

土壤湿度对近红外光谱反演剖面有机质精度的影响

周 鹏¹, 孔一诺¹, 郝珊珊¹, 印 祥¹, 肖新清³, 金诚谦^{2*}

(1. 山东理工大学农业工程与食品科学学院, 淄博 255000; 2. 农业农村部南京农业机械化研究所, 南京 210014;
3. 中国农业大学工学院, 北京 100083)

摘 要: 为深入分析土壤湿度对近红外光谱反演剖面土壤有机质 (soil organic matter, SOM) 精度的影响, 该研究依据水分张力这一指标, 将土壤划分为风干状态、1.500、0.330、0.100、0.033 MPa 和饱和状态共 6 种湿度水平, 在所选 16 个地点分别采集深度约 150 cm 剖面土壤芯柱为研究对象, 采用 7 种方法对所测剖面土壤光谱吸光度进行光谱预处理, 选择较佳的预处理方法。同时, 采用连续投影算法 (successive projection algorithm, SPA) 和竞争性自适应重加权-连续投影算法 (competitive adaptive reweighting-successive projection algorithm, CARS-SPA) 筛选特征波长。构建基于全谱及特征波长的 SOM 近红外光谱反演模型, 并将其与标准正态变量变化 (standard normal variate, SNV) 预处理方法相结合。结果表明: 1) SPA-PLSR 模型和 CARS-SPA-PLSR 模型的精度均优于基于全谱的 PLSR 模型; 2) SNV-SPA-PLSR 模型在饱和、风干状态下预测效果更好, 而 SNV-CARS-SPA-PLSR 模型在水分张力分别为 0.033、0.100、0.330 和 1.500 MPa 时预测精度更高; 3) 不同土壤湿度水平近红外光谱“一对一”式预测 SOM 模型难以满足实际应用, 经过对比研究, 选用水分张力为 1.500 MPa 时构建的 SNV-CARS-SPA-PLSR 模型分别预测 6 组土壤湿度水平和混合样本集中 SOM 取得效果最好。该研究结果对各湿度水平下估算 SOM 含量有一定的指导作用, 并为提高不同土壤湿度水平剖面 SOM 近红外光谱反演模型的适用性提供参考。

关键词: 土壤湿度; 有机质; 近红外光谱; 水分张力; 竞争性自适应重加权-连续投影算法

doi: 10.11975/j.issn.1002-6819.202311069

中图分类号: S153.621

文献标志码: A

文章编号: 1002-6819(2024)-16-0113-11

周鹏, 孔一诺, 郝珊珊, 等. 土壤湿度对近红外光谱反演剖面有机质精度的影响[J]. 农业工程学报, 2024, 40(16): 113-123. doi: 10.11975/j.issn.1002-6819.202311069 <http://www.tcsae.org>
ZHOU Peng, KONG Yinuo, HAO Shanshan, et al. Influence of soil moisture on the inversion accuracy of near-infrared spectra of organic matter[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(16): 113-123. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202311069 <http://www.tcsae.org>

0 引 言

土壤有机质 (soil organic matter, SOM) 是土壤的重要组成部分, 决定了土壤肥力和质量, 与土壤生产力密切相关, 因此预测 SOM 含量是了解土壤肥力从而科学精准施肥, 保持较高耕地质量的重要途径^[1]。

传统的 SOM 含量测量方法虽然精度较高^[2], 但操作步骤繁琐、耗时长, 不适合大面积测量^[3]。国内外学者利用近红外光谱技术, 对某一地区或某种土壤开展了大量土壤有机质的预测研究^[4-6]。其中一些学者的研究发现土壤水分在近红外区域有很强的特征性吸收, 会对 SOM 检测造成很大的干扰^[7]。XIA 等^[8]认为土壤水分对光谱反射率的影响较大并且会掩盖其他物质的光谱特性。

王世芳等^[9]通过二维同步相关光谱图分析指出, 水分中的 O-H 键会掩盖 600 nm 和 1 660 nm 波长处涵盖的土壤有机质信息, 对 SOM 检测造成干扰。土壤水分对光谱技术预测 SOM 影响的研究集中在土壤干与湿这 2 种状态。WANG 等^[10]指出, 基于接近饱和的湿润状态下土壤样本建立的偏最小二乘回归 (partial least squares regression, PLSR) 模型在 SOM 预测方面的准确性略高于风干样本。而另一些研究则表明, 随着土壤湿度水平的增加, 可见-近红外 (visible-near infrared, Vis-NIR) 光谱技术预测 SOM 的准确性逐步降低。CHANG 等^[11]研究表明使用田间湿润样品的 PLSR 模型对 SOM 的预测效果比风干样品要差。MARAKKALA 等^[12]认为湿润状态下土壤中的自由水会干扰对 SOM 预测重要的光谱特征, 因此土壤在水分张力下的预测结果优于湿润状态。JI 等^[13]通过直接标准化 (direct standardization, DS) 算法去除样品中水分对光谱的影响, 显著提高了土壤有机质的预测精度。

上述研究中土壤水分对有机质含量预测结果不一致, 这可能与所选取土壤样品的质地、表面粗糙度和室外温度等条件存在差异有关^[14]。但样品覆盖的土层深度不同也是不可忽略的因素。现有土壤光谱数据库 (LUCAS 数据库^[15], Global Vis-NIR 数据库^[16]等)、便携式土壤有机质检测仪^[17]及人为采集样品^[18-19]多集中于对表层土壤

收稿日期: 2023-11-09 修订日期: 2023-12-28

基金项目: 国家自然科学基金青年基金项目 (62305197); 国家重点研发计划项目 (2021YFD200050205); 山东省重点研发计划 (科技示范工程) 项目 (2022SFGC0201); 山东理工大学中青年骨干教师海内外访学计划项目 (1240010304)

作者简介: 周鹏, 博士, 讲师, 研究方向为电子信息技术在农业中的应用。

Email: zhoupeng@sdut.edu.cn

*通信作者: 金诚谦, 研究员, 研究方向为大田作物收获机械化与智能化技术。Email: 412114402@qq.com

(通常定义为土壤芯柱样品的顶部 5~30 cm^[20]) 的有机质含量预测研究。但是由植被类型、土壤质地、气候条件和粗糙度等因素^[21]引起的空间异质性会使表层土壤特征难以维持稳定,因此仅对该土层深度的 SOM 含量预测存在局限性^[22]。考虑到植物残体、根系分泌物、生物扰动和有机肥料等生物化学因素^[23],以及农机作业对田间土壤造成的机械压实等人为因素^[24]共同耦合作用导致 SOM 含量的纵向不均一性^[25],有必要考虑土壤剖面深度。

将土壤湿度差异粗略分为干与湿 2 种状态进行研究存在局限性,需要划分更为详细的土壤湿度水平。通过对土壤水分含量的分组,可以降低水分与 SOM 的相互作用对光谱的影响。其中以何种标准进行分组是研究的关键^[26]。以往研究水分对 SOM 预测的影响常通过水分与烘干土样的质量比来配制不同含水率的土壤样品^[27]或依据土壤水分含量采用因子判别分析 (factorial discriminant analysis, FDA) 对样品进行分类^[28]。除此之外,也利用测得的光谱特征构建归一化土壤水分指数 (normalized soil moisture index, NSMI)^[29]、相对吸收深度 (relative absorption depth, RAD)^[30]、线性回归 (normalized index of nswir domain for soil moisture content estimation from linear regression, NINSOL) 和非线性回归估计土壤水分含量的归一化指数 (normalized index of nswir domain for soil moisture content estimation from non-linear regression, NINSON)^[31]及水分吸收指数 (moisture absorption correction index, MACI)^[32]等光谱参数对样品分类后间接得到土壤水分等级。按质量比配制土壤含水率耗费时间长、自动化程度低,不适合于样品数量多且测量范围广的情况^[27],而光谱参数分类法会使各水分等级中土壤样品数量不均匀^[33]。为了弥补上述方法的不足,本研究依据水分张力这一指标,将土壤划分为风干状态、1.500、0.330、0.100、0.033 MPa 和饱和状态共 6 种湿度水平^[34],以美国伊利诺伊州玉米带 16 个观测站土壤为研究对象,样品取样深度为表层到地下 150 cm 左右,利用近红外反射传

感器^[35]采集原始光谱并进行预处理,选用不同的波长选择算法筛选得特征波长,再对各土壤湿度水平下全波长及特征波长进行 PLSR 建模分析,对比不同湿度水平下 SOM 预测精度的差异,选择效果最优的湿度水平下建立的校准模型应用于不同数据集上,以提高各土壤湿度水平模型的准确度及适用性。

1 材料与方法

1.1 土壤样品的收集和制备

本研究使用的土壤光谱数据由 HUMMEL 等^[34]先前使用并完整描述。具体来说,土壤芯柱样品来自美国中部伊利诺伊州玉米带的 16 个地点 (表 1)。所采集土壤样本的确切位置、土壤系列、纹理、质地和总孔隙度等信息见文献^[36]。土壤质地为粉砂壤土或粉砂质黏壤土。

表 1 所选 16 个地点剖面土壤的平均有机质含量

Table 1 Average organic matter content of profiled soils (SOM) at the 16 selected sites

