

DOI: 10.11766/trxb201606210174

基于组合模型的庐山森林土壤有效铁光谱间接反演研究*

谢文 赵小敏[†] 郭熙 叶英聪 李伟峰 汪晓燕 张佳佳

(江西农业大学江西省鄱阳湖流域农业资源与生态重点实验室/林学院, 南昌 330045)

摘要 铁是植物生长的重要微量营养元素之一, 土壤有效铁含量对林地环境起着重要的影响, 利用土壤光谱预测技术获取土壤有效铁含量信息具有重要意义。而要通过土壤光谱直接预测土壤有效铁含量是难以实现的, 因此提出利用土壤有机质含量与有效铁含量之间的相关性, 探讨间接估算土壤有效铁含量的可行性。以庐山森林土壤样本为研究对象, 研究基于偏最小二乘回归 (PLSR) 和径向基函数 (RBF) 神经网络的组合模型预测土壤有机质含量的适用性, 并且通过构建有机质含量与有效铁含量的二项式线性模型, 对土壤有效铁含量进行间接反演, 探讨不同权重下的最优组合模型。结果表明, 组合模型的预测效果优于偏最小二乘回归和RBF神经网络单个模型, 并且熵值组合为最优组合模型, 其中, 土壤有机质的反演模型验证的决定系数 (R^2) 为0.81, 均方根误差 ($RMSE_p$) 为 11.54 g kg^{-1} , 测定值标准差与标准预测误差的比值 (RPD) 为2.18; 有效铁的间接反演模型 R^2 为0.70, $RMSE_p$ 为 21.60 mg kg^{-1} , RPD为1.77。通过土壤有机质构建土壤有效铁含量的光谱反演间接模型, 在光谱反演模型中, 组合模型能较大限度地利用各种预测样本信息, 能有效减少单个预测模型中随机因素的影响, 增强预测稳定性, 提高模型的预测能力。因此, 组合模型可对土壤有机质含量的光谱预测及土壤有效铁的间接预测发挥更好的作用。

关键词 土壤光谱; 有效铁预测; RBF神经网络; 偏最小二乘回归 (PLSR); 组合模型

中图分类号 S127; O657.3 **文献标识码** A

铁是植物生长不可或缺的微量元素, 在土壤元素中含量相对较高^[1], 土壤是植物所需有效铁的主要供给源, 植物缺铁会影响叶绿素的合成, 土壤有效铁含量高会影响植物的生长发育^[2-4]。因此, 研究土壤中有效铁含量对评价庐山生态环境及植被健康生长状况非常必要。

化学成分是决定土壤反射光谱特征的最重要因素, 高光谱遥感数据具有丰富的光谱特征, 并且利用光谱分析技术可以快速、无损、低成本实现土壤参数实时在线监测, 为土壤化学成分的快速、无损测定提供新的方法, 因此, 高光谱遥感具有获取土

壤化学成分的研究潜力^[5-8]。目前已有不少学者在利用实验室和野外测量的反射光谱数据分析土壤Fe元素的特征波段、研究土壤Fe的光谱特性等方面取得了一定成果, 例如Chodak等^[9]利用线性回归分析法研究了森林覆盖区域的土壤Fe与反射光谱间的关系; Kemper和Sommer^[10]提出土壤重金属与铁含量的相关性是反演重金属的主要机理; Siebielec等^[11]发现中红外光谱较近红外光谱预测土壤中Fe含量效果更好; Ren等^[12]利用标准归一化处理预测Fe含量效果较好; 李巨宝等^[13]利用偏最小二乘法构建了土壤Fe含量的反射光谱预测模型; 彭杰

* 国家自然科学基金项目 (41361049) 和土壤与农业可持续发展国家重点实验室项目 (0812201202) 资助 Supported by the National Natural Science Foundation of China (No.41361049) and The Project of State Key Laboratory of Soil and Sustainable Agricultural (No.0812201202)

