<80	0.54
低 Low	65~100	5.07	10~15	4.16	80~150	3.18
中 Middle	100~180	61.71	15~30	25.66	150~220	8.06
高 High	180~240	28.33	30~40	13.53	220~350	25.26
极高 Very high	≥240	4.63	≥40	54.47	≥350	63.96

2.2 土壤光谱曲线特征

为探究不同 N、P、K 含量土壤光谱曲线特征的差异, 依据表 2 中 TN、TP 和 TK 的分级标准, 统计不同等级 TN、TP 和 TK 含量的平均光谱吸光度曲线 (图 2)。

图 2a 为可见-近红外光谱区域的吸收曲线, 随波长的增加, 在可见光波段土壤光谱吸光度迅速降低; 在近红外波段光谱吸光度较低, 随波长的增大吸光度变化平缓, 逐渐缓慢上升。各样本的光谱曲线形态和吸收特征位置大致相似, 均在吸收光谱的



注：全氮含量仅有四个等级，不存在极低含量。Note: There are only four levels of TN content and no very low level.

图2 不同氮磷钾含量的吸收光谱曲线

Fig. 2 Absorption spectral curves of different N, P and K contents

1 400、1 900、2 200 nm 附近出现不同幅度的吸收峰，这主要是由于水和黏土矿物引起的特征吸收带^[20]。分析可见-近红外光谱吸收曲线变化规律发现，TN 含量升高，光谱反射率降低，这与土壤有机质的光谱特性类似，其原因在于 TN 和有机质之间密切相关，相关系数达到 0.86，与陈颂超等^[21]的研究结果一致；对于 TP 而言，TP 含量最低时 ($<0.5 \text{ g}\cdot\text{kg}^{-1}$) 所对应的吸光度也最低，但含量在 $1.0\sim 1.5 \text{ g}\cdot\text{kg}^{-1}$ 的平均吸光度要略高于 TP 含量最高的等级 ($\geq 1.5 \text{ g}\cdot\text{kg}^{-1}$)；对于 TK 而言，不同等级 TK 含量对应的吸收度平均值较为接近，差异不如 TN 和 TP 明显，但 TK 含量最低时 ($<10.0 \text{ g}\cdot\text{kg}^{-1}$) 所对应的吸光度最高，含量最

高时 ($\geq 25.0 \text{ g}\cdot\text{kg}^{-1}$) 所对应的吸光度最低。

图2b为中红外光谱区域的吸收曲线，在 $4\ 000\sim 1\ 250 \text{ cm}^{-1}$ 范围内，光谱的吸光度变化较为平缓，在 $3\ 300 \text{ cm}^{-1}$ 和 $1\ 630 \text{ cm}^{-1}$ 附近出现小幅度的吸收峰，吸光度在 $1\ 000 \text{ cm}^{-1}$ 附近快速升高后急剧下降，在 910 cm^{-1} 和 810 cm^{-1} 附近形成反射谷后进行小幅度波动，最后在 780 cm^{-1} 附近形成吸收峰，在 730 cm^{-1} 附近形成反射谷。从已有研究可知， $3\ 300 \text{ cm}^{-1}$ 附近波段变化可能与 C-H 的伸缩振动和 N-H 的弯曲振动有关； $1\ 630 \text{ cm}^{-1}$ 附近的吸收峰可能与土壤中的束缚水有关； $1\ 000 \text{ cm}^{-1}$ 附近的吸收峰可能与黏土矿物相关； 910 cm^{-1} 和 810 cm^{-1} 附近的波段变化可能受到

黏土矿物(高岭石、蒙脱石)和石英的影响; 780 和 730 cm^{-1} 附近波段变化可能由 N-H 弯曲振动引起^[22]。相比于可见-近红外光谱, 不同 N、P、K 含量等级下, 中红外光谱吸收曲线的变异较小, 尤其针对 TK, 其不同含量的吸光度曲线几乎重合, 难以从视觉上区分; TN 和 TP 在 3 900~2 600 和 1 700~1 300 cm^{-1} 范围内

各含量对应光谱吸光度有较大的变异。

2.3 土壤养分预测模型的构建与验证

表 3 列举了基于可见-近红外和中红外光谱建立的 PLSR 和 SVM 模型的预测结果, 本研究将比较不同光谱及不同模型的预测精度, 以期筛选预测不同养分的最优预测模型。

