农田土壤有机碳高光谱特征及定量监测研究

杨林婧[a, b]，杨　莎[a, b]，张圣杨[a, b]，张文颜[a, b]，左圆琅[a, b]，闫碧瑶[a, b]，杨武德[a, b*]

（山西农业大学 a. 农学院；b. 智慧农业学院，太谷 030801）

摘　要：为实现对农田土壤有机碳含量的快速定量监测，采用 9 种常规预处理分析方法，优化原始光谱信息并分析各预处理光谱与小麦土壤有机碳含量之间的关系，利用连续投影算法（SPA）提取土壤有机碳的光谱特征，并建立基于全谱和光谱特征波段的两类土壤有机碳光谱监测模型。研究表明，与原始光谱相比，预处理光谱可以显著提高其与小麦土壤有机碳的相关性。同时，采用 SPA 方法提取并证实了光谱区域 400 ~ 450 nm、510 ~ 620 nm、1 010 ~ 1 060 nm、2 000 ~ 2 400 nm 的土壤有机碳含量等重要信息。对比两类模型表现可知，相同预处理光谱条件下，连续投影算法 - 多元线性回归（SPA-MLR）模型优于偏最小二乘回归（PLSR）模型，其中，基于多元散射校正（MSC）预处理光谱条件下的农田土壤有机碳含量估算模型整体表现最好（R_v^2=0.726、$RMSE_v$=0.109、RPD_v=1.956），且具有实践上的应用潜力。本研究证实，光谱预处理在一定程度上可以提高光谱反射率与小麦土壤有机碳含量的相关性，且影响监测模型表现，同时模型构建方法可能对模型估算精度产生更为积极的效果。本研究结果可以为农田土壤有机碳含量检测提供一定的理论依据和实践探索。

关键词：农田土壤；有机碳；预处理；高光谱；响应特性

中图分类号：S-3　　**文献标志码：**A　　**DOI：**10.3969/j.issn.1007-7146.2024.04.004

Hyperspectral Characteristics and Quantitative Monitoring of Soil Organic Carbon in Farmland

YANG Linjing[a, b], *YANG Sha*[a, b], *ZHANG Shengyang*[a, b],
ZHANG Wenyan[a, b], *ZUO Yuanquan*[a, b], *YAN Biyao*[a, b], *YANG Wude*[a, b*]

（Shanxi Agricultural University a. College of Agricultural; b. College of Smart Agriculture, Taigu 030801, China）

Abstract: In order to realize the rapid and quantitative monitoring of soil organic carbon content in farmland, nine conventional pretreatment analysis methods were used to optimize the original spectral information and analyze the relationship between each pretreatment spectrum and wheat soil organic carbon content, and the continuous projection algorithm (SPA) was used to extract the spectral characteristics of soil organic carbon, and two types of soil organic carbon spectral monitoring models based on full spectrum and spectral characteristic bands were established. The results showed that the correlation between the pretreatment spectrum and soil organic carbon in wheat could be significantly improved compared with the original spectrum. In the meantime, the SPA method was used to extract and confirm the important information of soil organic carbon content in the spectral regions of 400 ~ 450 nm, 510 ~ 620 nm, 1010 ~ 1 060 nm and 2 000 ~ 2 400 nm. Comparing the performance of the two types of models, it can be seen that the continuous projection algorithm-multiple linear regression (SPA-MLR) model is better than the partial least squares regression (PLSR) model under the same pretreatment spectra, and the overall performance of the soil organic carbon content estimation model based on multivariate scattering correction (MSC) pretreatment spectra is the

收稿日期：2024-01-25；**修回日期：**2024-03-25。

基金项目：2022 年大学生创新创业训练计划项目（202210113144）；山西省基础研究计划项目（202203021211275）；山西省研究生教育创新项目专项资金（2022Y312）；现代农产技术研究体系专项基金项目（2023CYJSTX02-23）；山西省关键技术研发计划项目（201903D211002）；国家自然科学基金项目（31871571，31371572）。

*** 通信作者：**杨武德，教授，主要从事 3S 技术与作物生产的研究。E-mail: sxauywd@126.com。

best (R_v^2=0.726, $RMSE_v$=0.109, RPD_v=1.956), and it has practical application potential. This study confirms that spectral pretreatment can improve the correlation between spectral reflectance and soil organic carbon content in wheat to a certain extent, and affect the performance of the monitoring model, and the model construction method may have a more positive effect on the accuracy of model estimation. The results of this study can provide a theoretical basis and practical exploration for the detection of soil organic carbon content in farmland.

Key words: farmland soil; organic carbon; pretreatment; hyperspectral; responsive features

（*Acta Laser Biology Sinica*, 2024, 33(4): 316-325）

农田土壤有机碳库因常年受到人为干扰而易发生剧烈变化，探索碳库形成及周转机制研究是实现我国“碳达峰、碳中和”战略的重要途径[1]。测定农田土壤有机碳库，不仅可明确农田土壤碳库的储量，评估其固碳潜力，也是联合国气候变化框架公约中确定各国温室气体净排放通量的一项具体工作[2-3]。基于光谱遥感技术开展土壤有机碳含量的定量估算被认为是未来进行实时、大尺度、快速估算的有效方法[4]。卫星遥感技术具有较强的空间性及连续性，但是访问周期短、空间分辨率低，而高光谱技术具有快速、高效、无损获取目标属性等优点，可以弥补这个缺陷，使分辨率分别达到厘米级和纳米级，但该技术也存在光谱信息冗余的缺点[5-6]。随着光谱技术的发展，光谱信息测量仪器不断更新，这一技术在农业领域也得到了广泛的应用。赵隽宇等[7]通过高光谱分析技术结合正交偏最小二乘回归分析（orthogonal partial least squares discriminant analysis, OPLS-DA）对桉树黄化叶片有了一定的预测。唐子竣等[8]基于地面高光谱遥感技术构建了大豆产量估算模型。此外，He 等[9]的研究表明，挖掘和提取目标属性的高光谱敏感特征波段有助于揭示和解释目标属性的高光谱响应机理，同时实现光谱数据降维，也是建立高精度模型和实践应用性模型的主要思路。

