

DOI: 10.11766/trxb201612040396

基于EPO算法去除水分影响的土壤有机质高光谱估算*

洪永胜^{1, 2} 于雷^{1, 2†} 朱亚星^{1, 2} 吴红霞^{1, 2} 聂艳^{1, 2} 周勇^{1, 2}
QI Feng³ 夏天^{1, 2}

(1 华中师范大学地理过程分析与模拟湖北省重点实验室, 武汉 430079)

(2 华中师范大学城市与环境科学学院, 武汉 430079)

(3 School of Environmental and Sustainability Sciences, Kean University, NJ 07083, USA)

摘要 野外进行土壤有机质的光谱快速预测时需考虑土壤含水量的影响。在室内设计人工加湿实验分别获取9个土壤含水量梯度(0~32%, 间隔4%)的土壤光谱数据, 分析土壤含水量变化对光谱的影响, 再利用外部参数正交化法(external parameter orthogonalization, EPO)进行湿土光谱校正, 并结合偏最小二乘回归和支持向量机回归分别建立土壤有机质预测模型。结果表明, 土壤光谱反射率随着土壤含水量的增加呈非线性降低趋势, 偏最小二乘回归模型的预测偏差比为1.16, 模型不可用; 经EPO算法校正后, 各土壤含水量梯度之间的光谱差异性降低, 能够实现土壤有机质在不同土壤含水量梯度的有效估算, 偏最小二乘回归和支持向量机回归模型的预测偏差比分别提高至1.76和2.15。研究结果可为田间快速预测土壤有机质提供必要参考。

关键词 土壤光谱; 有机质; 含水量; 外部参数正交化法; 支持向量机回归; 江汉平原

中图分类号 S127; TP79 **文献标识码** A

土壤有机质是土壤理化特性的一个重要参数, 是指导农业科学施肥的重要依据^[1]。快速、准确、低成本地获取大面积土壤有机质信息, 是实现农业现代化管理的重要基础。近年来, 以可见近红外光谱为代表的土壤近地传感技术凭借独特的优点(快速、方便、环保、无损等)为获取土壤有机质信息提供了新的途径, 逐渐成为动态监测土壤有机质信息的理想手段^[2-3]。

与室内土壤光谱测量相比, 野外原位监测可以有效提高数据的分析效率, 节约土样采集与制备(风干、研磨、过筛等)的工作时间。但是, 野外田间光谱观测极易受外部环境因素的干扰(如土壤湿度、环境温度、粗糙表面等), 其数据的模

型精度通常要低于室内相对稳定环境下的模拟结果^[4]。特别是土壤含水量变化对土壤有机质的光谱表征及定量预测^[5]。因此, 克服土壤含水量对土壤光谱的影响至关重要。

外部参数正交化法(external parameter orthogonalization, EPO)由Roger等^[6]首先提出并应用于校正温度变量对光谱预测水果中糖分含量的影响。Minasny等^[7]将该方法借鉴至土壤, 研究结果显示EPO-偏最小二乘回归方法可以实现湿土状态下的土壤有机碳的精确估算。Ge等^[8]、Ji等^[9]、Ackerson等^[10]、Liu等^[11]进一步考察了EPO法在光谱估算土壤各种理化属性研究中的去水效果。目前, 利用EPO算法校正后的转换光谱的建模方法多

* 国家自然科学基金项目(41401232)、中央高校基本科研业务费专项资金项目(CCN15A05006)、湖北省自然科学基金面上项目(2016CFB558)、华中师范大学研究生教育创新资助项目(2016CXZZ15)共同资助 Supported by the National Natural Science Foundation of China (No. 41401232), the Fundamental Research Funds for the Central Universities (No. CCNU15A05006), the General Project of Natural Science Foundation of Hubei Province (No. 2016CFB558), the Education Innovation Projects for Graduates of Central China Normal University (No. 2016CXZZ15)

† 通讯作者 Corresponding author, E-mail: yulei@mail.ccnu.edu.cn

作者简介: 洪永胜(1990—), 男, 安徽六安人, 硕士研究生, 主要从事土壤高光谱研究。E-mail: ccnuhys@sina.com

收稿日期: 2016-12-04; 收到修改稿日期: 2017-04-24; 优先数字出版日期(www.cnki.net): 2017-05-25

集中于线性方法——偏最小二乘回归，与非线性方法的结合并不多见。此外，EPO算法中的关键参数维度 g 对光谱的校正转换有重要影响，定量分析其对光谱的转换效果十分关键。

鉴于此，本研究在室内设计人工加湿土壤样本实验来模拟土壤含水量变化过程，1) 定性考察土壤含水量变化对土壤光谱的影响；2) 定量评估EPO算法结合线性和非线性模型对土壤有机质的预测精度。

1 材料与方 法

1.1 土样采集制备与样本集划分

2010—2015年，以江汉平原的公安县和潜江市为采样区域，采集耕层（0~20 cm深度）土壤样品。清除土样中小砾石、表层枯枝落叶、动物残体等外来侵入物后，将土样带回实验室进行风干、研磨、过2 mm筛等处理，并建立样品库（234个样本）。样品充分混均后分为两份，分别用于土壤农化分析测试和光谱数据采集。其中，土壤有机质含量的测定采用重铬酸钾容量法-外加热法^[12]。

从样品库中随机选取122份组成集合S0；然后选取取自潜江市竹根滩镇的潮土样本95份，其土壤母质为石灰质近代河流冲积物，土壤表层质地以砂质壤土为主；按照土壤有机质含量进行升序排列，间隔2或3个样本挑选出60份土样组成S1，剩余35份组成S2。

1.2 土壤含水率实验设计

本文采用土壤质量含水率作为土壤含水量，共设计9个土壤含水量梯度（0、4%、8%、12%、16%、20%、24%、28%、32%）。每份土样称取150 g烘干土（土壤含水量 ≈ 0 ），放置于直径7.5 cm、深度5 cm规格的圆柱形黑盒中进行光谱测量，得到干土光谱数据（干土样本）；而后以每个土样增加4%的土壤含水量进行湿水处理（用喷雾器均匀喷洒约6 g水），并迅速密封以防水分蒸发；静置8h使水分被土壤充分吸收后再测量湿土光谱数据，并称重以计算实际土壤含水量；重复加水步骤直至完成所有梯度实验。