表 3 基于可见-近红外和中红外光谱的 PLSR 和 SVM 模型的构建与验证

Table 3 Construction and validation of PLSR and SVM models based on VNIR and MIR spectra

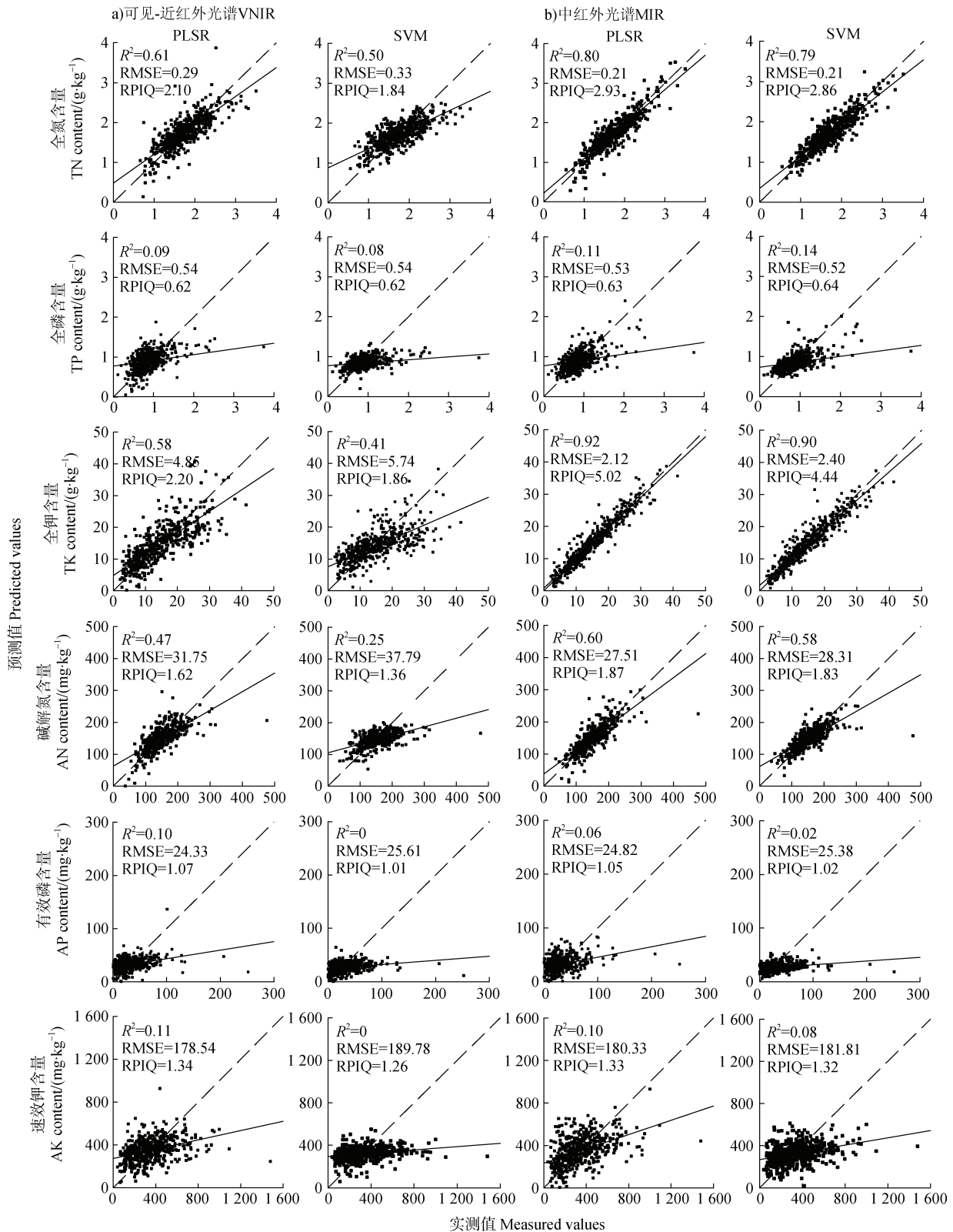
光谱 Spectra	模型 Model	精度指标 Precision index	全氮	全磷	全钾	碱解氮	有效磷	速效钾
			TN/ ($\text{g}\cdot\text{kg}^{-1}$)	TP/ ($\text{g}\cdot\text{kg}^{-1}$)	TK/ ($\text{g}\cdot\text{kg}^{-1}$)	AN/ ($\text{mg}\cdot\text{kg}^{-1}$)	AP/ ($\text{mg}\cdot\text{kg}^{-1}$)	AK/ ($\text{mg}\cdot\text{kg}^{-1}$)
VNIR	PLSR	R^2	0.61	0.09	0.58	0.47	0.10	0.11
		RMSE	0.29	0.54	4.85	31.75	24.33	178.5
		RPIQ	2.10	0.62	2.20	1.62	1.07*	1.34*
	SVM	R^2	0.50	0.08	0.41	0.25	0	0
		RMSE	0.33	0.54	5.74	37.79	25.61	189.8
		RPIQ	1.84	0.62	1.86	1.36	1.01	1.26
MIR	PLSR	R^2	0.80	0.11	0.92	0.60	0.06	0.10
		RMSE	0.21	0.53	2.12	27.51	24.82	180.3
		RPIQ	2.93*	0.63	5.02*	1.87*	1.05	1.33
	SVM	R^2	0.79	0.14	0.90	0.58	0.02	0.08
		RMSE	0.21	0.52	2.40	28.13	25.38	181.8
		RPIQ	2.86	0.64*	4.44	1.83	1.02	1.32