序号 No.	地点 Location	SOM/%	序号 No.	地点 Location	SOM/%
1	Brownstown	1.47	9	Springfield	1.97
2	Bondville	3.05	10	Oak Run	0.64
3	Dekalb	3.93	11	Perry	1.09
4	Dixon Springs	1.29	12	Olney	0.67
5	Freeport	2.18	13	Carbondale	1.08
6	Belleville	1.13	14	Stelle	2.89
7	Peoria	1.24	15	Monmouth	2.12
8	Ina	1.01	16	Martinsville	0.68

HOLLINGER 等具体描述了未受干扰的土壤芯柱样品的收集与制备过程^[37],总结为图 1:在所选的 16 个地点处,分别以 120°的间隔收集了 3 个直径为 5.56 cm 且深度不超过 150 cm 的土壤芯柱样品。为了防止水分从土壤芯柱样品流失或重新分配,每个采集的样品在现场被切割成 10 或 20 cm 的子样品,盖上土样储存容器的盖子后放于袋中封装。在实验室中,子样品又被进一步分成 2.5 cm 的样品,以便于测定土壤质量含水率和土壤有机质含量。

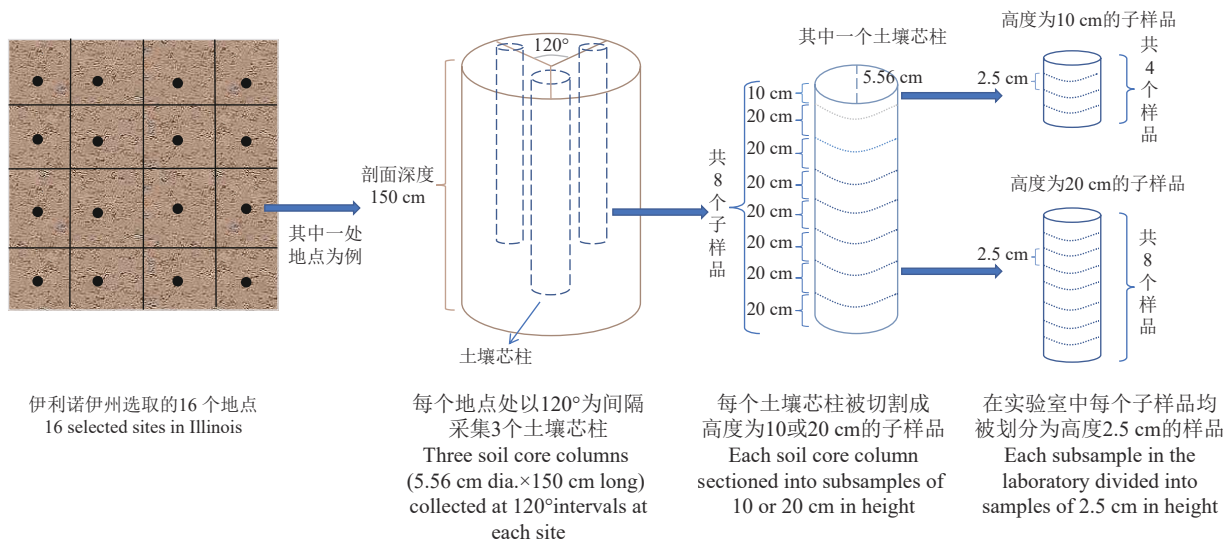


图 1 土壤样品的收集与制备流程图

Fig.1 Flowchart of collection and preparation of soil samples

如果所选 16 个地点都按图 1 所示采集 3 个深度为 150 cm 的土壤芯柱, 每个土壤芯柱又被划分为 8 个子样品, 则共获得 384 个子样品。但是在田间实际采集样本的过程中, 受土壤结构的影响, 不同地点处采样深度受限, 使得每个地点采集 3 个土壤芯柱的高度无法均达到 150 cm, 则无法将每个土壤芯柱都如图 1 的 c 中所示划分为 8 个子样品, 经统计 16 个地点处由每个土壤芯柱得到的子样品共计 313 个。

表 1 表示各个地点中剖面土壤的平均有机质含量 (%), 其中处于地点 Oak Run 的含量最低, 仅为 0.64%。而处于地点 DeKalb 的平均 SOM 含量最高, 达到 3.93% 明显高于其余 15 个地点的 SOM 平均含量。由此可以得出, 不同地区 SOM 的含量受该地区气候变化、植被覆盖、土地利用方式、土壤质地和酸碱反应等因素的影响产生了明显的空间变异性^[23-24]。

1.2 指标测定和计算

1.2.1 土壤水分张力

将样品湿润到预定水分张力的具体操作过程如下: 所选 16 个地点采集的土壤芯柱, 其子样品分割后所得 2.5 cm 的样品依次被放置多孔陶瓷压力板上, 每个样品由透明塑料管插入物的一个环包住^[37]。蒸馏水加入到

陶瓷板中, 通过毛细作用使每个样品饱和。从样品表面小心翼翼地抬起切割过程中残余的塑料管芯碎片。陶瓷板和湿润后的样品被放置在压力容器中, 经压力调节模块在其中施加适当的压力以获得所需的水分张力。

当样品水分达到平衡时, 压力被移除, 打开容器后用抹刀将每个样品从陶瓷板上抬起, 放置在近红外传感器^[35]下, 扫描后获得多个样品反射率结果并存储在数据采集计算机中。样品称质量后放在陶瓷板上将其送回压力容器, 在下一个较高的水分张力水平上进行平衡。重复这一过程, 直到在样品中建立 0.033、0.100、0.330 和 1.500 MPa 这 4 组土壤湿度水平, 并获得了相应的反射率扫描结果。上述具体操作过程见图 2。

土壤水分张力简单理解就是土壤对水的吸力, 土壤愈湿, 土壤质量含水率的平均值越大但对水的吸力就愈小所对应的水分张力数值越小; 反之则水分张力数值越大。土壤质量含水率的测量过程为: 首先把样品重新湿润至饱和, 并收集反射率扫描结果和样品质量。然后让样品在实验室环境条件下平衡, 得到风干样品并收集其反射率扫描结果和质量。最后将样品从塑料管插入环中取出, 在 105 °C 的烘箱中干燥 24 h, 并进行称质量, 根据样品的质量计算出每个样品在各土壤湿度水平下的质量含水率。

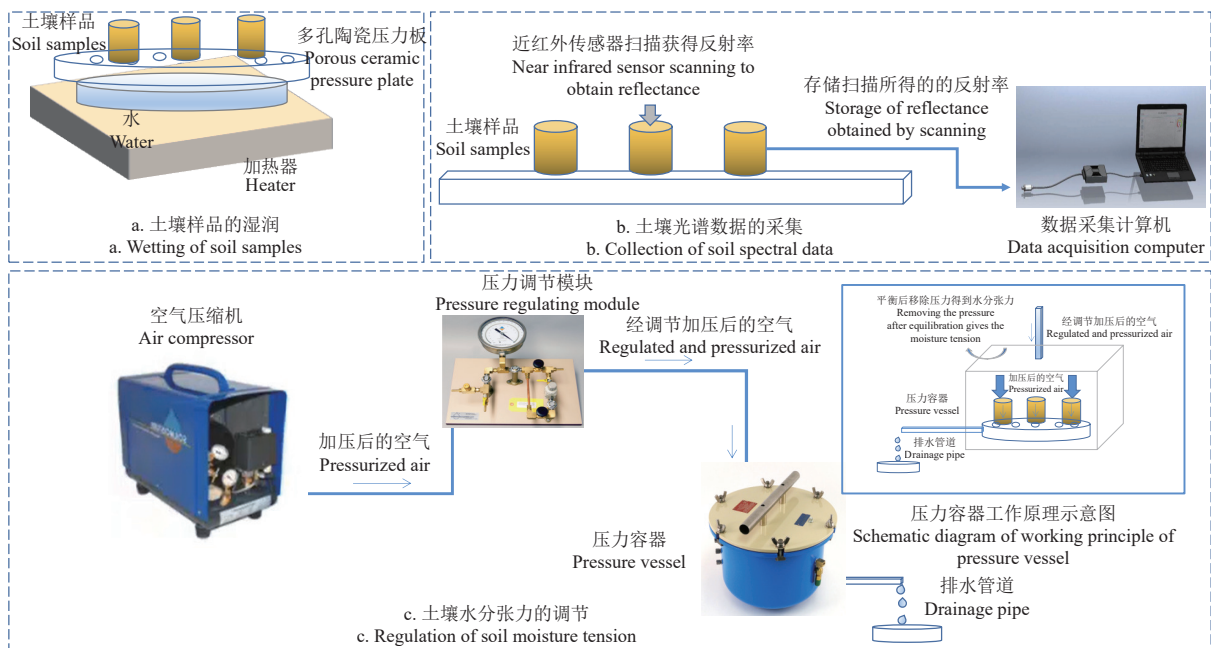


图 2 土壤水分张力水平的具体操作过程及压力容器工作原理示意图

Fig.2 The specific operation process of acquiring soil moisture tension and the schematic diagram of the working principle of the pressure vessel