1.3 光谱数据采集

采用美国Analytical Spectral Devices公司生产的ASD FieldSpec3地物光谱仪测定土壤可见近红外

光谱反射率。仪器波谱范围涵盖350~2 500 nm，采样间隔为1.4 nm（350~1 000 nm）和2 nm（1 000~2 500 nm）。配备50 W卤素灯和5°视场角光纤探头。光源到土壤表面距离（ L ）为50 cm，光源入射角度（ A ）为30°，探头到土壤表面距离（ H ）为15 cm^[13]。光谱测量在暗室内进行。测试之前，先对仪器进行标准白板校正，每个土样测量4个方向（转动3次，每次90°），每个方向上保存5条光谱曲线，共20条，作算术平均后得到土样实际的反射光谱数据。

S0共122条光谱数据（122个样本），S1共540条光谱数据（60个样本，9个土壤含水量梯度），S2共315条光谱数据（35个样本，9个土壤含水量梯度）。

选用2阶11点Savitzky-Golay（SG）滤波平滑法对3个样本集的原始土壤光谱反射率进行平滑去噪^[14]。每份土样保留400~2 400 nm的光谱数据用于处理分析，对全光谱数据的每10个相邻波段进行平均。

1.4 土壤含水量对土壤光谱反射率影响的分析方法

采用4种方法来分析土壤含水量对土壤光谱反射率的影响：差值光谱、包络线去除、二维相关光谱、主成分分析。二维相关光谱：能反映分子间相互作用，可以提供比一维光谱更多的结构信息，能够清晰显示出重叠峰^[15]，本研究采用二维同步相关光谱。主成分分析：通过变换保留方差大、包含信息量多的组分，舍弃信息量少的组分，其计算采用非线性迭代偏最小二乘算法，计算步骤简单，运算速度快^[16]。

1.5 EPO算法

通常，受土壤含水量影响的湿土光谱可以用矩阵表示为：

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

式中， P 为 X （湿土光谱矩阵）中对土壤有机质有用信息的投影矩阵； Q 为无用信息（受土壤含水量的影响）的投影矩阵； R 为残差矩阵。EPO算法的思想就是从 X 中提取有用光谱信息的投影矩阵 P ，采用S1开发投影矩阵 P ，具体算法描述如下：

（1）干土样本（ $n \times m$ ），计算干土样本平均光谱（ $1 \times m$ ）；（2）湿土样本（ $N \times m$ ），计算不同水分梯度样本平均光谱（ $h \times m$ ）；（3）计算干土样本与湿土样本的差值 D （ $h \times m$ ）；（4）对

$D^T D$ 进行主成分分析,得到矩阵 $V(m \times m)$;
(5)定义EPO的维度 g ,计算 V 矩阵的一个子集 V_s
($m \times g$);(6)计算 $Q=V_s V_s^T$;(7)计算投影
矩阵 P , $P=I-Q$, I 为一个单位矩阵。式中, $n=60$
(S1包含样本的总量), $h=8$ (土壤含水量梯度
个数), $N=n \times h=480$ (湿土样本总数), $m=201$
(波长变量个数), g (EPO维度)。

1.6 模型构建与精度评价

偏最小二乘回归是一种多变量统计分析方法,
是利用光谱技术估算土壤属性参数的常见方法之
一,属于线性建模技术。本研究中偏最小二乘回归
模型采用Leave-one-out交叉验证法来确定回归模
型中最佳因子数,防止过拟合的发生^[17]。

支持向量机回归在建模过程中将原来的样本空
间中低维非线性的输入映射到高维线性的输出,并
在高维特征空间进行线性回归,属于非线性建模技
术。研究采用 ϵ -支持向量机回归算法和RBF径向基
核函数。通常, ϵ -支持向量机回归算法有两个参数
需要优化确定:惩罚参数 C 和RBF核参数 γ ,本研
究中为与EPO维度 g 一起进行参数寻优,将RBF核
参数 γ 进行固定^[18]。

采用决定系数 R^2 、均方根误差RMSE和预测偏
差比(ratio of prediction to deviation, RPD)值作
为模型精度的评价指标。其中, R^2 越接近于1,说
明模型的稳定性越好、拟合程度越高;RMSE越接
近0,说明模型估算预测能力越好。当 $RPD < 1.4$
时,模型不可用; $1.4 \leq RPD < 2$ 时,模型效果一
般,只能进行粗略估计; $RPD \geq 2$ 时,模型具有较
好的预测能力^[19]。

1.7 模型建立与优选过程

第一步,利用S0建立偏最小二乘回归模型(S0
的122个样本的光谱数据为自变量,土壤有机质数
据为因变量),基于偏最小二乘回归模型将S2的35
个样本的315条光谱数据作为输入变量预测土壤有
机质。

第二步,利用S1构建EPO转换矩阵 P 。

第三步,将 P 矩阵应用于处理S0、S2中各个样
本的光谱数据,构建S0的转换样本集 $S0^*$ 和S2的
转换样本集 $S2^*$ 。

第四步,利用 $S0^*$ 建立EPO-偏最小二乘回归
模型和EPO-支持向量机回归模型($S0^*$ 的122个样
本的光谱数据为自变量,土壤有机质数据为因变
量),基于两个模型将 $S2^*$ 的35个样本的315条光
谱数据作为输入变量分别预测土壤有机质。

第五步,对比偏最小二乘回归模型和EPO-偏
最小二乘回归模型的预测精度,以检验土壤含水
量的消除效果;对比EPO-偏最小二乘回归模型和
EPO-支持向量机回归模型的预测精度,优选出最
佳的估算模型。

2 结果

2.1 土壤有机质的描述性统计特征

对3个样本集(S0, S1和S2)的土壤有机质统
计特征进行分析描述(表1),三者样本集之间的
土壤样本不重复。与S1、S2相比,S0具有较大的
标准差和变异系数,适合反演模型的构建,所建模
型将具有很好的代表性。

表1 3个样本集(S0, S1和S2)的土壤有机质描述统计特征

Table 1 Descriptive statistics of the soil organic matter in the three subsets (S0, S1 and S2)

样本集 Datasets	样本数 No. of samples	光谱数量 No. of scans	最小值 Minimum (g kg ⁻¹)	最大值 Maximum (g kg ⁻¹)	平均值 Mean (g kg ⁻¹)	标准差 Standard deviation (g kg ⁻¹)	变异系数 Coefficient of variation (%)
S0	122	122	9.17	27.92	17.52	3.99	22.76
S1	60	540	9.56	26.55	17.52	3.06	17.46
S2	35	315	9.86	21.85	17.19	3.01	17.54

