

可见光-近红外、中红外光谱的土壤有机质组分反演

罗德芳¹, 彭杰^{1*}, 冯春晖¹, 柳维扬¹, 纪文君², 王楠³

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

3. 浙江大学环境与资源学院, 浙江杭州 310058

摘要 土壤有机质是土壤肥力的物质基础, 其含量的高低是评价土壤肥力的重要标志。土壤有机质组分根据其溶解性可分为胡敏素(HM)、胡敏酸(HA)、富里酸(FA), 不同组分的肥力特性差异显著, 因此, 土壤有机质组分数据可更加全面、客观的反映土壤肥力状况。传统土壤有机质及组分的测定工序繁杂, 效率低下且时效性差, 大量研究表明高光谱技术能有效提高土壤属性的检测效率并降低测试成本, 但关于可见光-近红外、中红外光谱检测土壤有机质组分的报道鲜见。为了探索中红外光谱及可见光-近红外-中红外组合光谱对土壤有机质组分检测的可行性, 并对比有机质单一光谱模型与有机质不同组分的组合光谱模型的预测精度, 以南疆地区农田土壤为例, 在阿克苏及和田地区共采集 93 个土样, 进行有机质、胡敏素、胡敏酸、富里酸含量及光谱数据的测定。其次, 利用可见-近红外(VNIR)、中红外(MIR)及其组合光谱(VNIR-MIR)三种光谱数据集, 采用偏最小二乘(PLSR)、支持向量机(SVM)、随机森林(RF)三种建模方式对土壤有机质、胡敏素、胡敏酸、富里酸含量进行组合模型分析预测。结果表明: (1)土壤有机质及各组分均与光谱反射率有较好的相关性, 土壤有机质及组分在 MIR 谱段的特征波段数量明显多于 VNIR 谱段。(2)有机质最优预测模型的模式为 VNIR-MIR-RF, 该模型的决定系数 R^2 为 0.90; 胡敏素与胡敏酸最优预测模型的模式均为 VNIR-RF 模型, R^2 均为 0.92; 富里酸最优预测模型的模式为 MIR-RF 模型, R^2 为 0.94。(3)基于胡敏素、胡敏酸和富里酸的有机质组合光谱模型的预测精度明显高于有机质单一光谱模型, 两种模型的 R^2 分别为 0.93 和 0.90。实现了土壤有机质组分的高效快速反演, 且基于有机质组分的组合模型提高了土壤有机质预测精度, 为南疆地区大尺度土壤肥力的鉴定与精准施肥提供重要的参考价值。

关键词 土壤有机质组分; 光谱反射率; 偏最小二乘; 支持向量机; 随机森林; 反演模型

中图分类号: S153.6 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)10-3069-08

引言

土壤有机质(soil organic matter, SOM)是土壤肥力的决定性因素, 在提高植物营养、改善土壤肥力等方面具有重要作用, 它的含量与土壤肥力特性如作物产量等直接相关, 高肥力土壤生产的食物营养品质高于低肥力土壤。新疆是我国主要发展绿洲农业的地区, 优越的地理环境和成熟的栽培技术, 使棉花成为当地的优势特色种植作物, 同时棉花也是新疆主要的经济作物之一。近年来因过度开垦及化肥、农药等的不合理使用, 使棉田土壤肥力退化严重, 对棉花产量造成极大影响^[1]。快速诊断及改良土壤肥力, 是南疆农业可持续

发展的关键。土壤有机质包括胡敏素(humin, HM)、胡敏酸(humic acid, HA)和富里酸(fulvic acid, FA)三个组分, 有机质不同组分矿质化难易程度和养分释放的能力具有明显差异^[2]。在有机质组分中, 胡敏素的相对含量最高, 但其不易矿质化, 养分很难释放出来供植物吸收利用, 胡敏酸的相对含量和矿质化难易程度为其次, 富里酸的相对含量最低, 但其最易矿质化而释放出养分供植物吸收利用。

探明土壤有机质及其组分含量对于准确掌握土壤的潜在和有效肥力具有重要意义。传统土壤有机质测定过程繁琐, 具有采样成本高、分析周期长等缺点, 难以胜任大尺度范围土壤有机质的监测, 高光谱技术因其快速、便捷、无损、可重复分析技术等优点被广泛应用于土壤有机质测定。可见

收稿日期: 2020-09-04, 修订日期: 2020-12-28

基金项目: 国家重点研发计划项目(2018YFE0107000), 国家自然科学基金项目(41361048)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: pjzky@163.com

光-近红外(VNIR)和中红外(MIR)光谱能够对土壤性质进行光谱特征分析,操作简便、不破坏土壤结构,是近年来国内外学者研究的热点。Vasat 等^[3]研究表明,土壤有机质对土壤反射光谱曲线的形状及性质有显著影响,是反射光谱法可预测的主要土壤性质之一。Jiang 等^[4]利用 VNIR 光谱法研究表明在不同土壤水分条件下,土壤盐分含量及广义最小二乘加权(GLSW)结合偏最小二乘回归(PLSR)模型可以有效地估算农田土壤中有机碳的含量。陈思明等^[5]利用线性波谱分解技术重建土壤光谱,预测了土壤有机质的含量。Bao 等^[6]通过分析不同地形土壤有机质含量与光谱反射率之间的相关关系,发现一种新的光谱指数($R_{2.294\text{ nm}}/R_{2.286\text{ nm}}$)可用于土壤有机质的估算,并建立偏最小二乘-支持向量机(PLS-SVM)回归模型预测土壤有机质。以上研究主要基于土壤近红外光谱的有机质物理性质或建模方法等,对土壤中红外光谱及土壤有机质组分含量的预测研究还很罕见。