当前，应用较为广泛的变量选择方法和模型校正方法主要有连续投影算法（successive projections algorithm, SPA）[10]、遗传算法（genetic algorithm, GA）[11]、多元线性回归（multiple linear regression, MLR）[12]、逐步多元线性回归（stepwise multiple linear regression, SMLR）[13]、偏最小二乘回归（partial least squares regression, PLSR）[14]、无信息变量消除（uninformative variable elimination, UVE）和支持向量机（support vector machine, SVM）[15]等。也有部分学者根据化学计量学方法的不同功能，同时结合多种方法进行光谱特征信息提取和模型优化，如 UVE+SPA[16]、SPA-LS-SVM[17]、PLSR-SMLR[18]等。李阳阳等[19]的基于高光谱的大叶女贞叶片水分定量测定试验表明，通过 SPA，仅利用全波段 2.4% 的信息就能够达到与全波段建模相当的精度，大大简化了模型，预测平方相关系数为 0.97，预测均方根误差仅为 4.4×10^{-5}。邓志伟等[20]采用 PLSR 对稀土元素进行定量分析，结果表明，PLSR 定量分析模型的标准值和预测值相关曲线 R^2 得到了提高，该方法可以建立稳健的定量分析模型。徐冬等[21]在基于多光谱检测技术研究茄子果皮花青素含量的检测的试验中，证明了 MLR 校正模型具有较高的预测精度，且模型准确率和可靠性均较高。此外，对光谱数据进行预处理可以提高光谱的信噪比。马玮键等[22]对光谱数据进行了平滑（Savitzky-Golay, SG）、基线校正（baseline）、去趋势处理（detrend）、归一化（normalize）、多元散射校正（multivariate scattering correction, MSC）和标准正态变换（standard normal variate, SNV）6 种预处理，发现基于 SG 预处理并结合 PLSR 模型的估测效果最佳。因此，不同化学计量方法和高光谱技术的结合运用在土壤有机碳含量估测的方面展现了较为广阔的应用，为准确估测土壤有机碳提供了坚实的基础。

综上，土壤的差异、不同的光谱预处理方法、波长选择方法以及估测模型方法不同等，都会使得土壤有机碳含量的估测效果不同。所以，为了实现对区域农田土壤有机碳含量的监测和预测，本试验以小麦小区试验土壤为研究对象，采用 9 种不同的常规预处理分析方法，优化原始光谱信息并分析各预处理光谱与土壤有机碳含量之间的关系，利用 SPA 提取土壤有机碳的光谱特征，建立并比较基于全谱和光谱特征波段的两种土壤有机碳光谱监测模型，旨在监测农田土壤有机碳含量的高光谱特征及响应特性，挖掘和确定土壤有机碳影响因子的高光谱特征信息，构建和优化光谱模型。

1 材料与方法

1.1 试验设计

本试验地点位于山西农业大学农学院实验站（E 112°34′12″, N 37°25′21″），试验地是暖温带大陆性气候，年平均气温为 9.8℃，无霜期为 175 d，年降雨量约为 460 mm。试验地内每块试验区的面积均为 9 m^2（3 m × 3 m），试验小麦（*Triticum aestivum*）品种为长 6878（旱）、长 4738（水），总灌溉量为 350 mm。1 个试验区同时种植 2 个品种，行距为 20 cm。每个试验区在播前施肥，在不追肥的情况下施用氮肥 227.85 kg/hm^2、钾肥 655.5 kg/hm^2、磷肥 115.5 kg/hm^2。根据田间持水量的百分比控制水分，并进行随机试验，于拔节期 ~ 抽穗期、开花期 ~ 灌浆期 2 个生长阶段设置 1 个对照（control check，CK）。根据田间持水量对土壤水分进行控制，田间持水量的 65% 为轻度，55% 为中度，45% 为重度，以田间持水量的 75% 为对照，共 7 个处理，3 次重复。试验各处理如表 1 所示。

1.2 样本数据采集

1.2.1 土壤数据采集

采集 0 ~ 10 cm、10 ~ 20 cm、20 ~ 40 cm、40 ~ 60 cm 4 个土层的土壤样本，装入密封袋带回实验室。去除杂质后放在室内阴凉通风处进行风干，风干后过 2 mm 孔筛。将过筛土壤样本充分混合后，在室内测定每个样本的土壤光谱。研究表明，重铬酸钾外加热容量法测定的总有机碳值与土壤原有机碳值非常接近[23]，因此，本研究使用该方法测定土壤样本有机碳的含量。

表 1 水分处理表

Tab. 1 Moisture treatment table

Treatments	Jointing stage/% ~ Heading stage/%	Flowering stage/% ~ Filling stage/%
T1	45 ± 5	75 ~ 80
T2	55 ± 5	75 ~ 80
T3	65 ± 5	75 ~ 80
T4	75 ~ 80	45 ± 5
T5	75 ~ 80	55 ± 5
T6	75 ~ 80	65 ± 5
CK	75 ~ 80	75 ~ 80

1.2.2 光谱数据采集

使用美国 Analytical Spectral Devices 公司生产的 FieldSpec Pro FR250 型便携式光谱仪采集光谱数据，波长范围为 350 ~ 2 500 nm，其中，350 ~ 1 000 nm 光谱采样间隔为 1.4 nm，分辨率为 3 nm；1 000 ~ 2 500 nm 光谱采样间隔为 2 nm，分辨率为 10 nm。每个土壤样本分别采集 50 条光谱数据，进行算术平均后得到该样本的最终光谱反射率数据。获取土壤碳含量及光谱数据后剔除异常值及缺失值，共得到有效样本 164 个。采用建模集与验证集 3:1 的比例进行划分，得到建模集样本 123 个，验证集样本 41 个。

1.3 光谱数据处理

1.3.1 数据预处理

对原始光谱数据进行去噪和平均处理后得出原始光谱数据。在光谱数据采集过程中，光谱数据容易受到光谱测量仪器和光谱测量环境的影响，可能会降低后期建模的准确性，因此需要对其进行预处理。由于在光谱边缘有较大的噪声影响，故剔除 350 ~ 399 nm 和 2 451 ~ 2 500 nm 波段范围内的光谱数据。对光谱数据进行以下预处理：SG、一阶导数（first derivative，FD）、二阶导数（second derivative，SD）、倒数（reciprocal，RE）、平方（square，SQ）、三次方（cube，CU）、RE+FD、MSC 和 SNV。

SG 的基本原理：首先设定一个窗口，将该窗口中全部的数据相加，再除以窗口中的数据点的数量来计算其平均值，其实质是一种通过加权平均法来强调中心点的中心作用[24]。导数变换[25]是较为常用的光谱预处理方法，可去除不同的背景噪声。根据前人研究，FD 能有效减小部分线性或近线性干扰的影响[26]；SD 对于去除背景干扰、提高模型分析精度有重要作用[27]；RE、SQ 和 CU 处理这些简单数学变换可以在一定程度上提高数据与光谱的相关性[28]；MSC 后获得的光谱数据能够有效地去除因

散射强度变化引起的光谱差异，提高数据与光谱的关联程度[29]；SNV 一般用于克服和消除固体颗粒大小、表面散射及光程变化造成的光谱干扰[30]。

1.3.2　特征波段选择

SPA 是一种使矢量空间共线性最小化的前向变量选择算法，其目的是处理模型的共线性问题，可有效减少数据变量，提高运算效率，改善建模预测效果[31]。

1.4　模型构建

有研究指出，对目标属性的高光谱特征波段的挖掘和提取，有助于揭示和诠释目标属性的高光谱响应特性，并可以实现光谱数据降维[32]。然而，由于全谱模型中变量众多、信息量大，包含大量的无效信息，从而导致模型估计的准确性下降，同时计算效率也给实际应用带来了一些挑战。虽然利用光谱特性建立的模型能够很好地解决上述问题，但其准确性还需要进一步提高[33]。基于光谱特征的估算模型虽可以克服以上缺点，但是模型精度有待进一步提升。因此，本试验选用全波段 PLSR 和 SPA-MLR 两种方法建立模型。

