

DOI: 10.11766/trxb201904160060

国佳欣, 赵小敏, 郭熙, 徐喆, 朱青, 江叶枫. 基于 PLSR-BP 复合模型的红壤有机质含量反演研究[J]. 土壤学报, 2020, 57(3): 636-645.  
GUO Jiaxin, ZHAO Xiaomin, GUO Xi, XU Zhe, ZHU Qing, JIANG Yefeng. Inversion of Organic Matter Content in Red Soil Based on PLSR-BP Composite Model[J]. Acta Pedologica Sinica, 2020, 57(3): 636-645.

## 基于 PLSR-BP 复合模型的红壤有机质含量反演研究\*

国佳欣<sup>1, 2</sup>, 赵小敏<sup>1, 2†</sup>, 郭熙<sup>1, 2</sup>, 徐喆<sup>1, 2</sup>, 朱青<sup>1, 2</sup>, 江叶枫<sup>1, 2</sup>

(1. 江西农业大学国土资源与环境学院, 南昌 330045; 2. 江西省鄱阳湖流域农业资源与生态重点实验室, 南昌 330045)

**摘要:** 对红壤地区土壤有机质进行快速预测, 以满足智慧农业与精准施肥的需要。以江西省奉新县北部为研究区域, 采用 1 km × 1 km 标准格网划分研究区进行采样, 共得到红壤样本 248 个。对土壤光谱进行了包含分数阶导数在内的 3 种数学变换方法, 将经过  $P=0.01$  显著性检验的波段用于模型的构建, 选用偏最小二乘回归 (PLSR) 和 BP 神经网络建立土壤有机质含量预测模型。结果表明: 当对红壤光谱数据进行 1.5 阶导数变换后再使用 PLSR-BP 复合模型对土壤有机质含量进行预测时的结果为最优, 训练集  $R^2=0.89$ ,  $RMSE=4.68 \text{ g}\cdot\text{kg}^{-1}$ , 验证集  $R^2=0.87$ ,  $RMSE=5.55 \text{ g}\cdot\text{kg}^{-1}$ ,  $RPD=2.75$ 。1.5 阶导数对红壤光谱数据的变换能够更好地突出与有机质相关的特征信息, 有助于其含量预测。PLSR-BP 复合模型预测精度优于单一模型, 能够较好地预测红壤有机质含量, 为精准农业快速监测红壤有机质含量提供了新的途径。

**关键词:** 红壤; 有机质; 分数阶导数; 偏最小二乘回归; BP 神经网络

**中图分类号:** S158.2/TP7 **文献标志码:** A

## Inversion of Organic Matter Content in Red Soil Based on PLSR-BP Composite Model

GUO Jiaxin<sup>1, 2</sup>, ZHAO Xiaomin<sup>1, 2†</sup>, GUO Xi<sup>1, 2</sup>, XU Zhe<sup>1, 2</sup>, ZHU Qing<sup>1, 2</sup>, JIANG Yefeng<sup>1, 2</sup>

(1. College of Land Resources and Environment, Jiangxi Agricultural University, Nanchang 330045, China; 2. Key Laboratory of Poyang Lake Watershed Agricultural Resources and Ecology of Jiangxi Province, Nanchang 330045, China)

**Abstract:** 【Objective】 The purpose of this study is to explore how to rapidly predict soil organic matter in red soil so as to meet the needs of smart agriculture and precision fertilization. 【Method】 This paper took the northern part of Fengxin County in the Northwest Jiangxi Province as its research area and used the 1 km×1 km standard grid method to divide the study area for soil sampling. A total of 248 red soil samples were collected and dried for spectral measurement. Three different mathematical transformation methods, including fractional order derivatives, were used to analyze the soil spectra. In the tests the 350~399 nm and 2 451~2 500 nm bands were removed because they were very susceptible to environmental noises. And noises in the remaining bands were removed with Daubechies (DB) wavelet. Then samples were collected from the pretreated spectral bands at 10 nm intervals to form a 205-band so as to reduce data dimensions and data redundancy. The 800~1 000 nm band, which was

\* 国家重点研发计划项目 (2017YFD0301603)、江西省赣鄱英才“555”领军人才项目 (201295) 资助 Supported by the National Key R&D Program of China (No.2017YFD0301603), the Gan Po“555”Talent Research Funds of Jiangxi Province (No.201295)

† 通讯作者 Corresponding author, E-mail: zhaoxm889@126.com

作者简介: 国佳欣 (1996—), 男, 江西南昌人, 硕士研究生, 主要从事农业遥感与信息技术研究。E-mail: ncguojiaxin@163.com

收稿日期: 2019-04-16; 收到修改稿日期: 2019-07-25; 优先数字出版日期 (www.cnki.net): 2019-09-12

liable to the impact of iron oxide, was ruled out of the experiment. The bands used to construct the model were filtered by the  $P=0.01$  significance test. A model was built up with the partial least squares regression (PLSR) in combination with BP neural network for prediction of soil organic matter content. And the model was tested. 【Result】 Results show that the prediction using the PLSR-BP composite model was the best after the soil spectral data was transformed with the 1.5 order fractional derivative, with  $R^2=0.89$  and  $RMSE=4.68 \text{ g}\cdot\text{kg}^{-1}$  for the training dataset and  $R^2=0.87$ ,  $RMSE=5.55 \text{ g}\cdot\text{kg}^{-1}$  and  $RPD=2.75$  for the validation dataset. 【Conclusion】 The transformation of red soil spectral data with the 1.5 order fractional derivative better highlights characteristics of organic-matter-related information, which is helpful for prediction of organic matter contents. And the PLSR-BP composite model is higher than any single models in prediction accuracy, and can be used to predict organic matter content in red soil very well. So it can also serve as a new approach to predicting quickly organic matter content in red soil for precision agriculture.