因此,以南疆地区为研究区,基于近红外(VNIR)、中红外(MIR)以及组合光谱(VNIR-MIR),利用偏最小二乘回归(PLSR)、支持向量机(SVM)以及随机森林(RF)三种建模方法对土壤有机质、胡敏素、胡敏酸以及富里酸含量进行分析预测,比较有机质与各组分在不同反射光谱下的建模预测精度,为南疆地区土壤肥力的监测与精准施肥提供参考。

1 实验部分

1.1 土样采集及化学测定

新疆位于中国西北地区,分为北疆和南疆两个区域。土壤样品采集于南疆地区,包括南疆阿克苏地区的温宿县($40^{\circ}52' - 42^{\circ}15'N$, $79^{\circ}28' - 81^{\circ}30'E$)及阿瓦提县($39^{\circ}31' - 40^{\circ}50'N$, $79^{\circ}45' - 81^{\circ}05'E$);和田地区的和田县($34^{\circ}22' - 38^{\circ}27'N$, $78^{\circ} - 80^{\circ}30'E$)。阿克苏地区位于天山山脉和塔里木盆地之间,是冲积平原中心,年降水量约 42.4~94.4 mm,蒸发量约 1 200~1 500 mm,水资源丰富,为暖温带干旱型气候。和田地区位于南塔里木盆地边缘与北昆仑山之间,年降水量较低(约 35 mm),年蒸发量较高(约 2 480 mm),属于干旱荒漠性气候。

采样时间在 2014 年 11 月,采样方法为五点梅花状取样法,采集土壤类型温宿县为淤黄泥、阿瓦提县为灌淤土,和田县为水稻土,每个采样区各采集 31 个土壤样品,共采集 93 个 0~20 cm 的表层土壤样本。将土样带回实验室经风干、研磨、过筛后分为两份,一份用于土壤有机质、胡敏素、胡敏酸以及富里酸含量的测定,一份用于可见-近红外光谱及中红外光谱数据测定。土壤有机质采用重铬酸钾-滴定法进行测定,土壤胡敏素、胡敏酸、富里酸采用焦磷酸钠提取重铬酸钾法进行测定。

描述性统计包括平均值,标准差,最小值,最大值,峰度和变异系数,土壤有机质及其组分含量描述统计表如表 1 所示。由表 1 可得,土壤有机质含量最大值为 $68.02\text{ g} \cdot \text{kg}^{-1}$,最小值为 $12.52\text{ g} \cdot \text{kg}^{-1}$;土壤有机质及其组分含量的变异系数在 50%左右,属强变异性,因此本研究结果适用范围较广。

1.2 数据采集

VNIR 光谱利用美国生产的 Field SpecPro FR 型光谱仪进行测量^[7],测量范围为 350~2 500 nm,在 350~1 000 nm 采样间隔为 1.4 nm,光谱分辨率为 3 nm;在 1 000~2 500 nm 采样间隔为 2 nm,光谱分辨率为 10 nm。重采样间隔为 1 nm,输出波段数 2 151 个。光谱测量在可控光条件的暗室进行,光源为功率 50 W 的卤素灯,光源距离土壤样品 70 cm,天顶角 30° 。将样品装入直径 10 cm、深度 2 cm 的器皿中,表面用直尺刮平。传感器距土样表面 15 cm,探头为 25° 视场角,测试之前进行白板校正,每个土样进行 10 次重复测量,取平均值作为该土样实际反射可见-近红外光谱数据。

MIR 光谱利用 Agilent Technologies(美国)生产的 Agilent 4300 手持式 FTIR 光谱仪^[8]进行测量。光谱范围为 $650 \sim 4\,000\text{ cm}^{-1}$,采样间隔为 0.47 cm^{-1} ,光谱分辨率 4 cm^{-1} ,每个频谱都是 32 次内部扫描的结果。测量前将土壤样品在 45°C 条件下烘干 24 h,降低土壤样品中的水分含量之后再行 MIR 光谱测量。与可见-近红外光谱测量方法相似,白板校正之后,每个样本测量 10 次光谱,取平均值作为土壤中红外光谱数据用于分析处理。

表 1 土壤有机质及组分含量统计

Table 1 Statistics of soil organic matter and fractions

测定指标	单位	样本数	平均值/ ($\text{g} \cdot \text{kg}^{-1}$)	标准差/ ($\text{g} \cdot \text{kg}^{-1}$)	最大值/ ($\text{g} \cdot \text{kg}^{-1}$)	最小值/ ($\text{g} \cdot \text{kg}^{-1}$)	峰度/ ($\text{g} \cdot \text{kg}^{-1}$)	变异 系数/%
有机质	($\text{g} \cdot \text{kg}^{-1}$)	93	27.24	12.50	68.02	12.52	1.43	45.87
胡敏素	($\text{g} \cdot \text{kg}^{-1}$)	93	21.81	10.87	59.50	9.44	2.61	49.86
富里酸	($\text{g} \cdot \text{kg}^{-1}$)	93	1.65	1.00	4.90	0.05	1.68	60.56
胡敏酸	($\text{g} \cdot \text{kg}^{-1}$)	93	2.52	1.48	8.59	0.72	2.63	58.75

1.3 光谱数据预处理

在测量过程中,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。在分析可见-近红外光谱数据时需去除边缘 350~399 和 2 401~2 500 nm 噪声较大的土壤样品的初始反射率,保留 400~2 400 nm 波段的土壤原始光谱反射率进行分析处理。对中红外光谱进行重采样处理,去除边缘

