## 可见-近红外光谱原位估测土壤有机质含量中 测试点和采样位置导致的误差分析

张昊丹1,孙孝林1,2\*,王晓晴1,王会利3

1. 中山大学地理科学与规划学院,广东省城市化与地理环境空间模拟重点实验室,广东广州 510275

2. 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 江苏南京 210008

3. 广西壮族自治区林业科学研究院, 广西南宁 530002

摘 要 由于土壤的非均质性,同一样品的多个光谱测试点位置之间以及同一土壤发生层的多个样品之间, 均存在着不同程度的光谱差异,进而导致光谱估测的土壤属性具有误差。截至目前,还没有揭示这种误差的 研究报道。在我国南方典型丘陵区林地的小区内,用环刀重复2次采集了80个土壤剖面发生层和38个土壤 表层的样品,再分别用原位光谱法和传统实验室方法测定土壤有机质含量,从而分析因光谱测试点和样品 采集位置不同导致的土壤有机质含量估测误差。结果表明:每个样品在共计18个测试点位置上的光谱差异 为0.12°~8.13°,均值为1.55°;在共计2个重复采样位置之间的光谱差异为0.18°~3.65°,均值为0.88°。 因测试点和采样位置不同导致的土壤有机质估测误差用标准偏差表示,分别为0.92~14.66和0.0057~ 11.46g•kg<sup>-1</sup>,占实测有机质的3.8%~428%和0.017%~92%。同时,这两种误差会随着实测有机质值含 量的增大而增加。并且,测试点位置不同导致的误差比偏最小二乘回归模型所导致的误差还要大。研究认 为,今后的研究中需注意测试点和采样位置不同引起的误差,尤其是有机质含量较高的土壤。

关键词 可见-近红外光谱; 土壤有机质; 偏最小二乘回归; 光谱差异分析; 误差分析 中图分类号: O433.4 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2020)11-3499-09

## 引 言

土壤是一种复杂综合体,其反射光谱是由其组成成分及 结构等多种内在性质共同决定的结果<sup>[1-3]</sup>。因此,土壤光谱 作为土壤理化性状的综合指标,在研究中常用于估测土壤属 性,特别是可见-近红外(visible-near infrared, Vis-NIR)光 谱。与传统化学分析土壤属性相比,室内测定土壤光谱虽已 大大提高了效率,但仍需对样品进行预处理(如风干、研磨、 过筛等)。相较之下,原位测定土壤光谱的效率更高,还具有 快速、无损、可多次重复等优点。因此,越来越多的研究致 力于发展原位土壤光谱估测土壤属性的技术,并取得了一定 进展。例如,Viscarra Rossel 等<sup>[4]</sup>、Li等<sup>[5]</sup>使用原位土壤 Vis-NIR光谱估测土壤颜色、矿物组成、粘粒含量和有机 碳。

这些研究也指出,利用原位土壤光谱估测土壤属性的准 确性受诸多因素影响,包括两方面。一是土壤方面,主要是

土壤非均质性所导致的水分条件、粒径差异、质地条件等; 二是光谱测定方面,主要有光谱分辨率、测量时的外部环 境、几何条件、土样表面处理方法、光谱处理技术等。前一 方面的影响因素与室内测定土壤光谱影响因素的研究结果都 较一致。史舟等印已对此做了比较全面、系统的总结。后一 方面的因素在很多光谱估测土壤属性的研究中得到了深入分 析。例如,刘焕军等研究了光谱分辨率对黑土有机质预测模 型的影响,结果表明:黑土有机质含量高,土壤有机质的光 谱范围宽(445~1 380 nm),且光谱预测模型精度随光谱分 辨率降低呈先增后减的趋势,最优模型的光谱分辨率为50 nm; 侯燕平等<sup>[7]</sup>研究了土样表面处理方法对光谱测定的影 响,具体表现为表面刮平处理方式优于压平与摇平处理方 式; Stenberg<sup>[8]</sup>研究了实验室内土壤样品预处理和标准化复 湿对 Vis-NIR 光谱预测粘粒和土壤有机碳的影响,发现将土 壤样品重新润湿到体积标准化水平可显著提高土壤有机碳的 估测效果,粘粒含量估测效果也有改进但不如前者明显。

