

# 土壤水分去除算法的田间原位光谱反演棉田有机质

罗德芳<sup>1</sup>, 柳维扬<sup>1\*</sup>, 彭杰<sup>1</sup>, 冯春晖<sup>1</sup>, 纪文君<sup>2</sup>, 白子金<sup>1</sup>

1. 塔里木大学植物科学学院, 新疆阿拉尔 843300

2. 中国农业大学土地资源管理学院, 北京 100083

**摘要** 田间原位可见-近红外光谱(VIS-NIR)能够有效的提高土壤属性的检测效率,但由于原位土壤中水分因素的影响,土壤属性的预测精度很难达到预期。如何有效去除土壤中的水分对土壤其他属性光谱预测的影响,是利用田间原位光谱高精度预测土壤属性所面临的难题,也是土壤光谱技术由室内转向田间的突破口。该问题的有效解决,可减除土壤样品的采集与室内预处理等过程,实现土壤属性的田间原位光谱测定。以新疆南部地区阿拉尔垦区十二团棉田为研究区,采用网格采样法共采集了116个0~20 cm深度的表层土壤样品,剔除1个异常值样品,得到115个有用样品,利用SR-3500型便携式地物光谱仪采集了231个样点的田间原位光谱数据,土样经风干、研磨和过筛等处理后测定其室内光谱和有机质含量。利用Kennard-Stone算法将115个土样分为69个转换子集及46个预测集,采用外部参数正交化法(EPO)、光谱直接转换法(DS)及光谱间接转换法(PDS)三种去除水分算法结合原位光谱反射率( $R$ )、反射率一阶微分( $R'$ )、反射率对数( $\text{LOG}(R)$ )以及反射率倒数( $1/R$ )四种数学变换方式,运用随机森林(RF)模型进行不同组合模型的构建及精度评价。结果表明:(1)土壤有机质含量越高,土壤光谱反射率越低。土壤田间原位光谱反射率低于土壤室内光谱反射率;(2)室内光谱反射率与土壤有机质含量之间的相关性大于田间原位光谱,室内光谱经一阶微分变换后与土壤有机质含量之间的相关性显著提升。(3)土壤室内光谱反射率模型预测精度( $R^2=0.86$ ,  $\text{RPD}=2.08$ ,  $\text{RMSE}=1.55 \text{ g} \cdot \text{kg}^{-1}$ ,  $\text{MAPE}=0.14$ )高于田间原位光谱反射率模型( $R^2=0.71$ ,  $\text{RPD}=1.49$ ,  $\text{RMSE}=2.17 \text{ g} \cdot \text{kg}^{-1}$ ,  $\text{MAPE}=0.20$ )。在去除水分算法模型中,以EPO一阶微分模型去除水分效果最好,决定系数 $R^2$ 由0.71提高到0.83,RPD由1.49提高到2.04, RMSE由 $2.17 \text{ g} \cdot \text{kg}^{-1}$ 降低至 $1.58 \text{ g} \cdot \text{kg}^{-1}$ , MAPE由0.20降低至0.14。本研究实现了去除土壤水分因素的影响,提高了田间原位光谱预测土壤有机质的精度,为南疆棉田大尺度土壤有机质的预测及土壤肥力的评价提供了重要的参考。

**关键词** 土壤有机质;外参数正交化(EPO);光谱直接转换法(DS);光谱间接转换法(PDS);随机森林(RF)

**中图分类号:** S151.9 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)01-0222-07

## 引言

土壤是由多种成分组成的,如有机、无机矿物质、微生物和水。土壤有机质(soil organic matter, SOM)约为土壤有机碳的1.72倍,由不同分解阶段的动植物残余物组成,与土壤肥力和稳定性有关<sup>[1]</sup>。根据土壤有机质含量分布,绘制土壤肥力分布图,对精准施肥具有重要意义,因此快速准确的监测土壤有机质含量在智慧农业中尤为重要。但SOM测定仍以传统的室内分析法(重铬酸钾一外加热法)为主,该方法在测定土壤有机质时存在化学试剂消耗大、有毒废液量大、

分析费时等问题,不能满足智慧农业需求且会造成环境污染<sup>[2]</sup>。大量研究已经充分证明了利用可见光和近红外(VIS-NIR)反射光谱能够高效、低成本和高精度的预测土壤有机质、黏土矿物等土壤性质。史舟等<sup>[3]</sup>利用可见-近红外光谱分类方法结合偏最小二乘(PLSR),建立的光谱分类-局部模型有效的预测了不同地区土壤有机质含量,分类之后的土壤有机质模型精度显著提升,预测模型 $R^2$ 和RPD分别从0.70和1.82提高到0.90和3.16。

但室内光谱测定中存在采样及工序繁琐等缺点,用野外原位光谱替代室内光谱进行土壤属性测定可显著提高工作效率。Allory等<sup>[4]</sup>在原位和实验室条件下利用可见-近红外光