注: VNIR: 可见-近红外光谱; MIR: 中红外光谱; PLSR: 偏最小二乘回归; SVM: 支持向量机; R^2 : 决定系数; RMSE: 均方根误差; RPIQ: 性能与四分位间距之比。该表仅展示各土壤养分预测模型的精度指标数值, 可见-近红外光谱中 SVM 模型预测 AP 和 AK 的预测精度为 $0 < R^2 < 0.005$ 。“*”表示不同养分含量所取得的最高精度预测模型的 RPIQ, 表中的预测精度均为留一法交叉验证精度。下同。Note: PLSR: partial least squares regression; SVM: support vector machine; R^2 : coefficient of determination; RMSE: root mean square error; RPIQ: ratio of performance to interquartile distance. The table only shows the accuracy index values of each soil nutrient prediction model. The prediction accuracy of the SVM model for AP and AK in visible-near infrared spectroscopy was $0 < R^2 < 0.005$. The RPIQ of the prediction model with the highest accuracy obtained by different nutrient contents was shown in the table “*”, and the prediction accuracy in the table was the accuracy of the leave-one-out cross-validation. The same as below.

2.3.1 可见-近红外光谱的建模 对于 TN 和 TK, PLSR 模型的预测精度 RPIQ 分别为 2.10、2.20, SVM 模型的预测精度 RPIQ 分别为 1.84、1.86, PLSR 模型的预测能力更优; 对于 TP, PLSR 和 SVM 模型的预测精度均较低, RPIQ 接近 0.62。比较 N、P、K 有效态的预测精度, 其中以 AN 的预测精度最高, PLSR 模型的预测精度高于 SVM 模型的预测精度, RPIQ 分别为 1.62、1.36; 对于 AP 和 AK, PLSR 和 SVM 模型的预测结果相似, RPIQ 分别为 1.01~1.07、1.26~1.34。

基于可见-近红外光谱所构建的模型中, N、P、

K 全量和有效态的预测精度由高到低依次为: $\text{TK} > \text{TN} > \text{AN} > \text{AK} > \text{AP} > \text{TP}$, PLSR 模型能够对 TN 和 TK 进行可靠的预测 ($\text{RPIQ} > 2.00$), 但 TP 以及 N、P、K 有效态的预测效果均不佳 ($\text{RPIQ} < 1.62$)。有机质在可见-近红外波段内存在不同的倍频峰和合频峰, 具有很强的光谱相应特征, 本研究中 TN 和有机质相关性较高, 这应该是 TN 预测效果较好的原因^[21]。李颖等^[23]利用可见-近红外光谱对 TN 和 TK 的预测取得了相似的结果, Debaene 等^[24]对 AP 和 AK 的预测精度也均较低, R^2 不超过 0.5。依据散点图 3a, PLSR 模型的拟合曲线相比于 SVM 模型更接



注: 实线是散点图拟合线, 虚线是实测值和预测值之间的相等线。Note: The solid line is the fitted line of the scatter plot, and the dashed line is the line of equality between the measured and predicted values.