6 种湿度水平下土壤质量含水率的平均值分别为 49.86%、36.67%、31.85%、27.30%、21.93% 和 4.74% (表 2), 本研究使用的土壤水分张力下对应的土壤水分含量变化比 SUDDUTH 等^[38]的要大, 在选取的 6 组土壤湿度水平下, 土壤含水率的平均值变化较为明显, 其中饱和状态下土壤含水率的平均值比风干状态提升了 45.12 个百分点。而土壤湿度水平处于饱和状态时, 意味着此时土壤湿度增大到所有孔隙充满水, 对应的土壤水

张力将降为 0。尽管各种土壤的饱和含水率是不同的, 但对土壤水分张力而言却是一致的都为 0。

本研究中之所以使用这种广泛的湿度水平范围, 是因为在现场数据收集过程中, 同一地点深层的湿度水平可能与表层的湿度水平有很大差异。当地表处于干燥条件允许进行现场作业时, 地下土壤可能会出现饱和状态, 因此需要采取由饱和到风干状态划分的 6 组土壤湿度水平对全层土壤进行分析。

表 2 不同湿度水平下土壤质量含水率统计

Table 2 Statistics of water content at different soil moisture levels

水分张力等级 Moisture tension level	样本数 Sample size	土壤含水率 Soil moisture/%			
		平均值 Mean	最大值 Maximum	最小值 Minimum	中位数 Median
饱和状态 Saturation state	301	49.86	141.71	24.27	49.63
水分张力 0.033 MPa	304	36.67	70.06	16.63	36.16
水分张力 0.100 MPa	306	31.85	72.16	15.73	31.87
水分张力 0.330 MPa	304	27.30	77.51	14.38	27.26
水分张力 1.500 MPa	303	21.93	59.87	8.59	22.10
风干状态 Air-dry state	302	4.74	14.27	0.87	4.49

1.2.2 土壤有机质测定

由每份子样品切割所得高度为 2.5 cm 土壤样品, 经图 2 中压力容器调节至不同水分张力水平后, 将样品分别混合、过筛得到各子样品并且为分析有机质含量而进行再次取样。NELSON 等^[2]建议使用测定总有机碳来衡量有机质含量以弥补直接测定 SOM 含量所获得的结果不够准确的缺点。本研究, 通过在 LECO Model HF10 感应炉中将每种样品进行干燃烧来确定总土壤有机碳

(soil organic carbon, SOC) 含量。式 (1) 中有机碳值乘以 1.72, 得到土壤有机质数据值。

$$M_{\text{SOM}} = 1.72M_{\text{SOC}} \quad (1)$$

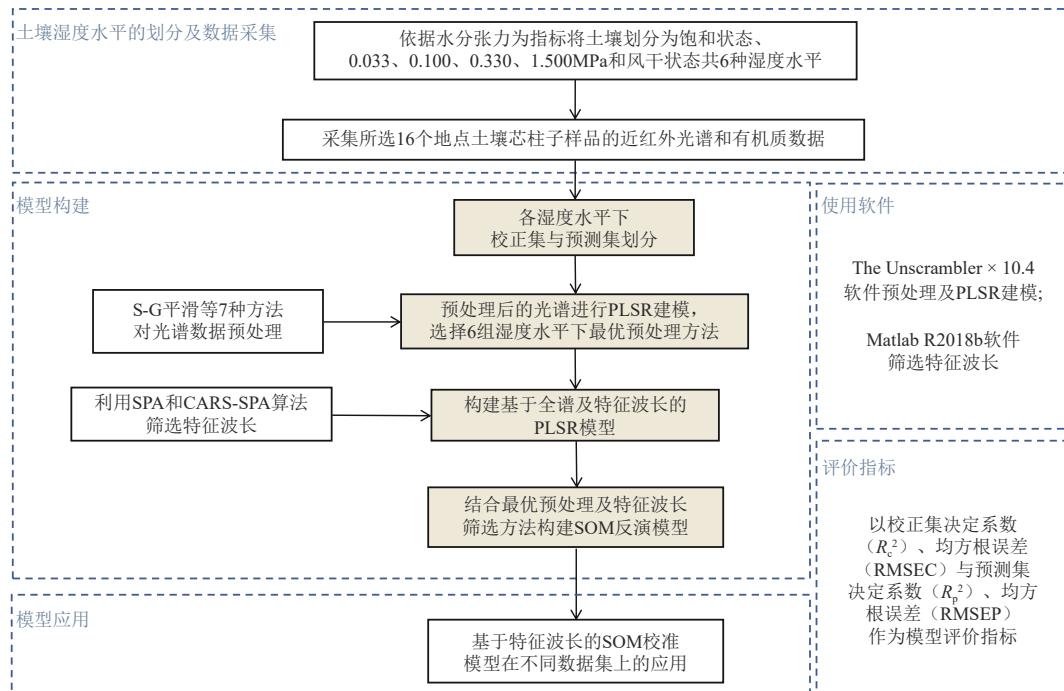
式中 M_{SOM} 为土壤有机质, %; M_{SOC} 为土壤有机碳, %。

1.2.3 土壤光谱反射率的测定

采用 SUDDUTH 等^[35]所设计的具有 45 nm 带宽的便携式近红外反射传感器, 进行土壤光谱反射率的测定, 该传感器的波长不稳定性最小并且可以在 10 s 内从专用微处理器在线获得样品反射率数据, 在 SUDDUTH 等以前的研究报告^[38]中详细说明该传感器的使用流程。

1.3 数据处理

本研究流程如图 3 所示, 其概述了土壤湿度水平的划分、近红外光谱和有机质数据的采集、模型构建及应用等步骤。光谱反射率数据归一化后再转换为吸光度, 该转换有助于降低噪声以突出吸收特性的边缘, 并增强光谱和 SOM 含量之间的线性关系^[39]。



注: PLSR 为偏最小二乘回归; S-G 为 Savitzky-Golay。SPA 为连续投影算法; CARS-SPA 为竞争性自适应重加权-连续投影算法。

Note: PLSR refers to partial least squares regression; S-G refers to Savitzky-Golay. SPA refers to successive projection algorithm; CARS-SPA refers to competitive adaptive reweighting-successive projection algorithm.

图 3 基于近红外光谱的剖面土壤有机质反演模型构建

Fig.3 Construct the inversion models of near-infrared spectra of profile soil organic matter

最初的光谱数据范围为 1 603~2 598 nm, 但由于光谱始末端的信噪比较低, 所以只包括 1 623~2 467 nm 范围内的数据 (以 6.6 nm 为间距, 共包含 129 个波长)^[34], 此范围内的光谱对土壤湿度变化的敏感性较高^[18]。在本研究中, 共选取的 6 种土壤湿度水平 (水分张力) 依次为饱和状态、0.033、0.100、0.330、1.500 MPa 和风干状态, 在试验进行过程中, 受人力、采集设备等因素限制的影响^[36], 这 6 组中土壤湿度水平下所对应的土壤子样品总数目不完全一致, 依次为 301、304、306、304、

303 和 302 个。在本研究中依据控制变量原则, 每份土壤样本仅以土壤湿度水平不同, 所选取的土壤位置, 深度与方向均保持一致则其校正集中包含的样本个数为 221 个, 预测集中所包含的样本个数为 80~85 个。选用恰当的光谱预处理方法能有效减轻信噪比低、杂散光和谱峰重叠等问题对光谱曲线的影响^[1]。对从饱和到风干状态的 6 组湿度水平下的土壤光谱数据分别进行 Savitzky-Golay (S-G) 平滑、一阶导数 (first derivative)、移动平滑 (moving average)、基线校正 (baseline)、

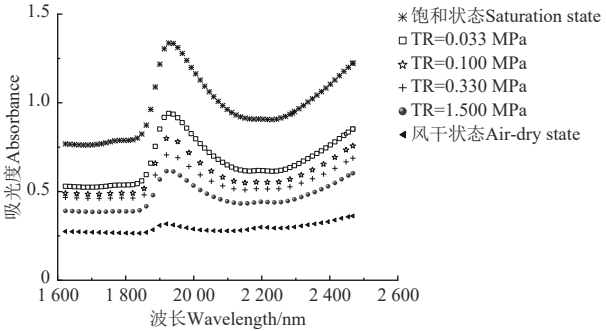
归一化 (normalize)、标准正态变量变化 (standard normal variate, SNV) 与多元散射校正 (multiple scattering correction, MSC) 这 7 种光谱预处理方法处理, 处理后的光谱数据再进行 PLSR 建模, 以校正集决定系数 (coefficient of determination in calibration, R_c^2) 以及均方根误差 (root mean square error of the calibration set, RMSEC) 作为模型评价指标。通过对光谱进行特征波长筛选可以有效解决光谱信息量大、数据冗杂等问题, 提高模型预测准确度和鲁棒性^[5]。在本研究共选取了 6 种土壤湿度下的光谱数据, 首先采用连续投影算法 (successive projection algorithm, SPA) 来进行特征波长筛选, 指定波长变量数为 1~39, 根据校正集的内部交叉验证均方根误差 (root mean square error of cross validation, RMSECV) 确定最佳特征波长数^[5]。

2 结果与分析

2.1 初步数据分析

图 4 表示选取相同的地点、深度和采集方向, 有机质含量均为 3.49% 的土壤样品在饱和状态、水分张力分别在 0.033、0.100、0.330、1.500 MPa 和风干状态这 6 组土壤湿度水平下得到的吸光度曲线。

由图 4 可知, 不同湿度水平下土壤吸光度曲线的变化趋势相似。吸光度随土壤湿度的增加而增大, 其中饱和状态与风干状态下土壤的光谱曲线差异最为显著, 此现象与 SEIDEL 等^[40]所得到的结论一致。导致这一现象的原因是, 当土壤样品处于湿润状态时, 水作为薄层吸附在粒子表面, 自由液态水充满孔隙空间, 将土壤颗粒周围介质的相对折射率由空气折射率 (λ 约为 1.00) 改变为水的折射率 (λ 约为 1.33), 其折射率之差减少导致更多的前向散射, 该散射过程比之前的路径长度更长^[10]。此外, 附着在土壤颗粒周围的水膜引起了散射过程光的额外反射, 这导致更大一部分光传播到土壤深处从而吸光度增大^[18]。



注：有机质含量为 3.49%；TR 为水分张力。
Note: The organic matter content is 3.49%; TR refers to moisture tension.