收稿日期:2020-11-29,修订日期:2021-02-23

基金项目:国家重点研发计划项目(2018YFE01070006)资助

作者简介:罗德芳,1994年生,塔里木大学植物科学学院硕士研究生

e-mail: zkyldf@163.com

\* 通讯作者 e-mail: lwyzy@163.com

谱分别预测了城市土壤的有机碳含量, 结果表明实验室条件下土壤有机碳的预测精度更高。土壤光谱反射率对外部环境条件相当敏感, 颗粒大小、温度和土壤湿度, 都影响土壤有机质预测精度。

为了克服土壤水分对可见-近红外光谱预测土壤有机质的影响, 国内外学者们提出了多种方法来提高田间原位光谱条件下土壤有机质的预测精度。洪永胜等<sup>[5]</sup>在室内进行了水分梯度实验, 对比干湿土壤光谱反射率并进行外部参数正交化法(EPO)校正, 在 EPO 算法校正后, PLSR 的模型预测偏差比由 1.16 提升到 1.76, 实现了不同含水量梯度条件下土壤有机质含量的有效估算。Ji 等<sup>[6]</sup>对浙江省水稻田土壤分别进行了野外原位光谱及室内光谱测定, 经光谱直接转换法(DS)校正后的野外原位 PLSR 模型精度  $R^2$  由 0.25 提高到 0.69、RPD 由 0.35 提高到 1.61, 结果表明 DS 能有效地去除水和环境因素对土壤光谱的影响, 提高土壤有机质的预测精度。虽已有众多关于去除土壤水分影响的有机质光谱预测的相关报道, 但目前针对南疆干旱区膜下滴灌棉田土壤有机质的原位光谱反演研究甚少, 南疆土壤普遍存在一定程度的盐渍化, 且以旱作为主, 土壤质地以沙土和沙壤土为主, 水分含量相对较低, 而目前的报道多集中于土壤质地较粘、土壤水分含量较高且无盐渍化危害的区域。因此, 现有去除水分的算法是否能对南疆这种特殊环境土壤的有机质原位光谱测定产生理想的效果, 尚未有明确结论。同时, 南疆棉花是机械化作业程度最高的一种作物, 解决水分影响土壤有机质原位光谱反演精度的问题, 有助于光谱仪与农业机械的协同作业, 实现智慧农机在线检测土壤有机质含量。

鉴于以上分析, 选择面积在南疆占主导地位的棉田为实验区, 旨在研究以随机森林(random forest, RF)建立的机器学习模型来预测通过外部参数正交化法(EPO)、光谱直接转换法(DS)及光谱间接转换法(PDS)三种去除水分算法后的棉田土壤有机质含量, 以期提高新疆南部地区大尺度范围内棉田土壤有机质的光谱预测精度, 为精准施肥提供数据支撑。

## 1 实验部分

### 1.1 研究区概况

试验区位于新疆维吾尔自治区阿拉尔垦区的一师十二团内。垦区属暖温带极端大陆性干旱荒漠气候, 垦区内太阳辐射年均  $133.7 \sim 146.3 \text{ kJ} \cdot \text{cm}^{-2}$ , 年均日照  $2\ 556.3 \sim 2\ 991.8 \text{ h}$ , 日照率为 5869%, 雨量稀少, 冬季少雪, 地表蒸发强烈, 年均降水量为  $40.1 \sim 82.5 \text{ mm}$ , 年均蒸发量  $1\ 876.6$

$\sim 2\ 558.9 \text{ mm}$ <sup>[7]</sup>。

试验区面积为  $400 \text{ m} \times 200 \text{ m}$ , 种植作物为棉花, 连作时间大于 20 年, 灌溉方式为膜下滴灌。采用等间距网格采样法进行土壤样品及田间原位光谱数据采集, 点间距为 20 m, 样点分布为 21 行  $\times$  11 列, 各样点均采集光谱数据, 共采集 231 个田间原位光谱数据, 每行样点间隔采集土壤样品, 共采集 116 个土壤样品。试验区中心地理坐标为北纬  $40^\circ 29' 22.194''$ , 东经  $81^\circ 19' 12.688''$ , 土壤类型为沙壤土, 试验区示意图如图 1 所示。