**Key words:** Red soil; Organic matter; Fractional order derivative; Partial least squares regression; BP neural network

土壤有机质是指土壤中以任何形式存在的含碳有机化合物，它是土壤的重要组成部分。传统的测量土壤有机质的方法主要是化学分析方法，尽管测定结果可靠，但存在着费时费力的问题，难以满足快速监测土壤有机质含量的需求<sup>[1]</sup>。近些年，土壤高光谱技术的出现为土壤有机质含量的快速预测提供了手段，以其极高的光谱分辨率来获取反映土壤特性的信息，可节省大量的人力物力，也为精准农业提供了重要的监测手段<sup>[2-4]</sup>。

光谱预处理在可见光/近红外光谱分析中具有重要的作用，传统的土壤高光谱的研究主要是采用原始光谱反射率及其 1 和 2 阶导数、倒数和对数等方式对光谱数据进行数学变换来构建土壤有机质预测模型。然而仅采用传统的 1、2 阶导数来对土壤光谱数据处理时，两者相差过大，会导致中间过渡信息的遗漏<sup>[5]</sup>。徐继刚等<sup>[5]</sup>研究汽油近红外光谱时得出分数阶导数的最优结果并不都在整数阶导数处，而是在 0 和 1 阶导数之间或 1 和 2 阶导数之间。随着导数阶数的增加会提高光谱分辨率，同时降低光谱信号的强度<sup>[6]</sup>。Tong 等<sup>[6]</sup>采用分数阶 Savitzky-Golay 导数法 (fractional order Savitzky-Golay derivation, FOSGD) 对近红外光谱模型进行优化，从而提供更好的方法来平衡分辨率和信号强度之间的矛盾。王敬哲等<sup>[7]</sup>在对荒漠土壤有机碳进行预测时，得出经过分数阶导数预处理后的模型精度较整数阶导数有较大提升，其研究结果表明采用分数阶导数对土壤光谱数据进行潜在信息挖掘是可取的。以上研究充分说明了对于土壤光谱数据进行导数变换时，不应拘泥于传统的整数阶导数，也要考虑分数阶导数变

换在数据预处理上的作用。

国内外众多研究表明，运用可见光-近红外波段进行线性的偏最小二乘回归 (partial least-squares regression, PLSR) 和非线性的 BP 神经网络等单一模型建模的研究较多<sup>[8-9]</sup>。洪永胜等<sup>[10]</sup>和 Conforti 等<sup>[11]</sup>通过偏最小二乘回归单一方法构建了土壤有机质含量的预测模型。丁国香<sup>[12]</sup>通过 BP 神经网络方法研究了土壤中的有机质与可见光/近红外光谱之间的关系，并建立相应的预测模型。然而线性模型仅适用于变量较少且具有很大大线性关系的情况，非线性模型在输入变量过多时容易出现“过拟合”现象，因此可以考虑线性模型与非线性模型相结合是否能够提高有机质含量的预测精度。

本文以奉新县北部为研究区，基于 PLSR、BP 神经网络和 PLSR-BP 复合模型，在对红壤光谱数据进行 0~2 阶分数阶导数、倒数的对数和对数的导数等数学变换的基础上，对红壤有机质含量预测模型的建立进行探索，以明确最优数学变换和模型的选取，以期为南方红壤地区土壤有机质含量的快速预测提供参考依据。

## 1 材料与方 法

### 1.1 研究区概况

研究区位于江西省西北部的奉新县北部， $115^{\circ}08' \sim 115^{\circ}40'E$ ， $28^{\circ}68' \sim 28^{\circ}80'N$ ，总面积为  $2.0 \times 10^4 \text{ hm}^2$ 。研究区属中亚热带湿润气候，年平均气温为  $17.3^{\circ}\text{C}$ ，年均降雨量为  $1\ 612 \text{ mm}$ ，年均相对湿度为 79%。如图 1 所示，本文所选研究区域主要包括

奉新县赤岸镇、会埠镇、冯川镇、干洲镇和罗市镇的部分村,是以昌德高速(S40)和两条省道(S308和S226)围成的闭合区域。从DEM图中能够得出,研究区海拔介于31 m至133 m之间,处于平原地带。低海拔区域主要分布在东南部地区,而较高海拔区

域则分布在研究区北部和西部地区。经实地调查及土地利用现状图统计得出,研究区内主要包括林地 $1.1 \times 10^4 \text{ hm}^2$ ,水田 $5.0 \times 10^3 \text{ hm}^2$ ,园地 $1.0 \times 10^3 \text{ hm}^2$ ,其他用地 $3.0 \times 10^3 \text{ hm}^2$ ,分别占研究区总面积的55%、25%、5%和15%。