波段,保留 $3\,600 \sim 650\text{ cm}^{-1}$ 波段范围光谱数据。光谱数据重采样在 ENVI 软件中进行,采样后对 VNIR 光谱以及 MIR 光谱进行 Savitzky-Golay(SG)平滑处理。

1.4 建模方法与精度验证

将 93 个土壤样品按有机质及组分含量从小到大进行排序,三分之二用于建模,三分之一用于验证,即 62 个建模集

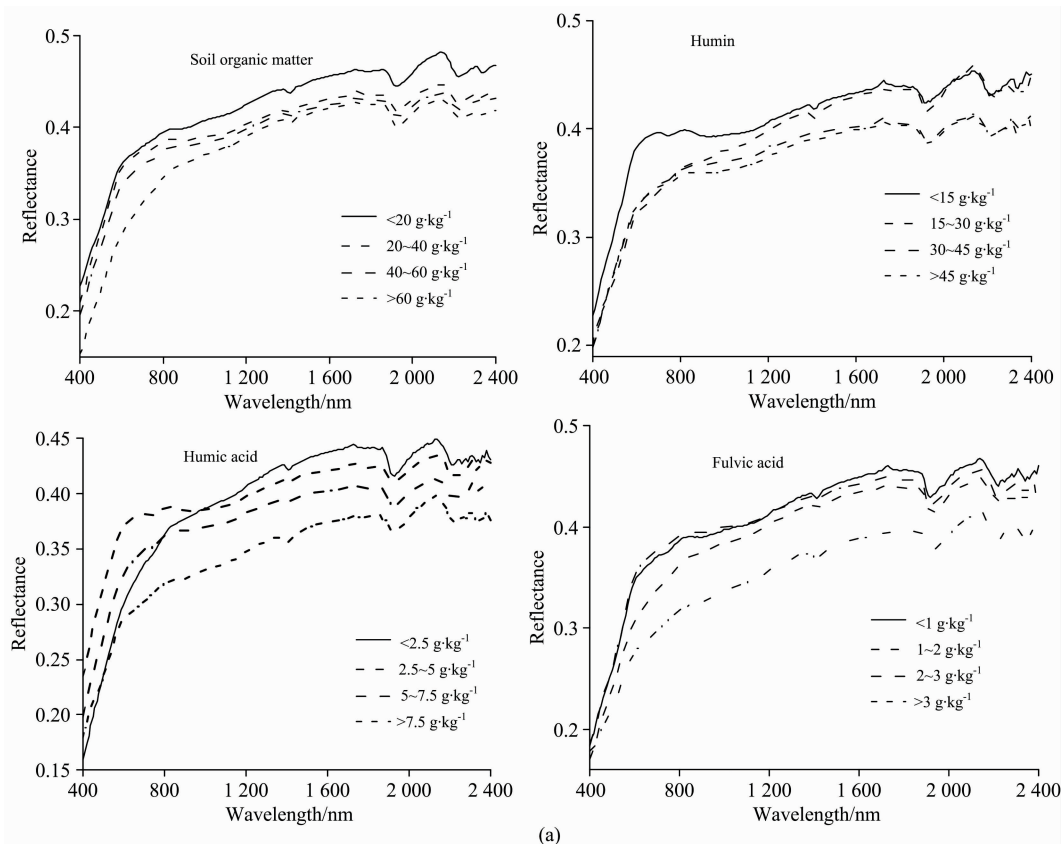
样本和 31 个预测集样本。采用偏最小二乘回归(PLSR)、支持向量机(SVM)回归与随机森林(RF)三种建模方式。PLSR 是一种常见的、简单的、易于使用的多元回归方法,能够从土壤光谱中准确估计土壤的物理和化学性质^[9]。SVM 主要用于回归和分类分析,在解决非线性以及高维模式识别中有其独特的优势^[10]。RF 是基于决策树的一种经典的机器学习方法,对于很多数据集表现良好,精确度比较高,不容易发生过拟合,在数据建模中具有一定的优势^[11]。PLSR 与 SVM 模型在 The Unscrambler X 10.5.1 软件中实现,RF 在 R 软件中实现。

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)、均方根误差(RMSE)、样本观测值三四分位数 Q3 与一四分位数 Q1 之差与 RMSE 的比值(RPIQ)。预测值与实测值的线性回归 R^2 量化了模型所解释的可变性水平,其余的变化归因于随机误差。RMSE 测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。RPIQ 是一个模型效度的度量,它同时考虑了预测误差和测量值的变化。RPIQ 越高,模型的预测能力越好。相对于残差预测偏差(RPD),RPIQ 更能表现出模型的精度,它对测量值的正态性不作任何假设。因此,一个高精度的预测模型应该具有较高的 R^2 和 RPIQ 值以及较低的 RMSE。