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

作者简介: 谢文 (1978—), 女, 江西上高人, 博士研究生, 讲师, 主要从事土壤遥感与信息技术研究。E-mail: 183385413@qq.com

收稿日期: 2016-06-21; 收到修改稿日期: 2016-11-03; 优先数字出版日期 (www.cnki.net): 2017-02-13

等^[14-15]发现利用土壤线参数反演氧化铁含量具有可行性,但在622~851nm波段范围内,氧化铁对土壤有机质的高光谱特征具有明显的抑制作用,并通过不同含量氧化铁对有机质光谱的影响分析,发现氧化铁含量高于20 g kg⁻¹或氧化铁含量与有机质含量比值高于2.0时,有机质定量反演精度会降低等。在光谱反演模型方面,多元线性回归、主成分回归、偏最小二乘回归和神经网络等线性和非线性方法在土壤各组分含量估算中应用较多,并且均得到了较好的预测精度^[5],而神经网络模型中应用于土壤光谱反演的主要是BP神经网络模型,但BP神经网络模型存在学习效率低、收敛速度慢的缺陷,而径向基函数(RBF)能提高BP的学习速度,在非线性和非线性拟合方面也有很好的优势^[16],但由于传统单一预测模型自身存在局限性,为解决单一模型误差的问题,Bates和Granger^[17]提出综合多个单一模型构建组合预测模型。综上所述,虽然有不少学者探讨过土壤中Fe含量的光谱反演,但由于土壤中Fe含量不高,存在形式多样,因此,反演精度均不理想;此外大多数土壤养分含量反演采用的均为单一模型,造成在利用土壤光谱反演土壤养分含量的研究中,存在一定局限性,因此,本文以庐山森林土壤为研究对象,将RBF神经网络模型与偏最小二乘回归(PLSR)模型进行组合,对比不同的加权最优组合预测模型,先进行土壤有机质含量的光谱反演,再通过有机质含量与有效铁含量的二项式线性模型对有效铁含量进行间接反演研究,探

讨利用高光谱技术进行土壤有效铁含量间接反演的可行性。

1 材料与方法

1.1 样品采集

土壤样品采集于亚热带典型红壤区的庐山,位于江西省北部,其土壤具有明显的垂直地带性及养分分布规律性,土壤类型随海拔的升高,依次为山地红壤、山地黄壤、山地黄棕壤、山地棕壤和山地草甸土^[18]。本文主要采集的是庐山林地表层0~20 cm的土壤,按照1 km×1 km的范围进行选点,共布设192个样点,去除两个异常样点,共有190个有效样点,其中,山地红壤143个、山地黄壤8个、山地黄棕壤2个、山地黄棕壤34个和山地棕壤3个。对土样进行风干、碾磨、过0.25 mm筛处理后,分成两份,分别用于土壤理化分析和光谱测定。土壤有机质含量采用重铬酸钾容量法-外加热法测定^[19],土壤有效铁含量采用DTP浸提,原子吸收光谱法测定^[20]。

由于样本海拔跨度较大,土壤有效铁含量差异很大,最小值仅为0.04 mg kg⁻¹,最大值为167.5 mg kg⁻¹,平均含量为46.31 mg kg⁻¹,土壤有机质含量的差异性也较大,最小值为4.80 g kg⁻¹,最大值为150.9 g kg⁻¹,土壤有效铁含量离散程度要高于土壤有机质含量(表1)。

根据Kennard-Stone光谱主成分空间欧式距离

表1 土壤有机质及有效铁含量统计特征

Table 1 Statistics of available iron and organic matter contents in soil

样本集 Dataset	样本数 Number of samples	有机质 Organic matter					有效铁 Soil available iron				
		均值	最小值	最大值	标准差	变异系数	均值	最小值	最大值	标准差	变异系数
		Mean (g kg ⁻¹)	Minimum (g kg ⁻¹)	Maximum (g kg ⁻¹)	SD	CV (%)	Mean (mg kg ⁻¹)	Minimum (mg kg ⁻¹)	Maximum (mg kg ⁻¹)	SD	CV (%)
全体样本 ^①	190	32.8	4.8	150.9	25.7	78.4	46.3	0.04	167.5	36.53	78.87
建模样本 ^②	143	32.82	4.8	150.9	26.0	79.2	45.9	1.25	167.5	36.1	78.7
验证样本 ^③	47	32.73	6.05	122.8	25.2	76.9	47.7	0.04	151.4	38.17	80.06