2.2 土壤含水量变化对土壤光谱反射率的影响

计算S1和S2(共95个样本,下同)在9个土壤
含水量梯度的平均土壤光谱反射率(图1a),在整

个光谱区域内(400~2400 nm),随着土壤含水
量的增加,光谱反射率整体呈降低趋势,这主要是
因为土壤水分造成光线在土壤颗粒间的折射增多,

导致光谱反射率降低；当土壤含水量 < 16% 时，土壤含水量变化对土壤光谱反射率变化的影响相对较大，当土壤含水量 ≥ 16% 时，土壤含水量变化对土壤光谱反射率变化的影响降低，土壤光谱反射率降低速率趋缓，尤其是 400 ~ 800 nm 范围内的光谱反射率基本重叠^[20-21]。

计算S1和S2不同土壤含水量梯度湿土样本与干土样本平均光谱反射率的差值（图1b），1 940 nm 附近的光谱吸收谷是全波段区域最大的吸收谷，要大于其他波段；土壤含水量对土壤近红外波段光谱反射率的影响要明显强于可见光波段。

计算S1和S2不同土壤含水量梯度湿土样本与干土样本平均光谱反射率的包络线去除（图1c），在1 450、1 940和2 200 nm波段附近存在明显的吸收谷；在1 450 nm附近，随着土壤含水量的增加，光谱吸收带向右移动（红外方向），而该现象在1 940 nm附近则不明显。

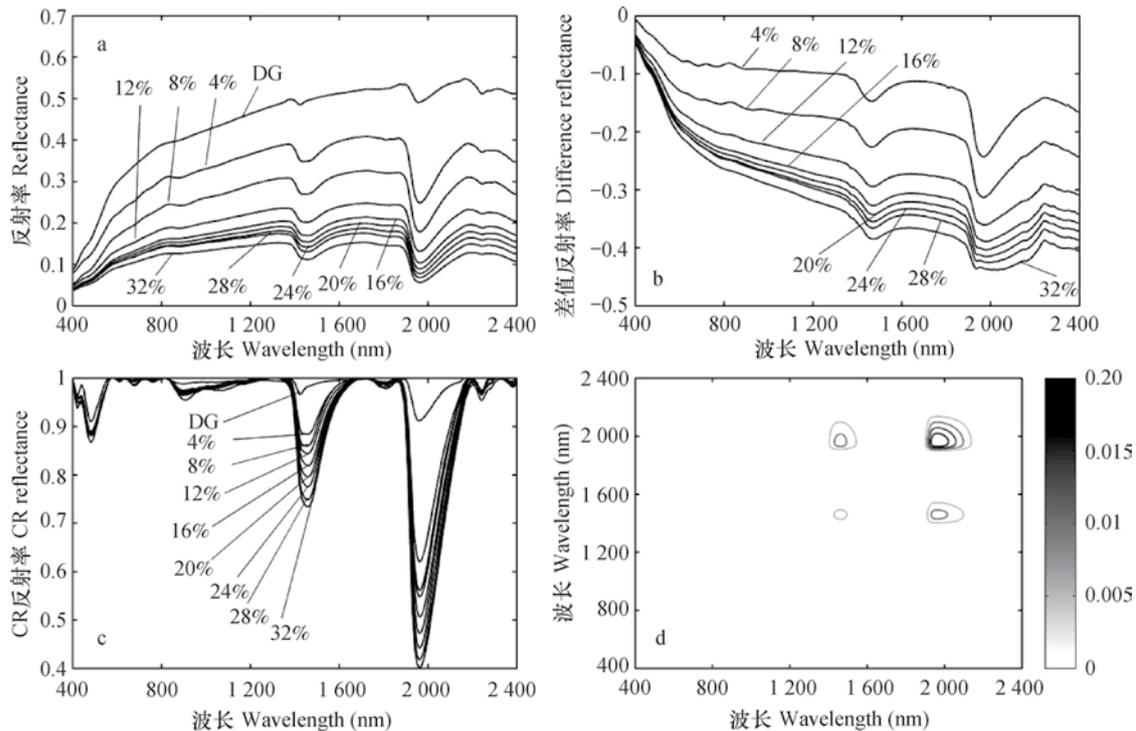
基于上述包络线去除结果进行二维相关光谱分

析，结果见图1d（为便于对比，将colorbar中0附近的值设置显示为白色）。在1 450 nm和1 940 nm附近出现了自相关峰，且1 940 nm附近的自相关峰要强于1 450 nm附近；1 940 nm附近以层间水为主的H₂O谱带对水分微扰较为敏感，1 450 nm附近的以羟基为主的羟基（-OH）带谱则相比不敏感^[22]。

已有研究表明，在1 450、1 940和2 200 nm附近存在有机质的响应波段^[23-24]，但是，由以上分析可知，土壤含水量对土壤光谱反射率的影响十分明显，尤其是1 940nm附近强烈的水分吸收谷甚至部分遮盖了2 200 nm附近的吸收谷，从而可能影响土壤有机质的预测。因此，必须采用特定土壤含水量消除算法对湿土光谱进行分析校正，降低土壤含水量对有机质估算的影响。

2.3 不同EPO维度g的投影矩阵P对比分析

选取维度g等于2、4、6、8、10时，分析投影矩阵P的变化（图2）。当维度g=2、4时，投影矩阵P的变化比较均匀、较为光滑；而维度g=6、8、



注：DG为干土，下同。a.不同土壤含水量梯度的平均光谱；b.研磨与加湿状态的差值平均光谱；c.不同土壤含水量梯度上的包络线去除平均光谱；d.包络线去除平均光谱的二维同步相关谱图

Note: DG stands for dry soil sample, the same as below. a. Mean spectra along the soil moisture gradient; b. Mean difference spectra between the rewetted and dry ground samples; c. Mean CR spectra along the moisture gradient; d. A 2-dimensional synchronous correlation spectra of the mean CR spectra