图 4 不同土壤湿度水平下土壤样品吸光度的原始曲线
Fig.4 Original absorbance curves of soil samples at different soil moisture levels

图 4 中不同土壤湿度水平下吸收峰的宽度与深度不完全一致, 但其均位于 1 940 nm 附近, 这是由于水中的 O-H 官能团的伸缩振动和变角振动的合频使其 1 940 nm 附近有强吸收波段, 其吸光系数在近红外区域是最大的, 尽管也存在着 C-H 键、N-H 键和 C=O 键等其他官能团的吸收, 但这些官能团的吸收系数要远小于水分的 O-H 键, 使 1 940 nm 波长处对水分吸收十分强烈, 则导致吸光度较大^[11-12]。而在 2 200 nm 附近有较弱的吸收峰, 该吸收峰由土壤有机化合物中的 O-H 官能团振动引起^[32]。因此当土壤中水分含量减少时, 与水分相关的 1 940 nm 附近吸收峰高度与宽度明显减少, 而 2 200 nm 附近的吸收峰不受水分含量变化的影响变得明显^[18]。并且在图 4 中也可以明显观察出在较长的波长范围内, 2 个相邻的土壤湿度水平之间吸光度曲线的差别会比在较短波长范围大^[10]。

2.2 原始光谱及光谱预处理

经比较后得到的 6 组土壤湿度水平下最优预处理方法如表 3 所示。

表 3 每种光谱预处理方法在 6 组土壤湿度水平下的 R_c^2 和 RMSEC
Table 3 R_c^2 and RMSEC of each spectral preprocessing method under six groups of soil moisture levels

水分张力水平 Moisture tension level	评价指标 Evaluation index	S-G 平滑	FD	Moving Average	Baseline	Normalize	SNV	MSC
饱和状态 Saturation state	R_c^2	0.696	0.709	0.696	0.690	0.731	0.727	0.725
	RMSEC/(g·kg ⁻¹)	0.914	0.893	0.914	0.923	0.860	0.866	0.869
0.033 MPa	R_c^2	0.747	0.790	0.722	0.750	0.719	0.793	0.791
	RMSEC/(g·kg ⁻¹)	0.822	0.749	0.863	0.818	0.867	0.744	0.748
0.100 MPa	R_c^2	0.754	0.820	0.754	0.790	0.711	0.820	0.819
	RMSEC/(g·kg ⁻¹)	0.822	0.704	0.822	0.759	0.891	0.702	0.704
0.330 MPa	R_c^2	0.754	0.796	0.740	0.782	0.757	0.813	0.810
	RMSEC/(g·kg ⁻¹)	0.822	0.748	0.844	0.773	0.817	0.717	0.723
1.500 MPa	R_c^2	0.759	0.784	0.759	0.752	0.674	0.735	0.748
	RMSEC/(g·kg ⁻¹)	0.813	0.769	0.813	0.824	0.944	0.853	0.831
风干状态 Air-dry state	R_c^2	0.796	0.835	0.796	0.807	0.737	0.835	0.844
	RMSEC/(g·kg ⁻¹)	0.748	0.674	0.749	0.729	0.849	0.673	0.654

注：S-G 为 Savitzky-Golay；FD 为一阶导数；SNV 为标准正态变量变化；MSC 为多元散射校正。 R_c^2 为校正集决定系数；RMSEC 为校正集均方根误差。
Note: S-G refers to Savitzky-Golay; FD refers to first derivative; SNV refers to standard normal variate; MSC refers to multiple scattering correction. R_c^2 refers to the coefficient of determination in calibration; RMSEC refers to the root mean square error of the calibration set.

由表 3 可知, 因共有 0.033、0.100 和 0.330 MPa 3 个水分张力下对应的土壤样品光谱数据选用 SNV 预处理效

果最好, 为了对不同湿度下的土壤样品的预处理方法进行统一, 故选择 SNV 方法对土壤光谱数据来进行预处理。

2.3 筛选特征波长

SPA 算法所筛选出的特征波长如图 5 中所示。图 5 中在饱和、0.033、0.100、0.330、1.500 MPa 和风干状态各湿度水平下经 SPA 算法筛选后的特征波长数目分别为 13、15、16、9、9 和 13 个。为了弥补只采用 SPA 算法筛选出特征波长可能包含一些无效或干扰信息使所得特征后数量较多的不足^[3], 在本文研究中设定蒙特卡罗采样次数为 500, 采用竞争性自适应重加权 (competitive adaptive reweighted sampling, CARS) 算法对全光谱波长初选剔除不重要变量后再采用 SPA 算法进行特征波长的选择^[5], 所筛选出的特征波长如表 4 中所示。经 CARS-SPA 算法筛选后的特征波长数目饱和状态下最高为 8 个, 风干状态下最少仅为 6 个, 各土壤湿度水平相比于全谱 129 个波长来说数目明显减少。

表 4 不同土壤湿度水平下 CARS-SPA 算法筛选后的特征波长

Table 4 Characteristic wavelengths filtered by CARS-SPA algorithm under different soil moisture levels

水分张力水平 Moisture tension level	波长数量 Number of wavelengths	CARS-SPA 算法筛选后的特征波长 Characteristic wavelength filtered by CARS-SPA algorithm/nm
饱和状态 Saturation state	8	1 629.6、1 807.8、2 038.8、2 137.8、2 210.4、2 309.4、2 349、2 441.4
0.033 MPa	7	1 629.6、1 781.4、2 032.2、2 111.4、2 309.4、2 362.2、2 461.2
0.100 MPa	7	1 629.6、1 807.8、1 873.8、2 091.6、2 289.6、2 362.2、2 461.2
0.330 MPa	7	1 636.2、1 755、1 966.2、1 986、2 098.2、2 309.4、2 335.8
1.500 MPa	7	1 629.6、1 755、2 118、2 296.2、2 335.8、2 388.6、2 408.4
风干状态 Air-dry state	6	1 636.2、1 735.2、1 939.8、2 012.4、2 302.8、2 342.4

2.4 土壤有机质含量预测建模与分析

2.4.1 基于全波长及特征波长的 PLSR 建模结果分析

对于本研究中所选取的 6 组土壤湿度水平下的光谱数据, 在 1 623~2 467 nm 范围内, 分别建立 PLSR、SPA-PLSR 和 CARS-SPA-PLSR 模型以 R_c^2 与 RMSEC 作为评价模型预测效果的指标, 其对比如表 5 所示。

表 5 不同土壤湿度水平下 PLSR、SPA-PLSR 和 CARS-SPA-PLSR 模型比较

Table 5 Comparison of PLSR、SPA-PLSR and CARS-SPA-PLSR under different soil moisture levels

水分张力 Moisture tension level	评价指标 Evaluation index	PLSR	SPA-PLSR	CARS-SPA-PLSR
饱和状态 Saturation state	R_c^2	0.698	0.709	0.731
	RMSEC/(g·kg ⁻¹)	0.911	0.893	0.860
0.033 MPa	R_c^2	0.723	0.701	0.783
	RMSEC/(g·kg ⁻¹)	0.862	0.894	0.763
0.100 MPa	R_c^2	0.753	0.749	0.775
	RMSEC/(g·kg ⁻¹)	0.823	0.830	0.786
0.330 MPa	R_c^2	0.754	0.795	0.825
	RMSEC/(g·kg ⁻¹)	0.822	0.750	0.693
1.500 MPa	R_c^2	0.760	0.785	0.784
	RMSEC/(g·kg ⁻¹)	0.812	0.767	0.770
风干状态 Air-dry state	R_c^2	0.797	0.811	0.831
	RMSEC/(g·kg ⁻¹)	0.747	0.720	0.681