1.4.1　PLSR

PLSR 是 16 世纪末 H. Wold 在计量经济学中提出的。它基于最小二乘法原理，在抽取含有更多自变量的因素时，能够确保所抽取的因子和因变量之间的相关程度达到最大化。简化光谱自变量和消除多余噪声的影响可以有效地对土壤养分信息进行解释和筛选，避免在利用光谱构建模型过程中对数据的过拟合现象，最终所构建的模型具有更好的精度及稳定性[34]。

1.4.2　SPA-MLR

MLR 是一种分析若干个自变量如何直接影响一个因变量及他们之间存在的相关关系的数理统计方法，以线性方程的形式建立响应变量与观测光谱变量之间的关系，但当变量数大于样本数时，模型容易受到变量间共线性的影响。因此，运用 MLR 方法建模，一般会与特征波段提取算法相结合[35]。

1.5　模型评价

使用决定系数（R^2）、标准偏差（root mean square error，*RMSE*）和相对预测偏差（relative percentage difference，*RPD*）作为模型的评价指标。R^2 越大，*RMSE* 越小，表示模型预测效果越好。一般认为，*RPD* 小于 1.40 时，模型很难有效地预测样本；*RPD* 介于 1.40 ~ 2.00 时，模型可以对样本进行粗略地预测；*RPD* 大于等于 2.00 时，表明模型具有较好的预测精度和稳健度；*RPD* 大于 3.00 时，表明模型具有非常好的预测效果。各验证参数计算公式如下：

$$R^2=1-\frac{\sum_{i=1}^{n}(Y_i-Y_i')/(n-p-1)}{\sum_{i=1}^{n}(Y_i-Y_i')^2/(n-1)} \quad (1)$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i-Y_i')^2} \quad (2)$$

$$PRD=\frac{m}{RMSE} \quad (3)$$

式中，n 为样本个数，Y_i' 是样本的预测值，Y_i 是样本的实测值；m 为样本实测值的标准差。

1.6　数据处理软件

使用 View Spec Pro 6.0 提取土壤光谱数据，使用 The Unscrambler X 10.4 对光谱数据进行预处理，使用 Excel 2019 以及 SPSS 20 对数据进行整理与分析，使用 MATLAB 2010 构建模型，使用 Origin 2021 制图。

2　结果与分析

2.1　土壤有机碳的描述性统计

对土壤有机碳含量进行统计分析，结果如表 2 所示。由表 2 可得，所采集的土壤样品有机碳含量的数值范围为 0.32 ~ 1.65 g/kg，试验中偏度和峰度系数接近 0，表明各数据集整体趋于正态分布，符合统计学的要求。数据变异程度为中低等，变异较大。总数据集、校正集和验证集的变异系数分别为 17.20%、17.85% 和 17.04%，平均值分别为 1.15、1.16、1.15 g/kg，三者数据变异程度和平均值接近，证明建模集数和验证集数划分较为合理。

2.2　土壤有机碳光谱响应特性

按照四分位法将小麦土壤有机碳含量由小到大排列并分成四等份，每份分别包含总样本数据的 0% ~ 25%、25% ~ 50%、50% ~ 75%、75% ~ 100%，计算每部分内土壤样本对应的土壤光谱数据算术平均值，得到不同含量水平下的光谱反射率，结果如图 1 所示。

从整体来看，小麦土壤有机碳含量不同，其光谱反射强度也有差异。土壤光谱反射率随着土壤有机碳含量的增加而降低，二者呈负相关关系。在不同的有机碳水平下，土壤的光谱反射率曲线具有相似性，变化趋势一致；在 400 ~ 850 nm 范围内，光谱反射率随波长的增加呈单调上升的趋势，曲线

有较高的斜率，并迅速向长波方向偏移；在 850 nm 后，光谱曲线的斜率较小，光谱反射率变化平缓；在 1 414 nm、1 912 nm 和 2 210 nm 附近处出现吸收谷，其中 1 912 nm 处的吸收谷较大，吸收谷附近的光谱反射率急速下降后又快速上升；在 2 210 nm 附近之后的波峰和波谷交替出现，且在 2 158 nm 附近，土壤有机碳光谱反射率达到最大。

2.3 土壤有机碳与不同预处理光谱的相关性

对土壤有机碳含量与原始光谱数据以及经过不同变换处理后的光谱数据做相关性分析，并进

表 2 土壤有机碳描述性统计

Tab. 2 Descriptive statistics of soil total nitrogen

Dataset	*n*	Min/(g·kg^{-1})	Max/(g·kg^{-1})	Mean/(g·kg^{-1})	*m*	Skweness	Kurtosis	CV/%
Date set	164	0.32	1.65	1.15	0.198	–0.706	1.687	17.20
Calibration set	123	0.53	1.56	1.16	0.207	–0.994	1.795	17.85
Validation set	41	0.32	1.65	1.15	0.196	–0.612	1.804	17.04

注：CV为变异系数。

Note: CV means coefficient of variation.

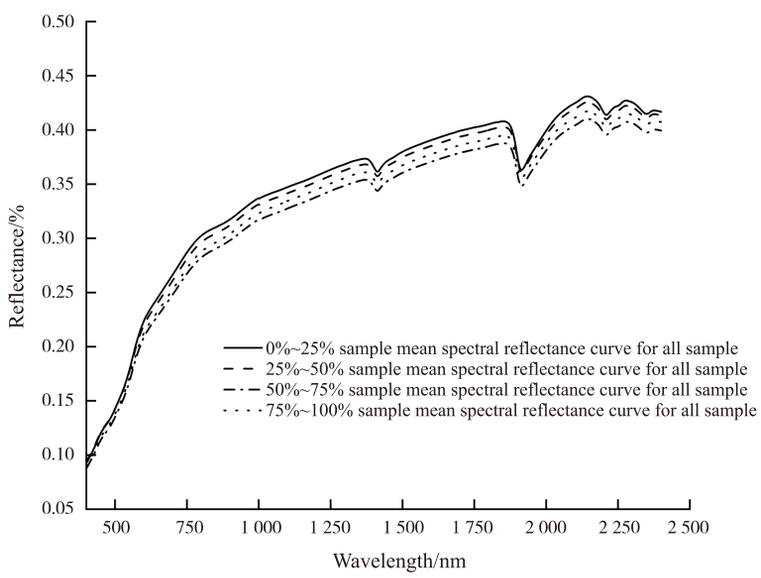

图 1 不同土壤有机碳含量水平下的光谱曲线

Fig. 1 Reflection curve of soil organic carbon at different content levels

行 0.05 显著性水平检验，结果如图 2 所示。由图 2 可知，不同变换光谱均与土壤有机碳含量存在一定的相关性，其中，RE 变换光谱与有机碳含量呈正相关，SG、SQ 变换光谱与有机碳含量为负相关关系，其余变换光谱均与有机碳含量既存在正相关关系，又存在负相关关系。SQ 预处理和 CU 预处理与原始光谱与土壤有机碳的相关性曲线一致，两者相关系数的差异较小；FD、SD 预处理与 SG 相比，相关性得到显著提高，达到 0.05 显著性水平的波段明显增多，但 FD 预处理下，在多数波段范围内，土壤有机碳与光谱反射率的相关性有所增高，最高相关系数达到 0.516；MSC、SNV 预处理下，土壤光谱与土壤有机碳的相关性有所提高，在部分波段都达到了 0.05 显著性水平。总体来看，不同预处理都增强了土壤光谱与土壤有机碳的相关性。