图1 四种光谱分析方法

Fig. 1 Four spectral analysis methods

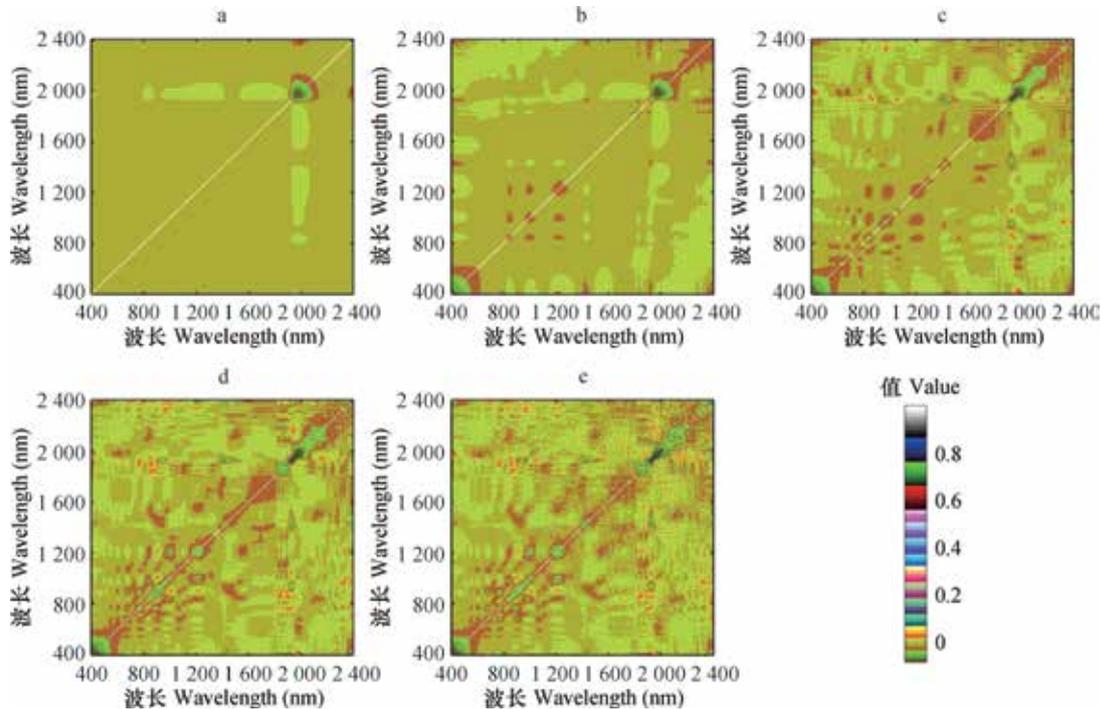


图2 2 (a)、4 (b)、6 (c)、8 (d)、10 (e) 个EPO维度的投影矩阵 P 视觉图

Fig. 2 Visual diagrams of EPO projection matrix P with two (a), four (b), six (c), eight (d) and ten (e) EPO dimensions

10时, 投影矩阵 P 的变化较错乱、“毛刺”较多、噪声明显。这主要是由于, 较大的EPO维度包含着由 $D^T D$ 矩阵分解时产生大量的特征向量, 子空间 Q 矩阵会引入外部光谱噪音, 进而导致投影矩阵 P ($P=I-Q$) 也有更多的噪音信息。因此, 不同EPO维度 g 对转换光谱的模型精度有重要影响, 选取合适的维度 g 至关重要。

2.4 模型参数优化

EPO-偏最小二乘回归模型中有两个参数需要优化确定: EPO维度 g 和偏最小二乘回归最佳因子数 k , 以模型中的 $RMSE_{cv}$ 值最小作为确定最佳参数组合的标准。EPO-偏最小二乘回归模型中维度 g 优化范围为1~10, 因子数 k 优化取值范围为1~20, 利用S1建立土壤有机质的偏最小二乘回归估算模型。当因子数 $k \geq 14$, $RMSE_{cv}$ 的值变化趋缓; 当维度 $g=4$ 时, 可以获得最小的 $RMSE_{cv}=1.48$ 值(图3a)。因此, $g=4, k=14$ 为最佳参数组合, 用于后续数据建模分析。

同理, EPO-支持向量机回归模型中EPO维度 g 和惩罚参数 C 需要优化确定。维度 g 优化范围为1~10, 固定支持向量机回归算法中的RBF核参数 γ , 惩罚参数 C 的变化范围为8~64(步长为8), 以模型中的 $RMSE_{cv}$ 最小值作为确定最佳参数组合

(g, C)的标准。当维度 $g=2, C=64$ 时, 可以获得最小的 $RMSE_{cv}=1.17$ 值(图3b)。因此, $g=2, C=64$ 为最佳参数组合, 用于后续数据建模分析。

2.5 EPO算法校正前、后的前两个主成分得分对比

对EPO算法校正前、后(EPO-偏最小二乘回归模型的维度 $g=4$ 、EPO-支持向量机回归模型的维度 $g=2$)的S2光谱数据集的样本进行均值中心化(mean centre)后, 再利用主成分分析法计算每个样本在前两个主成分得分(图4)。EPO算法校正前的样本点分布总体上相对分散(图4a), 相同土壤含水量梯度的样本点聚集程度高, 湿土样本点与干土样本点之间相互分离程度大, 不同土壤含水量梯度的湿土样本点之间空间重叠程度很低, 反映出不同土壤含水量梯度样本点之间具有明显差异; 这表明土壤含水量的变化对土样样本的光谱数据具有明显的影响。而EPO算法校正后的湿土样本点在主成分得分空间的位置基本与干土样本点相同(图4b、图4c), 说明因土壤含水量造成的光谱差异基本消除。

2.6 EPO算法校正前、后的模型精度对比

EPO算法校正前(表2), 偏最小二乘回归模型的 R_c^2 为0.75, S2的 R_p^2 、RPD分别为0.28、1.16, 模型精度较差, 无法实现土壤有机质在不同土壤含