由表 5 和图 6 观察可得, 在校正集与预测集所建立的 PLSR 模型中, 同一土壤湿度水平下 RMSEC 和预测集均方根误差 (root mean square error of the prediction set, RMSEP) 差别不明显, 即可说明 PLSR 模型稳定性较高, 未出现过拟合状态。当水分张力由 1.500 MPa 降低至 0.033 MPa 时, 土壤中水分含量逐渐增大, 此时 R_c^2 与预测集决定系数 (coefficient of determination in prediction,

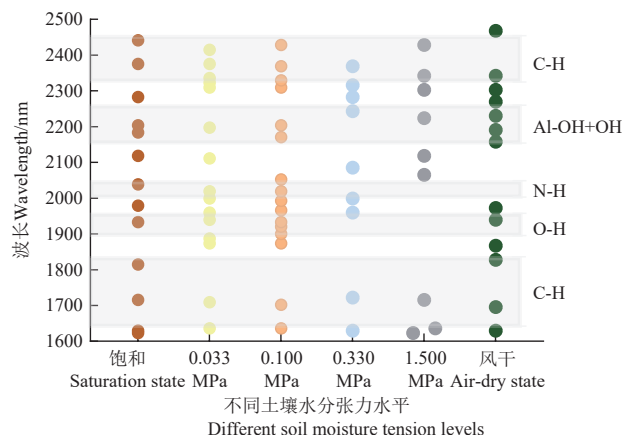


图 5 不同土壤湿度水平下 SPA 算法筛选后的特征波长

Fig.5 Characteristic wavelengths after screening by SPA algorithm at different soil moisture levels

R_p^2) 越来越小并且 RMSEC 和 RMSEP 越来越大; 水分张力为 1.500 MPa 时, R_c^2 和 R_p^2 分别为 0.760 和 0.808, RMSEC 和 RMSEP 则是 0.812 和 0.801 g/kg, 当土壤逐步湿润到 0.033 MPa 时, R_c^2 和 R_p^2 依次为 0.723 和 0.682, 均方根误差分别是 0.862 和 0.990 g/kg。即可以说明水分张力为 1.500 MPa 下对 SOM 的预测效果比水分张力 0.033 MPa 时要好。这是因为水分张力由 1.500 MPa 逐步到 0.033 MPa 的过程中, 土壤含水率增加, 土壤结构及各组分比例发生变化, 从而水分对 SOM 的光谱信息影响增加^[26], 而当土壤中水分含量继续增加, 使其由水分张力 0.033 MPa 湿润至饱和状态时, 水分中 O-H 键信号覆盖了有机质中 C-H 键等化学键的信号, 同时水分也和其他物质相互作用引起光谱的差异导致预测效果变差^[12], 此状态下 R_c^2 和 R_p^2 分别为 0.698 和 0.706 远低于风干状态下 $R_c^2=0.797$ 、 $R_p^2=0.796$, 同时 RMSEC 和 RMSEP 为 0.911、0.889 g/kg 明显高于风干状态下 RMSEC=0.747 g/kg、RMSEP=0.776 g/kg。

由表 5 可以得出, SPA-PLSR 模型在饱和、0.330、1.500 MPa 与风干状态这 4 种湿度水平下的 R_c^2 比 PLSR 模型更接近于 1, RMSEC 更小; CARS-SPA-PLSR 模型在 6 种湿度水平下的 R_c^2 均高于对应湿度水平下 PLSR 模型的 R_c^2 并且 RMSEC 更低; 只有当水分张力为 1.500 MPa 时, CARS-SPA-PLSR 模型的 R_c^2 和 RMSEC 与 SPA-PLSR 模型中的对应指标十分接近, 而在其他 5 种土壤湿度水平下, 均是 CARS-SPA-PLSR 模型精度要优于 SPA-PLSR 模型。综上, 以 R_c^2 和 RMSEC 为模型评价指标, 在 6 种湿度水平下 SPA-PLSR 模型和 CARS-SPA-PLSR 模型的精度均优于 PLSR 模型。其中采用 CARS-

SPA-PLSR 模型不仅可以使筛选出的特征波长数目相比于仅采用 SPA 算法明显减少, 而且也提高了精确度。

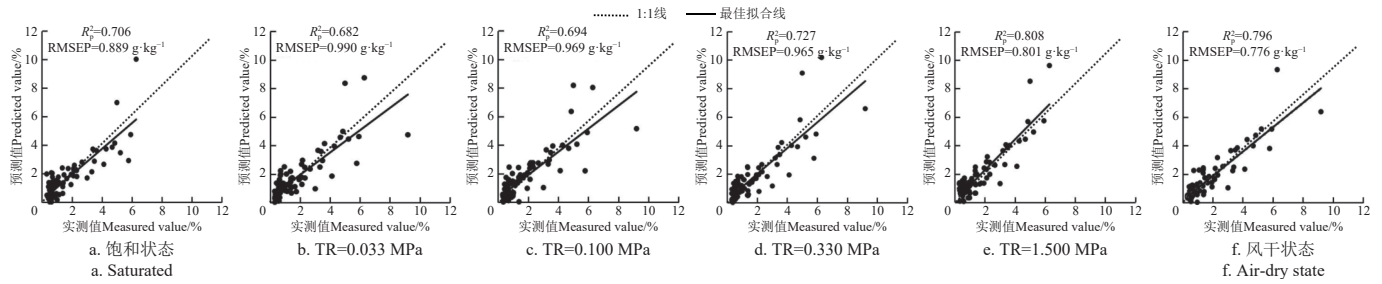


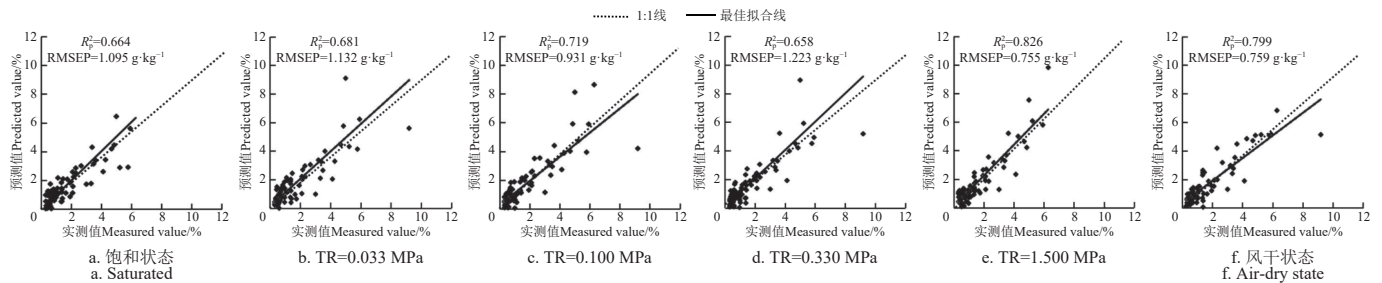
图 6 不同土壤湿度水平下土壤有机质 PLSR 预测相关性分析

Fig.6 Correlation analysis of PLSR prediction of soil organic matter under different soil moisture levels

2.4.2 结合预处理方法的模型构建及验证

在实际应用中, 选用恰当的预处理与筛选特征波长方法可以消除噪声进而对光谱进行修正^[12], 并且以更少的波长达到对 SOM 更优的预测效果。但不同光谱预处理方法对变量选择算法可能会造成影响^[41], 因此本研究

预处理及筛选特征波长均对原始光谱进行处理后再选用 2.4 节中 SPA 算法和 CARS-SPA 算法同 2.3 节中选用的 SNV 预处理方法相结合, 在图 7 和图 8 中分别在 6 种土壤湿度水平下建立 SNV-SPA-PLSR 模型和 SNV-CARS-SPA-PLSR 模型。



注: R_p^2 为预测集决定系数; RMSEP 为预测集均方根误差。

Note: R_p^2 refers to the coefficient of determination in prediction; RMSEP refers to the root mean square error of the prediction set.