2 结果与讨论

2.1 土壤有机质及其组分的光谱特征分析

不同有机质及其组分含量土样的 VNIR 和 MIR 反射光谱曲线如图 1(a)和(b)所示,图中的光谱曲线为不同有机质及其组分含量范围内多个土样的平均光谱。由图 1 可以看出,在不同土样的 VNIR 和 MIR 光谱曲线中,土壤有机质及其组分含量越高,土样光谱反射率越低。在 VNIR 光谱曲线中,不同有机质及其组分含量的土样光谱反射曲线特征大致相同,在 400~700 nm 波段增长速率较快,700~1 900 nm 范围内反射率缓慢增加,在 1 900~2 100 nm 范围内大幅度减小,此后直到 2 400 nm 反射率呈上升趋势。所有光谱曲线在 1 400, 1 900 和 2 200 nm 附近均有三个明显的光谱吸收谷,在 1 400 和 1 900 nm 处, VNIR 光谱反射率受 H_2O 和 OH^- 的影响,在 2 200 nm 处 VNIR 光谱反射率受土壤有机分子的影响^[12]。但有机质各组分的光谱反射率有一定细节变化,在 400~700 nm 波段范围内,土壤胡敏素含量越低,光谱反射率增长速度越大,几乎呈 90° 直线上升,在 700 nm 之后,不同胡敏素含量的土样反射率趋于一致;在 400~900 nm 波段范围内,胡敏酸含量越低,土样光谱反射率越低,但低胡敏酸含量的土样光谱反射率增加速度最大,在 900 nm 之后,胡敏酸含量越低,土样光谱反射率越大。



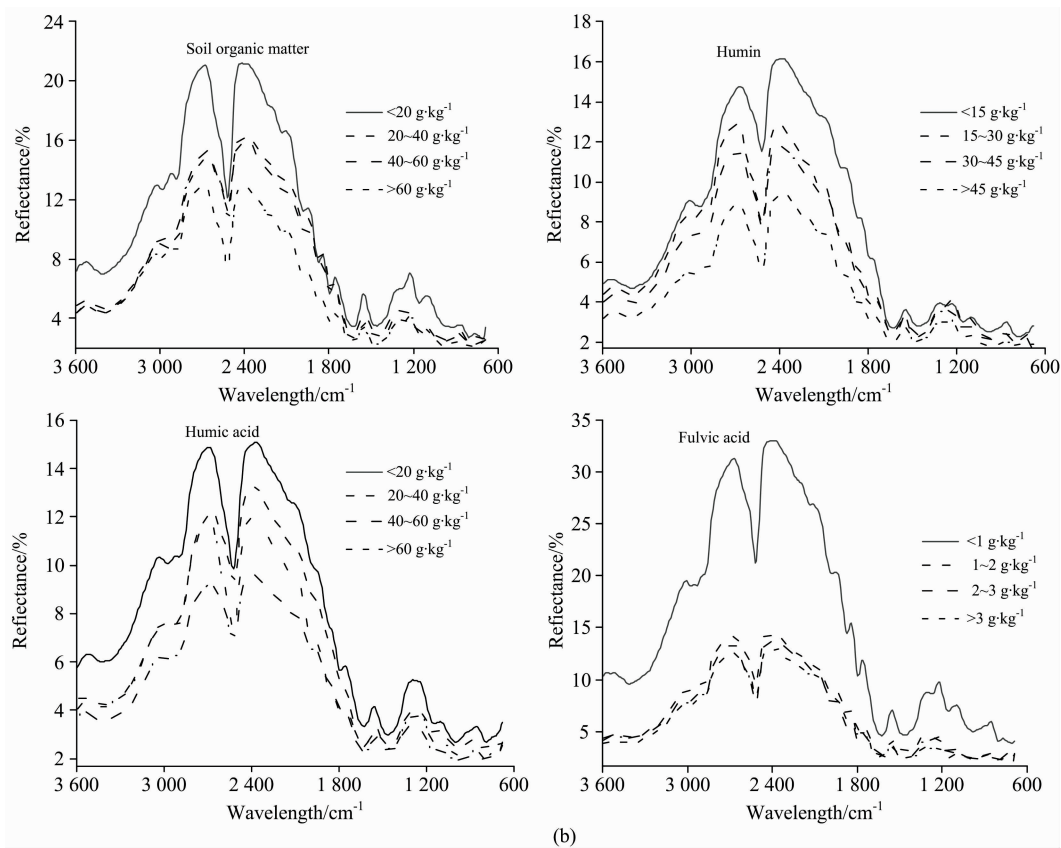


图 1 土壤有机质及其组分的可见-近红外(a)和中红外(b)反射光谱特性

Fig. 1 VNIR (a) and MIR (b) reflectance spectra of soil organic matter and fractions

MIR 光谱反射曲线表现出更加明显的特征波段, 出现更多的反射峰和吸收谷, 在 $3\ 600\sim 3\ 500\ \text{cm}^{-1}$ 处呈增加的趋势, 随后下降至 $3\ 300\ \text{cm}^{-1}$ 波长后反射率大幅度增加至 $2\ 900\ \text{cm}^{-1}$ 处, 在 $2\ 900\sim 2\ 600\ \text{cm}^{-1}$ 处迅速下降后又迅速上升至 $2\ 200\ \text{cm}^{-1}$, 之后下降直到 $1\ 200\ \text{cm}^{-1}$ 处出现波峰之后到 $600\ \text{cm}^{-1}$ 处都持续下降。MIR 光谱反射曲线的吸收谷出现在 $3\ 500, 2\ 520, 1\ 800$ 和 $2\ 500\ \text{cm}^{-1}$ 左右, 吸收深度和面积与土壤有机质及其组分含量直接相关。与吸收谷相比, 反射峰更加明显可见, 在 $2\ 900, 2\ 200$ 以及 $1\ 200\ \text{cm}^{-1}$ 附近波段处, 土壤有机质及其组分含量越高, 反射深度和反射面积越大^[13]。在不同含量有机质组分的土样反射光谱曲线中, 富里酸含量越小, 土样光谱反射率越大, 特征峰更明显。