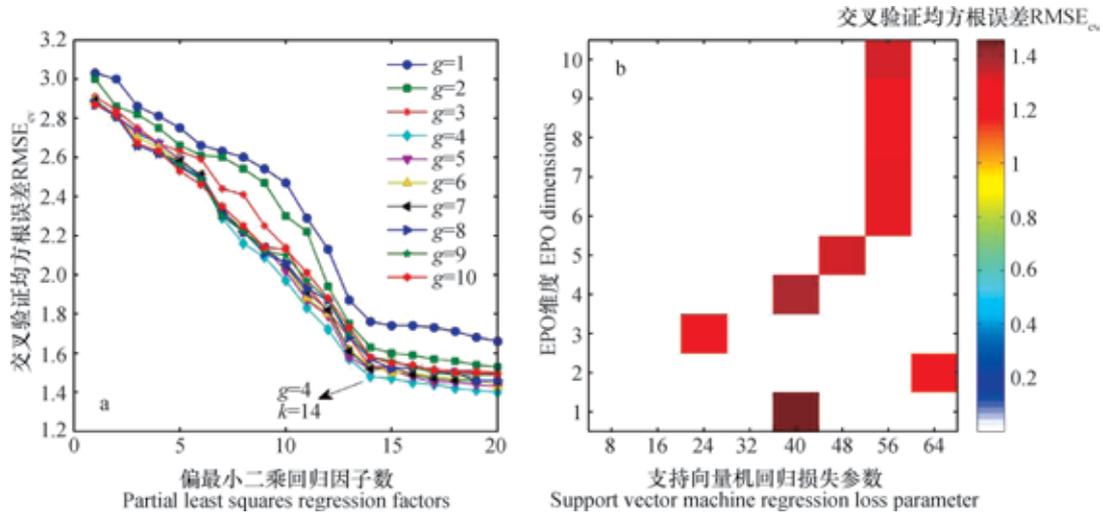


图3 EPO训练集S1建立EPO-偏最小二乘回归 (a)、EPO-支持向量机回归 (b) 的土壤有机质估算模型中不同参数组合所产生的RMSE_{cv}值

Fig. 3 RMSE_{cv} values yielded by the soil organic matter prediction models, EPO-partial least squares regression (a) and EPO-support vector machine regression (b) , different in parameter composition built for EPO training subset

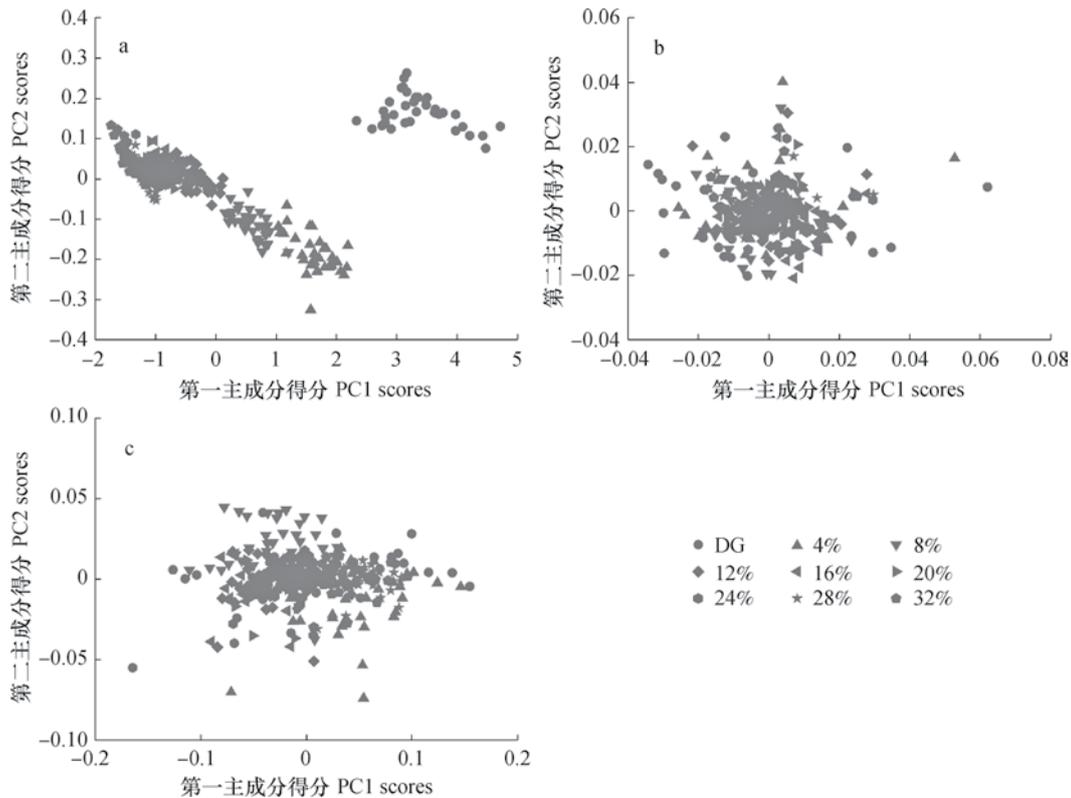


图4 EPO算法校正前 (a)、后的验证样本集S2光谱的前两个主成分得分图 (EPO-偏最小二乘回归 (b) , EPO-支持向量机回归 (c))

Fig. 4 Scatter diagrams of the scores obtained by the first and second principal components in the principal component analysis of the spectra of Subset S2 before (a) and after EPO calibration (EPO-partial least squares regression (b) , EPO-support vector machine regression (c))

表2 EPO算法校正前、后的干土模型(S0建模)验证S2的土壤有机质预测结果

Table 2 Verification of predictions of soil organic matter in soil samples of S2 with the S0 based model calibrated with or without EPO

模型 Models	建模集Calibration sets		预测集Prediction sets		
	R^2_c	均方根误差RMSE _c (g kg ⁻¹)	R^2_p	均方根误差RMSE _p (g kg ⁻¹)	预测偏差比 RPD
偏最小二乘回归 ^①	0.75	2.01	0.28	2.59	1.16
外部参数正交化法-偏最小二乘回归 ^②	0.80	1.78	0.68	1.71	1.76
外部参数正交化法-支持向量机回归 ^③	0.88	1.37	0.78	1.40	2.15

①Partial least-squares regression, ②External parameter orthogonalization-partial least-squares regression, ③External parameter orthogonalization-support vector machine regression

水量梯度的有效估算(图5)。经EPO算法校正后(表2), EPO-偏最小二乘回归模型的 R^2_c 为0.80, S2*的 R^2_p 、RPD分别为0.68、1.76, 模型精度一般, 可以实现土壤有机质在不同土壤含水量梯度的粗略估算(图5); 相比校正前, 预测精度有所提升, S2*的验证 R^2_p 提高了0.4, RMSE_p降低了0.88, RPD值从1.16提高至1.76, 模型从不可用级别提高至可用级别, 表明EPO算法对于去除土壤含水量

影响、提高土壤有机质的估测精度具有很好的可行性。

EPO-支持向量机回归模型的 R^2_p 、RPD分别为0.78、2.15(表2), 可以实现土壤有机质在不同土壤含水量梯度的精确估算, 在给定5%显著性水平, 计算得到95%预测置信区间(图5), 说明模型精度较好、偏差小, 解释能力较强; 相比EPO-偏最小二乘回归模型, S2*的 R^2_p 、RPD值分别提高