为了削弱土壤的非均质性对原位土壤光谱估测土壤属性

#### 收稿日期: 2019-11-13,修订日期: 2020-04-07

作者简介:张昊丹,1993年生,中山大学地理科学与规划学院硕士研究生 e-mail: zhanghd25@mail2. sysu. edu. cn

基金项目:国家自然科学基金项目(41771246),土壤与农业可持续发展国家重点实验室基金项目(Y20160004)资助

的准确性的影响,原位土壤光谱的测定一般是多点测定后取 平均值,再用于土壤属性估测。如 Morgan 等<sup>[9]</sup>在土芯上每 隔 2.5 cm 各测定 4 个位置的光谱,并取平均值用于估测有 机碳和无机碳; Lobsey 和 Viscarra Rossel<sup>[10]</sup>测定了每个环刀 土样两面各 5 个位置的光谱并取平均值用于土壤水分估测; Li等<sup>[5]</sup>测定了环刀样品上 5 个随机位置的光谱后取平均值, 用于土壤有机碳分析。此外,广泛的土壤学研究中,一般对 同一层的土壤,如土壤发生层或同一深度层,选取多个位置 的样品充分混合。因此,同一层的不同采样位置也会影响原 位土壤光谱估测土壤属性的准确性。然而,迄今的大量研究 主要集中在数据预处理与预测模型方面<sup>[11]</sup>,鲜有研究分析 原位土壤光谱测试点位置以及同一土层不同采样位置对土壤 属性估测的影响。因此,这两个因素导致的光谱差异及土壤

为此,以我国南方典型丘陵区林地上的土壤样品为例, 首先分析同一原状土样品不同测试点位置上以及同一土层不 同采样位置上的光谱差异,再按照原位土壤光谱估测土壤属 性的一般模式,用偏最小二乘回归(partial least square regression, PLSR)方法建立样品的平均光谱与有机质含量之 间的模型,接着使用该模型估测不同光谱测试点位置和采样 位置上的有机质含量,从而评价测试点和采样点位置不同导 致的有机质含量估测误差。研究目的在于定量揭示光谱测试 点位置、土壤采样点位置的不同导致的土壤属性估测误差, 为后续进一步研究如何降低这些误差指明方向,有助于未来 更好地开展原位土壤光谱估测土壤属性的研究和应用。

1 实验部分

#### 1.1 研究区概况

研究区选在广西壮族自治区南宁市北郊丘陵地带的高峰 林场(108°20′57″—108°21′54″E,22°57′8″N—22°58′41″N), 面积约3.03 km²,高程约120~300 m。该研究区位于南亚热 带湿润季风气候区,年均温为21.6 ℃,年降雨量1304.2 mm;土壤母质主要为泥岩、泥质页岩、砂质页岩等沉积岩 系,均风化发育为赤红壤;植被以人工桉树林为主,林下灌 草植物以木姜(Litsea pungens)、毛桐(Mallotus barbatus)、 盐肤木(Rhus chinenesis)、半边旗(Pteris semi pinnata)、五节 芒(Miscanthus floridulus)、铁芒箕(Dicranopteris dichotoma)等为优势树种。该研究区历史上一直为林地,21世纪初 开始种植桉树<sup>[12]</sup>。

#### 1.2 土壤采样

从研究区的数字高程模型中提取出高程、坡向、坡度、 剖面曲率、平面曲率和地形湿度指数,并在这些地形参数的 基础上,采用条件拉丁超立方方法选取了 20 个不同地形特 征的采样点。在这些采样点位置上,挖掘土壤剖面,深至母 质层,当母质层深度超过 140 cm 时深至 140 cm。每个剖面 划分发生层后,用环刀在每层的不同位置重复采样 2 个,共 采集 160 个环刀样品。另外,在该研究区内用网格法选取 38 个样点,每个样点重复采集表层不同位置上的环刀样 2 个, 共采集 76 个环刀样品。因此,共采集 236 个环刀样品。



Fig. 1 Distribution of sampling points in the study area

#### 1.3 土壤测试

原状土壤光谱由美国 ASD 公司生产的 FieldSpec4 型便 携式高分辨率地物光谱仪在室内测得,测量光源为杯状光 源,内置12°天顶角的卤素灯,配有接触式光纤探头。该仪器 波长范围是 350~2 500 nm, 350~1 000 nm 内采样间隔为 1.4 nm, 1 001~2 500 nm 内采样间隔为 1.1 nm, 光谱分辨 率为 3 nm@700 nm, 6 nm@1 400/2 100nm, 输出总波段数 是 2 151。测量时,将杯状光源置于环刀土壤样品上方,光纤 探头直接接触样品。每个样品的上、下两面分别均匀洗取 9 个点进行测量,每个点测得10条光谱。由于光谱在起始波段 (350~400 nm)和长波近红外波段(2 451~2 500 nm)受杂散 光、样品背景、测量仪器系统等因素的影响而混有噪声,故 本文用 401~2 450 nm 波段范围的光谱数据进行下一步分 析<sup>[5]</sup>。同时,为避免因数据冗余出现过度拟合现象,参考 Shepherd<sup>[13]</sup>和 Stenberg<sup>[8]</sup>的研究, 分别以 5 和 10 nm 采样间 隔对 401~1 000 和 1 001~2 450 nm 两个波段进行重采样, 共输出 265 个波段。