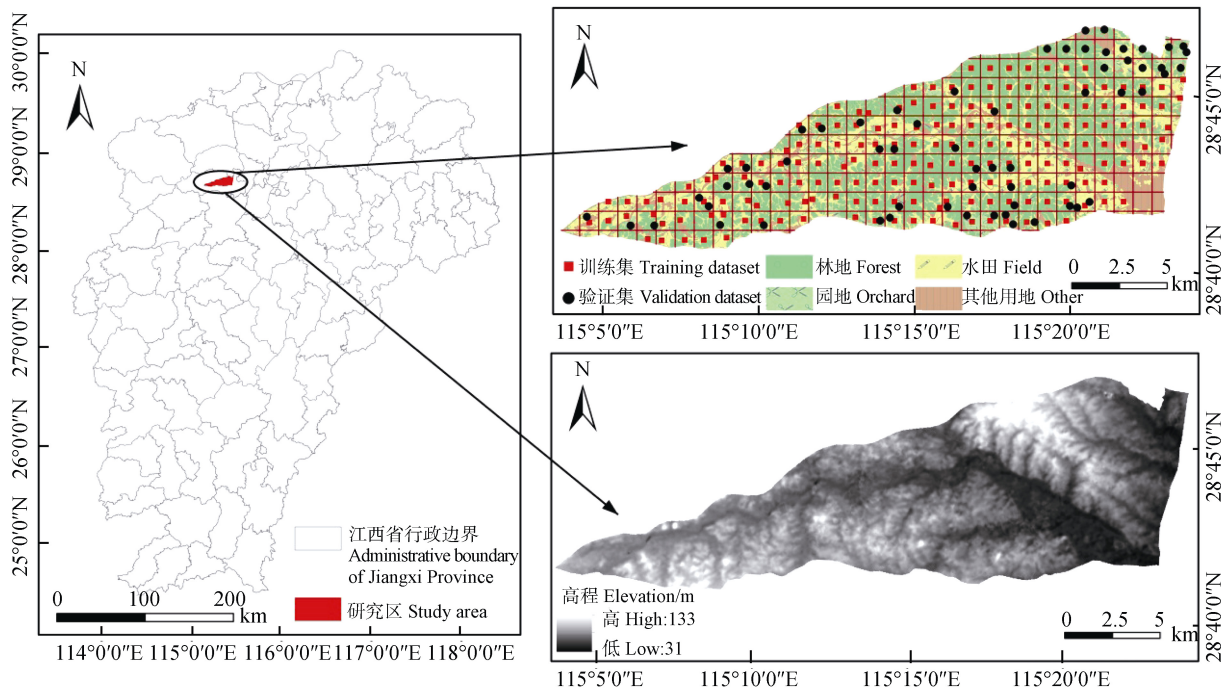


图1 研究区位置、DEM与采样点分布

Fig. 1 Location of the study area and the distribution of DEM and sampling points

## 1.2 土壤样本采集

样本的采集时间为2018年7月23日至8月11日,采用 $1 \text{ km} \times 1 \text{ km}$ 规则格网划分研究区,在各格网内随机选取采样点,综合考虑地势、植被覆盖、土地利用类型及道路可达性,对于个别地理环境较为复杂的区域,进行了采样点的加密,以保证数据的代表性。在采样点附近5 m范围内不同方向选取四个重复样点,混合均匀后,用四分法得到最终样本,剔除其中的植物根系和石块等,同时使用手持GPS仪器读取采样点的位置信息。采集样本覆盖园地、林地、水田三种土地利用类型,其中园地57个、林地93个、水田98个,园地、林地的采样深度为30 cm,水田为20 cm。样本于实验室风干、研磨,过2 mm筛,并将其均分为两部分,分别用于土壤光谱和有机质的测定。土壤有机质含量采用重铬酸钾容量法测定<sup>[13]</sup>。

## 1.3 光谱测量

采用美国ASD FieldSpec4地物光谱仪进行土壤光谱反射率的测量。光谱采集范围为350~2500 nm,光谱采样间隔为1.4 nm(350~1000 nm)和2 nm(1001~2500 nm),重采样间隔为1 nm,共输出2151个波段。在暗室内进行光谱的采集,以避免外界干扰。将土样置于直径6 cm深2 cm的黑色盛样皿中,盛满并用直尺将表面刮平。使用MugLite仪器中自带的内置光源进行测量,将盛样皿置于仪器顶部槽中,每次采集数据之前对仪器进行暗电流和标准白板校正,每个样本采集5条光谱数据,取其算术平均值作为该样本的光谱曲线。

## 1.4 光谱预处理

由于环境和仪器自身的影响会对测量光谱的边缘波段造成较大的噪声,因此去除350 nm~399 nm及2451 nm~2500 nm波段。通过Daubechies6

(DB6) 小波进行三层分解, 采用软阈值法对高频系数进行去噪处理, 去除测量过程中产生的噪声影响<sup>[14-15]</sup>。研究采用 10 nm 间隔进行重采样, 得到由 205 个波段组成的光谱曲线, 以降低数据维数, 减小数据冗余。

除上述预处理外, 本文还对光谱数据进行下述数学变换, 包括光谱反射率 (reflectance, R), 分数阶导数: 0.5 阶导数 (fractional order derivative, FOD (0.5)), 1 阶导数 (FOD (1)), 1.5 阶导数 (FOD (1.5)), 2 阶导数 (FOD (2)), 倒数的对数 (inverse-log reflectance, ILR) 和对数的导数 (log-derivative reflectance, LDR)。已有研究表明这些变换在土壤光谱研究中有广泛的应用, 有助于突出光谱特征, 为有机质反演提供更高的模型精度<sup>[8, 16-17]</sup>。其中分数阶导数变换通过 MatlabR2017b 编程实现。

分数阶导数 (FOD) 目前广泛应用于建模、信号分析等领域<sup>[18-19]</sup>, 有三种主要类型的算法, 分别是 Riemann-Liouville (R-L), Grünwald-Letnikov (G-L) 和 Caputo<sup>[20]</sup>, 其中 G-L 分数阶导数是由整数阶导数的定义推广而来。分数阶导数有助于光谱信息的增强, 在一定程度上减小了噪声对数据的干扰。本文采用 G-L 算法求出分数阶导数:

$$\frac{d^{\nu} f(x)}{dx^{\nu}} \approx f(x) + (-\nu)f(x-1) + \frac{(-\nu)(-\nu+1)}{2} f(x-2) + \dots + \frac{\Gamma(-\nu+1)}{n!\Gamma(-\nu+n+1)} f(x-n) \quad (1)$$

式中,  $\nu$  为阶数,  $\Gamma(x)$  为 Gamma 函数,  $n$  为导数上下限之差<sup>[21-23]</sup>。

### 1.5 模型的建立与评价参数

本文分别采用线性模型 PLSR 和非线性模型 BP 神经网络进行红壤光谱有机质含量估测, 再将 PLSR 与 BP 神经网络模型进行复合, 用于有机质含量预测模型的构建。

PLSR 是目前较为常用的一种线性多元回归分析方法, 它能够分析预测矩阵  $X$  (即自变量) 与响应矩阵  $Y$  (即因变量) 之间的关系, 将初始输入的数据投影到一个潜在的空间中, 利用正交结构提取出大量潜变量, 找出这些新变量与  $Y$  之间的线性关系<sup>[24]</sup>。采用留一法 (Leave-one-out) 交叉验证来确定提取出的潜变量个数, 建立 PLSR 模型<sup>[8]</sup>。

BP 神经网络是人工神经网络中一种应用较为广泛的非线性建模方法, 适用于数据预测<sup>[9]</sup>, 其网络结构由输入层、输出层和隐含层构成。学习过程由前向传播和反向传播两方面组成, 在前向传播过程中, 输入数据由输入层经由隐含层向输出层逐步处理, 如果输出层得到的数据误差不在允许范围内, 则进行误差反向传播, 通过梯度下降法逐层调整各神经元的权重, 直至误差符合指定要求<sup>[25]</sup>。