图3 基于可见-近红外 (a) 和中红外 (b) 光谱的氮磷钾预测模型 PLSR 和 SVM 结果比较

Fig. 3 Comparison of PLSR and SVM models for predicting N, P and K based on VNIR (a) and MIR (b) spectra

近 1:1 线, PLSR 模型的预测能力整体优于 SVM 模型, 该结论可能与样本集的划分法或光谱的预处理方法有关^[2], 何积秀^[22]也发现 PLSR 模型预测 AN 要优于 SVM 模型。

2.3.2 中红外光谱的建模 基于中红外光谱所构建的土壤养分预测模型中, N、P、K 全量的预测精度有大幅提升, TK 预测精度最高, PLSR 和 SVM 模型的预测精度 RPIQ 分别为 5.02、4.44; 其次为 TN, 预测精度 RPIQ 分别为 2.93、2.86; TP 的预测精度最低, PLSR 和 SVM 的预测精度相近, RPIQ 分别为 0.63、0.64。对于 N、P、K 有效态含量的预测, AN 预测精度最高, PLSR 和 SVM 模型的预测精度相近, RPIQ 分别为 1.87、1.83; 对于 AP 和 AK, 中红外光谱的预测精度未见提升, PLSR 和 SVM 的预测结果相似, RPIQ 分别为 1.02~1.05、1.32~1.33。

N、P、K 全量和有效态的预测精度由高到低依次为: TK>TN>AN>AK>AP>TP, 基于中红外光谱, PLSR 和 SVM 模型对 TN 和 TK 均能进行可靠的预测, 对 AN 的预测较为可靠 ($2.0>RPIQ>1.80$), 对 TP 和 N、P、K 有效态的预测效果较差。依据散点图 3b, PLSR 模型的拟合曲线相比于 SVM 模型更接近 1:1 线, 与可见-近红外光谱的预测结果类似, PLSR 模型的预测能力也整体优于 SVM 模型。

2.3.3 可见-近红外与中红外光谱模型的比对 无论基于可见-近红外光谱还是中红外光谱, PLSR 模型的预测精度整体优于 SVM 模型。因此, 以下对可见-近红外光谱与中红外光谱预测能力的比较将仅针对 PLSR 模型。

针对 TN、TK 和 AN, 相比于可见-近红外光谱, 中红外光谱的预测精度均有显著提升; 其中, TK 的预测精度提升幅度最大, RMSE 由 $4.85 \text{ g}\cdot\text{kg}^{-1}$ 减少至 $2.12 \text{ g}\cdot\text{kg}^{-1}$, R^2 由 0.58 提高至 0.92, RPIQ 由 2.20 提高至 5.02; 其次为 TN, RMSE 由 $0.29 \text{ g}\cdot\text{kg}^{-1}$ 减少至 $0.21 \text{ g}\cdot\text{kg}^{-1}$, R^2 由 0.61 提高至 0.80, RPIQ 由 2.10 提高至 2.93; AN 的预测精度提升幅度最小, RMSE 由 $31.75 \text{ mg}\cdot\text{kg}^{-1}$ 减少至 $27.51 \text{ mg}\cdot\text{kg}^{-1}$, R^2 由 0.47 提高至 0.60, RPIQ 由 1.62 提高至 1.87。中红外光谱可能由于 CO、CH 等官能团及光谱活跃成分 (有机质、碳酸盐、黏土矿物等) 较多, 预测 TN、TK 和 AN 的模型性能得到明显改善, 这与 Ng 等^[25]发现中红外光谱预测土壤属性的能力更优的结论一致。对于 TN 和 TK, 可见-近红外和中红外光谱均能可靠

地预测 ($RPIQ>2.00$); 对于 AN, 中红外光谱的预测精度有所提高, 说明土壤中红外光谱对 AN 含量进行预测是有可能的 ($RPIQ>1.80$)。

对于 TP、AP 和 AK, 可见-近红外和中红外光谱的预测效果均较差, 预测精度的变化幅度不大。相比于可见-近红外光谱, 中红外光谱对 TP 的预测精度有提升, RMSE 由 $0.54 \text{ g}\cdot\text{kg}^{-1}$ 减少至 $0.53 \text{ g}\cdot\text{kg}^{-1}$, R^2 由 0.09 提高至 0.11, RPIQ 由 0.62 提高至 0.63; 对 AP 和 AK 的预测精度略微降低, AP 的 RMSE 由 $24.33 \text{ mg}\cdot\text{kg}^{-1}$ 提高至 $24.82 \text{ mg}\cdot\text{kg}^{-1}$, R^2 由 0.10 减少至 0.06, RPIQ 由 1.07 减少至 1.05; AK 的 RMSE 由 $178.5 \text{ mg}\cdot\text{kg}^{-1}$ 提高至 $180.3 \text{ mg}\cdot\text{kg}^{-1}$, R^2 由 0.11 减少至 0.10, RPIQ 由 1.34 减少至 1.33。