光谱测定后的环刀样品经过风干、研磨等一系列处理 后,用重铬酸钾容量法测得有机质含量。

#### 1.4 土壤光谱曲线之间的差异计算

不同的光谱曲线具有不同的形状和幅度。本文用光谱角 度 θ<sup>[1]</sup>来评价它们之间的差异性,计算公式如式(1)

$$\theta = \arccos\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}\right) \frac{180}{\pi}, \ \theta \in [0, \ 90]$$
(1)

式(1)中, $x_i$ 和 $y_i$ 分别为两个光谱曲线x和y在波段i处对应的反射率值,n表示波段数。 $\theta$ 值越小,表明曲线x和y之间的差异越小,反之则差异越大。

为消除仪器因素引起的误差,以每个测试点10条光谱 的平均作为该测试点的光谱,并称为测试点光谱(共4248 条),以便区分其他光谱。进一步地,计算每个环刀样品上、 下两面共18个测试点光谱的平均值,称为环刀样光谱(共 236条);计算每个土层两个环刀样品共36个测试点光谱的 平均值,称为土层光谱(共118条)。然后,计算每个环刀样 品18个测试点光谱与对应环刀样光谱之间的光谱角度,以 评价光谱测试点位置不同引起的光谱测定差异。相似地,计 算每个土层2个环刀样光谱与对应土层光谱之间的光谱角 度,以评价采样位置不同引起的光谱测定差异。

#### 1.5 土壤光谱估测土壤属性的模型建立与验证

使用 PLSR 建立土壤光谱与土壤属性之间的模型。 PLSR 在土壤近地传感研究中应用最广泛,应用效果良 好<sup>[3-5]</sup>。为避免过度拟合 PLSR 模型,使用留一交叉验证法 (leave-one-out cross validation, LOOCV)选择最合适的潜变 量(latent variable, LV)个数来建立 PLSR 模型。然后,再采 用随机独立样本对模型进行验证:全部样本(236个)被随机 分为建模集和验证集。为避免样本数不同给验证结果带来的 影响,验证集的样本数取值为1~40,取剩后的样本全部用 于建模。同时,为了降低随机采样的不均匀性对验证结果的 影响,对每个验证集样本数重复进行了100次,并取100次 的平均结果为该模型的验证结果。验证指标包括:平均偏差 (mean error, ME)、均方根误差(root mean square error, RMSE)、决定系数 R2 和相对分析误差(residual prediction deviation, RPD),计算公式如式(2)—式(5)

$$ME = \frac{1}{n} \sum_{i=1}^{n} (obs_i - pred_i)$$
(2)

$$\text{RMSE} = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{n} (\text{obs}_i - \text{pred}_i)^2}$$
(3)

$$R^{2} = r^{2}, r = \frac{1}{(n-1)} \frac{\sum_{i=1}^{n} (\operatorname{pred}_{i} - \operatorname{pred}) (\operatorname{obs}_{i} - \overline{\operatorname{obs}})}{\sqrt{\sigma_{\operatorname{pred}}^{2}} \sqrt{\sigma_{\operatorname{obs}}^{2}}}$$
(4)  

$$RPD = \frac{SD_{y}}{RMSE}$$
(5)

式中,obs;为采样点*i*上的实测值,pred,是该点上的估测 值, $\sigma_{\text{pred}}^2$ 和 $\overline{\text{pred}}$ 分别表示估测值的方差和均值, $\sigma_{\text{obs}}^2$ 和 $\overline{\text{obs}}$ 分别 表示观测值的方差和均值,SD,为样本观测值的标准差。ME 指示估测值的总体误差;RMSE 受误差的影响较大,因而能 反应估测值的精密度; R<sup>2</sup> 指示估测值与实际值的拟合优度; RPD 是样本标准差与均方根误差 RMSE 的比值。ME 和 RMSE 越小而 R<sup>2</sup> 越大,则模型的误差越小,准确度越高。此 外,当 RPD>2 时,模型的预测能力极佳;当 1.4<RPD<2 时,模型的预测能力较为一般,可做粗略估计;当 RPD<1.4 时,模型的预测能力较差,无法对样本进行有效预测<sup>[14]</sup>。

#### 1.6 误差分析

将每个环刀样品的 18 条测试点光谱输入到 PLSR 模型 中,得到 18 个有机质含量估测值。在此基础上,计算这些估 测值的平均值、标准偏差和变异系数。用其中的标准偏差表 示因测试点位置不同导致的光谱估测土壤有机质的误差。