在利用 BP 神经网络进行建模时, 减少输入变量的数量可以减小数据之间的相关性。因此采用 PLSR 与 BP 神经网络相结合的方法, 将 PLSR 提取出的潜变量作为 BP 神经网络的输入层数据, 这些新变量能反映原变量的绝大部分信息以达到减少数据量降低维度的目的, 从而避免“过拟合”现象的发生<sup>[25]</sup>。

采用决定系数 ( $R^2$ )、均方根误差 (RMSE)、预测偏差比 (RPD) 作为精度评价指标。 $R^2$  和 RPD 越大说明模型越好, RMSE 越小说明预测效果好<sup>[26]</sup>。当  $RPD < 1.5$  时, 表明模型无法对样本进行预测; 当  $1.5 < RPD < 2$  时, 表明模型只能对样本进行粗略预测; 当  $2 < RPD < 2.5$  时, 表明模型有较好的预测样本的能力; 当  $RPD > 2.5$  时, 表明模型有很好的预测样本的能力。式 (2) 中  $n$  为数量,  $y_m$  和  $y_p$  分别为有机质实测值与预测值。式 (3) 中 STDEV 计算的是验证集样本实测值的标准偏差。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_m - y_p)^2} \quad (2)$$

$$RPD = \frac{STDEV}{RMSE} \quad (3)$$

常规数据统计分析软件使用软件 ArcGIS 10.5、IBM SPSS Statistics 22.0、OriginPro 2016 和 Microsoft Excel 2010, PLSR 使用 The Unscrambler X 10.4, BP 神经网络使用 Matlab R2017b。

## 2 结 果

### 2.1 红壤有机质描述性统计特征

获取的 248 个样本, 其有机质含量范围为 5.27~64.00  $\text{g}\cdot\text{kg}^{-1}$ 。由于异常值的存在会影响建模

精度,因此采用拉依达准则<sup>[27]</sup>对样本进行检验,发现数据无异常值。将 248 个土壤样本选用 K-S (Kennard-Stone) 算法<sup>[28]</sup>按照样本间的欧氏距离以 3:1 的比例分为两组,其中训练集包含 186 个样本,

验证集包含 62 个样本,用于模型的精度检验。全集、训练集和验证集的有机质含量均值分别为 33.97、35.04 和 30.78  $\text{g}\cdot\text{kg}^{-1}$ , 变异系数分别为 43.27%、41.07% 和 49.64%, 属于中等强度变异(图 2)。

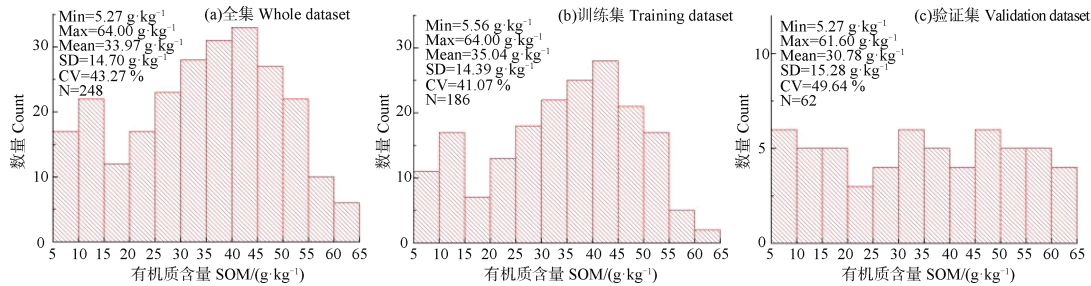


图 2 红壤样本有机质含量描述性统计

Fig. 2 Descriptive statistics of soil organic matter contents in red soil samples

## 2.2 红壤光谱特征

图 3 为经预处理后的红壤光谱曲线,在可见光部分呈陡坎型<sup>[15]</sup>。从中可以发现在 900 nm 左右有较明显的氧化铁吸收谷,因此在建模时去除 800~1000 nm 波段以减小氧化铁对光谱的影响。而在 1400 nm、1900 nm、2200 nm 处有明显的水分吸收谷<sup>[29]</sup>,考虑到样品已经经过了风干处理,对于有机质含量建模影响较小,不作处理。

研究将有机质含量按高低划分为<15、15~25、25~35、35~45、45~55、 $\geq 55$   $\text{g}\cdot\text{kg}^{-1}$  六组,每个组别内求取其光谱曲线的平均值。从图 3 中可以看出,随着有机质含量的增加,在可见光波段内,不同含量的样本光谱曲线相差不大,而在近红外波段,可以看出有机质对光谱的影响较为明显,有机质含量与光谱反射率呈现负相关的现象。500~800 nm 部分数据存在交叉现象,可能是由于在可见光部分土壤反射率数值相接近,平均之后相差不大。

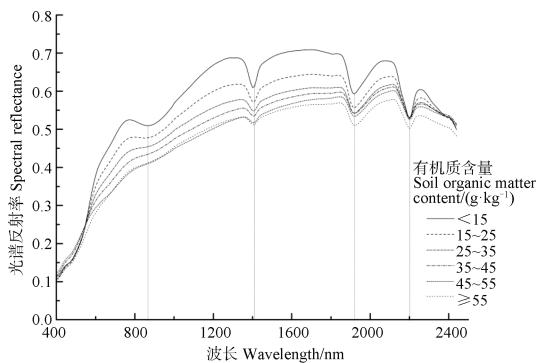


图 3 不同有机质含量红壤光谱曲线

Fig. 3 Spectral curves of the red soil samples relative to content of organic matter

## 2.3 红壤有机质含量 PLSR 建模

运用经过数学变换后的光谱数据(400~800 nm、1000~2450 nm)中经过  $P=0.01$  显著性检验后的波段分别建模的自变量,土壤有机质含量作为因变量。从表 1 可以看出,PLSR 建模时,R、FOD(2)、ILR 和 LDR 模型的 RPD 均未达到 2.0,模型效果一般,只能对样品有机质含量进行粗略估测;而 FOD(0.5)、FOD(1) 及 FOD(1.5) 模型的 RPD 分别为 2.19、2.23 和 2.34,均在 2.0 以上,说明模型对有机质含量有较好的预测能力。对 R 进行数学变换后,仅有 LDR 的验证集精度下降, $R^2$  虽然升高了 0.04,但是 RMSE 升高了 0.49  $\text{g}\cdot\text{kg}^{-1}$ ,RPD 下降了 0.09。表中数据充分说明数学变换有效地提高了红壤有机质含量的预测能力,其中 FOD(1.5) 模型的训练集和验证集的  $R^2$  最高,分别达到了 0.88 和 0.83, RMSE 分别为 4.98  $\text{g}\cdot\text{kg}^{-1}$  和 6.62  $\text{g}\cdot\text{kg}^{-1}$ ,预测能力最为显著,其 RPD 达到了 2.31。