利用可见-近红外和中红外光谱预测 TP 和 AP 的精度均较低, 可能的原因是磷在这两类光谱区间均无特定的光谱吸收带。Mouazen 等^[26]和 Wijewardane 等^[27]分别发现磷在可见-近红外和中红外光谱中无直接的光谱响应, 需要借助与其相关性高的光谱活性成分 (有机物、碳酸盐、黏土等) 建立间接的光谱预测模型, 但在不同的研究中, 与磷相关的因子并不固定。因此, 磷不易像 TN 和 AN 那样能被较好地预测。TK 在可见-近红外和中红外光谱中预测精度均较高, 而 AK 的预测精度均较低, 预测结果差异较大的原因可能是土壤 TK 和 AK 的组成和含量不同^[28]。

2.4 土壤养分的可见-近红外和中红外光谱预测重要波段

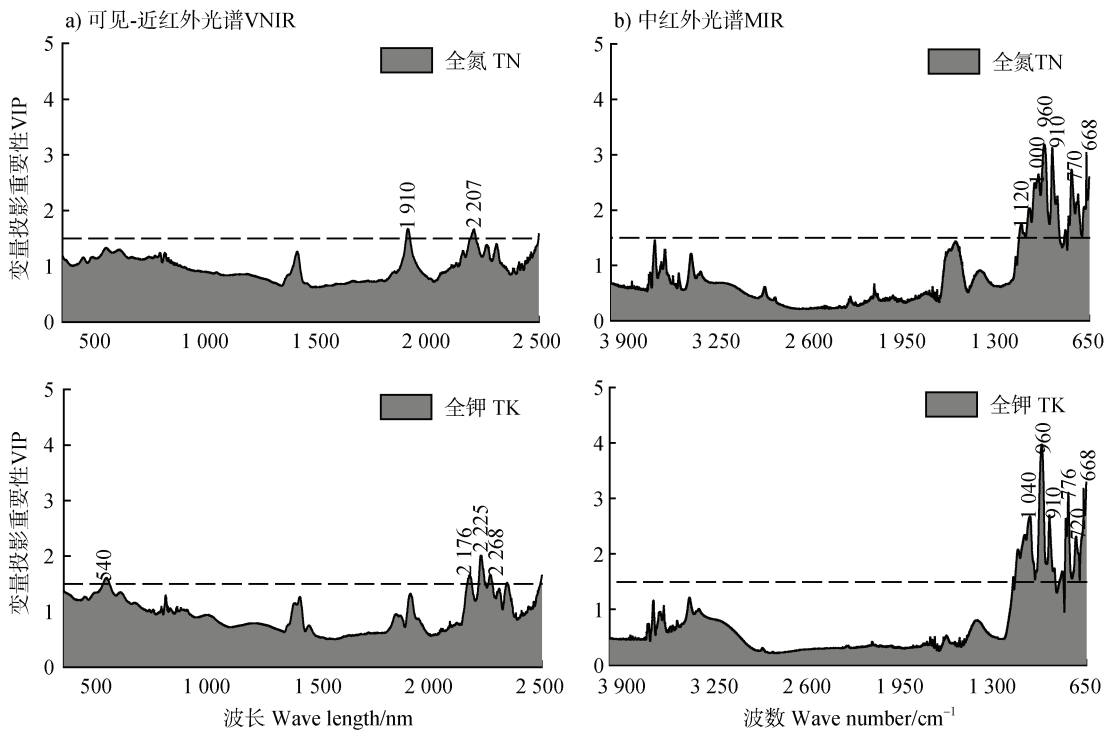
在本研究中 TN 和 TK 的预测精度较高, 而 TP 和 N、P、K 有效态的预测精度低。因此 PLSR 模型的重要波段分析以 TN 和 TK 为例。如图 4 所示, 当 $VIP>1$ 时, TK 和 TN 在可见-近红外和中红外光谱区域内的重要波段入选较多, 而当 $VIP>1.5$ 时, 其重要波段减少, 但对于预测贡献更大, 因此本研究将 VIP 得分的阈值提高为 1.5^[29]。