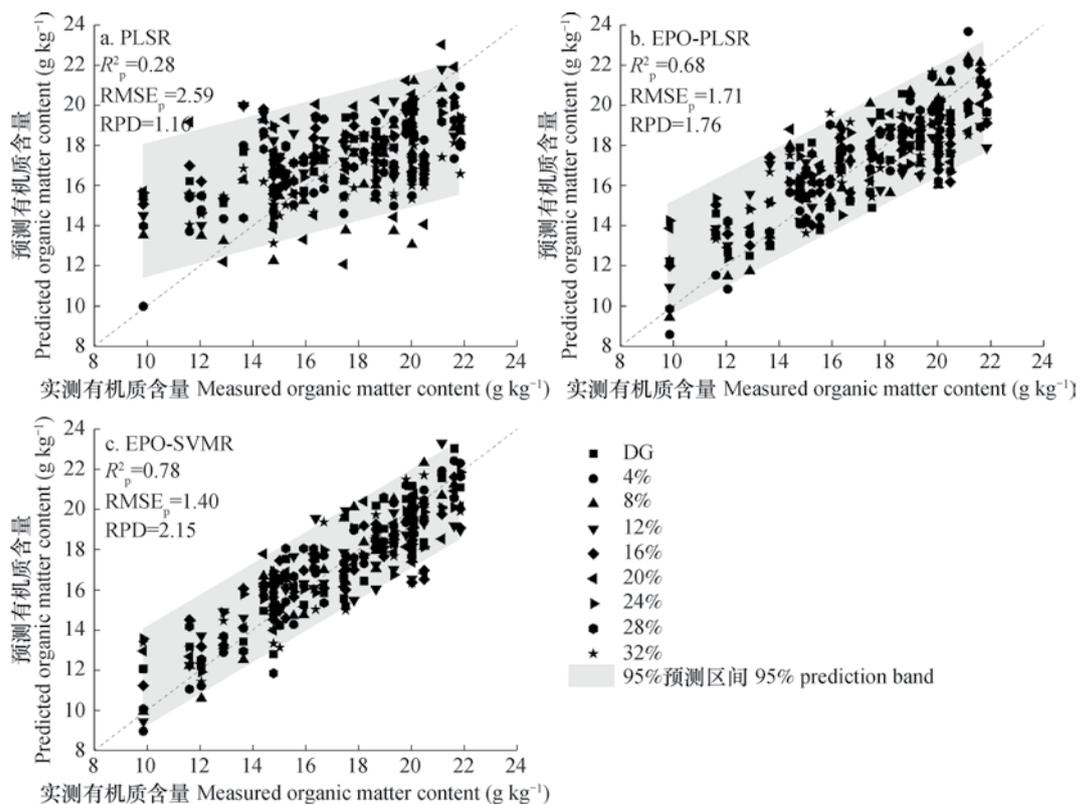


图5 EPO算法校正前(a)、后的验证样本集S2的土壤有机质预测散点图(EPO-偏最小二乘回归(b), EPO-支持向量机回归(c))

Fig.5 Scatter plots of predicted soil organic matter in the soil samples in the S2 validation dataset before (a) and after EPO calibration (EPO-partial least squares regression (b), EPO-support vector machine regression (c))

了0.1、0.39，模型从可用级别提高到较好级别，表明EPO算法与非线性支持向量机回归建模方法结合可以有效处理不同土壤含水量梯度中非线性影响因素，提高有机质估测精度。

3 讨 论

本研究以潮土为试验载体，通过在室内人工设置不同土壤含水量梯度，模拟野外环境土壤含水量的影响，获取9个梯度土壤含水量（0~32%）的土壤光谱数据。Nocita等^[25]设计了0~25%的土壤含水量变化范围（5%的土壤含水量梯度间隔），采用主成分分析法研究表明当土壤含水量>15%时光谱对变化土壤含水量的敏感度有所降低。这与本研究的结果具有很好的相似性（本研究设置了4%的土壤含水量梯度间隔）：土壤含水量的增加会导致光谱反射率整体降低，当土壤含水量<16%时，土壤含水量变化对土壤光谱反射率变化的影响相对较大，而当土壤含水量 \geq 16%时，土壤含水量变化对土壤光谱反射率变化的影响降低，土壤光谱反射率降低速率趋缓，继续提高土壤含水量至32%（小于田间最大持水量）过程中土壤光谱反射率的变化逐渐趋于稳定。因此，本文设置0~32%范围的土壤含水量，为揭示土样在不同土壤含水量条件的光谱响应规律提供了基础。

通过对比EPO算法校正前、后的土样在不同土壤含水量梯度的光谱主成分分析散点图，发现EPO算法能够有效去除土壤含水量对土壤光谱的影响，与Ge等^[8]、Ackerson等^[10]、Liu等^[11]研究较为一致。但不同EPO维度 g 对转换光谱有重要影响（图2），在实际应用中，应注意选取合适的维度 g 。此外，由于本研究已经在室内获取了湿土光谱的土壤含水量消除转换矩阵 \mathbf{P} ，野外实践时不必掌握土壤含水量的变化情况（但小于田间最大持水量），而且在实际应用操作时简单方便，只需将转换矩阵 \mathbf{P} 与湿土光谱数据进行相乘。

Morgan等^[26]利用田间湿土光谱直接进行土壤有机碳预测，其RPD可达1.45，但相比室内过筛研磨后的干土光谱的模型精度（RPD=1.71），发现田间原位光谱的土壤有机碳预测精度要低；Stenberg^[27]研究表明土样加湿对于提升土壤有机碳建模精度具有积极作用，利用Vis-NIR建立湿土

光谱时模型的验证 R^2 最大可提升0.15，但必须在特定的土壤含水量状态下（建模和验证样本需要来自相似的土壤含水量）才可以减小土壤含水量对土壤有机碳的光谱预测影响。因此，在保证其他外界干扰因素（环境光、砾石、植被残积物等）差异最小化的前提下，需要采用特定的土壤含水量消除算法结合非线性建模技术用以进一步优化模型性能。本研究中，相比偏最小二乘回归建模方法，采用EPO算法与支持向量机回归方法的结合可以在消除土壤含水量影响的基础上提高模型精度，这与Li等^[28]将bagging-LS-SVM方法应用到原位光谱监测中较为相似。