2.4 光谱特征提取

利用 SPA 对原始光谱数据及经过不同变换处理的光谱数据进行重要波长提取，其结果如图 3 所示。不同预处理下所提取的重要波长数量及位置存在差异。基于 SPA 提取的重要波长，SG 与 FD 预处理下有 405 nm 这个相同的重要波长；SG、SQ、CU 和 SNV 预处理下均有 2 400 nm 这个相同的重要波长；SG 与 SQ 预处理下有 2 个相同重要波长在 1 068、2 204 nm 处，有 2 个相近重要波长在 2 345 nm

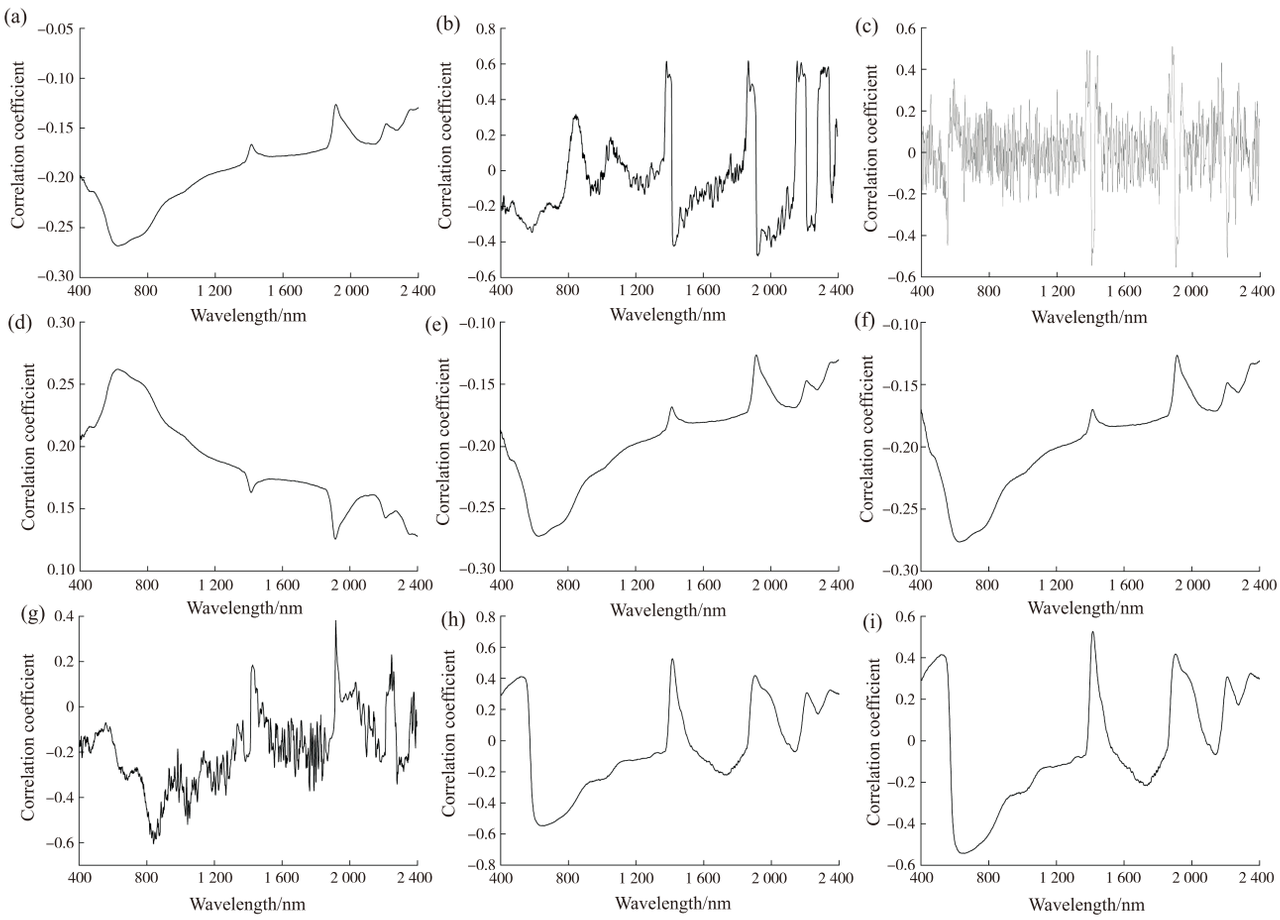

图 2　不同预处理光谱与土壤有机碳的相关关系

Fig. 2　Correlation between soil organic carbon and soil spectrum based on different spectral pretreatments

(a) SG; (b) FD; (c) SD; (d) RE; (e) SQ; (f) CU; (g) RE +FD; (h) MSC; (i) SNV.

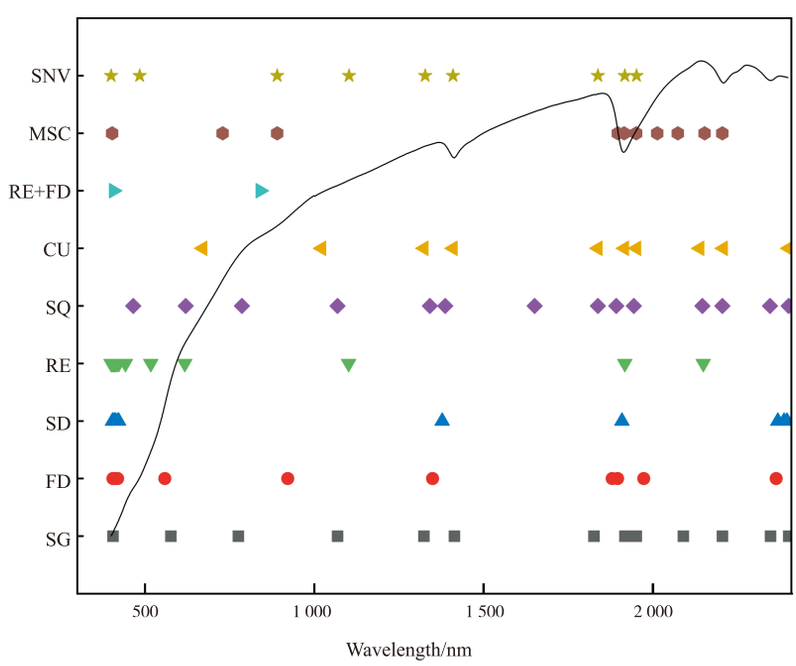

图 3　不同光谱预处理下基于 SPA 所提取土壤有机碳重要波长的分布

Fig. 3　Distribution of spectral important wavelengths of soil organic carbon extracted by SPA based on different spectral pretreatments