同样的,将每个土层的2条环刀样光谱输入到PLSR模型中,得到2个有机质含量估测值,接着计算这些估测值的 平均值、标准偏差和变异系数,并用其中的标准偏差表示因 采样位置不同而导致的光谱估测土壤有机质的误差。

### 2 结果与讨论

#### 2.1 采样土壤的有机质含量统计特征

表1列出了采样获得的土壤有机质含量的统计特征。所 有236个环刀样的土壤有机质含量的变异性较大,变异系数 达61%,属中等变异。偏度与峰度值说明样品有机质含量呈 向右偏离正态分布,且峰态平缓。

表1同时还列出了每个土层2个重复样的土壤有机质含 量统计特征;图2展示了它们之间的对比。表1中的结果表 明,重复样本之间的统计特征差别很小,且都接近所有环刀 样的统计特征。例如,两次重复的均值分别为20.68和 19.81g•kg<sup>-1</sup>,标准偏差分别为12.23和12.51g•kg<sup>-1</sup>, 与所有环刀样本的均值20.25g•kg<sup>-1</sup>和标准差12.36g• kg<sup>-1</sup>都很接近。由图2可见,当有机质含量低于25g•kg<sup>-1</sup> 时,每个土层的两次重复实测有机质含量接近1:1线,而当 有机质含量高于25g•kg<sup>-1</sup>时,两次重复与1:1线有一定 程度的偏离。这种偏离说明,高有机质含量的土层中,土壤 有机质在层内分布不均匀,例如有机质含量丰富的土壤表层 过渡到有机质含量较低的下层时,有机质含量并不均匀,而 是快速或慢速降低。这一结果说明研究土壤采样位置不同导 致的误差具有重要意义。

表 1 土壤有机质含量  $(g \cdot kg^{-1})$  的统计特征 Table 1 Statistics of soil organic matter content  $(g \cdot kg^{-1})$ 

					0	0	8		
	样本数	均值	最小值	最大值	中值	标准差	变异系数 CV/%	偏度	峰度
所有环刀样	236	20.25	1.20	48.64	19.12	12.36	61	0.31	-1.15
重复1	118	20.68	2.37	46.87	18.99	12.23	59	0.29	-1.22
重复 2	118	19.81	1.20	48.64	19.12	12.51	63	0.34	-1.09
土层	118	20.25	1.79	45.10	20.32	12.10	60	0.20	-1.36

此外,表1还列出了每个土层的有机质含量,即该土层 2个重复样的平均值的统计特征。可见,土层的有机质含量 统计特征与上述所有环刀样及2个重复样中有机质含量的统 计特征基本一致。

#### 2.2 光谱曲线之间的差异分析

本文以所有样品中有机质含量最低(1.20g•kg<sup>-1</sup>)、最





高(48.64 g•kg<sup>-1</sup>)和最接近平均值(20.57 g•kg<sup>-1</sup>)所对应 土层的样品为例,展示本研究测得的光谱,如图 3 所示。由 于有机质含量的较大差异,图 3(a),(b)和(c)中的光谱反射 率差异较大,表现为有机质含量越高的土壤,其反射率 低<sup>[6]</sup>。不同环刀样品的光谱在 700~1 000 nm 范围内有明显 的波形差异,尤其是图 3(a)中 900 nm 附近可见明显吸收谷。 因 690~930 nm 波段为铁的氧化矿物对光谱的主要吸收 区<sup>[16]</sup>,故该波段反射光谱主要受土壤有机质和氧化铁的共 同影响。本研究采集的土壤类型为赤红壤,含丰富的氧化 铁,尤其是有机质含量少的下层土壤中。因此,在图 3(a)中 表现出了 900 nm 附近的明显吸收谷。

在图 3 中, 灰色和浅绿色分别表示同一土层 2 个环刀样 品各 18 条测试点光谱(共 36 条); 蓝色表示 2 个环刀样光 谱; 红色表示土层光谱。可以看到, 同一环刀样品上不同测 试点位置处的测试点光谱(图 3 中灰色和浅绿色曲线)在形态 上大体一致, 但在不同的波段上, 反射率值存在着不同程度 的差异。较大的差异发生在 600~1 350, 1 400~1 850 和 1 900~2 500 nm 三个波段范围内。在 1 400, 1 900 和 2 200 nm 三个波段附近光谱曲线差异较小, 分别对应于水分吸收 特征明显的波段<sup>[3, 15]</sup>。

图 3 还反映出同一土层不同取样位置之间,即环刀样光 谱(图 3 中蓝色曲线)之间,也存在着一定程度的光谱差异, 并与土层光谱(图 3 中红色曲线)有一定差异。这些差异与上 述测试点光谱之间的差异相比非常小。