近年来，随着全球、洲际、国家、区域等各个不同尺度的土壤光谱数据库的建立与完善，积累了大量的存量土壤数据，形成了非常有价值的先验知识库，这为研究结合已有土壤光谱库的先验信息来消除野外动态变化的土壤含水量对土壤属性参数的反演提供了数据基础^[29-30]。因此，在利用光谱数据库对田间湿土光谱进行预测时，对湿土光谱数据进行校正处理是十分必要的。在本研究中，利用前期采样的干土样本作为EPO算法校正的建模集，再对不同土壤含水量梯度的土壤有机质进行预测，其中EPO-支持向量机回归模型的RPD为2.15，模型精度稳定可靠，该思路为今后野外监测土壤属性参数提供了理论基础和技术参考。

4 结 论

随着土壤含水量的增加，土壤光谱反射率呈非线性降低趋势；土壤高光谱受土壤含水量的影响非常严重，掩盖了土壤有机质的光谱吸收信息。当EPO维度 $g=2、4$ 时，投影矩阵 \mathbf{P} 的变化比较均匀；而EPO维度 $g=6、8、10$ 时，投影矩阵 \mathbf{P} 的变化比较错乱，噪音明显，优选出最佳的维度 g 对转换光谱的模型精度至关重要。EPO算法校正后的湿土光谱数据得到的主成分得分的样本点空间分布与干土样本位于基本相同的区域，干土、湿土样本之间的光谱数据相似性很高，EPO能够有效消除土壤含水量对土壤光谱的影响。EPO-支持向量机回归模型的验证RPD值为2.15，可以更好地处理土壤光谱与土壤有机质之间的非线性关系，实现土壤有机质的有效估算。

参 考 文 献

- [1] 方少文, 杨梅花, 赵小敏, 等. 红壤区土壤有机质光谱特征与定量估算——以江西省吉安为例. 土壤学报, 2014, 51 (5) : 1003—1010
Fang S W, Yang M H, Zhao X M, et al. Spectral characteristics and quantitative estimation of SOM in red soil typical of Ji'an County, Jiangxi Province (In Chinese). Acta Pedologica Sinica, 2014, 51 (5) : 1003—1010
- [2] Xu S X, Shi X Z, Wang M Y, et al. Effects of subsetting by parent materials on prediction of soil organic matter content in a hilly area using Vis-NIR spectroscopy. Plos One, 2016, 11 (3) : e0151536
- [3] Shi Z, Ji W J, Viscarra Rossel R A, et al. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library. European Journal of Soil Science, 2015, 66 (4) : 679—687
- [4] Ji W J, Li S, Chen S C, et al. Prediction of soil attributes using the Chinese soil spectral library and standardized spectra recorded at field conditions. Soil & Tillage Research, 2016, 155 (SI) : 492—500
- [5] Ji W J, Viscarra Rossel R A, Shi Z. Improved estimates of organic carbon using proximally sensed vis-NIR spectra corrected by piecewise direct standardization. European Journal of Soil Science, 2015, 66 (4) : 670—678
- [6] Roger J M, Chauchard F, Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemometrics and Intelligent Laboratory Systems, 2003, 66 (2) : 191—204
- [7] Minasny B, McBratney A B, Bellon-Maurel V, et al. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma, 2011, 167/168: 118—124
- [8] Ge Y F, Morgan C L S, Ackerson J P. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma, 2014, 221: 61—69
- [9] Ji W J, Viscarra Rossel R A, Shi Z. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. European Journal of Soil Science, 2015, 66 (3) : 555—565
- [10] Ackerson J P, Demattê J A M, Morgan C L S. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. Geoderma, 2015, 259: 196—204
- [11] Liu Y, Pan X Z, Wang C K, et al. Predicting soil salinity with Vis-NIR spectra after removing the effects of soil moisture using external parameter orthogonalization. Plos One, 2015, 10 (10) : e0140688
- [12] 张甘霖, 龚子同. 土壤调查实验室分析方法. 北京: 科学出版社, 2012: 1—54
Zhang G L, Gong Z T. Soil survey laboratory methods (In Chinese). Beijing: Science Press, 2012: 1—54
- [13] 洪永胜, 于雷, 耿雷, 等. 应用DS算法消除室内几何测试条件对土壤高光谱数据波动性的影响. 华中师范大学学报 (自然科学版), 2016, 50 (2) : 303—308
Hong Y S, Yu L, Geng L, et al. Using direct standardization algorithm to eliminate the effect of laboratory geometric parameters on soil hyperspectral data fluctuate characteristic (In Chinese). Journal of Central China Normal University (Natural Sciences), 2016, 50 (2) : 303—308
- [14] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry, 1964, 36 (8) : 1627—1639
- [15] Noda I. Generalized two-dimensional correlation method applicable to infrared, raman, and other types of spectroscopy. Applied Spectroscopy, 1993, 47 (9) : 1329—1336
- [16] 褚小立. 化学计量学方法与分子光谱分析技术. 北京: 化学工业出版社, 2011
Chu X L. Molecular spectroscopy analytical technology combined with chemometrics and its applications (In Chinese). Beijing: Chemical Industry Press, 2011
- [17] Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 2001, 58 (1) : 109—130
- [18] Gao Y, Cui L J, Lei B, et al. Estimating soil organic carbon content with Visible-Near-Infrared (Vis-NIR) Spectroscopy. Applied Spectroscopy, 2014, 68 (7) : 712—722
- [19] 于雷, 洪永胜, 周勇, 等. 高光谱估算土壤有机质含量的波长变量筛选方法. 农业工程学报, 2016, 32 (13) : 95—102
Yu L, Hong Y S, Zhou Y, et al. Wavelength variable selection methods for estimation of soil organic matter content using hyperspectral technique (In Chinese). Transactions of the CSAE, 2016, 32 (13) : 95—102
- [20] Lobell D B, Asner G P. Moisture effects on soil reflectance. Soil Science Society of America Journal, 2002, 66 (3) : 722—727