2.2 土壤有机质及其组分对不同光谱的响应

将土壤有机质及其组分含量与土样 VNIR 和 MIR 光谱反射率分别作相关性曲线。由图 2 可得, 土壤有机质及其组分含量与光谱反射率均呈负相关, 相关系数在 $-0.55\sim 0$ 之间。较 VNIR 来说, MIR 与土壤有机质及其组分相关性曲线中表现出更多的响应特征波段, 相关性更高。在 VNIR 波段中, 特征点集中于 $400, 600, 900, 1\ 400, 1\ 700$ 和 $2\ 200\ \text{nm}$ 等波段处, 在 $400\ \text{nm}$ 处, 土壤有机质及其组分相关性都最高, 土壤有机质、胡敏素、胡敏酸、富里酸的相关系数分别为 $-0.799, -0.792, -0.653$ 和 -0.708 。总体来看, 不同有机质及其组分含量与 VNIR 光谱反射率相关性从大到小为

有机质 > 胡敏素 > 胡敏酸 > 富里酸, 但在 $900\sim 1\ 200\ \text{nm}$ 处, 土壤胡敏酸的相关性大于土壤有机质、胡敏素及富里酸的相关性, 出现一个波峰, 在之后波段的相关性曲线中, 胡敏酸中出现的波峰或波谷都会提前于其他几种土壤性质, 且出现的波峰及波谷均明显大于其他几种土壤性质。

MIR 波段中, 特征点出现在 $3\ 000, 2\ 300, 1\ 800, 1\ 500, 1\ 200, 900$ 和 $600\ \text{cm}^{-1}$ 等波段处。土壤富里酸含量土样光谱反射率的相关性几乎全部为极显著性相关, 而土壤胡敏酸仅在 $1\ 700\sim 2\ 000\ \text{cm}^{-1}$ 波段处存在显著性相关。土壤有机质及组分相关性在 $1\ 800\ \text{cm}^{-1}$ 波段处最高, 土壤有机质、胡敏素、胡敏酸、富里酸的相关系数分别为 $-0.835, -0.834, -0.630$ 和 -0.836 。不同有机质及其组分含量与 MIR 光谱反射率相关性从大到小为富里酸 > 有机质 > 胡敏素 > 胡敏酸。通过对比分析, 土壤有机质及其组分在 VNIR 中相关性最好的是土壤有机质, 最大相关系数为 -0.799 ; 在 MIR 中土壤富里酸的相关性最高, 相关系数为 -0.836 。

2.3 土壤有机质及其组分反演模型的构建与验证

利用 PLSR, SVM 和 RF 模型结合 VNIR, MIR 和 VNIR-MIR 光谱数据集对土壤有机质及组分进行模型构建与验证, 分别得到土壤有机质及组分的 9 种预测模型, 如表 2 所示。

建模集中, 有机质及其组分建立的模型建模精度都较好。在土壤有机质、胡敏素、胡敏酸、富里酸中, 以土壤富里

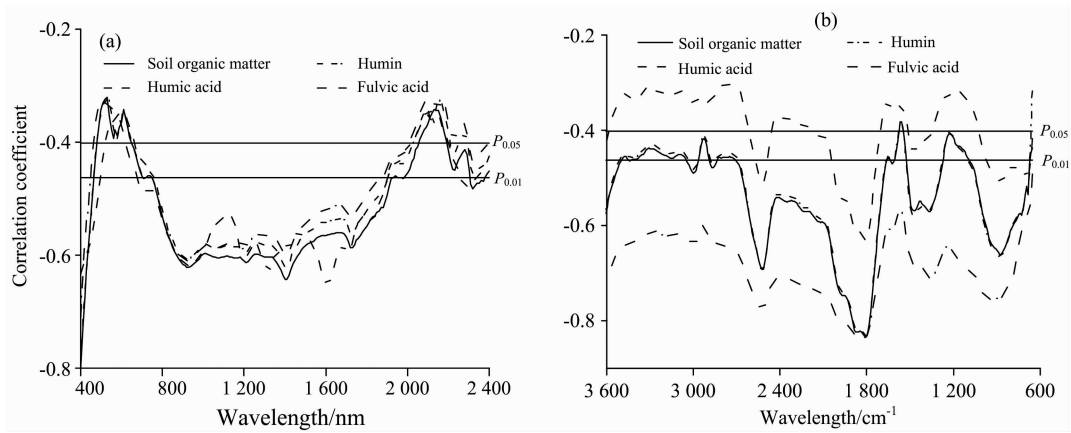


图 2 不同土壤有机质及组分与可见-近红外 (a) 和中红外 (b) 反射率的相关关系

Fig. 2 Correlations between soil organic matter content and visible-near infrared (a), mid-infrared (b) reflectivity

表 2 基于不同光谱的 PLSR, SVM 和 RF 模型构建与验证

Table 2 Construction and verification of PLSR, SVM, and RF models based on different spectra