## 2.4 红壤有机质含量 BP 神经网络建模

BP 神经网络模型是将训练集各波段光谱数据作为输入层数据,土壤有机质含量作为输出层数据进行构建。从表 2 中可以看出,基于光谱反射率数据(R)的 BP 神经网络模型的训练集和验证集精度均不高且 RPD 仅为 0.89,无法对样本有机质进行预测。FOD(0.5)、FOD(1)、FOD(1.5)、FOD(2)、ILR 和 LDR 模型的训练集  $R^2$  均达到了 0.90 以上, RMSE 均小于 5.00  $\text{g}\cdot\text{kg}^{-1}$ ;验证集  $R^2$  基本达到 0.80, RMSE 基本在 7.00  $\text{g}\cdot\text{kg}^{-1}$  以下,RPD 均在 2.10 以上,

具有较好的预测有机质含量的能力。其中 ILR 模型训练集  $R^2$  最高达到了 0.95, RMSE 为  $3.14 \text{ g}\cdot\text{kg}^{-1}$ , 但验证集中 FOD (1) 模型  $R^2$  最高为 0.84, RMSE

为  $6.55 \text{ g}\cdot\text{kg}^{-1}$ , RPD 为 2.34。结合表 1 和表 2 可以看出, 与 PLSR 相比, BP 神经网络 R 模型的 RPD 下降了 0.86, 其余变换 RPD 均有所提升。

表 1 红壤有机质含量 PLSR 模型精度

Table 1 Precision of the PLSR model in predicting red soil SOM content

数学变换 Mathematical transformation	潜变量 (个) Potential variable	训练集 Training dataset		验证集 Validation dataset		
		$R^2$	RMSE	$R^2$	RMSE	RPD
R	5	0.76	7.02	0.67	8.74	1.75
FOD (0.5)	6	0.81	6.22	0.80	6.97	2.19
FOD (1)	6	0.84	5.78	0.81	6.86	2.23
FOD (1.5)	6	0.88	4.98	0.83	6.62	2.31
FOD (2)	5	0.78	6.78	0.76	7.57	1.99
ILR	6	0.81	6.24	0.73	8.16	1.87
LDR	6	0.85	5.59	0.71	9.23	1.66

注: R, 光谱反射率; FOD (0.5), 0.5 阶导数; FOD (1), 1 阶导数; FOD (1.5), 1.5 阶导数; FOD (2), 2 阶导数; ILR, 倒数的对数; LDR, 对数的导数; 下同。Note: R stands for reflectance; FOD(0.5) for fractional order derivative (0.5); FOD(1) for fractional order derivative (1); FOD (1.5) for fractional order derivative (1.5); FOD (2) for fractional order derivative (2); ILR for inverse-log reflectance; and LDR for log-derivative reflectance. The same below.

表 2 红壤有机质含量 BP 神经网络模型精度

Table 2 Precision of the BP neural net model in predicting red soil SOM content

数学变换 Mathematical transformation	训练集 Training dataset		验证集 Validation dataset		
	$R^2$	RMSE	$R^2$	RMSE	RPD
R	0.54	13.26	0.30	17.22	0.89
FOD (0.5)	0.90	4.72	0.81	6.66	2.29
FOD (1)	0.94	3.39	0.84	6.55	2.34
FOD (1.5)	0.92	4.07	0.81	6.57	2.33
FOD (2)	0.91	4.51	0.80	6.88	2.22
ILR	0.95	3.14	0.81	6.76	2.26
LDR	0.93	4.50	0.79	7.11	2.15

## 2.5 PLSR-BP 复合建模

在进行 PLSR-BP 复合建模时, 将 PLSR 建模中提取出的潜变量作为自变量进行土壤有机质预测模型的构建。表 3 为 PLSR-BP 复合建模结果, 与表 1、表 2 相比较, 可以看出 PLSR-BP 复合建模的 R 模型 RPD 较单一的 PLSR 和 BP 模型而言提升到了 1.96, 但仍只能对样本有机质含量进行粗略估测。ILR 模型的 RPD 为 1.98, 较 PLSR 建模提升了 5.9%, 但相比 BP 神经网络降低了 12.5%, 只能粗略估计样本有机

质含量。FOD (0.5)、FOD (2) 及 LDR 模型的 RPD 分别为 2.41、2.27、2.46, 均具有较好的估测土壤有机质含量的能力。FOD (1) 和 FOD (1.5) 模型的 RPD 分别为 2.63 和 2.75, 具有很好的预测能力。除 R 模型, 其余模型训练集  $R^2$  均低于 BP 神经网络模型, 验证集  $R^2$  有所提升。其中 FOD (1.5) 优于 0.5、1、2 阶导数模型, 相比经过 FOD (1.5) 变换的 BP 神经网络模型, 训练集  $R^2$  下降了 3.2%, RMSE 上升了 15.0%, 验证集  $R^2$  上升了 7.4%, RMSE 下降了 15.5%,

表 3 红壤有机质含量 PLSR-BP 模型精度

Table 3 Precision of the PLSR-BP model in predicting red soil SOM content

数学变换 Mathematical transformation	训练集 Training dataset		验证集 Validation dataset		
	$R^2$	RMSE	$R^2$	RMSE	RPD
$R$	0.83	6.38	0.77	7.79	1.96
FOD (0.5)	0.85	5.71	0.84	6.34	2.41
FOD (1)	0.87	5.32	0.86	5.81	2.63
FOD (1.5)	0.89	4.68	0.87	5.55	2.75
FOD (2)	0.86	5.81	0.81	6.74	2.27
ILR	0.84	5.79	0.75	7.70	1.98
LDR	0.89	4.73	0.84	6.20	2.46