- [21] 刘洋, 丁潇, 刘焕军, 等. 黑土土壤水分反射光谱特征定量分析与预测. 土壤学报, 2014, 51 (5) : 1021—1026
- Liu Y, Ding X, Liu H J, et al. Quantitative analysis of reflectance spectrum of black soil as affected by soil moisture for prediction of soil moisture in black soil (In Chinese). *Acta Pedologica Sinica*, 2014, 51 (5) : 1021—1026
- [22] 宋海燕, 程旭. 水分对土壤近红外光谱检测影响的二维相关光谱解析. 光谱学与光谱分析, 2014, 34 (5) : 1240—1243
- Song H Y, Cheng X. Analysis of the effect of moisture on soil spectra detection by using two-dimensional correlation near infrared spectroscopy (In Chinese). *Spectroscopy and Spectral Analysis*, 2014, 34 (5) : 1240—1243
- [23] Viscarra Rossel R A, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 2010, 158 (1/2) : 46—54
- [24] 史舟, 王乾龙, 彭杰, 等. 中国主要土壤高光谱反射特性分类与有机质光谱预测模型. 中国科学: 地球科学, 2014, 44 (5) : 978—988
- Shi Z, Wang Q L, Peng J, et al. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations (In Chinese). *Science China: Earth Sciences*, 2014, 44 (5) : 978—988
- [25] Nocita M, Stevens A, Noon C, et al. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma*, 2013, 199 (SI) : 37—42
- [26] Morgan C L S, Waiser T H, Brown D J, et al. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma*, 2009, 151 (3/4) : 249—256
- [27] Stenberg B. Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on VIS-NIR predictions of clay and soil organic carbon. *Geoderma*, 2010, 158 (1/2) : 15—22
- [28] Li S, Shi Z, Chen S C, et al. In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai-Tibet Plateau. *Environmental Science & Technology*, 2015, 49 (8) : 4980—4987
- [29] Viscarra Rossel R A, Behrens T, Ben-Dor E, et al. A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 2016, 155: 198—230
- [30] 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. 土壤学报, 2015, 52 (2) : 312—320
- 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 (In Chinese). *Acta Pedologica Sinica*, 2015, 52 (2) : 312—320

Removing the Effect of Soil Moisture on Prediction of Soil Organic Matter with Hyperspectral Reflectance Using External Parameter Orthogonalization

HONG Yongsheng^{1, 2} YU Lei^{1, 2†} ZHU Yaxing^{1, 2} WU Hongxia^{1, 2} NIE Yan^{1, 2} ZHOU Yong^{1, 2}
 QI Feng³ XIA Tian^{1, 2}

(1 Key Laboratory for Geographical Process Analysis & Simulation, Hubei Province, Central China Normal University, Wuhan 430079, China)

(2 College of Urban & Environmental Science, Central China Normal University, Wuhan 430079, China)

(3 School of Environmental and Sustainability Sciences, Kean University, NJ 07083, USA)

Abstract 【Objective】 Soil organic matter is an important index of soil properties, because it is vital to crop growth and soil quality. The technology of hyperspectral analysis is a rapid, convenient, low-cost and alternative method and exhibits an increasingly remarkable development potential in estimation of soil organic matter. However, when hyperspectral reflectance is used in the field, there are several external environmental factors, including soil moisture content, temperature, and surface of the soil that may affect soil spectra. Especially soil moisture content, a major limit to field hyperspectral survey, might mask the

absorption features of soil organic matter, and hence dramatically lower accuracy of the prediction of soil organic matter. Therefore, it is essential to find a method capable of removing the impact of soil moisture content on spectral reflectance, so as to improve the accuracy of quantitative prediction of soil organic matter. In this paper, the EPO (external parameter orthogonalization) algorithm was introduced for that purpose.

【Method】 A total of 217 soil samples were collected from the 0~20 cm soil layer in the Jiangnan Plain. In the laboratory, the soil samples were air-dried and ground to pass a sieve with mesh < 2 mm. Then the soil samples were analyzed separately for soil organic matter content with the potassium dichromate external heating method. The total of 217 soil samples were further divided into three non-overlapping subsets: a model calibration set (S0), consisting of 122 samples and dedicated to development of a multivariate model for soil organic matter; an EPO development subset (S1) consisting of 60 samples for EPO development; and a validation subset (S2) consisting of 35 samples for independent EPO validation. Then, the samples in S1 and S2 were rewetted in line with the following procedure: from each soil sample 150 g oven-dried soil was weighed out, put in a black cylindrical box and rewetted along the gradient of soil moisture content increment with interval being 4% each, making up a total of 9 treatments in soil moisture content along the gradient i.e. 0, 4%, 8%, 12%, 16%, 20%, 24%, 28% and 32%. An spectrometer was used to acquire hyperspectral reflectances of the samples of three subsets (S0, S1 and S2, including the rewetting samples) on 350 to 2500 nm. And then influences of the soil moisture content on the soil spectra were analyzed, and the scores of the first two principal components in the principal component analysis were used for comparison to determine performance of EPO algorithm in removing the effects of soil moisture content on spectral reflectance of the wet samples. In the end, modeling for the S0 subset was done using the partial least squares regression and support vector machine regression, and the S2 subset of wet samples were used as external validation set before and after calibration with EPO. The coefficient of determination (R^2), root mean squared error (RMSE) and the ratio of prediction to deviation (RPD) between the predicted and measured values of soil organic matter were used to compare the 3 models in performance: High R^2 , RPD and low RMSE were indicators of optimal models for partial least squares regression (before EPO calibration), EPO-partial least squares regression and EPO-support vector machine regression.

【Result】 Results show that (1) Soil moisture content does have obvious influence on spectral reflectance, and the reflectance decreases in value across the entire wavelength domain with increasing soil moisture content, making it more challenging to identify useful features of soil organic matter with spectra; (2) For Subset S2 before EPO calibration, no spectral overlaps are observed between the wet and dry samples, and spectra of the wet sample cluster in spaces free from those of the dry sample (mutual independent space). However, after EPO calibration of Subset S2 set, the spectra of the wet sample appear almost in the same positions as those of the dry sample do within the eigen space, demonstrating that the two groups of spectra are highly similar; (3) Before EPO calibration, the partial least squares regression model is the poorest in prediction accuracy (the validation RPD=1.16). EPO calibration has improved prediction accuracy of the model up to an acceptable level (the validation RPD=1.76). And EPO-support vector machine regression model performs better than the other two with validation R^2 reaching 0.78, and RPD = 2.15, which indicates that the effects of soil moisture content on spectra are successfully eliminated.

【Conclusion】 In the future, this approach will facilitate rapid measurement of soil organic matter for this study area.

Key words Soil spectra; Soil organic matter; Moisture content; External parameter orthogonalization; Support vector machine regression; Jiangnan Plain

(责任编辑: 檀满枝)