图 7 不同土壤湿度水平下土壤有机质 SNV-SPA-PLSR 模型预测相关性分析

Fig.7 Correlation analysis of soil organic matter prediction by SNV-SPA-PLSR models under different soil moisture levels

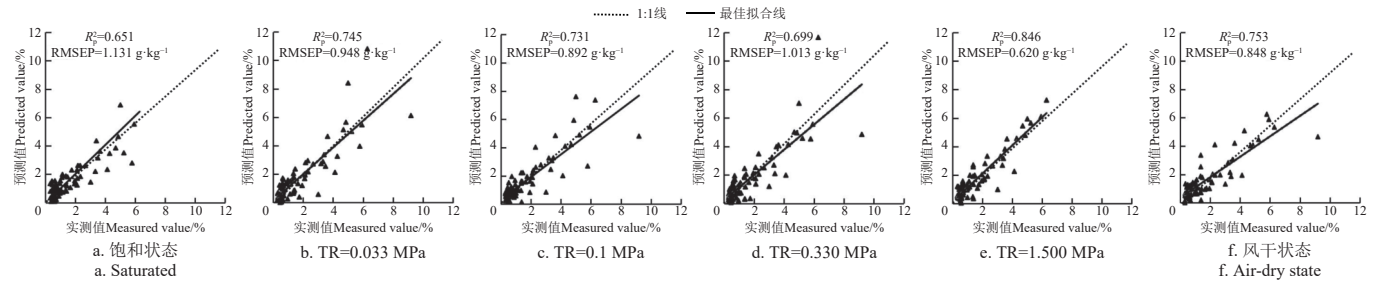


图 8 不同土壤湿度水平下土壤有机质 SNV-CARS-SPA-PLSR 模型预测相关性分析

Fig.8 Correlation analysis of soil organic matter prediction by SNV-CARS-SPA-PLSR models under different soil moisture levels

在不同土壤湿度下, 在图 7 和图 8 中将得到的 SNV-SPA-PLSR 模型和 SNV-CARS-SPA-PLSR 模型中的 R_p^2 与 RMSEP 进行比较, 当水分张力由 0.033 MPa 增加到 1.500 MPa 时, SNV-CARS-SPA-PLSR 模型的 R_p^2 比 SNV-SPA-PLSR 模型更高且对应的 RMSEP 逐渐降低。而对于没有采用水分张力作为土壤湿润程度衡量标准的饱和状态与风干状态来说, SNV-SPA-PLSR 模型 R_p^2 分别为 0.664、0.799, 高于 SNV-CARS-SPA-PLSR 模型对应的 R_p^2 为 0.651 和 0.753 且 SNV-CARS-SPA-PLSR 模型中 RMSEP 更高。因此在对水分张力分别为 0.033、0.100、0.330 和 1.500 MPa 的土壤有机质含量预测时, 选用 SNV-CARS-SPA-PLSR 模型效果最佳。而对于土壤处于饱和与干燥状态时则选用 SNV-SPA-PLSR 模型。但无论 SNV-SPA-

PLSR 模型还是 SNV-CARS-SPA-PLSR 模型, 均是土壤 在风干状态下的 SOM 预测精度远高于饱和状态。这也与图 6 中仅采用 PLSR 模型进行 SOM 预测得到的结论一致。

2.5 基于特征波长的 SOM 校准模型在不同数据集上的应用

传统的不同土壤湿度水平近红外光谱“一对一”式预测 SOM 模型已经满足不了实际生产应用, 一组土壤湿度水平对应一个模型对于操作者来说过于繁琐, 同时繁多又复杂的模型种类也不利于仪器中模型的调用^[42]。由表 6 得出, SNV-CARS-SPA-PLSR 模型在 3 种湿度水平下预测效果均优于其他 2 种模型, 而 PLSR 模型和 SNV-SPA-PLSR 模型分别在 2 种与 1 种湿度水平下取得最优结果。因此往往选择一组土壤湿度水平下有机质预

测精度最高的 SNV-CARS-SPA-PLSR 模型来预测其余 5 组土壤湿度水平下的有机质。

表 6 不同土壤湿度水平下 PLSR、SNV-SPA-PLSR 和 SNV-CARS-SPA-PLSR 模型比较

Table 6 Comparison of PLSR, SNV-SPA-PLSR and SNV-CARS-SPA-PLSR under different soil moisture levels

水分张力水平 Moisture tension level	PLSR		SNV-SPA-PLSR		SNV-CARS-SPA-PLSR	
	R_p^2	RMSEP/(g·kg ⁻¹)	R_p^2	RMSEP/(g·kg ⁻¹)	R_p^2	RMSEP/(g·kg ⁻¹)
饱和状态 Saturation state	0.706	0.889	0.664	1.095	0.651	1.131
0.033 MPa	0.682	0.990	0.681	1.132	0.745	0.948
0.100 MPa	0.694	0.969	0.719	0.931	0.731	0.892
0.330 MPa	0.727	0.965	0.658	1.223	0.699	1.013
1.500 MPa	0.808	0.801	0.826	0.755	0.846	0.620
风干状态 Air-dry state	0.796	0.776	0.799	0.759	0.753	0.848

注: R_p^2 为预测集决定系数; RMSEP 为预测集均方根误差。

Note: R_p^2 refers to the coefficient of determination in prediction; RMSEP refers to the root mean square error of the prediction set.

从表 7 中对 4 种土壤湿度水平下的 R_p^2 和 RMSEP 比较可以得出, 使用水分张力为 1.500 MPa 时对应的 SNV-SPA-CARS-PLSR 模型分别预测选取 6 组土壤湿度水平下的有机质时, 所得到 R_p^2 分别为 0.765、0.762、0.766、

0.738、0.846 和 0.760, 除水分张力为 0.100 MPa 的其余 5 种湿度水平下, 水分张力为 1.500 MPa 明显高于使用其他土壤湿度水平建立 SNV-SPA-CARS-PLSR 模型的预测效果。

表 7 6 组土壤湿度水平下的 SNV-CARS-SPA-PLSR 模型分别对 6 组湿度水平下土壤有机质预测所得 R_p^2 和 RMSEP

Table 7 The SNV-CARS-SPA-PLSR model for each of the six soil moisture levels predicted R_p^2 and RMSEP for soil organic matter at each of the six moisture levels

SNV-CARS-SPA-PLSR 模型 Evaluation index	评价指标 Evaluation index	水分张力 Moisture tension level						混合集 Mixed set
		饱和状态 Saturation state	0.033 MPa	0.100 MPa	0.330 MPa	1.500 MPa	风干状态 Air-dry state	
饱和状态模型 Saturation state model	R_p^2	0.651	0.661	0.676	0.596	0.722	0.707	0.166
	RMSEP/(g·kg ⁻¹)	1.131	1.698	1.936	2.697	3.665	11.536	4.874
0.033 MPa 模型 0.033 Mpa model	R_p^2	0.686	0.745	0.746	0.675	0.801	0.738	0.413
	RMSEP/(g·kg ⁻¹)	0.932	0.948	1.006	1.307	1.557	5.344	2.287
0.100 MPa 模型 0.100 Mpa model	R_p^2	0.564	0.677	0.731	0.645	0.772	0.718	0.393
	RMSEP/(g·kg ⁻¹)	1.011	0.996	0.892	1.136	1.416	3.962	1.768
0.330 MPa 模型 0.330 Mpa model	R_p^2	0.619	0.747	0.774	0.699	0.772	0.747	0.311
	RMSEP/(g·kg ⁻¹)	0.935	0.902	0.821	1.013	1.126	7.438	3.463
1.500 MPa 模型 1.500 Mpa model	R_p^2	0.765	0.762	0.766	0.738	0.846	0.760	0.468
	RMSEP/(g·kg ⁻¹)	1.016	1.053	0.952	0.912	0.620	2.750	1.593
风干状态模型 Air-dry state model	R_p^2	0.398	0.493	0.374	0.435	0.217	0.753	0.156
	RMSEP/(g·kg ⁻¹)	3.458	3.254	3.082	2.869	2.618	0.848	2.679

同样使用该模型预测包含所有湿度水平的混合样本集有机质时所得到的 R_p^2 仅为 0.468, RMSEP 高达 1.593 g/kg, 远差于对单独一种土壤湿度水平下的 SOM 预测效果。综上, 1.500 MPa 下的 SNV-SPA-CARS-PLSR 模型对混合土壤湿度样本集 SOM 的预测效果仍优于其他湿度水平。

结合表 4 中水分张力为 1.500 MPa 时 CARS-SPA 算法筛选出的特征波长所对应的化学键分析可知, 当土壤样品受到 NIR 波段的光照射时, SOM 中包含的 O-H、C-H、N-H、C=O 等基团会在相关波段产生基准振动 (含有伸缩振动与变角振动) 的倍频或者合频振动的吸收^[12]。1 629.6~1 807.8 nm 区域为 C-H 键第一倍频的伸缩振动吸收, 2 118 nm 附近处与 N-H 键不对称伸缩振动有关, 2 296.2 nm 附近处为 N-H 键伸缩振动与 C=O 键伸缩振动吸收, 2 335.8 nm 处 C-H 键伸缩振动与变形振动, 2 388.6 nm 处为 O-H 键的第二倍频的变形振动吸收, 水分张力为 1.500 MPa 时 CARS-SPA 算法筛选出的特征波长正好同 SOM 中含有的多数官能团有直接相关性, 这可能会使得在 1.500 MPa 下 SNV-SPA-CARS-PLSR 模型

对 SOM 预测精度较高。但是在田间同一深度处, 各个地点的土壤湿度水平难以保证完全一致, 因此所选用的模型也应该尽可能提高在不同土壤湿度水平数据集之间的模型可转移性, 在 Vis-NIR 范围内已经引入了一些技术, 如 WIJEWARDANE 等^[43]提出的全局水分模型 (global moisture modelling, GMM) 是在校正集中包含更多不同含水量的样本后重新校准模型。GMM 并不总是比外部参数正交化 (external parameter orthogonalization, EPO)^[6]和 DS 算法^[13]效果好, 但其在所有的湿度水平上稳定性最高。JIANG 等^[44]通过对比 8 组不同水分含量对土壤有机碳预测的影响, 得出广义最小二乘加权 (generalized least squares weighting, GLSW) 方法结合 PLSR 在去除水分干扰方面的效果最佳并且该模型可以提高不同湿度下预测模型的可移植性。因此如果要建立一个适用于更广泛的土壤湿度下对有机质含量预测精度更高的模型仍需要结合斜率偏置校正 (slope bias correction, SB)^[43]、正交信号校正 (orthogonal signal correction, OSC)^[44]与分段直接标准化 (piecewise direct standardization, PDS)^[45]等算法或者将预测集中具有代表性的部分样品并入校正