处；FD、SD、RE 预处理下都有在 400 ~ 422 nm 范围内的重要波长，且有较多是相同的，位置相近且相对集中；SNV 和 MSC 预处理下有 2 个相同的重要波长在 890、2 138 nm 处，有 2 个相近的重要波长在 1 950 nm 处。所有预处理下，在可见光和近红外波段都有重要波长。综上，不同预处理下所提取的重要波长在全波段范围内较为分散，重要波长在可见光波段范围内分布较多，基于 SPA 所提取的重要波长主要分布在了 400 ~ 422 nm 波段范围内。

2.5 不同光谱预处理下土壤有机碳的估测模型

采用 PLSR 和 SPA-MLR 多元统计分析方法构建土壤有机碳的定量估算模型，模型表现结果如表 3。

由表 3 可得，基于 SPA-MLR 线性模型，在 MSC 处理下的估测效果表现最好（R_c^2=0.724、$RMSE_c$=0.117、RPD_c=1.903、R_v^2=0.726、$RMSE_v$=0.109、RPD_v=1.956），在 FD、SNV 预处理下也取得了较好的估测效果，但估测精度略低于 SQ 处理，在 RE+FD 处理下模型效果表现较差。基于全光谱的 PLSR 模型，在 SNV 处理下估测效果表现最好（R_c^2=0.623、$RMSE_c$=0.124、RPD_c=1.291、R_v^2=0.604、$RMSE_v$=0.160、RPD_v=1.310），在 SG 和 SQ 处理下模型效果表现较差。在 FD、SD、RE、SQ、CU、RE+FD 的预处理下，PLSR 模型估测效果较好，在 SG、MSC、SNV 处理下，SPA-MLR 模型估测效果较好。在多数光谱的预处理下，对土壤有机碳含量的估测模型效果为：SPA-MLR＞PLSR。对比两种模型可以看出，基于 MSC 预处理的 SPA-MLR 模型对土壤有机碳含量的估测效果最好。

为了更加直观地比较这两种模型的效果，本研究绘制了两种估测效果最好的模型土壤有机碳含量实测值与预测值的 1:1 拟合图（图 4）。从图 4 中可以看出，基于 MSC 处理下所构建的 SPA-MLR 模型建模集和验证集的拟合线相较于其他模型更加

表 3 基于不同预处理下的 PLSR、SPA-MLR 模型表现

Tab. 3 PLSR, SPA-MLR model performance based on different preprocessing

Model	Pretreatment	Calibration set			Validation set		
		R^2	*RMSE*	*RPD*	R^2	*RMSE*	*RPD*
PLSR	SG	0.646	0.132	1.495	0.374	0.170	0.996
	FD	0.649	0.125	1.427	0.612	0.141	1.354
	SD	0.614	0.134	1.281	0.474	0.155	1.151
	RE	0.649	0.127	1.380	0.533	0.161	1.391
	SQ	0.594	0.142	1.390	0.470	0.180	1.327
	CU	0.580	0.135	1.383	0.583	0.167	1.053
	RE+FD	0.746	0.103	1.710	0.501	0.180	0.966
	MSC	0.607	0.129	1.250	0.596	0.163	0.908
	SNV	0.623	0.124	1.291	0.604	0.160	1.310
SPA-MLR	SG	0.526	0.147	1.452	0.503	0.163	1.419
	FD	0.527	0.139	1.454	0.531	0.176	1.461
	SD	0.454	0.164	1.354	0.322	0.166	1.215
	RE	0.464	0.149	1.366	0.489	0.177	1.399
	SQ	0.558	0.137	1.504	0.457	0.173	1.358
	CU	0.502	0.150	1.417	0.510	0.159	1.428
	RE+FD	0.408	0.159	1.299	0.414	0.184	1.306
	MSC	0.724	0.117	1.903	0.726	0.109	1.956
	SNV	0.674	0.126	1.742	0.639	0.124	1.586

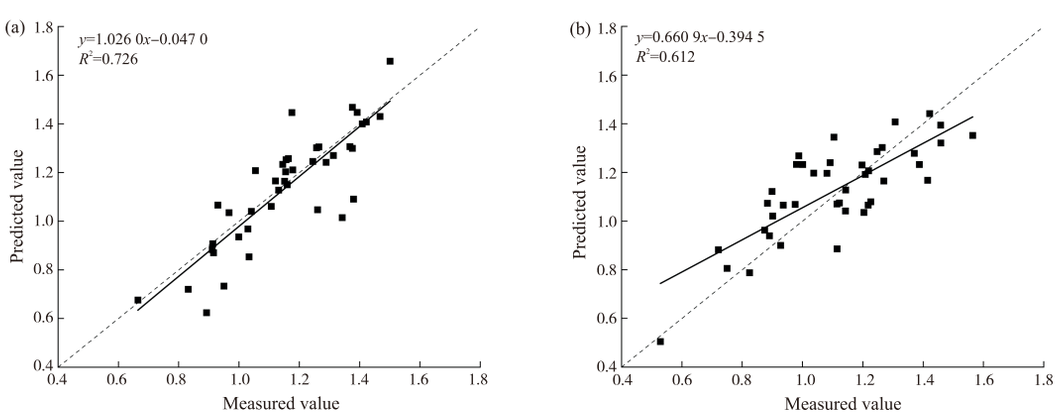

图 4 预测模型实测值与预测值 1:1 拟合图

Fig. 4 The measured value of the prediction model is fitted to the predicted value 1:1

(a) MSC 变换光谱的 SPA-MLR 模型；(b) FD 预处理下的 PLSR 模型。

(a) SPA-MLR model with MSC transform spectra; (b) PLSR model under FD preprocessing.