(a): 有机质含量最低; (b): 有机质含量最接近均值; (c): 有机质含量最高

Fig. 3 Soil spectral curves of soil horizons with different organic matter contents

(a): Horizon with the lowest soil organic matter content;(b): Horizon with the organic matter content close to average value;(c): Horizon with the highasu soil organic matter content

表 2 列出了测试点光谱与环刀样光谱之间、环刀样光谱 与土层光谱之间角度的统计特征。偏度值和峰度值说明两组 角度的分布比较相近,均向右偏离正态分布,峰态陡峭。前 者的均值、最大值、中值都大于后者,说明测试点光谱与环 刀样光谱之间的差异大于后者。这一结果表明,土壤光谱测 试点不同导致的光谱差异远远大于采样位置不同导致的光谱 差异。前者比后者大约 76%。这可能是因为每一个土壤样品 的 18 个光谱测试点分布在环刀样品的上、下两面(每个面各 9 个点,而环刀深度为5 cm,即上下两面相距5 cm),因而测 试点位置实际上已包含了采样位置不同引起的差异。同时, 环刀样光谱是测试点光谱的平均值,因而已消除了部分差 异,使得环刀样光谱之间差异缩小。另外,环刀样光谱之间 的差异较小也可能与样本数有关,本文中每个土层仅有 2 个 环刀样光谱。

表 2 光谱曲线与参考谱线之间的角度差异统计特征(单位:度) Table 2 Statistics of differences in degree between the sample spectrum and reference spectrum

θ	参考谱线	样本数	均值	最小值	最大值	中值	标准差	偏度	峰度
测试点光谱	环刀样光谱	4 248	1.55	0.12	8.13	1.28	1.00	1.68	6.93
环刀样光谱	土层光谱	236	0.88	0.18	3.65	0.69	0.60	1.50	5.77

图 4 列出了土壤光谱之间的角度与环刀样有机质含量之 间的对比关系。该图反映出,光谱之间的角度与有机质含量 无明显的相关关系。进一步的相关性分析表明,测试点光谱 与环刀样光谱之间的角度与有机质含量的最大负相关和正相 关分别为一0.061 和 0.079,最小负相关和最小正相关分别 为一0.028 和 0.008。环刀样光谱与土层光谱之间的角度与 有机质含量之间的相关性系数为一0.02。可见,这些相关性 并未达到统计学上的显著性水平,说明它们之间的相关性不 明显。



(a): The angle between spectrum of each test point and spectrum of each sample; (b): The angle between spectrum of each sample and spectrum of each horizon

#### 2.3 土壤原位光谱估测有机质含量的 PLSR 模型与验证

用所有(236 个)环刀样光谱和有机质含量建立 PLSR 模型。图 5 显示了这些光谱与有机质含量之间的相关性。可见, 在大部分波段上,光谱反射率与土壤有机质含量之间呈现负 相关,最大负相关系数为-0.918,对应波长为1 215.5 nm; 只在极小波段范围内(2 196~2 226,2 326~2 406 nm)呈正 相关,最大正相关系数为0.163,对应波长为2 205.5 nm。这 一结果与 Shepherd<sup>[13]</sup>的研究结果相似,说明可以合理地利 用光谱建立估测有机质含量的 PLSR 模型。

使用一系列不同潜变量个数的 PLSR 建模和 LOOCV 验证的结果(图 6)表明:当潜变量个数小于 23 时,估测值的误差平方和呈现总体下降趋势,仅在个别潜变量个数上略有微小增加的趋势;当潜变量个数为 23 时,估测值的误差平方和达到最低;当潜变量个数大于 23 时,估测值的误差平方和有微小增加,但并没有太大改变,趋于平稳。因此,使用潜变量个数为 23 来建立 PLSR 模型。LOOCV 的结果表明,该模型的 ME 是-0.002 6 g·kg<sup>-1</sup>, RMSE 是 3.57 g·kg<sup>-1</sup>,分别占表 1 中所有样本有机质含量平均值的-0.013%和 18%。其中,RMSE 的结果说明该模型具有一定误差。该模型的决定系数 R<sup>2</sup> 是 0.92, RPD 是 3.46,与其他研究中的 R<sup>2</sup>





图 5 土壤有机质含量与环刀样光谱反射率的相关性分析 Fig. 5 Correlogram of soil organic matter content to spectral reflectance of each sample



(a):验证:预测集均方根误差(使用 236 个样本进行留一交叉验证);(b):训练:%累积贡献率

#### Fig. 6 Modeling results of PLSR

(a): Validation: RMSEP (Cross-validated using 236 leave-one-out segments);(b): Training: % variance explained