集中重新建模以提高模型通用性^[46]。

3 结 论

为更好地研究土壤湿度对近红外光谱反演剖面土壤有机质 (soil organic matter, SOM) 精度的影响, 该研究通过控制水分张力这一指标, 构建基于全谱及特征波长的 SOM 近红外光谱反演模型, 并将其与预处理方法相结合。选择效果最优的湿度水平下建立的校准模型应用于不同数据集上。结论如下:

1) 土壤吸光度随土壤湿度的增加而增大, 不同湿度水平下土壤吸光度曲线的吸收峰均位于 1 940 和 2 200 nm 附近, 但其宽度与深度不完全一致。

2) 基于连续投影算法 (successive projection algorithm, SPA) 和竞争性自适应重加权-连续投影算法 (competitive adaptive reweighting-successive projection algorithm, CARS-SPA) 构建的偏最小二乘回归 (partial least squares regression, PLSR) 模型效果均优于 PLSR 模型。将这 2 种模型与选定标准正态变量变化 (standard normal variate, SNV) 预处理方式结合, 得到 SNV-CARS-SPA-PLSR 模型在水分张力分别为 0.033、0.100、0.330 和 1.500 MPa 下的建模效果明显优于仅采用 SNV-SPA-PLSR 的模型; 但是在土壤处于饱和与风干状态时则选用 SNV-SPA-PLSR 模型更好。选用 PLSR 模型、SNV-SPA-PLSR 模型和 SNV-CARS-SPA-PLSR 模型均是土壤在风干状态下的 SOM 预测精度明显优于饱和状态。

3) 在田间同一深度处, 各地点的土壤湿度难以保证完全一致, 导致传统的不同土壤湿度水平近红外光谱“一对一”式预测 SOM 模型难以满足实际应用。本文对光谱模型在不同土壤湿度水平间的适用性研究表明, 采用水分张力为 1.500 MPa 时所建立的 SNV-CARS-SPA-PLSR 模型来分别预测风干状态、0.330、0.100、0.033 MPa 与饱和状态这 5 组湿度水平和其混合样本集中的土壤有机质取得效果最好。这可能是由于水分张力为 1.500 MPa 时筛选出的特征波长同 SOM 中含有的多数官能团有直接相关性。

致谢: 感谢美国 Hummel J W、Sudduth K A 和 Hollinger S E 等为本研究提供数据支持, 特别感谢他们在土壤数据采集、光谱测量和实验室土壤养分测定等方面做的大量工作。Sudduth 教授在本研究论文初稿撰写等方面也提供了指导, 在此表示感谢。

[参 考 文 献]

- [1] WEI L, YUAN Z, WANG Z, et al. Hyperspectral inversion of soil organic matter content based on a combined spectral index model[J]. *Sensors*, 2020, 20(10): 2777.
- [2] NELSON D W, SOMMERS L E. Total carbon, organic carbon, and organic matter[M]//KLUTEO. Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties. Hoboken: Wiley, 1986.
- [3] 胡国田, 尚会威, 谭瑞虹, 等. 不同土壤类型的有机质含量的可见-近红外光谱检测模型传递方法研究[J]. *光谱学与光谱分析*, 2022, 42(10): 3148-3154.
- HU Guotian, SHANG Huiwei, TAN Ruihong, et al. Research on model transfer method of organic matter content estimation of different soils using VNIR spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(10): 3148-3154. (in Chinese with English abstract)
- [4] CONFORTI M, CASTRIGNANÒ A, ROBUSTELLI G, et al. Laboratory-based Vis-NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content[J]. *Catena*, 2015, 124: 60-67.
- [5] 钟翔君, 杨丽, 张东兴, 等. 砂壤潮土有机质含量可见-近红外光谱预测[J]. *光谱学与光谱分析*, 2022, 42(9): 2924-2930.
- ZHONG Xiangjun, YANG Li, ZHANG Dongxing, et al. Prediction of organic matter content in sandy fluvo-aquic soil by visible-near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2022, 42(9): 2924-2930. (in Chinese with English abstract)
- [6] DE SANTANA F B, DE GIUSEPPE L O, DE SOUZA A M, et al. Removing the moisture effect in soil organic matter determination using NIR spectroscopy and PLSR with external parameter orthogonalization[J]. *Microchemical Journal*, 2019, 145: 1094-1101.
- [7] ZHANG J, LIU T, WANG J, et al. A simple but effective evaluation criterion for parameters optimization of EPO and its application to moisture insensitive prediction of soil organic matter[J]. *Chemometrics and Intelligent Laboratory Systems*, 2023, 236: 104794.
- [8] XIA K, XIA S, SHEN Q, et al. Moisture spectral characteristics and hyperspectral inversion of fly ash-filled reconstructed soil[J]. *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy*, 2021, 253(3): 119590.
- [9] 王世芳, 程旭, 宋海燕. 水分对土壤有机质检测影响的光谱特性分析及抗水分干扰模型建立[J]. *光谱学与光谱分析*, 2016, 36(10): 3249-3253.
- WANG Shifang, CHENG Xu, SONG Haiyan. Analysis of the effect of moisture on soil organic matter determination and anti-moisture interference model building based on Vis-NIR spectral technology[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(10): 3249-3253. (in Chinese with English abstract)
- [10] WANG Y P, LEE C K, DAI Y H, et al. Effect of wetting on the determination of soil organic matter content using visible and near-infrared spectrometer[J]. *Geoderma*, 2020, 376: 114528.
- [11] CHANG C W, LAIRD D A, HURBURGH JR C R. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties [J]. *Soil Science*, 2005, 170(4): 244-255.
- [12] MARAKKALA MANAGE L P, GREVE M H, KNADEL M, et al. Visible-Near-Infrared Spectroscopy prediction of soil characteristics as affected by soil-water content[J]. *Soil Science Society of America Journal*, 2018, 82(6): 1333-1346.
- [13] JI W, LI S, CHEN S, et al. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions[J]. *Soil and Tillage Research*, 2016, 155: 492-500.
- [14] 于雷, 洪永胜, 朱亚星, 等. 去除土壤水分对高光谱估算土壤有机质含量的影响[J]. *光谱学与光谱分析*, 2017, 37(7): 2146-2151.
- YU Lei, HONG Yongsheng, ZHU Yaxing, et al. Removing the effect of soil moisture content on hyperspectral reflectance for the estimation of soil organic matter content[J]. *Spectroscopy*