接近 1:1 直线，数据点较为集中，说明预测值更加接近于实测值。

3　讨论

3.1　土壤有机碳的光谱响应特性

土壤光谱可同时反映各种土壤信息，了解土壤反射光谱特征对于人们快速获取土壤信息具有重大意义。本研究对小麦土壤有机碳含量各四分位数范围内的反射光谱分别计算算术平均值，得到土壤有机碳四分位均值光谱反射曲线。在可见光波段范围内，不同有机碳水平下，土壤光谱反射率无明显差异且均急剧升高；在近红外波段范围内，光谱反射率随着有机碳含量的增加而降低，二者呈负相关关系。这与牛方鹏[36]基于高光谱数据的博斯腾湖西岸湖滨绿洲不同土壤有机碳含量的光谱曲线特征一致。本试验在可见光波段范围内，利用两种波长选择算法筛选出的重要波长数较多，土壤光谱反射率随着波长的增加而增加，前波与后波之间具有较好的正相关性，在 400 ~ 850 nm 波段反射率增加较快，在 800 ~ 2 450 nm 波段反射率增加缓慢。在 1 414、1 912、2 210 nm 处有显著的光谱吸收谷，且在 1 912 nm 处光谱吸收变化更为强烈。

对于全波段和连续投影变换这两种波长的选择，本研究证明了基于 SPA 的 MLR 模型效果优于基于全谱的 PLSR 模型。这是因为 SPA 可以有效减少模型的输入变量，具有降低重复信息等优点[19]。但张月等[37]研究发现，基于全谱的 PLSR 模型效果优于 SPA-MLR 模型。这可能是因为全谱信息量大，且信息之间存在一定的过拟合现象，但 PLSR 模型可以克服数据共线性的作用。同时，光谱预处理方法对于重要波长变量选择影响较大，所选的重要波长因预处理方法的不同而有所变化，也会因变量筛选方法的不同而有所不同[38]。综合比较发现，土壤有机碳重要波长区域主要位于 400 ~ 450 nm、510 ~ 620 nm、1 010 ~ 1 060 nm、2 000 ~ 2 400 nm。

3.2　不同光谱预处理对土壤有机碳估测模型效果的影响

不同光谱预处理对土壤有机碳估测模型效果有一定的影响。光谱测量过程中会受到测量仪器、测量环境等的影响，使得光谱曲线发生一定的变形，对实际的光谱信息造成一定的误差，而对光谱数据的预处理是光谱分析中的重要步骤。前人的研究表明，光谱数据预处理可以消除光谱仪器的噪声和背景干扰，在一定程度上可以提高光谱数据与待测物的相关性[39]。在本试验中，农田土壤有机碳与原始光谱的相关系数较低，经过不同的预处理后，土壤有机碳与光谱相关性有了一定的提高，并且本研究在多数预处理下，PLSR 与 SPA-MLR 都具有较好的估测模型效果。由此可以证明，预处理可以在一定程度上提高模型的估测效果，这与前人的研究结果一致。

此外，张倩倩等[40]研究发现，两种预处理方法配合使用的效果优于单个预处理方法。但是在本试验中，MSC 单个预处理方法使用的效果最佳。这是因为两种预处理方法组合时的顺序会影响模型的精度[41]，而单个预处理可以有效避免这一问题，从而增强光谱与数据之间的相关性。相比于其他的预处理，土壤有机碳在 MSC 的预处理下估测效果较好。这与宋金鹏等[42]认为的 MSC 处理能较好地提高模型估测效果的观点一致。本研究发现，MSC 与 SNV 单个预处理方法构建的两种模型对于土壤有机碳具有较好的估测效果。

3.3　多元统计分析方法对土壤有机碳模型效果的影响

本文研究结果表明，在 FD、SD、RE、SQ、CU、RE+FD 的预处理下，PLSR 模型估测效果较好，而在 SG、MSC、SNV 处理下，SPA-MLR 模型估测效果较好。此外，基于重要波长的 MLR 线性模型表现与 PLSR 非线性模型相比较效果更佳。而陶培峰[43]的研究表明，非线性模型对于土壤养分具有更好的估测效果。这可能是因为相比于全波段构建模型，连续投影变换可以有效剔除光谱冗余数据，减少较多的特征变量。

综上所述，光谱反射率与土壤有机碳含量呈负相关关系，在 400 ~ 450 nm、510 ~ 620 nm、1010 ~ 1 060 nm、2 000 ~ 2 400 nm 波段范围内光谱响应明显，其中以 1 912 nm 附近最为明显。预处理光谱可以提高小麦土壤光谱与土壤属性的相关性，对模型的反演结果产生较好的影响，且 MSC 单个预处理方法使用的效果最佳。基于 MSC 的 SPA-MLR 线性模型可以实现对农田土壤有机碳含量的准确估算（R_c^2=0.724、$RMSE_c$=0.117、RPD_c=1.903、R_v^2=0.726、$RMSE_v$=0.109、RPD_v=1.956）。本研究结果可以为农田土壤有机碳含量检测提供一定的

理论依据和实践探索。

参考文献(References):

[1] 李佳珍, 董文旭, 陈拓, 等. 增温施肥对农田土壤有机碳和全氮含量及δ13C、δ15N值的影响[J]. 中国生态农业学报, 2022, 30(5): 842-850.
LI Jiazhen, DONG Wenxu, CHEN Tuo, *et al.* Effects of warming and fertilization on soil organic carbon and total nitrogen contents, and δ13C and δ15N in farmland[J]. Chinese Journal of Eco-agriculture, 2022, 30(5): 842-850.

[2] 张丽华, 刘瀚阳. 全球气候治理嵌套式机制互动研究——以《联合国气候变化框架公约》与粮农组织为例[J]. 阅江学刊, 2023, 15(5): 47-57, 169.
ZHANG Lihua, LIU Hanyang. Research on the interaction of nested mechanisms of global climate governance tasking UNFCCC and FAO as example[J].Yuejiang Journal, 2023, 15(5): 47-57, 169.

[3] 翟茂彤. 基于原位vis-NIR高光谱的鄱阳湖湿地土壤有机碳预测研究[D]. 南昌: 江西财经大学, 2020.
ZHAI Maotong. Prediction of soil organic carbon based on *in situ* vis-NIR hyperspectra data in Poyang lake wetland[D]. Nanchang: Jiangxi University of Finance and Economics, 2020.

[4] 聂鹏程, 钱程, 覃锐苗, 等. 天空地一体化信息感知与融合技术发展现状与趋势[J]. 智能化农业装备学报, 2023, 4(2): 1-11.
NIE Pengcheng, QIAN Cheng, QIN Ruimiao, *et al.* Development status and trends of air-ground integrated informationsensing and fusion technology[J]. Journal of Intelligent Agricultural Equipment, 2023, 4(2): 1-11.

[5] DAINING D, HUICHUN Y, YONG Y, *et al.* Determination of chlorophyll and hardness in cucumbers by raman spectroscopy with successive projections algorithm (SPA) -extreme learning machine (ELM)[J]. Analytical Letters, 2023, 56(8): 1216-1228.

[6] 李通. 基于无人机高光谱技术的松材线虫病监测技术研究[J]. 测绘技术装备, 2023, 25(4): 19-21.
LI Tong. Monitoring technology of pine wood nematode disease based on UAV hyperspectral technology[J]. Geomatics Technology and Equipment, 2023, 25(4): 19-21.