接着用随机抽取的 1~40 个样本对该模型进行了验证, 每个随机抽取样本数重复 100 次并取其平均值作为验证结 果。如图 7 所示,当随机抽取的样本数大于 3 时,验证指标 值趋于稳定。而且,由于样本数较小时很可能出现较大偏差 而使得评价结果失真,例如图 7(c)中 R<sup>2</sup> 值在样本数为 1 时 其值为 1,显然是因为样本数仅为 1 造成的。因此,以随机抽 取的样本数大于 3 时的结果来评价模型的准确性。ME 在 -0.52~0.34 g•kg<sup>-1</sup>之间,占表 1 中所有样本平均值的 -2.57%~1.68%,可见总体误差非常小。然而,这可能是





因为正、负误差相互抵消的原因。相对应地,图 7(b)中的 RMSE 相对较高,在 3.15~3.70 g·kg<sup>-1</sup>之间,占表 1 中所 有样本平均值的 16%~18%。这一结果说明该模型具有一定 的误差。图 7(c)中的决定系数  $R^2$  在 0.91~0.93 之间,高于 Li等统计文献中曾报道的 0.68~0.92 之间。同时,RPD 在 3.38~4.38 之间。这些结果说明,该模型具有较高的准确 性。

### 2.4 测试点和采样位置导致的误差分析

图 8 展示了光谱估测的有机质含量及其平均值、标准偏 差与实测有机质含量的对比关系。从图 8(a)中可以看到,不 同的测试点位置,估测值有较大的变异,说明测试点位置不 同导致的误差比较大。用标准偏差表示该误差的结果也展示 在图 8(a)中,即蓝色点。进一步分析表明,该标准偏差与有 机质含量实测值的相关系数为 0.23,达到了 0.01 水平上的 显著性。这说明,实测值越大,标准偏差越大。同时,表 4 列 出了该标准偏差的统计数据。可以看到,该标准偏差的平均 值、最小值和最大值分别占表 1 中所有样品有机质含量平均 值的 17.98%,4.54%和 72.40%。可见,测试点位置不同导 致的误差并不小。鉴于前述标准偏差与实测有机质含量有显 著的相关性,进一步分析了标准偏差占对应有机质含量实测 值的百分数,统计结果也列于表4中。可以看到,测试点位 置不同导致的标准偏差占实测值的百分数平均可达31%,最 小值为3.8%,而最大值可达428%。可见,该标准偏差比 2.3节中模型的误差还要大(在2.3节中,模型的 RMSE 占 样品有机质含量平均值的16~18%)。

在分析采样位置导致的误差中,取每个土层两个环刀样 实测有机质含量的平均值作为该土层有机质含量的实测值。 图 8(b)表明,同一土层上环刀采样位置的不同也会导致明显 不同的估测值,说明采样位置不同导致的误差比较明显[图 8 (b)中的蓝色点]。进一步的相关性分析表明,用标准偏差表 示的该误差与土层有机质实测值的相关系数为 0.30,达到了 0.01 水平上的显著性。因此,采样位置不同产生的标准偏差 也会随着实测值的增大而增加。表 4 列出了该标准偏差的统 计特征。可见,该标准偏差的平均值、最小值和最大值分别 占表 1 中所有土层实测有机质含量平均值的 11%,0.03%和 57%。因此,由于采样位置不同导致的估测标准偏差也较 大。进一步分析了该标准偏差占对应实测值的百分数,并将 结果列于表 4 中。该结果表明,该标准偏差占实测值的百分 数平均值为 15%。可见, 该标准偏差比 2.3 节中模型的误差 要小(在 2.3 节中, 模型的 RMSE 占样品有机质含量平均值 的 16%~18%)。

图 8 和表 4 的结果都表明,因测试点位置不同导致的标准偏差明显大于因采样位置不同导致的标准偏差。以表 4 中 两者的平均值来看,前者比后者大约 60%。这与前述土壤光 谱差异有关。如前所述,测试点不同导致的光谱差异比采样 位置不同导致的光谱差异大约76%。因此,在原位光谱估测 土壤属性的研究中,更应该注意测试点位置不同导致的误 差。迄今为止的文献大多采用多点测定和多样本采样取平均 值来应对这两种误差,并未定量分析它们的大小。然而,尽 管本文指明了这两种误差的大小,但仍然未能提出有效地降 低它们的方法。未来,除了研究利用不同的模型来降低误差 外,还需要研究更有效的方法来降低这些误差。



图 8 测试点光谱(a)和环刀样光谱(b)估测的有机质含量(黑色)、平均值(绿色)及标准偏差(蓝色)与实测有机质含量的对比; 在(b)中,光谱估测值与实测有机质含量进行对比,而其他值都与每个土层的两个环刀样的平均值进行对比