RPD 上升了 18.0%。图 4 可以看出 FOD (1.5) 模型的验证样本基本在 1:1 线附近, 预测能力为最优。总体而言, PLSR-BP 复合模型验证集的各项判定指标均优于 PLSR 或 BP 神经网络单独建模。

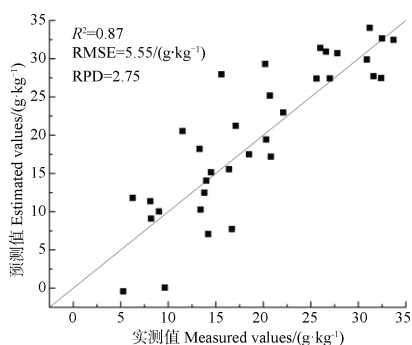


图 4 1.5 阶导数变换的有机质含量 PLSR-BP 复合模型实测值与预测值比较

Fig. 4 Comparison of measured SOM and the values predicted with the PLSR-BP model based on 1.5 order derivative transformation

### 3 讨论

红壤氧化铁含量较高, 主要是由于红壤在形成过程中的脱硅富铁铝化过程所造成的<sup>[31]</sup>。季耿善和徐彬彬<sup>[32]</sup>发现游离氧化铁在红壤光谱的 900 nm 附近吸收最强, 对光谱特性的影响极大。而在对红壤有机质含量估测模型的构建中, 刘磊等<sup>[33]</sup>通过对红壤原始光谱进行去包络线处理, 使用其特征吸收带 480~580 nm、820~950 nm、1 010~1 060 nm、1 360~1 500 nm、1 880~2 020 nm、2 160~2 240 nm 进行有机质含量建模; 谢文等<sup>[34]</sup>通过研究发现山地红壤有机质光谱特征波段位于 600~2 450 nm 处,

并选用全波段进行建模。他们在建立估测模型时均未考虑到氧化铁对于模型精度的影响。考虑到氧化铁在 900 nm 左右的吸收谷的干扰, 在波段选择上, 本研究去除 801~1 000 nm 波段, 选用 400~800 nm、1 001~2 450 nm 波段进行建模, 以减小氧化铁对于有机质含量预测的干扰。

在土壤光谱数据建模前, 对其进行各种数学变换是非常必要的, 有助于构建精度更高的模型<sup>[35,36]</sup>。常用的数学变换主要有 1 阶导数、2 阶导数、倒数的对数等, 本文在此基础上引入了目前研究较少的分数阶导数这一概念。分数阶导数的引入对整数阶导数的概念进行了扩展, 在阶数的选择上更加广泛, 同时也将隐含在光谱内的信息表现出来。张东<sup>[37]</sup>利用分数阶导数公式将 0~2 阶的阶数间隔细化至 0.1, 发现在 2 阶导数时模型的预测精度达到最优。本研究以 0.5 为阶数间隔对土壤光谱数据进行 0~2 阶的导数变换, 经过导数变换后, 发现采用 PLSR 模型建模时模型的 RPD, 即预测能力 FOD (1.5) > FOD (1) > FOD (0.5) > FOD (2); BP 神经网络建模时, 各模型的 RPD, FOD (1) > FOD (1.5) > FOD (0.5) > FOD (2); 基于 PLSR-BP 复合模型建模时的 RPD, FOD (1.5) > FOD (1) > FOD (0.5) > FOD (2), 与 PLSR 模型具有相同的趋势, 即在 0~2 的区间上对于有机质的预测能力呈现一个先升高后下降的趋势, 在 1.5 阶导数时得到最优预测模型。BP 神经网络建模时, 1 阶导数变换得到了较好的预测模型可能是由于隐藏层神经元等参数设置不同而导致了该结果。与张东<sup>[37]</sup>研究结果不一致的原因可能在于土壤含盐量和有机质二者对土壤光谱的敏感性不同。

土壤光谱的近红外波段往往含有数据冗余，会增加建模的复杂性<sup>[38]</sup>。PLSR 模型能够很好地提取土壤光谱中的信息，同时使其与有机质含量的相关程度达到最大。文中非线性的 BP 神经网络模型较线性的 PLSR 模型有更好的预测能力，其不足之处在于 BP 神经网络训练集虽然有很高的决定系数，但由于输入变量过多，网络规模过大，影响收敛速度，造成了“过拟合”的现象，这也导致验证集与训练集精度相差较大。因此本文使用 PLSR-BP 复合模型进行土壤有机质的预测，采用 PLSR 先对土壤光谱数据进行潜变量的提取，减少数据冗余，再对这些潜变量进行 BP 神经网络建模，这一方法可以有效避免使用单一的 BP 神经网络模型进行全波段拟合时出现共线性现象。结果表明，PLSR-BP 复合模型的 RPD 较单一模型高出了 0.12~1.07，说明 PLSR-BP 复合模型在对红壤有机质含量的预测中的实用性。

本文不足之处在于仅构建了一个综合模型对林地、园地和农田三种土地利用类型的土壤有机质含量进行预测，而并未深入研究对于不同土地利用类型是否能采用同一模型进行土壤有机质含量的预测，这也是今后需要进一步研究的方向。

## 4 结 论

分数阶导数是在传统的整数阶导数上的扩展，减少了有用信息的遗漏，有助于土壤有机质含量的预测。对于使用经过分数阶导数变换的红壤光谱而言，在 0~2 阶的区间上，对土壤有机质含量的预测能力呈现出先升高后下降的趋势，并在 1.5 阶处能够得到最优模型。在建模方法的选择上，偏最小二乘回归能够在保证土壤光谱与有机质含量相关性最大的基础上进行数据的压缩，减少数据冗余；BP 神经网络预测精度虽然较高，但由于输入变量过多易出现过拟合现象；偏最小二乘回归与 BP 神经网络结合可以综合二者的优点，提高模型的预测精度。