- and Spectral Analysis, 2017, 37(7): 2146-2151. (in Chinese with English abstract)
- [15] STEVENS A, NOCITA M, TÓTH G, et al. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy[J]. *PloS One*, 2013, 8(6): e66409.
- [16] ROSSEL R A V, BEHRENS T, BEN-DOR E, et al. A global spectral library to characterize the world's soil[J]. *Earth-Science Reviews*, 2016, 155: 198-230.
- [17] 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(S0): 323-328, 350.
CUI Yulu, YANG Wei, WANG Weichao, et al. Design and experiment of portable soil organic matter detector based on spectroscopy principle[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(S0): 323-328, 350. (in Chinese with English abstract)
- [18] NOCITA M, STEVENS A, NOON C, et al. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy[J]. *Geoderma*, 2013, 199: 37-42.
- [19] 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 Xiuquan, LI Zhiwei, ZHENG Decong, 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. (in Chinese with English abstract)
- [20] PEI X, SUDDUTH K A, VEUM K S, et al. Improving in-situ estimation of soil profile properties using a multi-sensor probe[J]. *Sensors*, 2019, 19(5): 1011.
- [21] LI S, SHI Z, CHEN S, et al. In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai-Tibet plateau[J]. *Environmental Science & Technology*, 2015, 49(8): 4980-4987.
- [22] RUMPEL C, KÖGEL-KNABNER I. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle[J]. *Plant and Soil*, 2011, 338: 143-158.
- [23] 郝翔翔, 邹文秀, 韩晓增. 长期不同利用方式对黑土剖面中有机质化学组成的影响[J]. 土壤学报, 2022, 59(5): 1228-1237.
HAO Xiangxiang, ZOU Wenxiu, HAN Xiaozeng. Impact of Long-term Land-Use Patterns on Chemical Composition of Soil Organic Matter in Mollisol Profile[J]. *Acta Pedologica Sinica*, 2022, 59(5): 1228-1237. (in Chinese with English abstract)
- [24] 付娟, 马仁明, 贾燕锋, 等. 机械压实对农田土壤性质及土壤侵蚀的影响研究进展[J]. 农业工程学报, 2022, 38(增刊): 27-36.
FU Juan, MA Renming, JIA Yanfeng, et al. Research progress in the effects of mechanical compaction on soil properties and soil erosion in farmland[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2022, 38(Suppl): 27-36. (in Chinese with English abstract)
- [25] 李硕, 李春莲, 陈颂超, 等. 基于野外可见近红外光谱和水分影响校正算法的土壤剖面有机碳预测[J]. 光谱学与光谱分析, 2021, 41(4): 1234-1239.
LI Shuo, LI Chunlian, CHEN Songchao, et al. Removing the Effects of Water From Visible-Near Infrared Spectra in Soil Profiles for the Estimation of Organic Carbon[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(4): 1234-1239. (in Chinese with English abstract)
- [26] 王超, 冯美臣, 杨武德, 等. 一种降低土壤水分对土壤有机质光谱监测精度的新方法[J]. 光谱学与光谱分析, 2015, 35(12): 3495-3499.
WANG Chao, FENG Meichen, YANG Wude, et al. A new method to decline the SWC effect on the accuracy for monitoring SOM with hyperspectral technology[J]. *Spectroscopy and Spectral Analysis*, 2015, 35(12): 3495-3499. (in Chinese with English abstract)
- [27] 胡晓艳, 崔旭, 韩小平, 等. 基于修正系数法的抗水分干扰土壤有机质近红外预测模型研究[J]. 光谱学与光谱分析, 2019, 39(4): 1059-1062.
HU Xiaoyan, CUI Xu, HAN Xiaoping, et al. Study on soil organic matter prediction model based on moisture correction algorithm and near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2019, 39(4): 1059-1062. (in Chinese with English abstract)
- [28] MOUAZEN A M, KAROUI R, DE BAERDEMAEKER J, et al. Classification of soils into different moisture content levels based on VIS-NIR spectra[C]//Proceedings of the 2006 ASABE Annual International Meeting. USA: ASABE, 2006: 9-12.
- [29] HAUBROCK S N, CHABRILLAT S, LEMMNITZ C, et al. Surface soil moisture quantification models from reflectance data under field conditions[J]. *International Journal of Remote Sensing*, 2008, 29(1): 3-29.
- [30] 武永峰, 董一威, 胡新, 等. 基于近红外漫反射光谱的农田原位表层土壤含水量定量建模方法比较[J]. 光谱学与光谱分析, 2015, 35(12): 3416-3421.
WU Yongfeng, DONG Yiwei, HU Xin, et al. Quantification of agricultural in-situ surface soil moisture content using near infrared diffuse reflectance spectroscopy: A comparison of modeling methods[J]. *Spectroscopy and Spectral Analysis*, 2015, 35(12): 3416-3421. (in Chinese with English abstract)
- [31] FABRE S, BRIOTTET X, LESAINOUX A. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4-2.5um domain[J]. *Sensors*, 2015, 15(2): 3262-3281.
- [32] ZHOU P, YANG W, LI M, et al. A new coupled elimination method of soil moisture and particle size interferences on predicting soil total nitrogen concentration through discrete nir spectral band data[J]. *Remote Sensing*, 2021, 13(4): 762.
- [33] YUE J, TIAN J, TIAN Q, et al. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 154: 216-230.
- [34] HUMMEL J W, SUDDUTH K A, HOLLINGER S E. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor[J]. *Computers and Electronics in Agriculture*, 2001, 32(2): 149-165.
- [35] SUDDUTH K A, HUMMEL J W. Near-infrared spectrophotometry for soil property sensing[C]//Proceedings of the Optics in Agriculture and Forestry. USA: SPIE, 1993: 14-25.
- [36] HOLLINGER S E, ISARD S A. A soil moisture climatology of Illinois[J]. *Journal of Climate*, 1994, 7(5): 822-833.
- [37] HOLLINGER S E, ISARD S A. A time- and labor-saving method of extracting undisturbed soil samples[J]. *Journal of Production Agriculture*, 1989, 2(4): 374-375.
- [38] SUDDUTH K A, HUMMEL J W. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer[J]. *Transactions of the ASAE*, 1993, 36(6): 1571-1582.
- [39] BINEY J K M, BORŮVKA L, CHAPMAN AGYEMAN P, et al. Comparison of field and laboratory wet soil spectra in the Vis-

- NIR range for soil organic carbon prediction in the absence of laboratory dry measurements[J]. *Remote Sensing*, 2020, 12(18): 3082.
- [40] SEIDEL M, VOHLAND M, GREENBERG I, et al. Soil moisture effects on predictive VNIR and MIR modeling of soil organic carbon and clay content[J]. *Geoderma*, 2022, 427: 116103.
- [41] 宋相中, 唐果, 张录达, 等. 近红外光谱分析中的变量选择算法研究进展[J]. *光谱学与光谱分析*, 2017, 37(4): 1048-1052. SONG Xiangzhong, TANG Guo, ZHANG Luda, et al. Research advance of variable selection algorithms in near infrared spectroscopy analysis[J]. *Spectroscopy and Spectral Analysis*, 2017, 37(4): 1048-1052. (in Chinese with English abstract)
- [42] LIU J, ZHANG D, YANG L, et al. Developing a generalized vis-NIR prediction model of soil moisture content using external parameter orthogonalization to reduce the effect of soil type[J]. *Geoderma*, 2022, 419: 115877.
- [43] WIJEWARDANE N K, GE Y, MORGAN C L S. Prediction of soil organic and inorganic carbon at different moisture contents with dry ground VNIR: A comparative study of different approaches[J]. *European Journal of Soil Science*, 2016, 67(5): 605-615.
- [44] JIANG Q, CHEN Y, GUO L, et al. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy[J]. *Remote Sensing*, 2016, 8(9): 755.
- [45] NAWAR S, ABDUL MUNNAF M, MOUAZEN A M. Machine learning based on-line prediction of soil organic carbon after removal of soil moisture effect[J]. *Remote Sensing*, 2020, 12(8): 1308.
- [46] ZHAO D, ARSHAD M, WANG J, et al. Soil exchangeable cations estimation using Vis-NIR spectroscopy in different depths: Effects of multiple calibration models and spiking[J]. *Computers and Electronics in Agriculture*, 2021, 182: 105990.

Influence of soil moisture on the inversion accuracy of near-infrared spectra of organic matter

ZHOU Peng¹, KONG Yinuo¹, HAO Shanshan¹, YIN Xiang¹, XIAO Xinqing³, JIN Chengqian^{2*}

(1. College of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China; 2. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China; 3. College of Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Soil organic matter (SOM) is one of the essential components of soil and moisture to interfere with the detection. The soil moisture content by mass ratio and spectral parameters are often utilized to classify the moisture levels. But previous studies on the prediction of SOM content have focused mostly on the surface soil without covering the profile of depth range. This study aims to explore the influence of soil moisture on the inversion accuracy of the SOM profile. This study analyzed previously obtained near-infrared spectra were collected from the samples of the soil core column in the surface layer to about 150 cm underground. Each core was divided into subsamples with a height of 10 or 20 cm. The samples were gradually moistened to 6 groups of levels of soil moisture. According to the index of moisture tension, the air-dry state was defined as 1.500, 0.330, 0.100, 0.033 MPa, and the saturated state in turn. The data was normalized and transformed to the absorbance. Each set of spectral data was processed by seven spectral preprocessing. Among them, the standard normal variate transformation (SNV) was achieved the best. Meanwhile, successive projection algorithms (SPA) and competitive adaptive reweighting-successive projection algorithms (CARS-SPA) were used to screen the characteristic wavelengths. The number of characteristic wavelengths was then reduced to 6-8 at each level of soil moisture. The inversion models of SOM were constructed to combine with SNV preprocessing using full spectrum and characteristic wavelengths. The results indicated that: 1) The model accuracies of the SPA-PLSR and CARS-SPA-PLSR models were better than that of the PLSR model at six levels of moisture. 2) The SNV preprocessing was also combined to determine the coefficient of determination in prediction (R^2_p) and root mean square error of the prediction set (RMSEP). Under the saturated state, R^2_p values of the SNV-SPA-PLSR and SNV-CARS-SPA-PLSR models were 0.664 and 0.651, respectively, while the RMSEP values were 1.095 and 1.131 g/kg, respectively. In the air-dry state, R^2_p values of the two models were 0.799 and 0.753, respectively, and RMSEP values were 0.759 and 0.848 g/kg, respectively, indicating the better prediction of the SNV-SPA-PLSR model. The SNV-CARS-SPA-PLSR model shared the higher accuracy of prediction when the moisture tension levels were 0.033, 0.100, 0.330, and 1.500 MPa. Specifically, R^2_p values increased from 0.699 to 0.846, whereas, the RMSEP values decreased from 1.013 to 0.620 g/kg. 3) However, it was difficult to guarantee the same soil moisture level at the same depth in different locations of the field. The SOM calibration model was applied to the characteristic wavelengths on different datasets. The SNV-CARS-SPA-PLSR model was selected at the moisture tension of 1.500 MPa. The best performance was achieved for the organic matter in the six groups of soil moisture levels and mixed samples. The models can be expected to estimate the SOM content of the profile at various moisture levels. The findings can also provide a strong reference to improve the applicability of near-infrared spectra inversion models for the organic matter content at different levels of soil moisture.

Keywords: soil moisture; organic matter; near-infrared spectroscopy; moisture tension; competitive adaptive reweighting-successive projection algorithm