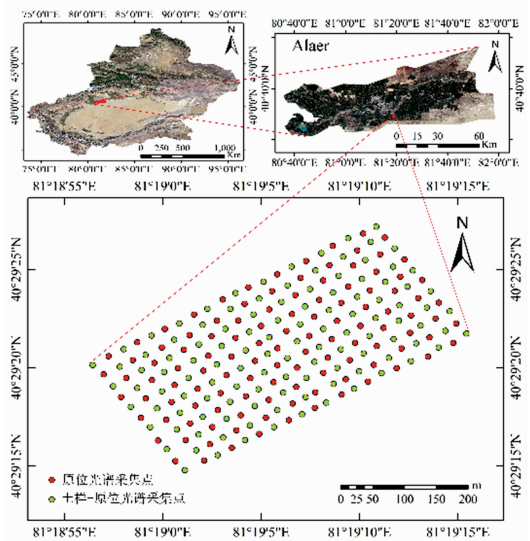


图 1 试验区示意图

Fig. 1 Schematic diagram of the study area

### 1.2 土壤采集与数据分析

土壤样品采集时间为 2018 年 11 月初, 均匀布局采样点位置, 同时记录样点坐标位置及地理环境。用土钻钻取 0~20 cm 表层土壤, 迅速装入自封袋中带回实验室, 剔除植物根系及砂砾等杂质, 取约 40 g 新鲜土样测定土壤质量含水量, 将剩余的土样风干研磨过 1 mm 筛后分别进行土壤有机质含量及室内光谱的测定, 共得到 116 个土壤有机质及室内光谱反射率数据, 数据分析后剔除 1 个异常值, 共得到 115 个有效土壤有机质及室内光谱数据。对土壤样本进行统计分析, 如表 1 所示。土壤有机质含量最大值为  $20.26 \text{ g} \cdot \text{kg}^{-1}$ , 最小值为  $4.23 \text{ g} \cdot \text{kg}^{-1}$ , 土壤水分含量最小值为 7.36%, 最大含水量 55.25%。根据变异系数等级划分原则, 变异系数小于 10% 为弱变异性; 变异系数在 10%~100% 为中等变异性; 变异系数大于 100% 为强变异性, 从表中可以看出, 土壤水分与有机质含量都属于中等变异。

表 1 土样统计分析

Table 1 Statistics of soil samples

项目	样本数	平均值	最大值	最小值	标准差	变异系数/%
土壤有机质/( $\text{g} \cdot \text{kg}^{-1}$ )	115	10.74	20.26	4.23	3.27	30.44
土壤水分/%	115	15.39	55.25	7.36	4.43	28.81

### 1.3 光谱测定及预处理

光谱测量仪器为美国 Spectral Evolution 制造的 SR-3500 型便携式地物光谱仪,其光谱测量范围在 350~2 500 nm 波段;光谱分辨率在 350~1 000 nm 为 3.5 nm,在 1 000~1 900 nm 为 10 nm,1 900~2 500 nm 为 7 nm;采样间隔为 1 nm。采集光谱样本使用设备配套的手枪式光纤手柄,光源为手柄内置光源;将农田土壤表面的植物组织、残膜及各类杂物清理后,探头紧贴平整地面测量,共采集 231 个样点的田间原位光谱,每个样点取 10 次测量结果的算术平均值为该样点原位反射光谱数据(Situ-spectra)。在进行室内风干土样的光谱测定时,将过 1 mm 筛的土样装入直径为 10 cm,深度为 2 cm 的器皿中,并用直尺刮平,用 SR-3500 便携式光谱仪测量土壤光谱,与田间测定方法一致,每 10 次测量取平均值为该土样的室内光谱数据(Dry-spectra),共得到了 116 个风干土样的室内光谱,剔除了 1 个异常值光谱,得到 115 条可用室内光谱。测量光谱数据时,每测量 1 个样点清理手柄镜面 1 次,每测量 10 个样点白板标定校准 1 次。本研究仅用与室内光谱数据相对应的 115 条田间原位光谱数据,剩余的 115 条土样的田间原位光谱数据用于后续的数字土壤制图研究,有关数字土壤制图研究,因此不做相应的处理和利用。

室内外光谱数据测定结束后,利用 DARWin SP 软件进行土壤光谱曲线校正,由于仪器受自身精度误差及大气影响会使光谱反射率有轻微偏移。在分析时需去除边缘(350~399 和 2 401~2 500 nm)噪声较大土壤样品的光谱反射率,保留 400~2 400 nm 波段的土壤原始光谱反射率进行分析处理。对田间原位及室内光谱分别做 Savitzky-Golay (SG)平滑处理,并进行反射率一阶微分( $R'$ )、反射率对数( $\text{LOG}(R)$ )以及反射率倒数( $1/R$ )数学形式变换。

### 1.4 土壤水分因素去除

#### 1.4.1 外部参数正交化(EPO)

外部参数正交化(external parameter orthogonalization, EPO)是将光谱投影到空间的正交面上进行预处理来降低光谱( $X$ )的维度<sup>[8]</sup>。一般  $X$  的列空间由两个子空间的和组成,其中只有一个包含对模型有用的信息,通过投影保留有用的子空间,EPO 算法主要用于改善现有模型校正的稳健性。通常  $n$  个样本  $m$  个波段的田间光谱  $\mathbf{X}_{(n \times m)}$  可以用矩阵的形式表示为

$$\mathbf{X} = \mathbf{XP} + \mathbf{XQ} + \mathbf{R} \quad (1)$$

式(1)中, $\mathbf{P}_{m \times m}$  为有用光谱信息(土壤属性)的投影矩阵; $\mathbf{Q}_{m \times m}$  为无用光谱信息(额外的环境因素)的投影矩阵; $\mathbf{R}_{n \times m}$  为残差矩阵。 $n$  为土壤样本数, $m$  为光谱波段数。