Fig. 8 Estimated soil organic matter content (black) by the spectrum of each test point (a) and sample (b), and their mean value (green), standard deviation (blue) against the measured. In (b), the estimated are compared with the measured soil organic matter contents, while the other values are compared with the average soil organic matter content of two samples in each horizon

		0			·				
	样本数	均值	最小值	最大值	中值	标准差	变异系数 CV/%	偏度	峰度
测试点光谱的 估测标准偏差	236	3.64	0.92	14.66	3.11	2.09	57	1.90	4.92
测试点光谱的估测标准 偏差占实测值百分比	236	0.31	0.038	4.28	0.18	0.46	149	5.94	46.19
环刀样光谱的 估测标准偏差	118	2.27	0.0057	11.46	1.54	2.10	92	1.54	3.19
环刀样光谱的估测标准 偏差占实测值百分比	118	0.15	0.00017	0.92	0.093	0.18	115	2.34	6.60

表 4 原位土壤光谱估测有机质含量的误差(用标准偏差表示)的统计特征 Table 4 Statistics of the error of soil organic matter content estimated by in situ soil spectra (by standard deviation)

## 3 结 论

以我国南方典型丘陵区林地中的土壤样品为例,定量分 析了光谱测试点和采样位置不同导致原位土壤光谱估测有机 质含量中的误差。结果表明,光谱测试点不同导致的土壤光 谱差异平均为1.55°,而采样位置不同导致的土壤光谱差异 则相对较小,平均为 0.88°。两种差异导致光谱估测土壤有 机质的误差平均值分别为 3.64 和 2.27 g•kg<sup>-1</sup>,分别占对 应实测值的 31%和 15%。并且,前者大于 PLSR 模型导致的 误差。结果表明,测试点和采样位置不同导致原位光谱估测 土壤属性的误差较大,在今后的研究中应当注意它们的影 响。未来研究还需要探讨如何降低这些误差,提高原位光谱 估测土壤属性的准确性。

#### References

[1] SHI Zhou(史 舟). Principle and Method of Soil Surface Hyperspectral Remote Sensing(土壤地面高光谱遥感原理与方法). Beijing:
 Science Press(北京:科学出版社), 2014.

- [2] XIONG Jing-ling, ZHU Xi-cun, GAO Hua-guang, et al(熊静玲, 朱西存, 高华光, 等). Acta Pedologica Sinica(土壤学报), 2018, 55 (6): 1.
- [3] Viscarra Rossel R A, Walvoort D J J, Mcbratney A B, et al. Geoderma, 2006, 131(1-2): 59.
- [4] Viscarra Rossel R A, Cattle S R, Ortega A, et al. Geoderma, 2009, 150(3-4): 253.
- [5] Li S, Shi Z, Chen S, et al. Environmental Science & Technology, 2015, 49(8): 4980.
- [6] JI Wen-jun, SHI Zhou, ZHOU Qing, et al(纪文君, 史 舟, 周 清, 等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2012, 31(3): 277.
- [7] HOU Yan-ping, LÜ Cheng-wen, XIANG Hong-liang, et al(侯燕平, 吕成文, 项宏亮, 等). Chinese Journal of Soil Science, 2015, 46 (2): 287.
- [8] Stenberg B. Geoderma, 2010, 158(1-2): 15.
- [9] Morgan C L S, Waiser T H, Brown D J, et al. Geoderma, 2009, 151(3-4): 249.
- [10] Lobsey C R, Viscarra Rossel R A. European Journal of Soil Science, 2016, 67(4): 504.
- [11] ZHANG Gan-lin, ZHU A-xing, SHI Zhou, et al(张甘霖,朱阿兴,史 舟,等). Progress in Geography(地理科学进展), 2018, 37(1):
   57.
- [12] WANG Xiao-qing, SUN Xiao-lin, WANG Hui-li(王晓晴,孙孝林,王会利). Acta Pedologica Sinica(土壤学报), 2019, 56(4); doi: 10. 117661trxb201808310246.
- [13] Shepherd K D, Walsh M G. Soil Science Society of America Journal, 2002, 66(3): 988.
- [14] Chang C W, Laird D A, Mausbach M J, et al. Soil Science Society of America Journal, 2001, 65(2): 480.
- [15] JI Geng-shan, XU Bin-bin(季耿善, 徐彬彬). Acta Pedologica Sinica(土壤学报), 1987, 24(1): 67.
- [16] SHA Jin-ming, CHEN Peng-cheng, CHEN Song-lin(沙晋明, 陈鹏程, 陈松林). Research of Soil and Water Conservation(水土保持研究), 2003, 10(2): 21.