[7] 赵隽宇, 石媛媛, 杨瑞青, 等. 基于高光谱分析技术的桉树叶片黄化识别[J]. 广西林业科学, 2024, 53(1): 94-99, 1006-1126.
ZHAO Junyu, SHI Yuanyuan, YANG Ruiqing, *et al.* Yellowing identification in eucalypt leaves based on hyperspectral analysis technology[J]. Guangxi Forestry Science, 2024, 53(1): 94-99, 1006-1126.

[8] 唐子竣, 张威, 黄向阳, 等. 基于地面高光谱遥感技术的大豆产量估算模型研究[J]. 农业机械学报, 2024, 55(1): 145-153, 240.
TANG Zijun, ZHANG Wei, HUANG Xiangyang, *et al.* Research on soybean seed yield estimation model based on ground hyperspectral remote sensing technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 145-153, 240.

[9] HE Z Q, ZHANG R, FANG S Y, *et al.* Research on the measurement method of printing ink content based on spectrum[J]. Optik, 2021, 243: 167389.

[10] 王润润, 张淑娟, 苏立阳, 等. 基于高光谱技术的杏品种判别[J]. 农产品加工, 2023(23): 20-22, 25.
WANG runrun, ZHANG Shujuan, SU Liyang, *et al.* Identification of apricot varieties based on hyperspectral technology[J]. Farm Products Processing, 2023(23): 20-22, 25.

[11] 韩秀枝, 曹源, 詹跃勇, 等. 基于响应面耦合遗传算法优化关山樱花黄酮提取工艺及其抗氧化、抗炎作用研究[J]. 食品安全质量检测学报, 2023, 14(24): 272-280.
HAN Xiuzhi, CAO Yuan, ZHAN Yueyong, *et al.* Optimization of Kanzan flower flavonoids extraction process based on response surfacecoupled genetic algorithm methodology and its antioxidant and anti-inflammatory effects[J]. Journal of Food Safety and Quality, 2023, 14(24): 272-280.

[12] 谢永凯, 冯美臣, 秦明星, 等. 基于多元统计分析的冬小麦干旱综合指标构建及监测研究[J]. 天津农业科学, 2023, 29(5): 17-24.
XIE Yongkai, FENG Meichen, QIN Mingxing, *et al.* Construction and monitoring of comprehensive indicators of winter wheat drought based on multivariate statistical analysis[J]. Tianjin Agricultural Sciences, 2023, 29(5): 17-24.

[13] 陈嘉伟, 周德强, 崔晨昊, 等. 近红外光谱的小麦粉粉质特性预测模型研究[J]. 光谱学与光谱分析, 2023, 43(10): 3089-3097.
CHEN Jiawei, ZHOU Deqiang, CUI Chenhao, *et al.* Prediction model of farinograph characteristics of wheat flour based on near infrarer spectroscopy[J]. Spectroscopy and Spectral Analysis, 2023, 43(10): 3089-3097.

[14] 梁奇峰, 曾令聪, 郭红卫, 等. 近红外光谱结合偏最小二乘法测定米粉中硼砂含量[J]. 广东化工, 2023, 50(16): 152-153, 184.
LIANG Qifeng, ZENG Lingcong, GUO Hongwei, *et al.* Determination of borax in rice flour by near infrared spectrum with partial least square method[J]. Guangdong Chemical Industry, 2023, 50(16): 152-153, 184.

[15] 郭发旭, 冯全, 杨森, 等. 基于无人机高光谱的马铃薯冠层叶片全氮含量反演[J]. 浙江农业学报, 2023, 35(8): 1904-1914.
GUO Faxu, FENG Quan, YANG Sen, *et al.* Inversion of leaf nitrogen content in potato canopy based on unmanned aerial vehicle hyperspectral images[J]. Journal of Zhejiang Agricultural Sciences, 2023, 35(8): 1904-1914.

[16] 章海亮, 刘雪梅, 何勇. SPA-LS-SVM检测土壤有机质和速效钾研究[J]. 光谱学与光谱分析, 2014, 34(5): 1348-1351.
ZHANG Hailiang, LIU Xuemei, HE Yong. Measurement of soil organic matter and available K based on SPA-LS-SVM[J]. Spectroscopy and Spectral Analysis, 2014, 34(5): 1348-1351.

[17] 柴红. 谷子田土壤颗粒有机碳氮的高光谱监测研究[D]. 晋中: 山西农业大学, 2019.
CHAI Hong. Hyperspectral monitoring of siol organic carbon and nitrogen in millet field the study[D]. Jinzhong: Shanxi Agricultural University, 2019.

[18] 李国傲, 陈雪, 孙建伶, 等. 土壤有机碳含量测定方法评述及最新研究进展[J]. 江苏农业科学, 2017, 45(5): 22-26.
LI Guoao, CHEN Xue, SUN Jianling, *et al.* Review of methods for determination of soil organic carbon content and recent research progress[J]. Jiangsu Agricultural Sciences, 2017, 45(5): 22-26.

[19] 李阳阳, 孙雨安, 王国庆, 等. 基于高光谱的大叶女贞叶片水分定量测定[J]. 河南师范大学学报 (自然科学版), 2017, 45(6): 47-51.
LI Yangyang, SUN Yu'an, WANG Guoqing, *et al.* Determination of ligustrum leaf water content based on hyperspectral[J]. Journal of Normal University (Natural Science Edition), 2017, 45(6): 47-51.

[20] 邓志伟, 郭苏彤, 徐智帅, 等. 基于双脉冲LIBS结合偏最小二乘回归的稀土元素定量分析[J]. 中国无机分析化学, 2024, 14(2): 183-190.
DENG Zhiwei, GUO Sutong, XU Zhishuai, *et al.* Quantitative analysis of fare earth elements based on double-pulse LIBS combined with partial least squares regression[J]. Chinese Journal of Inorganic Analytical Chemistry, 2024, 14(2): 183-190.

[21] 徐冬. 基于多光谱的茄子果皮花青素含量检测设备研制[D]. 咸阳: 西北农林科技大学, 2023.
XU Dong. Apparatus development for detecting anthocyanin content in eggplant peel based on multi-spectral technolog[D]. Xianyang: Northwest Agriculture and Forestry University, 2023.

[22] 马玮键, 邢泽炳, 韩春风, 等. 可见近红外光谱的山西玉米地土壤氮含量建模[J]. 山西农业科学, 2023, 51(7): 750-755.
MA Weijian, XING Zebing, HAN Chunfeng, *et al.* Modeling of soil nitrogen content in maize field of Shanxi province by visible near-infrared spectroscopy[J]. Shanxi Agricultural Sciences, 2023, 51(7): 750-755.

［23］ 史可. 重铬酸钾外加热法在施用生物炭土壤有机质含量测定中的应用改进［D］. 泰安: 山东农业大学, 2022.
SHI Ke. Application improvement of the potassium dichromate external heating method in the determination of siol organic matter content with biochar applied［D］. Taian: Shandong Agricultural University, 2022.