#### 1.4.2 光谱直接转换法(DS)

光谱直接转换法(direct standardization, DS)可通过测量田间光谱和室内光谱之间的变化规律,从田间光谱中去除水分,来表征土壤湿度或温度等环境影响<sup>[9]</sup>。室内光谱  $X$  是田间原位光谱  $X_w$  的线性组合,DS 转换存在着如式(2)的关系

$$X = X_w \mathbf{B} + e k_s^T \quad (2)$$

式(2)中, $X(m \times p)$  为转换样本的室内光谱数据; $X_w(m \times p)$  为转换样本的田间光谱数据; $m$  是转换样本的个数, $P$  为光谱波段数。 $\mathbf{B}_{p \times p}$  是转换矩阵,用以衡量室内土壤光谱与田间

土壤光谱之间的差异; $\mathbf{D}_{m \times p}$  是残差矩阵,用以校正田间不同于室内测量环境下所产生的基线偏移。

#### 1.4.3 光谱间接转换法(PDS)

光谱间接转换法(piecewise direct standardization, PDS)是为了改进 DS 方法所引入的算法<sup>[10]</sup>。DS 和 PDS 的主要区别在于转移矩阵的计算方式,PDS 是将相邻几个波长的原位光谱区域对应到室内光谱中的每个波长中,即在预测时,由相邻波长的光谱窗口  $O=2n+1$  大小的光谱段的数据重建室内光谱的第  $i$  个波长,PDS 算法存在如式(3)关系

$$X_i = X_w B_i \quad (3)$$

式(3)中, $X_i(m \times 1)$  是室内光谱中波长  $i$  所组成的列向量, $X_w(m \times O)$  是  $i$  相邻的  $2n+1$  大小的田间原位光谱组成的矩阵, $m$  是转换样本的数目, $O$  是窗口的宽度, $O=2n+1$ ;  $B_i(O \times 1)$  是对应  $i$  的转换系数,由 PLS 方法求出。以上三种去除水分算法中转换子集的选取均由 Kennard-Stone 算法得出。

### 1.5 模型构建与精度验证

随机森林(RF)是一种用于分类、回归和其他任务的集成学习方法,通过在训练时间内构建多个决策树并输出结果,RF 能够修正决策树的归纳偏好而产生的过拟合<sup>[11]</sup>。RF 对于很多数据集表现良好,精确度比较高,不易发生过拟合,在数据建模中具有一定的优势。故选用 RF 方法进行构建模型。

模型的稳定性与预测的精度评价指标主要包括决定系数( $R^2$ )、相对分析误差(RPD)、均方根误差(RMSE)以及平均绝对百分误差(MAPE)<sup>[12]</sup>。 $R^2$  是表征回归方程在多大程度上解释了因变量的变化以及方程对观测值的拟合程度,根据 RPD 的分类标准,RPD<1.5 时,表明模型的预测能力很差;1.5≤RPD<2 时,表明模型预测能力良好,尚可用于进行大致估算;RPD≥2.0,代表模型质量很好,可以用于土壤性质的定量预测。RMSE 可以度量预测值和实测值之间的差异,并通过比较不同模型的预测误差来量化预测的准确性。因此,RMSE 越小,模型精度越好。MAPE 的范围在  $[0, +\infty)$ ,MAPE 为 0% 表示完美模型,MAPE 大于 100% 则表示劣质模型。相比 RMSE,MAPE 相当于把每个点的误差进行了归一化,降低了个别离群点带来的绝对误差的影响。

## 2 结果与讨论

### 2.1 不同有机质含量土样的室内与田间原位光谱特征分析

将土壤样本按有机质含量分为 <5, 5~10, 10~15 和 >15 g·kg<sup>-1</sup> 四个等级,计算每个等级的所有土样光谱曲线的各波段反射率平均值并做图 2。不同土壤有机质含量的室内和田间原位的土壤光谱曲线形状整体趋势一致,土壤有机质含量越大,光谱反射率越小,且反射率光谱曲线的波谷向短波方向偏移。室内光谱反射率与田间原位光谱反射率的土壤类型及测定方式一致,唯一不同的是水分因素的影响。由图 2 可知,田间原位光谱反射率整体小于室内光谱反射率,但在 1 400 和 1 900 nm 波段处,田间原位光谱反射率下降的趋势更为明显,波谷更深。随着土壤含水率的增加,土壤光谱反

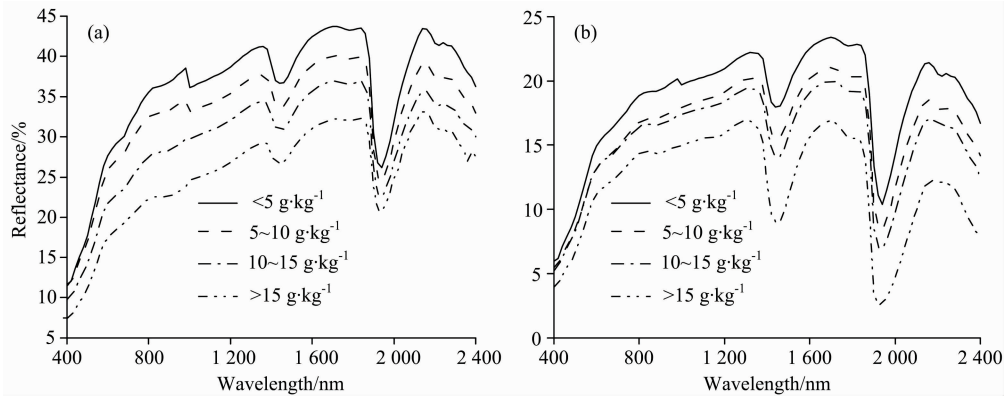


图 2 不同有机质平均含量的土壤室内光谱曲线 (a) 及田间原位光谱曲线 (b)