# Analyzing Errors due to Measurement Positions and Sampling Locations for In Situ Measurements of Soil Organic Matter Using Vis-NIR Spectroscopy

ZHANG Hao-dan<sup>1</sup>, SUN Xiao-lin<sup>1, 2\*</sup>, WANG Xiao-qing<sup>1</sup>, WANG Hui-li<sup>3</sup>

- Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China
- State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
- 3. Guangxi Zhuang Autonomous Region Forestry Research Institute, Nanning 530002, China

Abstract Due to the heterogeneity of soil, there are spectral differences between different measurement positions of a soil sample and among different samples of a soil genetic layer. As a result, an estimated value of a soil property using in situ spectra of soil inevitably has errors. However, these errors have not been revealed so far. In this paper, 80 soil profiles and 38 soil surface samples were collected twice with cutting ring from a small area of woodland in typical hilly areas in southern China. Then, the soil organic matter content was measured by in-situ spectrometry and traditional laboratory method, so as to analyze the estimation error of soil organic matter content caused by different spectral test points and sampling locations. The results show that, the spectral difference of each sample at a total of 18 test points ranges from 0.12° to 8.13°, with an average value of 1.55°. The spectral difference between two repeated sampling locations of each sample is  $0.18^{\circ} \sim 3.65^{\circ}$ , with an average value of  $0.88^{\circ}$ . The estimated error of soil organic matter due to the different positions of test points was  $0.92 \sim 14.66 \text{ g} \cdot \text{kg}^{-1}$ , accounting for  $3.8\% \sim 428\%$  of the measured organic matter content. The estimation error of soil organic matter caused by different sampling locations is 0.005 7  $\sim$  11.46 g  $\cdot$  kg<sup>-1</sup>, accounting for 0.017  $\% \sim$  92 % of the measured organic matter. Moreover, the error caused by the former is larger than that caused by the partial least squares regression model used in this paper, while the error caused by the latter is slightly smaller than that caused by this model. In addition, it is found that these two errors increase with the increase of measured organic matter content. Hence, this paper argues that the errors caused by different test points and sampling locations should be paid attention to in future studies, especially in soils with higher organic matter content, and calls for research on more effective methods to reduce these errors.

3507

**Keywords** Visible-near infrared spectrum; Soil organic matter; Partial least squares regression; Spectral difference analysis; The error analysis

\* Corresponding author

(Received Nov. 13, 2019; accepted Apr. 7, 2020)

## 《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学 共同承办的专业学术期刊。国内外公开发行,从 2004 年起为月刊,大 16 开本,2020 年仍为月刊,每期 332 页。《光谱学与光谱 分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并 培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

#### 栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以 8000 字(包括图表、参考文献、作者姓名、单位和中文、英文摘要, 下同)为宜。

2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过 5000 字。

3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过10000字。

4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过 5000 字。

5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以 3000~4000 字为宜。

#### 稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。

2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以 800 字为宜,英文摘要(建议经专业英语翻译机构润色)与中文摘要要对照;另附关键词。要求来稿应达到"齐、清、定",中文、英文文字通顺,方可接受送审。

3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊**只接收网上在线投稿**,不接收以邮寄方式或 e-mail 方式的 投稿,严禁"一稿多投",对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。

4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须 在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图 7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中 数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用 6 号字。电子文档中除实物图外,曲线图要用 Matlab,Excel, Visio或 Origin 等软件制作,稿件中图片的原图并转成相应(可编辑)的文件格式(.fig,.xls,.vsd,.opj),非".jpg"格式 的文档,随电子版修改稿一同发送到本刊的修改稿专用邮箱。

5. 文中出现的单位必须按"中华人民共和国计量标准"及有关 GB 标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。

6. 名词术语,请参照全国科学技术名词规定缮写。

7. 参考文献,采用顺序编码制,只列主要文献;以15~20条为宜。内部资料、私人通讯、未经公开发表的一律不能引用。 日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。

8. 请在投稿第一页左下角写明投稿联系人的电话和两个 e-mail,以便及时联系。

#### 稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修 改稿在 30 天内寄回编辑部,并作为作者最终定稿(当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动),若二个 月内编辑部没收到修改稿,将视为自行撤稿处理。

2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。

3. 来稿一经发表将酌致稿酬并送样刊2册。

4. 遵照《中华人民共和国著作权法》,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给《光谱学与光谱分析》期 刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一 律认为已获作者授权认可。

5. 修改稿请寄:100081 北京市海淀区学院南路 76 号(南院西北门),《光谱学与光谱分析》期刊社(收)

电话:010-62182998 或 62181070 传真:010-62181070

e-mail: chngpxygpfx@vip. sina.com; 修改稿专用邮箱: gp2008@vip. sina.com 网址: http://www.gpxygpfx.com