## 参考文献 (References)

- [ 1 ] He J L, Jiang J J, Zhou S L, et al. The hyperspectral characteristics and retrieval of soil organic matter content[J]. *Scientia Agricultura Sinica*, 2007, 40 (3): 638—643. [贺军亮, 蒋建军, 周生路, 等. 土壤有机质含量的高光谱特性及其反演[J]. *中国农业科学*, 2007, 40 (3): 638—643.]
- [ 2 ] Huang Y F, Liu T H. Relationship between soil spectral reflectance characteristics and soil properties—Taking the main soil in the south as an example[J]. *Chinese Journal of Soil Science*, 1989, 20 (4): 158—160, 176. [黄应丰, 刘腾辉. 土壤光谱反射特性与土壤属性的关系——以南方主要土壤为例[J]. *土壤通报*, 1989, 20 (4): 158—160, 176.]
- [ 3 ] Xu B B, Dai C D. The correlation between nanjiang soil spectral reflectance property and organic matter[J]. *Chinese Science Bulletin*, 1980, 25 (6): 282—284. [徐彬彬, 戴昌达. 南疆土壤光谱反射特性与有机质含量的相关分析[J]. *科学通报*, 1980, 25 (6): 282—284.]
- [ 4 ] Henderson T L, Szilagyi A, Baumgardner M F, et al. Spectral band selection for classification of soil organic matter content[J]. *Soil Science Society of America Journal*, 1989, 53 (6): 1778—1784.
- [ 5 ] Xu J G, Feng X L, Guan L, et al. Fractional differential application in reprocessing infrared spectral data [J]. *Control and Instruments in Chemical Industry*, 2012, 39 (3): 347—351. [徐继刚, 冯新沪, 管亮, 等. 分数阶微分在红外光谱数据预处理中的应用[J]. *化工自动化及仪表*, 2012, 39 (3): 347—351.]
- [ 6 ] Tong P J, Du Y P, Zheng K Y, et al. Improvement of NIR model by fractional order Savitzky-Golay derivation (FOSGD) coupled with wavelength selection[J]. *Chemometrics and Intelligent Laboratory Systems*, 2015, 143: 40—48.
- [ 7 ] Wang J Z, Tashpolat-Tiyip, Ding J L, et al. Estimation of desert soil organic carbon content based on hyperspectral data preprocessing with fractional differential[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32 (21): 161—169. [王敬哲, 塔西甫拉提·特依拜, 丁建丽, 等. 基于分数阶微分预处理高光谱数据的荒漠土壤有机碳含量估算[J]. *农业工程学报*, 2016, 32 (21): 161—169.]
- [ 8 ] Yu L, Hong Y S, Geng L, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31 (14): 103—109. [于雷, 洪永胜, 耿雷, 等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. *农业工程学报*, 2015, 31 (14): 103—109.]
- [ 9 ] Guo Y K, Liu N, Liu L, et al. Hyperspectral inversion of soil Cu content based on BP neural network model[J]. *Science of Surveying and Mapping*, 2018, 43 (1): 135—139, 152. [郭云开, 刘宁, 刘磊, 等. 土壤 Cu 含量高光谱反演的 BP 神经网络模型[J]. *测绘科学*, 2018, 43 (1): 135—139, 152.]
- [ 10 ] Hong Y S, Yu L, Zhu Y X, et al. Removing the effect of soil moisture on prediction of soil organic matter with



- hyperspectral reflectance using external parameter orthogonalization[J]. *Acta Pedologica Sinica*, 2017, 54(5): 1068—1078. [洪永胜, 于雷, 朱亚星, 等. 基于EPO算法去除水分影响的土壤有机质高光谱估算[J]. *土壤学报*, 2017, 54(5): 1068—1078.]
- [ 11 ] Conforti M, Buttafuoco G, Leone A P, et al. Studying the relationship between water-induced soil erosion and soil organic matter using Vis - NIR spectroscopy and geomorphological analysis: A case study in southern Italy[J]. *Catena*, 2013, 110: 44—58.
- [ 12 ] Ding G X. Studies on the retrieval of soil organic matter and total iron content from hyper-spectrum based on ANN[D]. Nanjing: Nanjing University of Information Science & Technology, 2008. [丁国香. 基于神经网络的土壤有机质及全铁含量的高光谱反演研究[D]. 南京: 南京信息工程大学, 2008.]
- [ 13 ] Wu C W, Xia J X, Duan Z R. Review on detection methods of soil organic matter[J]. *Soils*, 2015, 47(3): 453—460. [吴才武, 夏建新, 段峥嵘. 土壤有机质测定方法述评与展望[J]. *土壤*, 2015, 47(3): 453—460.]
- [ 14 ] Ma C H, Liu L Y. Spectral data processing based on wavelet analysis[J]. *Metallurgical Analysis*, 2012, 32(1): 34—37. [马翠红, 刘立业. 基于小波分析的光谱数据处理[J]. *冶金分析*, 2012, 32(1): 34—37.]
- [ 15 ] Shi Z. Principle and method of soil surface hyperspectral remote sensing[M]. Beijing: Science Press, 2014: 61—63. [史舟. 土壤地面高光谱遥感原理与方法[M]. 北京: 科学出版社, 2014: 61—63.]
- [ 16 ] 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[J]. *Geoderma*, 2011, 167/168: 118—124.
- [ 17 ] Xu M X, Zhou S L, Ding W, et al. Hyperspectral reflectance models for predicting soil organic matter content in coastal tidal land area, northern Jiangsu[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(2): 219—223. [徐明星, 周生路, 丁卫, 等. 苏北沿海滩涂地区土壤有机质含量的高光谱预测[J]. *农业工程学报*, 2011, 27(2): 219—223.]
- [ 18 ] Sun H G, Chang A L, Chen W, et al. Anomalous diffusion: Fractional derivative equation models and applications in environmental flows[J]. *Scientia Sinica: Physica, Mechanica & Astronomica*, 2015, 45(10): 8—22. [孙洪广, 常爱莲, 陈文, 等. 反常扩散: 分数阶导数建模及其在环境流动中的应用[J]. *中国科学: 物理学 天文学*, 2015, 45(10): 8—22.]
- [ 19 ] Kong F, Wang H, Xu J, et al. Wavlet-Galerkin based approach for determining power spectral density of a system endowed with fractional derivative damping[J]. *Journal of Vibration Engineering*, 2018, 31(4): 671—680. [孔凡, 王恒, 徐军, 等. 分数阶导数系统响应功率谱密度的小波-Galerkin方法[J]. *振动工程学报*, 2018, 31(4): 671—680.]
- [ 20 ] Benkhettou N, Brito da Cruz A M C, Torres D F M. A fractional calculus on arbitrary time scales: Fractional differentiation and fractional integration[J]. *Signal Processing*, 2015, 107: 230—237.
- [ 21 ] Wang W. Research in digital image restoration and enhancement based on differential[D]. Shanghai: Fudan University, 2012. [王伟. 基于微分的数字图像复原与增强技术研究[D]. 上海: 复旦大学, 2012.]
- [ 22 ] Tsai F, Philpot W. Derivative analysis of hyperspectral data[J]. *Remote Sensing of Environment*, 1998, 66(1): 41—51.
- [ 23 ] Hong Y S, Chen Y Y, Yu L, et al. Combining fractional order derivative and spectral variable selection for organic matter estimation of homogeneous soil samples by VIS-NIR spectroscopy[J]. *Remote Sensing*, 2018, 10(3): 479—498.
- [ 24 ] Panagou E Z, Mohareb F R, Argyri A A, et al. A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints[J]. *Food Microbiology*, 2011, 28(4): 782—790.
- [ 25 ] Wang K L, Xiong H G, Zhang F. PLSR-BP complex model-based hyper-spectrum retrieval of oasis soil pH[J]. *Arid Zone Research*, 2014, 31(6): 1005—1009. [王凯龙, 熊黑钢, 张芳. 基于PLSR-BP复合模型的绿洲土壤pH高光谱反演[J]. *干旱区研究*, 2014, 31(6): 1005—1009.]
- [ 26 ] 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[J]. *Science China: Earth Sciences*, 2014, 44(5): 978—988. [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[J]. *中国科学: 地球科学*, 2014, 44(5): 978—988.]
- [ 27 ] Zhang M, Yuan H. The Pau-Ta criterion and rejecting the abnormal value[J]. *Journal of Zhengzhou University of Technology*, 1997, 18(1): 84—88. [张敏, 袁辉. 拉依达(PauTa)准则与异常值剔除[J]. *郑州工业大学学报*, 1997, 18(1): 84—88.]
- [ 28 ] Nan F, Zhu H F, Bi R T. Hyperspectral prediction of soil organic matter content in the reclamation cropland of coal mining areas in the Loess Plateau[J]. *Scientia Agricultura Sinica*, 2016, 49(11): 2126—2135. [南锋, 朱洪芬, 毕如田. 黄土高原煤矿区复垦农田土壤有机质含量的高光谱预测[J]. *中国农业科学*, 2016, 49(11): 2126—2135.]
- [ 29 ] Zhao X M, Yang M H. Hyperspectral characteristics of major types of soils in red soil region of Jiangxi Province, China [J]. *Acta Pedologica Sinica*, 2018, 55(1): 31—42. [赵小敏, 杨梅花. 江西省红壤地区主要