［24］ 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用［J］. 化学进展, 2004(4): 528-542.
CHU Xiaoli, YUAN Hongfu, LU Wanzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique［J］. Progress in Chemistry, 2004(4): 528-542.

［25］ 孙亚楠, 李仙岳, 史海滨, 等. 基于高光谱数据的盐荒地和耕地土壤盐分遥感反演优化［J］. 农业工程学报, 2022, 38(23): 101-111.
SUN Yanan, LI Xianyue, SHI Haibin, *et al.* Optimizing the inversion of siol salt in salinized wasteland using hyperspectral data from remote sensing［J］. Transaction of the Chinese Society of Agricultural Engineering, 2022, 38(23): 101-111.

［26］ RAUL J L M, ALEJANDRO O S G, MANUEL J J C. Simultaneous quantification of acetylsalicylic acid and caffeine in tablets by first-order derivative spectroscopy［J］. International Journal of Applied Pharmaceutics, 2022, 14(4): 200-204.

［27］ 李鑫星, 周婧, 唐红, 等. 基于紫外光谱的水产养殖水质总氮含量快速检测研究［J］. 光谱学与光谱分析, 2020, 40(1): 195-201.
LI Xinxing, ZHOU Jing, TANG Hong, *et al.* Rapid determination of total nitrogen in aquaculture water based on ultraviolet spectroscopy［J］. Spectroscopy and Spectral Analysis, 2020, 40(1): 195-201.

［28］ 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.* Estimating of soil total nitrogen content in coastal wetland based on spectral transformation［J］. Soil, 2023, 55(4): 880-886.

［29］ BAO Y, MI C, WU N, *et al.* Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics［J］. Applied Sciences, 2019, 9(19): 4119.

［30］ 苗林林. 基于无人机遥感的柑橘果树叶面积指数反演研究［D］. 桂林: 桂林理工大学, 2023.
MIAO Linlin. Inversion of citrus fruit leaf area index based on UAV remote sensing［D］. Guilin: Guilin University of Technology, 2023.

［31］ HE Z Q, ZHANG R, FANG S Y, *et al.* Research on the measurement method of printing ink content based on spectrum［J］. Optik, 2021, 243: 167389.

［32］ SEQURIRA C H,WILLS S A, GRUNWALD S, *et al.* Development and update process of vnir-based models built to predict soil organic carbon［J］. Soil Science Society of America Journal, 2014, 78(3): 903-913.

［33］ 丁建军, 章盛, 孙超, 等. 基于 PLSR 的土壤有机碳预测模型中建模组与验证组最优数量关系［J］. 江汉大学学报: 自然科学版, 2018, 46(5): 404-408.
DING Jianjun, ZHANG Sheng, SUN Chao, *et al.* Optimal quantity relationship between modeling group and prediction group in soil organic carbon prediction model based on PLSR［J］. Journal of Jianghan University: Natural Science Edition, 2018, 46(5): 404-408.

［34］ 李晓涵. 基于多元线性回归分析模型的中国海洋经济发展与交通关系浅析［J］. 交通节能与环保, 2023, 19(2): 35-38.
LI Xiaohan. Analysis of the relationship between China's marine economic development and transportation based on multiple linear regression analysis model［J］. Energy Conservation and Environmental Protection in Transportation, 2023, 19(2): 35-38.

［35］ 孙晶京. 基于近红外光谱技术的小麦籽粒蛋白质含量模型构建及优化研究［D］. 晋中: 山西农业大学, 2021.
SUN Jingjing. Research on construction and optimization of wheat grain protein content model based on near infrared spectroscopy technology［D］. Jinzhong: Shanxi Agricultural University, 2021.

［36］ 牛芳鹏. 基于高光谱数据的博斯腾湖西岸湖滨绿洲土壤有机碳反演建模［D］. 乌鲁木齐: 新疆师范大学, 2022.
NIU Fangpeng. Inversion modeling of soil organic carbon in the Western Lakeside Oasis in Bosten Lake based on hyper-spectral data, China［D］. Wulumuqi: Xinjiang Normal University, 2022.

［37］ 张月, 张煊, 闫晓斌, 等. 干旱胁迫下冬小麦植株含水量的高光谱监测［J］. 山西农业科学, 2022, 50(12): 1638-1646.
ZHANG Yue, ZHANG Xuan, YAN Xiaobin, *et al.* Hyperspectral monitoring on plant water content in winter wheat under drought stress［J］. Shanxi Agricultural Sciences, 2022, 50(12): 1638-1646.

［38］ 孙嘉豪, 张伟, 施鉴芩, 等. 光谱数据预处理策略选择及应用［J］. 计量学报, 2023, 44(8): 1284-1292.
SUN Jiahao, ZHANG Wei, SHI Jianqin, *et al.* Selection and application of spectral data preprocessing strategy［J］. Journal of Metrology, 2023, 44(8): 1284-1292.

［39］ 姜雪芹, 叶勤, 林怡, 等. 基于谐波分析和高光谱遥感的土壤含水量反演研究［J］. 光学学报, 2017, 37(10): 300-310.
JIANG Xueqin, YE Qin, LIN Yi, *et al.* Inversion study of soil water content based on harmonic analysis and hyperspectral remote sensing［J］. Acta Optica Sinica, 2017, 37(10): 300-310.

［40］ 张倩倩, 谭琨. 基于高光谱的牛奶中真蛋白质含量反演［J］. 光谱学与光谱分析, 2015, 35(12): 3436-3439.
ZHANG Qianqian, TAN Kun. Inversion of true protein content in milk based on hyperspectral data［J］. Spectroscopy and Spectral Analysis, 2015, 35(12): 3436-3439.

［41］ 张朱珊莹, 蒙泳吉, 曹汇敏, 等. 血红蛋白定量分析的数据集划分及预处理方法研究［J］. 中南民族大学学报 (自然科学版), 2022, 41(4): 454-458.
ZHANG Zhushanying, MENG Yongji, CAO Huimin, *et al.* Dataset partitioning and pretreatment method of hemoglobin quantitative analysis［J］. Journal of South-Central University for Nationalities (Natural Science Edition), 2022, 41(4): 454-458.

［42］ 宋金鹏, 梁琨, 张驰, 等. 基于深度学习与可见-近红外光谱的患腥黑穗病小麦籽粒分类研究［J］. 分析测试学报, 2023, 42(7): 784-793.
SONG Jinpeng, LIANG Kun, ZHANG Chi, *et al.* Research on classification of commin bunt of wheat kernels based on visible-near infrared spectroscopy combined with deep learning algorithms［J］. Journal of Analysis and Testing, 2023, 42(7): 784-793.

［43］ 陶培峰. 基于航空高光谱遥感的黑土养分含量反演研究［D］. 北京: 中国地质大学, 2020.
TAO Peifeng. Estimation of nutrient content in black soil based on hyperspectral remote sensing［D］. Beijing: China University of Geosciences, 2020.