Fig. 2 Dry-spectral (a) and Situ-spectral (b) of soil with different average organic matter contents

射率下降, 说明水分因素主要在近红外波段影响土壤光谱反射率。

2.2 不同处理形式下模型精度对比

采用 Kennard-Stone 算法, 经过多次不同样本转换数的试验分析, 在 115 个土样中选取具有代表性且均匀分布的 69 个样本作为转换集样本, 剩余的 46 个样本作为验证集样本。对原位及室内光谱反射率分别做一阶微分、对数及倒数形式的处理, 并利用 RF 模型进行建模, 不同数学形式的土壤有机质模型如表 2 所示。其次, 筛选模型精度较好的田间原位光谱数据, 进行 EPO, DS 和 PDS 三种去除水分算法处理并建立 RF 模型, 表 3 为不同数学变换形式结合去除水分算法的土壤有机质预测精度表。

表 2 不同数学形式的土壤有机质模型精度

Table 2 Accuracy of soil organic matter models in different mathematical forms

模型	$R^2$	RMSE/ ( $g \cdot kg^{-1}$ )	RPD	MAPE
Dry spectra	0.87	1.59	2.38	0.15
Situ spectra	0.71	2.13	1.78	0.20
Dry- $R'$ -spectra	0.90	1.54	2.46	0.14
Situ- $R'$ -spectra	0.73	2.08	1.82	0.16
Dry-1/ $R$ -spectra	0.81	1.85	2.04	0.16
Situ-1/ $R$ -spectra	0.70	2.18	1.73	0.20
Dry-LOG( $R$ )-spectra	0.88	1.65	2.29	0.15
Situ-LOG( $R$ )-spectra	0.73	1.99	1.90	0.17

由表 2 可得, 经光谱反射率一阶微分和对数形式变换之后, 室内光谱和田间原位光谱模型精度均有所增加, 室内光谱模型以反射率一阶微分变换后的土壤有机质模型预测精度最优,  $R^2$  和 RPD 最大为 0.90 和 2.46, RMSE 和 MAPE 最小为  $1.54 g \cdot kg^{-1}$  和 0.14, 田间原位光谱模型以反射率对数变换后的土壤有机质模型精度最优,  $R^2$  和 RPD 最大为 0.73 和 1.90, RMSE 和 MAPE 最小为  $1.99 g \cdot kg^{-1}$  和 0.17。经反射率倒数变换之后, 室内光谱和田间原位光谱模型精度均降低, 在室内光谱有机质模型中, 模型精度  $R^2$  和 RPD 分别降低了 0.06 和 0.34, RMSE 和 MAPE 分别增加了  $0.26 g \cdot$

$kg^{-1}$  和 0.01, 田间原位光谱模型精度  $R^2$  和 RPD 分别降低了 0.01 和 0.05, RMSE 增加了  $0.05 g \cdot kg^{-1}$ , MAPE 没有变化。室内光谱与田间原位光谱模型精度相比较来说, 室内光谱的模型精度高于原位光谱模型精度, 原始室内光谱反射率有机质模型精度  $R^2$  可达到 0.87, RPD 为 2.38 大于 2.0, 能够精准的预测出土壤有机质含量。田间原位光谱模型精度  $R^2$  为 0.71, RPD 仅为 1.78, 只能大致估测出土壤有机质含量。

表 3 不同去除水分算法的土壤有机质预测模型精度

Table 3 Accuracy of soil organic matter prediction model under different treatment forms

模型	$R^2$	RPD	RMSE/ ( $g \cdot kg^{-1}$ )	MAPE
EPO-situ spectra	0.80	1.73	2.18	0.16
EPO- $R'$ -situ spectra	0.83	2.35	1.61	0.15
EPO-LOG( $R$ )-situ spectra	0.82	2.20	1.72	0.16
DS-situ spectra	0.81	2.09	1.75	0.17
RF DS- $R'$ -situ spectra	0.83	2.35	1.63	0.16
RF DS-LOG( $R$ )-situ spectra	0.81	2.17	1.74	0.17
PDS-situ spectra	0.76	2.08	1.84	0.19
PDS- $R'$ -situ spectra	0.78	1.90	1.99	0.19
PDS-LOG( $R$ )-situ spectra	0.80	2.06	1.84	0.17