光谱	模型	精度指标	有机质/(g · kg ⁻¹)		胡敏素/(g · kg ⁻¹)		胡敏酸/(g · kg ⁻¹)		富里酸/(g · kg ⁻¹)	
			建模	预测	建模	预测	建模	预测	建模	预测
VIS-NIR	PLSR	R ²	0.80	0.77	0.83	0.81	0.82	0.81	0.85	0.85
		RMSE	5.17	5.73	4.33	4.37	0.53	0.67	0.37	0.40
		RPIQ	2.39	2.37	2.73	2.72	2.92	2.87	3.46	2.81
	RF	R ²	0.83	0.81	0.93	0.92	0.93	0.92	0.80	0.76
		RMSE	4.14	3.96	2.44	2.66	0.34	0.34	0.35	0.36
		RPIQ	3.36	2.36	4.87	4.61	4.75	3.93	4.00	3.10
	SVM	R ²	0.75	0.73	0.80	0.77	0.71	0.70	0.83	0.80
		RMSE	4.59	5.74	4.11	4.15	0.62	0.65	0.32	0.39
		RPIQ	2.70	2.37	2.96	2.86	3.18	2.95	4.09	2.90
MIR	PLSR	R ²	0.78	0.75	0.82	0.81	0.74	0.74	0.90	0.88
		RMSE	5.42	5.55	4.41	4.45	0.60	0.66	0.30	0.40
		RPIQ	2.95	2.42	2.68	2.67	2.67	2.56	4.38	2.82
	RF	R ²	0.87	0.86	0.89	0.87	0.88	0.86	0.94	0.94
		RMSE	3.87	2.51	2.64	3.79	0.42	0.33	0.22	0.24
		RPIQ	3.89	1.91	4.51	2.75	3.62	4.56	6.08	4.41
	SVM	R ²	0.80	0.78	0.78	0.77	0.74	0.74	0.80	0.76
		RMSE	5.04	5.24	3.94	4.40	0.62	0.71	0.36	0.44
		RPIQ	2.46	2.59	2.99	2.70	2.49	2.71	3.64	2.59
VIS-NIR-MIR	PLSR	R ²	0.80	0.78	0.85	0.81	0.83	0.80	0.86	0.80
		RMSE	4.78	5.01	3.63	4.35	0.53	0.67	0.39	0.71
		RPIQ	2.79	2.71	3.25	2.73	2.83	2.87	3.63	2.07
	RF	R ²	0.92	0.90	0.88	0.87	0.89	0.89	0.87	0.87
		RMSE	2.90	4.08	2.96	2.84	0.41	0.33	0.31	0.31
		RPIQ	4.52	3.51	4.14	2.63	3.65	4.58	3.99	4.28
	SVM	R ²	0.86	0.85	0.87	0.85	0.84	0.83	0.85	0.83
		RMSE	3.81	4.07	3.60	3.69	0.44	0.52	0.30	0.41
		RPIQ	3.25	3.34	3.28	3.22	3.68	3.67	4.28	3.44

酸建立的模型精度最高, 决定系数 R^2 均大于 0.80, 最高可达到 0.94; 土壤有机质、胡敏素及胡敏酸建立的模型精度 R^2 都能大于 0.70, 最优土壤胡敏素及胡敏酸建模精度 R^2 均为 0.93; 土壤有机质的建模精度相对较差, 最大建模精度

R^2 为 0.92。在 PLSR, SVM 和 RF 三种模式建立的预测模型中, 以 RF 模式建立的模型精度明显高于以 PLSR 和 SVM 建立的模型精度, 模型精度 R^2 均大于 0.80; PLSR 模式建立的模型中有两个模型的精度 R^2 小于 0.80, 总体来说, PLSR

模型精度也可较好的预测土壤有机质及其组分含量；SVM 模型精度最差，模型精度 R^2 最低的仅为 0.71。在 VNIR、MIR 和 VNIR-MIR 三种光谱数据建立的模型中，土壤有机质以 VNIR-MIR 建立的模型精度最高，土壤胡敏素与胡敏酸以 VNIR 建立的模型精度最高，土壤富里酸以 MIR 建立的模型精度最高。

综上所述，土壤有机质在基于 VNIR-MIR 光谱的 FR 模型预测精度最高， $R^2=0.92$ ，RMSE=2.90，RPIQ=4.52；土壤胡敏素以及胡敏酸在基于 VNIR 光谱的 RF 模型中精度最高，胡敏素预测精度为 $R^2=0.93$ ，RMSE=2.44，RPIQ=4.87，胡敏酸预测精度为 $R^2=0.93$ ，RMSE=0.34，RPIQ=4.75；土壤富里酸以 MIR 的 RF 模型的预测精度最高，模型精度 $R^2=0.94$ ，RMSE=0.22，RPIQ=6.08。

土壤有机质及组分在模型精度验证中均以 RF 模型反演精度最高，除 VNIR-MIR 光谱 RF 模型预测富里酸时，模型决定系数 R^2 小于 0.8，其他 RF 模型 R^2 均大于 0.8。而 PLSR 模型与 SVM 模型相比，PLSR 模型反演模型精度整体

上高于 SVM 模型的反演模型精度，但在 VNIR-MIR 光谱模型中，SVM 模型精度大于 PLSR 模型精度。9 种模型的模型验证精度均小于模型预测精度。反演模型精度最高的为富里酸，模型 R^2 可达 0.94，RMSE 为 0.24，RPIQ 为 4.41。胡敏素和胡敏酸的反演模型 R^2 均为 0.92，RMSE 分别为 2.66 和 0.34，RPIQ 分别为 4.61 和 3.93。有机质模型预测精度低于各组分模型精度， R^2 为 0.90，RMSE 为 4.08，RPIQ 为 3.51。

2.4 土壤有机质单一光谱模型及组合模型的精度对比

土壤有机质由胡敏素、胡敏酸及富里酸 3 个组分组成，这三种组分含量之和等于有机质含量。筛选出上文中土壤有机质组分最优模型的预测结果，将各组分含量预测结果之和与土壤有机质实测值建立线性模型，图 3 为土壤有机质单一模型和组合模型与土壤有机质实测值的散点图。有机质组合模型决定系数 R^2 为 0.93，大于土壤有机质单一光谱模型精度 ($R^2=0.90$)。