在可见-近红外光谱区域 (图 4a), 预测 TN 的重要波段主要集中于 1 910 和 2 207 nm 附近, TK 的重要波段主要集中于 540、2 176、2 225 和 2 268 nm 附近。本研究中可见-近红外光谱区域 TN 的重要波段与陈颂超等^[21]认为的 1 400 和 2 200 nm 的结论相似, TK 的重要波段与张超等^[11]认为的 530、617、1 988 和 2 208 nm 的结论比较一致。重要波段 540 nm 可能与土壤颜色或赤铁矿有关^[19], 1 910、2 207、2 176 和 2 225 nm 可能与土壤水分子和黏土矿物有关^[30],

2 268 nm 可能与 C-H 伸缩振动有关^[29]。

在中红外光谱区域内 (图 4b), 预测 TN 和 TK 的重要波段多于可见-近红外光谱区域。其中, 预测 TN 的重要波段主要集中在 1 120、1 000、960、910、770 和 668 cm^{-1} 附近, TK 的重要波段主要集中于 1 040、960、910、776、720 和 668 cm^{-1} 附近。TN 和 TK 的重要波段重叠较多, 与何积秀^[22]和陈颂超

等^[29]认为有机质的重要波段在 1 220、1 030、810 cm^{-1} 的结论相近。重要波段 1120 cm^{-1} 可能与 C-O 伸缩振动有关^[29], 1 040、1 000 和 960 cm^{-1} 可能与黏土矿物中的 Si-O 伸缩振动有关^[31], 910 cm^{-1} 可能与高岭石、三水铝石等黏土矿物的 O-H 弯曲振动有关^[32], 776、770、720 和 668 cm^{-1} 可能与 C-H 弯曲振动有关或受 N-H 影响^[22, 33]



注: 灰色虚线 VIP=1.5, 为本研究设定阈值。Note: Black dotted line VIP = 1.5, set threshold for this study.

图 4 基于 PLSR 模型的土壤全氮和全钾的变量投影重要性 (VIP) 得分

Fig. 4 Variable in the projection (VIP) scores of soil TN and TK based on the PLSR model

3 结论

本文采用 PLSR 和 SVM 两种方法建立土壤 N、P、K 含量的光谱预测模型, 探讨可见-近红外和中红外光谱技术对预测模型精度的影响。结果发现利用可见-近红外和中红外光谱对 TN 和 TK 含量预测是可行的, 就建模结果而言, PLSR 预测模型整体上优于 SVM 预测模型, 中红外光谱预测 TN、TK 和 AN 精度均较可见-近红外光谱显著提升, 两类光谱均对 TN 和 TK 的预测精度较高, 其中 MIR-PLSR-TK 模型的预测精度最佳, 但仅有中红外光谱可较可靠地预测 AN。而对于 TP、AP 和 AK 的预测效果而言,

两类光谱的预测能力均较差。结合可见-近红外和中红外光谱的最优预测模型筛选预测 TN 和 TK 的重要波段, 两类光谱在中红外光谱区域预测 TN 和 TK 的重要波段多于可见-近红外光谱区域。因此, 建议一是采用中红外光谱技术结合 PLSR 模型来实现对土壤 TN 和 TK 的预测; 二是针对土壤 TP 和 N、P、K 有效态的预测精度较低的问题, 进一步尝试提取 N、P、K 主成分或特征波段, 通过光谱融合及探索其他模型等提升这些属性的预测精度。

参考文献 (References)

- [1] Gao X S, Xiao Y, Deng L J, et al. Spatial variability of soil total nitrogen, phosphorus and potassium in Renshou