表 3 为光谱反射率、反射率一阶微分、反射率对数与 EPO, DS 和 PDS 三种去除水分算法相结合建立的有机质模型精度评价表。经过去除水分影响后, 土壤有机质光谱模型精度均有所提高, PDS 以反射率对数变换后的去除水分效果最好, 模型精度  $R^2$  为 0.80, RPD 为 2.06, RMSE 为  $1.84 g \cdot kg^{-1}$ , MAPE 为 0.17, 但模型精度仅能大致估测出土壤有机质含量。EPO 和 DS 以反射率一阶微分变换后的模型去除水分效果最好, 模型精度较原位光谱模型精度  $R^2$  均提高了 0.12, RPD 均提高了 0.57, RMSE 分别降低了 0.52 和  $0.50 g \cdot kg^{-1}$ , MAPE 分别降低了 0.05 和 0.04, 均能精确地预测出土壤有机质含量。EPO 和 DS 一阶微分模型的  $R^2$  和 RPD 相同, 但 DS 模型的 RMSE 和 MAPE 均大于 EPO 模型。因此, EPO 模型能够更有效的去除水分因素的影响, 以

EPO 一阶微分去除水分模型进行土壤有机质预测, 精度最高, 预测效果最好。

### 2.3 最优反演模型的筛选

通过对三种反射率变换形式、三种去除水分算法结合并进行 RF 建模, 共得到 11 个土壤有机质预测模型, 分别对 46 个验证集样本进行模型精度验证。其中, 以土壤室内光谱预测的结果最好, 田间原位光谱预测的结果最差, 在去除水分算法中, 以反射率一阶微分进行的 EPO 和 DS 算法模型预测精度相对较好。利用室内光谱、田间原位光谱、反射率一阶微分进行的 EPO 和 DS 算法模型对土壤有机质进行预测的散点图如图 3 所示。

由图 3 可得, 田间原位光谱预测土壤有机质含量模型精度最差,  $R^2$  为 0.71, RPD 为 1.49, RMSE 为  $2.17 \text{ g} \cdot \text{kg}^{-1}$ ,

MAPE 为 0.20, 不能预测出土壤有机质的含量。土壤室内光谱预测模型精度最高,  $R^2$  为 0.86, RPD 为 2.08, RMSE 为  $1.55 \text{ g} \cdot \text{kg}^{-1}$ , MAPE 为 0.14, 能够精准的预测出土壤有机质的含量。经 EPO 一阶微分和 DS 一阶微分去除水分影响后的模型预测精度较田间原位光谱模型精度均有所提高,  $R^2$  和 RPD 分别提高了 0.12, 0.55 和 0.11, 0.54, RMSE 和 MAPE 分别降低了  $0.59 \text{ g} \cdot \text{kg}^{-1}$ , 0.06 和  $0.58 \text{ g} \cdot \text{kg}^{-1}$ , 0.06, EPO 一阶微分和 DS 一阶微分去除水分模型均能够精确的预测出土壤有机质含量, 以 EPO 一阶微分去除水分影响后建立的模型能够更精确地预测反演出土壤有机质含量, 本研究为南疆棉田土壤大尺度预测土壤肥力状况提供了新的思路。

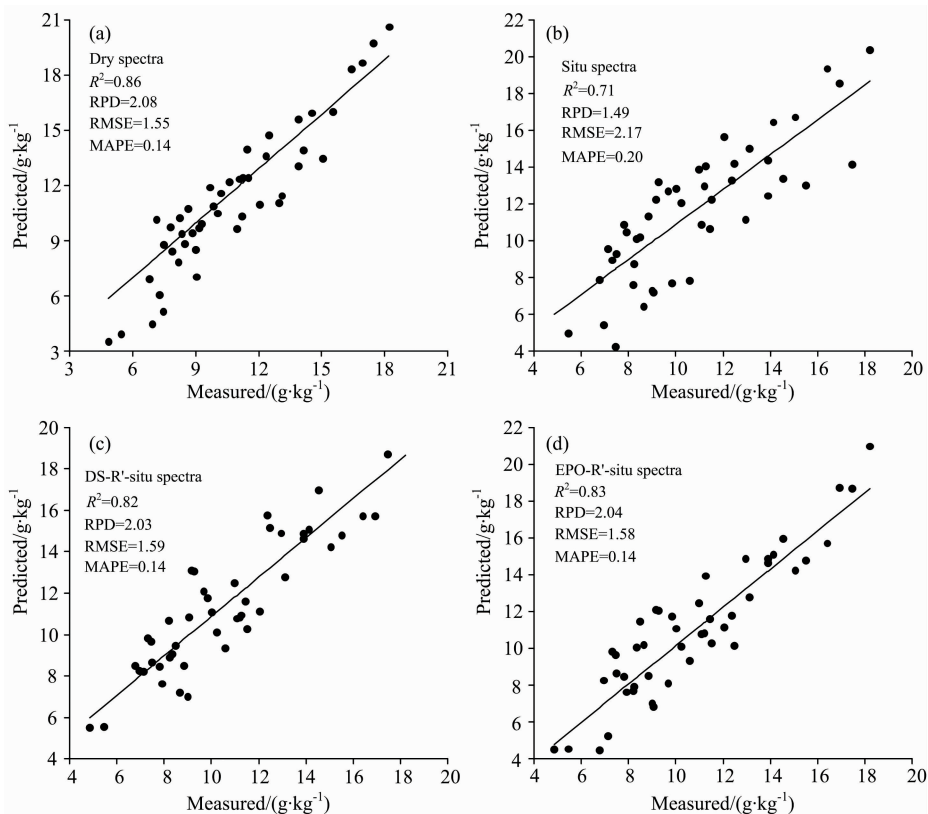


图 3 不同模型对土壤有机质含量预测散点图

Fig. 3 Scatter plots of soil organic matter content predicted by different models