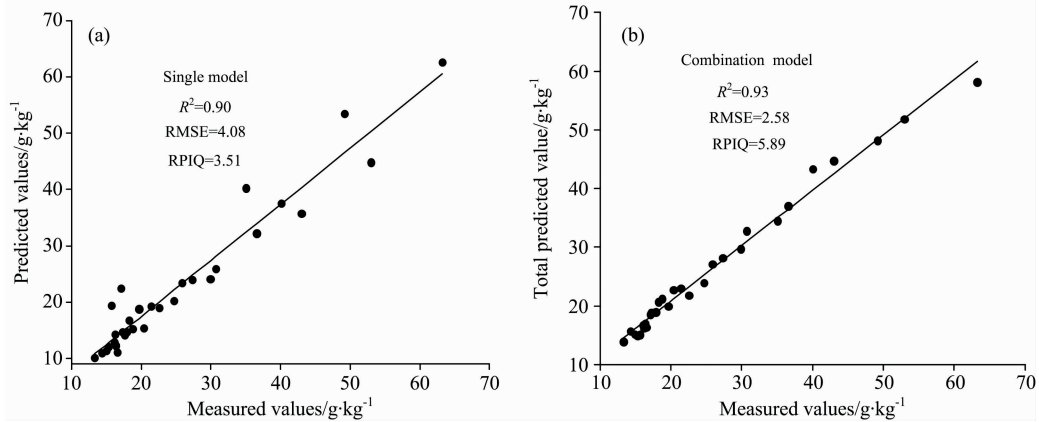


图 3 土壤有机质组合模型、单一模型与有机质实测值散点图

Fig. 3 Scatter plots of predicted vs. measured values for single model (a) and combination model (b)

土壤有机质的主要成分为腐殖质，根据其溶解性可分为胡敏素、胡敏酸和富里酸。其中，富里酸易溶于酸碱溶液，在土壤中最易被分解，胡敏素是在所有 pH 水平下仍不溶于水溶液的萃取部分，在土壤中难以分解，且与疏水性有机化合物紧密结合，并在其结构中呈现出不变的生物聚合物，如木质素和多糖，提取过程复杂，对土壤肥力性质的研究造成了一定的困难。光谱技术是应用于土壤属性测定的一种新的技术手段，已被证明能够有效的分析 SOM。Gu 等^[14]发现，土壤有机质含量大于 2% 会对土壤光谱影响较大，可能忽略其他因素的影响；当小于 2% 时，随着土壤有机质含量降低，对光谱的影响能力逐渐减小，也说明了土壤有机质含量越大，光谱反射率越小。Dinesh 等^[15]基于中红外光谱，建立 MNIRS-PLSR 模型，预测了再造林地土壤的总有机碳等指标，结果表明总有机碳的验证精度 R^2 可达到 0.96。现阶段大部分研究主要是对有机质的物理性质等方面进行研究，对有机质组分研究的报道还很少见。本研究分别做了有机质与其组分的预测模型，在有机质组分层面对土壤肥力更为深入

的研究发现，尽管在近红外光谱中，土壤有机质与光谱反射率的相关性最高，相关系数最大，但在实际建模预测过程中，有机质的预测精度指标要小于其组分的预测精度，富里酸的模型反演精度最好，预测精度 R^2 可达到 0.94，胡敏素与胡敏酸的模型反演精度 R^2 也可达到 0.92，均大于有机质的模型反演精度，说明有机质模型反演并不能完全反映出土壤肥力的高低，有机质组分的矿质化程度也会影响土壤有机质的预测精度。Machado 等^[16]研究发现，土壤有机质组分中胡敏酸的反射率会掩盖土壤反射率和土壤颜色等信息，当土壤中胡敏酸的含量增大时，土壤有机质的反射率会降低，与本研究得出的结论一致。此外，对比了胡敏素、胡敏酸和富里酸的有机质组合模型与有机质单一光谱模型的预测精度，有机质组合光谱模型预测精度高于有机质单一光谱模型，表明这种组合建模方法即可精确预测有机质各组分的含量，同时也提高了有机质的预测精度，为提高光谱预测有机质的模型精度，更精准的反映土壤肥力状况提供了新思路。

3 结 论

为探讨土壤有机质组分对土壤有机质光谱预测的影响,大范围尺度的更精准的研究土壤肥力状况,通过对南疆不同地区土壤有机质、胡敏素、胡敏酸以及富里酸进行反射光谱曲线及相关性分析,基于 VNIR, MIR 和 VNIR-MIR 三种不同光谱数据,结合 PLSR, SVM 和 RF 三种模型,建立基于不同光谱数据的土壤有机质及组分的预测模型,筛选最优模型,得出以下结论:

(1)土壤中有有机质及组分含量越大,土壤光谱反射率越小。VNIR 光谱中,在 1 400, 1 900 和 2 200 nm 波长附近出现吸收谷;MIR 光谱中,在 3 500, 2 520, 1 800 和 2 500 cm^{-1} 波长附近出现波谷,在 2 900, 2 200 以及 1 200 cm^{-1} 波长附近出现波峰。

(2)不同土壤有机质及组分与土壤光谱反射率相关性在 VNIR 和 MIR 光谱中均为负相关,土壤有机质及其组分在

VNIR 中相关性最高的是土壤有机质,最大相关系数为 -0.799;在 MIR 中土壤富里酸的相关性最高,相关系数为 -0.836。

(3)对各模型精度进行比较及分析,土壤富里酸模型精度最高,MIR-RF 模型精度 R^2 可达到 0.94, RMSE 为 0.24, RPIQ 为 4.41,胡敏素和胡敏酸最优预测模型为 VNIR-RF 模型,决定系数 R^2 可达到 0.92, RMSE 分别为 2.66 和 0.34, RPIQ 分别为 4.61 和 3.93。预测精度相对较差的是土壤有机质,最优模型为 VNIR-MIR-RF 模型,反演精度 R^2 为 0.90, RMSE 为 4.08, RPIQ 为 3.51。