- 土壤类型的高光谱特性研究[J]. 土壤学报, 2018, 55(1): 31—42.]
- [ 30 ] Zhong R L, Xiao X, Zhang P C, et al. Research on spatial variability statement methods of soil anti-erodibility in Jiangxi Province[J]. Journal of Yangtze River Scientific Research Institute, 2010, 27(5): 13—18. [钟壬琳, 肖潇, 张平仓, 等. 江西省土壤抗侵蚀性空间变异表述方法研究[J]. 长江科学院院报, 2010, 27(5): 13—18.]
- [ 31 ] Liu X, He X Y. Progress and achievement in the research on red soil in Jiangxi[J]. Acta Agriculturae Jiangxi, 1991, 31(1): 66—75. [刘勋, 贺湘逸. 江西红壤科研进展与成就[J]. 江西农业学报, 1991, 31(1): 66—75.]
- [ 32 ] Ji G S, Xu B B. Reflectance of soil clay minerals and its application in pedology[J]. Acta Pedologica Sinica, 1987, 24(1): 67—76. [季耿善, 徐彬彬. 土壤黏土矿物反射特性及其在土壤学上的应用[J]. 土壤学报, 1987, 24(1): 67—76.]
- [ 33 ] Liu L, Shen R P, Ding G X. Studies on the estimation of soil organic matter content based on hyper-spectrum[J]. Spectroscopy and Spectral Analysis, 2011, 31(3): 762—766. [刘磊, 沈润平, 丁国香. 基于高光谱的土壤有机质含量估算研究[J]. 光谱学与光谱分析, 2011, 31(3): 762—766.]
- [ 34 ] Xie W, Zhao X M, Guo X, et al. Spectrum based estimation of the content of soil organic matters in mountain red soil using RBF combination model[J]. Scientia Silvae Sinicae, 2018, 54(6): 16—23. [谢文, 赵小敏, 郭熙, 等. 基于 RBF 组合模型的山地红壤有机质含量光谱估测[J]. 林业科学, 2018, 54(6): 16—23.]
- [ 35 ] He T, Wang J, Cheng Y, et al. Spectral features of soil moisture[J]. Geomatics and Information Science of Wuhan University, 2006, 31(11): 975—979. [何挺, 王静, 程焯, 等. 土壤水分光谱特征研究[J]. 武汉大学学报(信息科学版), 2006, 31(11): 975—979.]
- [ 36 ] Yu S K. Study on the inversion of soil organic matter content based on hyper-spectrum[D]. Beijing: Chinese Academy of Agricultural Sciences, 2013. [于士凯. 基于高光谱的土壤有机质含量反演研究[D]. 北京: 中国农业科学院, 2013.]
- [ 37 ] Zhang D. Application research of fractional derivative on the remote sensing monitoring of soil salinization[D]. Urumqi: Xinjiang University, 2017. [张东. 分数阶微分在土壤盐渍化遥感监测中的应用研究[D]. 乌鲁木齐: 新疆大学, 2017.]
- [ 38 ] Vohland M, Ludwig M, Harbich M, et al. Using variable selection and wavelets to exploit the full potential of visible-near infrared spectra for predicting soil properties[J]. Journal of Near Infrared Spectroscopy, 2016, 24(3): 255—269.

(责任编辑: 檀满枝)