在进行田间测量土壤光谱时, 由于杂散光等自然条件的影响, 会使测量的土壤光谱曲线有一定的误差。本研究使用的是内置光源的手持式土壤光谱仪, 能够避免杂散光等影响, 保证在田间进行土壤光谱测定时的环境因素不受干扰, 但土壤自身的水分含量对土壤光谱有很大的影响。吴龙国等<sup>[13]</sup>在研究土壤水分高光谱的无损检测时发现土壤光谱反射率随着土壤含水率的增加而减小, 当超过田间持水率时, 光谱曲线的反射率会随着土壤含水率的增加而增大。因此田间原位土壤中水分含量较大时, 在进行田间光谱直接预测土壤有机质时会严重影响预测精度。前人研究提高土壤有机质预测精度的方法主要是进行光谱反射率预处理算法或者模型

方法的提高, 虽然能够提高模型的预测精度, 但却不能完全去除水分因素的影响, 在与室内光谱预测土壤有机质相比, 田间原位光谱的预测结果还是较低, 不能直接用来预测土壤有机质含量<sup>[14]</sup>。

对田间原位光谱及室内光谱原始光谱反射率作反射率一阶微分与反射率对数数学形式变换, 并与土壤有机质含量作相关性分析, 结果表明室内光谱反射率与土壤有机质的相关性大于田间原位光谱反射率。与原始光谱反射率相比, 经一阶微分及对数变换后, 光谱反射率的相关性有了很大的提高, 尤其是在特征性波段 1 400 和 1 900 nm 处, 光谱反射率的相关性明显增加。因此, 经一阶微分及对数变换后可在一

定程度上消除土壤水分的影响。本研究在选用 EPO, DS 和 PDS 三种去除水分的算法来校正土壤田间原位光谱中的水分影响因素的基础上, 还结合了一阶微分及对数数学形式预处理后的光谱反射率来进一步提高土壤有机质含量的预测精度。在模型选择中, 本研究选择建模精度较好的 RF 模型。结果表明, 基于 EPO 一阶微分去除水分影响后的 RF 模型能够更精确地预测反演出土壤有机质含量, 在土壤肥力鉴定中能有一定的借鉴意义。

### 3 结 论

田间原位光谱预测土壤有机质时水分因素会对有机质预测精度有一定的影响。本研究采用 EPO, DS 和 PDS 三种去除水分算法结合反射率一阶微分、倒数及对数三种数学变换方式, 利用 RF 模型进行模型预测, 提高了原位土壤有机质含量的预测精度, 得出以下结论:

(1) 土壤有机质含量越高, 土壤光谱反射率越低。土壤原位光谱反射率低于土壤室内光谱反射率, 且在 1 400 以及 1 900 nm 附近, 土壤原位光谱反射率出现更深的波谷。

(2) 田间原位光谱预测土壤有机质含量模型精度  $R^2$  为 0.71, RPD 为 1.49, RMSE 为  $2.17 \text{ g} \cdot \text{kg}^{-1}$ , MAPE 为 0.20, 不能预测出土壤有机质的含量。土壤室内光谱预测模型精度  $R^2$  为 0.86, RPD 为 2.08, RMSE 为  $1.55 \text{ g} \cdot \text{kg}^{-1}$ , MAPE 为 0.14, 能够精确地预测出土壤有机质的含量。

(3) 经 EPO 一阶微分去除水分 RF 模型精度最高, 决定系数  $R^2$  为 0.83, RPD 为 2.04, RMSE 为  $1.58 \text{ g} \cdot \text{kg}^{-1}$ , MAPE 为 0.14, 能够有效的去除土壤中水分因素的影响, 精确的预测出土壤有机质含量。本研究为南疆地区大尺度范围内进行田间原位光谱预测土壤有机质含量提供了参考, 也为快速监测土壤肥力, 精准灌溉及合理施肥提供了一定的研究意义。