(4)基于胡敏素、胡敏酸和富里酸的有机质组合光谱模型决定系数 R^2 为 0.93, RMSE 为 2.58, RPIQ 为 5.89,明显大于有机质单一光谱模型预测精度。本研究表明不同土壤光谱反射率能够快速简便准确的预测出土壤有机质及组分含量,实现迅速鉴定土壤肥力,对南疆地区土壤肥力的监测及改良提供一些必要理论和技术,也对南疆地区土壤肥力的治理具有一定的指导意义。

References

- [1] Peng J, Biswas A, Jiang Q S, et al. *Geoderma*, 2019, 337: 1309.
- [2] Hayes M H B, Mylotte R, Swift R S. *Advances in Agronomy*, 2017, 143: 47.
- [3] Vasat R, Kodesova R, Boruvka L, et al. *Geoderma*, 2014, 232.
- [4] Jiang Q H, Chen Y Y, Guo L, et al. *Remote Sensing*, 2016, 8(9): 755.
- [5] CHEN Si-ming, ZOU Shuang-quan, MAO Yan-ling, et al(陈思明, 邹双全, 毛艳玲, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(3): 912.
- [6] Bao N, Wu L X, Ye B Y, et al. *Geoderma*, 2017, 288: 47.
- [7] Wang S Q, Li W D, Li J, et al. *Soil Science*, 2013, 178(11): 626.
- [8] Peng J, Ji W J, Ma Z Q, et al. *Proximal Soil Sensing*, 2016, 152: 94.
- [9] Nowkandeh S M, Noroozi A A, Homae M. *Environmental Development*, 2018, 25: 23.
- [10] Zhao C H, Gao B, Zhang L J, et al. *Infrared Physics and Technology*, 2018, 95: 61.
- [11] Zhang Y, Sui B, Shen H, et al. *Computers and Electronics in Agriculture*, 2019, 160: 23.
- [12] Clark R N, King T V V, Klejwa M, et al. *Journal of Geophysical Research*, 1990, 95(12): 653.
- [13] Bernier M H, Levy G J, Fine P, et al. *Geoderma*, 2013, (209-210): 233.
- [14] Gu X H, Wang Y C, Sun Q, et al. *Computers and Electronics in Agriculture*, 2019, 167: 105053.
- [15] Dinesh B M, Baldock J A, Read Z J, et al. *Journal of Environmental Management*, 2017, 193: 290.
- [16] Machado W, Franchini J C, Guimaraes M G, et al. *Heliyon*, 2020, 6: e04078.

Inversion of Soil Organic Matter Fraction in Southern Xinjiang by Visible-Near-Infrared and Mid-Infrared Spectra

LUO De-fang¹, PENG Jie^{1*}, FENG Chun-hui¹, LIU Wei-yang¹, JI Wen-jun², WANG Nan³

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

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

3. College of Environment and Resources, Zhejiang University, Hangzhou 310058, China

Abstract Soil organic matter is the material basis of soil fertility, and its fraction is an important indicator to evaluate soil fertility. Soil organic matter fractions can be divided into humin (HM), humic acid (HA) and fulvic acid (FA) according to their solubility. The fertility characteristics of different fractions are significantly different. Therefore, the data of soil organic matter fractions can reflect the status of soil fertility more comprehensively and objectively. The traditional determination of soil organic matter and its fractions is complex, inefficient and time-effective. Many studies show that hyperspectral technology can effectively improve the detection efficiency of soil properties and reduce the testing cost, but the reports on the detection of soil organic matter fractions by visible-near infrared and mid infrared spectroscopy are rare. 93 soil samples were collected and analyzed to acquire the content and spectral information of SOC, HM, HA in Aksu and Hetian, southern Xinjiang, and to explore further the feasibility of mid-infrared spectroscopy and visible near infrared-mid infrared combined spectroscopy in detecting soil organic matter fractions and to comparing the prediction accuracy of a single spectral model for organic matter with that of a combined spectral model for different soil organic matter fractions. Secondly, three kinds of spectral data sets of visible near-infrared (VNIR), mid-infrared (MIR) and their combined spectra (VNIR-MIR) were used to analyze and predict the contents of soil organic matter, HM, HA and FA by using three modeling methods of partial least squares (PLSR), support vector machine (SVM) and random forest (RF). The results show that: (1) soil organic matter and its fractions had a good correlation with spectral reflectance, and the number of characteristic bands of soil organic matter and its fractions in Mir was significantly more than that in VNIR. (2) The optimal prediction model of organic matter is VNIR-MIR-RF with R^2 of 0.90; the optimal prediction model of HM and HA is VNIR-RF model with R^2 of 0.92; the optimal prediction model of FA is VNIR-RF model with R^2 of 0.94. (3) The prediction accuracy of the organic matter combination spectral model based on HM, HA and FA is significantly higher than that of the single spectral model. The R^2 of the two models is 0.93 and 0.90, respectively. The results of this study realized the efficient and rapid inversion of soil organic matter fractions, and the combined model based on organic matter fractions improved the prediction accuracy of soil organic matter and provided important reference value for large-scale soil fertility identification and precision fertilization in southern Xinjiang.

Keywords Soil organic matter; Spectral reflectance; Partial least squares regression (PLSR); Support vector machine regression (SVM); Random forest regression (RF); Inversion model

(Received Sep. 4, 2020; accepted Dec. 28, 2020)

* Corresponding author