- County of Sichuan Basin, China[J]. *Journal of Integrative Agriculture*, 2019, 18 (2): 279—289.
- [2] Santos U J D, Demattê J A D M, Menezes R S C, et al. Predicting carbon and nitrogen by visible near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy in soils of Northeast Brazil[J]. *Geoderma Regional*, 2020, 23: e00333.
- [3] Li W, Zhang S H, Zhang Q, et al. Rapid prediction of available N, P and K content in soil using near-infrared reflectance spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2007, 23 (1): 55—59. [李伟, 张书慧, 张倩, 等. 近红外光谱法快速测定土壤碱解氮、速效磷和速效钾含量[J]. *农业工程学报*, 2007, 23 (1): 55—59.]
- [4] Wu Q, Yang Y H, Xu Z L, et al. Applying local neural network and visible/near-infrared spectroscopy to estimating available nitrogen, phosphorus and potassium in soil[J]. *Spectroscopy and Spectral Analysis*, 2014, 34 (8): 2102—2105. [吴茜, 杨宇虹, 徐照丽, 等. 应用局部神经网络和可见/近红外光谱法估测土壤有效氮磷钾[J]. *光谱学与光谱分析*, 2014, 34 (8): 2102—2105.]
- [5] Yang J, Wang X, Wang R, et al. Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using vis-NIR spectroscopy[J]. *Geoderma*, 2020, 380: 114616.
- [6] Devianti D, Sufardi S, Bulan R, et al. Vis-NIR spectra combined with machine learning for predicting soil nutrients in cropland from Aceh Province, Indonesia[J]. *Case Studies in Chemical and Environmental Engineering*, 2022, 6: 100268.
- [7] Yang J B. Application of infrared spectroscopy in environmental monitoring[J]. *China Resources Comprehensive Utilization*, 2020, 38 (4): 148—150. [杨家宝. 红外光谱技术在环境监测中的应用[J]. *中国资源综合利用*, 2020, 38 (4): 148—150.]
- [8] Jiang L L, Zhang Y, Wang Y Y, et al. Fast determination of nutritional parameters in soil based on spectroscopic techniques[J]. *Journal of Zhejiang University (Agriculture & Life Sciences)*, 2010, 36 (4): 445—450. [蒋璐璐, 张瑜, 王艳艳, 等. 基于光谱技术的土壤养分快速测试方法研究[J]. *浙江大学学报(农业与生命科学版)*, 2010, 36 (4): 445—450.]
- [9] Greenberg I, Seidel M, Vohland M, et al. Performance of in situ vs laboratory mid-infrared soil spectroscopy using local and regional calibration strategies[J]. *Geoderma*, 2022, 409: 115614.
- [10] Hati K M, Sinha N K, Mohanty M, et al. Mid-infrared reflectance spectroscopy for estimation of soil properties of Alfisols from eastern India[J]. *Sustainability*, 2022, 14 (9): 4883.
- [11] Zhang C, Liu Y M, Sun Y N, et al. Hyperspectral prediction model of soil nutrient content in the loess hilly-gully region, China[J]. *Chinese Journal of Applied Ecology*, 2018, 29 (9): 2835—2842. [张超, 刘咏梅, 孙亚楠, 等. 黄土丘陵沟壑区土壤养分含量的高光谱预测模型[J]. *应用生态学报*, 2018, 29 (9): 2835—2842.]
- [12] Chen Q X, Lu X H, Zhao W X, et al. Spatial differentiation of main soil types in Guizhou Province under different topographical conditions[J]. *Jiangsu Agricultural Sciences*, 2022, 50 (7): 205—213. [陈清霞, 陆晓辉, 赵文轩, 等. 不同地形条件下贵州省主要土壤类型的空间分异[J]. *江苏农业科学*, 2022, 50 (7): 205—213.]
- [13] Zhang G L, Gong Z T. *Soil survey laboratory methods*[M]. Beijing: Science Press, 2012. [张甘霖, 龚子同. *土壤调查实验室分析方法*[M]. 北京: 科学出版社, 2012.]
- [14] Tiecher T, Moura-Bueno J M, Caner L, et al. Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet-visible, near- and middle-infrared spectroscopy[J]. *Geoderma*, 2021, 384: 114815.
- [15] de Santana F B, Otani S K, De Souza A M, et al. Comparison of PLS and SVM models for soil organic matter and particle size using vis-NIR spectral libraries[J]. *Geoderma Regional*, 2021, 27: e00436.
- [16] Chauhan R, Kumar R, Kumar V, et al. On the discrimination of soil samples by derivative diffuse reflectance UV-vis-NIR spectroscopy and chemometric methods[J]. *Forensic Science International*, 2020, 319: 110655.
- [17] Tang H T, Meng X T, Su X X, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37 (2): 105—113. [唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. *农业工程学报*, 2021, 37 (2): 105—113.]
- [18] 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.
- [19] Qi H J, Li S W, Karnieli A, et al. Prediction method of soil available phosphorus using hyperspectral data based on PLS-BPNN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49 (2): 166—172. [齐海军, 李绍稳, Karnieli A, 等. 基于 PLS-BPNN 算法的土壤速效磷高光谱回归预测方法[J]. *农业机械学报*, 2018, 49 (2): 166—172.]
- [20] Zhao M S, Xie Y, Lu L M, et al. Modeling for soil organic matter content based on hyperspectral feature indices[J]. *Acta Pedologica Sinica*, 2021, 58 (1): 42—54.