### References

- [ 1 ] ZHANG Wei-li, Kolbe H, ZHANG Ren-lian, et al(张维理, Kolbe H, 张认连, 等). Chinese Agricultural Science (中国农业科学), 2020, 53(2): 332.
- [ 2 ] Nawar S, Buddenbaum H, Hill J, et al. Soil and Tillage Research, 2016, 155: 510.
- [ 3 ] SHI Zhou, WANG Qian-long, PENG Jie, et al(史舟, 王乾龙, 彭杰, 等). Chinese Science: Earth Science(中国科学: 地球科学), 2014, 44(5): 978.
- [ 4 ] Allory V, Cambou A, Moulin P, et al. Science of the Total Environment, 2019, 686: 764.
- [ 5 ] HONG Yong-sheng, YU Lei, ZHU Ya-xing, et al(洪永胜, 于雷, 朱亚星, 等). Journal of Soil(土壤学报), 2017, 54(5): 1068.
- [ 6 ] Ji W, Viscarra R R A, Shi Z. European Journal of Soil Science, 2015, 66(3): 555.
- [ 7 ] Peng J, Biswas A, Jiang Q S, et al. Geoderma, 2019, 337: 1309.
- [ 8 ] Liu Y, Deng C, Lu Y Y, et al. Geoderma, 2020, 376: 114568.
- [ 9 ] Wijewardane N K, Ge Y F, Morgan C L S. Geoderma, 2016, 267: 92.
- [10] Ji W, Viscarra R R A, Shi Z. European Journal of Soil Science, 2015, 66(4): 670.
- [11] Zhang Y, Sui B, Shen H, et al. Computers and Electronics in Agriculture, 2019, 160: 23.
- [12] Zhang X, Sun WC, Cen Y, et al. Science of the Total Environment, 2019, 650: 321.
- [13] WU Long-guo, WANG Song-lei, HE Jian-guo, et al(吴龙国, 王松磊, 何建国). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(8): 2563.
- [14] CHEN Hong-yan, ZHAO Geng-xing, LI Yu-huan, et al(陈红艳, 赵庚星, 李玉环, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(12): 119.

# Field in Situ Spectral Inversion of Cotton Organic Matter Based on Soil Water Removal Algorithm

LUO De-fang<sup>1</sup>, LIU Wei-yang<sup>1\*</sup>, PENG Jie<sup>1</sup>, FENG Chun-hui<sup>1</sup>, JI Wen-jun<sup>2</sup>, BAI Zi-jin<sup>1</sup>

1. College of Plant Sciences, Tarim University, Alar 843300, China

2. College of Land Resources Management, China Agricultural University, Beijing 100083, China

**Abstract** Field in-situ visible-near infrared spectroscopy (VIS-NIR) can effectively improve the detection efficiency of soil properties, but due to the influence of in-situ soil moisture factors, the prediction accuracy of soil property models is difficult to reach expectations. How to effectively remove the influence of soil moisture on the spectrum prediction of other soil properties is a difficult problem for using field in-situ spectroscopy to predict soil properties with high precision. It is also a breakthrough for soil spectroscopy technology to shift from indoor to field. The effective solution to this problem can eliminate the process of soil sample collection and indoor pretreatment, and achieve field in-situ spectroscopy of soil properties. In this study, a total of 115 surface soil samples of 0~20 cm were collected by using the grid sampling method in the cotton field of the 12th group in the Alar reclamation area in southern Xinjiang. And use the SR-3500 portable ground object spectrometer to collect 231 sample points of field in-situ spectral data. The soil samples are air-dried, ground and sieved, and then their indoor spectrum and organic matter content are measured. 115 soil samples were divided into two subsets, i. e. the conversion subset (69 samples) and the prediction set (46 samples), with the Kennard-Stone algorithm. The external parameter orthogonalization (EPO), spectral direct standardization (DS) and piece-wise spectral direct standardization (PDS) were performed on spectra and the spectra with other three pretreatments including first-order differential of reflectance ( $R'$ ), logarithm of reflectance ( $\text{LOG}(R)$ ) and inverse reflectance ( $1/R$ ), respectively. The random forest (RF) model is used to construct different combination models and evaluate the accuracy. The results showed: (1) The higher the soil organic matter content, the lower the soil spectral reflectance. The in-situ spectral reflectance of soil in the field is lower than the indoor spectral reflectance of soil; (2) The correlation between indoor spectral reflectance and soil organic matter content is greater than that of field in-situ spectra. The correlation between indoor spectra and soil organic matter content is significantly improved after first-order differential transformation. (3) The prediction accuracy of the soil indoor spectral reflectance model ( $R^2=0.86$ ,  $\text{RPD}=2.08$ ,  $\text{RMSE}=1.55 \text{ g} \cdot \text{kg}^{-1}$ ,  $\text{MAPE}=0.14$ ) is higher than the field in-situ spectral reflectance model ( $R^2=0.71$ ,  $\text{RPD}=1.49$ ,  $\text{RMSE}=2.17 \text{ g} \cdot \text{kg}^{-1}$ ,  $\text{MAPE}=0.20$ ). Among the moisture removal algorithm models, the EPO first-order differential model has the best moisture removal effect, with the coefficient of determination  $R^2$  increased from 0.71 to 0.83,  $\text{RPD}$  increased from 1.49 to 2.04,  $\text{RMSE}$  decreased from 2.17 to 1.58  $\text{g} \cdot \text{kg}^{-1}$ , and  $\text{MAPE}$  decreased from 0.20 to 0.14. This study achieved the removal of the influence of soil moisture factors, improved the accuracy of field in-situ spectroscopy prediction of soil organic matter, and provided important reference value for the prediction of large-scale soil organic matter and the evaluation of soil fertility in southern Xinjiang cotton fields.

**Keywords** Soil organic matter; External parameter orthogonalization (EPO); Direct standardization (DS); Segmented direct standardization (PDS); Random forest (RF)

(Received Nov. 29, 2020; accepted Feb. 23, 2021)

\* Corresponding author