- [赵明松, 谢毅, 陆龙妹, 等. 基于高光谱特征指数的土壤有机质含量建模[J]. 土壤学报, 2021, 58 (1): 42—54.]
- [21] Chen S C, Feng L L, Li S, et al. Vis-NIR spectral inversion for prediction of soil total nitrogen content in laboratory based on locally weighted regression[J]. *Acta Pedologica Sinica*, 2015, 52 (2): 312—320. [陈颂超, 冯来磊, 李硕, 等. 基于局部加权回归的土壤全氮含量可见-近红外光谱反演[J]. 土壤学报, 2015, 52 (2): 312—320.]
- [22] He J X. Prediction of fertility factors in paddy soil using VNIR and MIR spectroscopy[D]. Hangzhou: Zhejiang University, 2015. [何积秀. 基于 VNIR 和 MIR 光谱的水稻田土壤肥力因子预测[D]. 杭州: 浙江大学, 2015.]
- [23] Li J, Zhang X C, Yuan Y W, et al. Analysis of soil nutrient content based on near infrared reflectance spectroscopy in Beijing region[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28 (2): 176—179. [李颖, 张小超, 苑严伟, 等. 北京典型耕作土壤养分的近红外光谱分析[J]. 农业工程学报, 2012, 28 (2): 176—179.]
- [24] Debaene G, Niedźwiecki J, Pecio A, et al. Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale[J]. *Geoderma*, 2014, 214/215: 114—125.
- [25] Ng W, Minasny B, Montazerolghaem M, et al. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra[J]. *Geoderma*, 2019, 352: 251—267.
- [26] Mouazen A M, Kuang B, de Baerdemaeker J, et al. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy[J]. *Geoderma*, 2010, 158 (1/2): 23—31.
- [27] Wijewardane N K, Ge Y, Wills S, et al. Predicting physical and chemical properties of US soils with a mid-infrared reflectance spectral library[J]. *Soil Science Society of America Journal*, 2018, 82 (3): 722—731.
- [28] Liu X Y, Shi Z Y, Chang Q R, et al. Hyperspectral model for estimation of soil potassium content in loessal soil[J]. *Acta Pedologica Sinica*, 2018, 55 (2): 325—337. [刘秀英, 石兆勇, 常庆瑞, 等. 黄绵土钾含量高光谱估算模型研究[J]. 土壤学报, 2018, 55 (2): 325—337.]
- [29] Chen S C, Peng J, Ji W J, et al. Study on the characterization of VNIR-MIR spectra and prediction of soil organic matter in paddy soil[J]. *Spectroscopy and Spectral Analysis*, 2016, 36 (6): 1712—1716. [陈颂超, 彭杰, 纪文君, 等. 水稻土可见-近红外-中红外光谱特性与有机质预测研究[J]. 光谱学与光谱分析, 2016, 36 (6): 1712—1716.]
- [30] He J L, Cui J L, Zhang S Y, et al. Hyperspectral estimation of heavy metal Cu content in soil based on partial least square method[J]. *Remote Sensing Technology and Application*, 2019, 34 (5): 998—1004. [贺军亮, 崔军丽, 张淑媛, 等. 基于偏最小二乘的土壤重金属铜含量高光谱估算[J]. 遥感技术与应用, 2019, 34 (5): 998—1004.]
- [31] Liu C Y, Zhang J R, Zeng T, et al. Determination of soil organic carbon and total nitrogen contents in aggregate fractions from Fourier transform infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2020, 40 (12): 3818—3824. [刘翠英, 张津瑞, 曾涛, 等. 傅里叶变换红外光谱的土壤团聚体有机碳和全氮含量估测[J]. 光谱学与光谱分析, 2020, 40 (12): 3818—3824.]
- [32] Hao X X, Han X Z, Zou W X. Studies on composition of soil organic matter by Fourier transform infrared spectroscopy differential analysis[J]. *Chinese Journal of Analytical Chemistry*, 2018, 46 (4): 616—622. [郝翔翔, 韩晓增, 邹文秀. 示差红外光谱在土壤有机质组成研究中的应用[J]. 分析化学, 2018, 46 (4): 616—622.]
- [33] Bernier M H, Levy G J, Fine P, et al. Organic matter composition in soils irrigated with treated wastewater: FT-IR spectroscopic analysis of bulk soil samples[J]. *Geoderma*, 2013, 209/210: 233—240.

(责任编辑: 陈荣府)