①Total dataset; ②Modeling dataset; ③Validation dataset

算法,选取总样本的75%作为建模样本(143个)构建模型,25%(47个)作为验证样本进行验证。从表1样本有机质和有效铁含量统计特征来看,建模样本与验证样本的标准差、变异系数与总样本接

近,具有代表性。

1.2 光谱数据采集与预处理

光谱测定是将风干、研磨、过筛处理后的土样采用美国ASD公司生产的ASD FieldSpec3地

物光谱仪进行土壤测试, 波长范围为350~2 500 nm, 采样间隔为1.4 nm (350~1 000 nm)、2 nm (1 000~2 500 nm), 重采样间隔为1 nm。光谱测量时, 在能控制光照条件的暗室内进行, 将土样平铺装满于盛样皿中, 利用50 W的卤素灯作为光源, 距土壤表面30 cm, 天顶角15°, 5°探头, 距离土样15 cm垂直观测, 测定前以白板定标, 并每间隔20~30 min进行一次标准白板校正, 每个土样测定5次, 取算术平均值作为每个土样的实际反射光谱数据。每个土样重采样间隔为1nm进行光谱输出, 测定得到350~2 500 nm波长、波段数为2 151的反射光谱数据。

由于光谱采集过程中存在背景及环境因子的影响, 为消除误差, 利用ASD光谱仪自带的ViewSpec Pro Version 6.0工具进行断点校正, 并且考虑光谱边缘波段噪声的影响, 去除边缘波段(350~399 nm, 2 451~2 500 nm), 保留400~2 450 nm波段的反射光谱数据。为研究不同处理方法对土壤有效铁含量的影响, 对原始光谱进行了倒数、一阶导数、对数、平方根、去包络线等处理, 对比选择原始光谱作为最优光谱数据进行研究。

1.3 建模方法

本研究首先通过分析土壤有机质含量与土壤有效铁含量之间的相关性, 建立两者之间的二项式线性模型, 然后通过对土壤有机质含量的光谱反演, 最后通过建立的二项式模型对土壤有效铁含量进行估测。

研究土壤属性与反射光谱之间关系常用的方法有多元逐步线性回归、偏最小二乘回归(PLSR)、人工神经网络、支持向量机等方法, 而偏最小二乘回归是集成了主成分分析、典型相关分析和线性回归分析的经典和最常用的一种线性模型方法^[21]。目前BP神经网络模型是应用最广神经网络模型之一, 但由于BP神经网络算法存在很多收敛速度慢等问题, 在处理较复杂的模型时, 径向基函数(RBF)神经网络模型在时间、非线性曲线拟合能力和学习速度等方面均优于BP神经网络模型, 因此, RBF网络模型在预测非线性问题中精度高、学习速度快, 值得推广^[16, 22-23]。而单一模型自身存在的局限性明显, 通过单一模型进行预测分析, 不能全面反映事物的信息, 会使预测结果误差较大, 组合预测模型较单个预测模型具有更高的预

测精度^[24]。

RBF神经网络是具有突出的函数逼近能力并且在模式识别、函数逼近和经济预测等领域有着广泛应用的一类局部逼近的前馈式神经网络, 是由输入层、隐含层和输出层组成的非线性函数, 多输入单输出RBF神经网络的输出为:

$$y = \sum_{i=1}^n \omega_i \varphi_i(x) = \sum_{i=1}^n \omega_i \exp\left(-\frac{\|x - c_i\|^2}{2\delta^2}\right)$$

式中, n 为隐含层神经元的个数, $x(x_1, \dots, x_m)$ 为输入向量, ω_i 为第 i 个隐含层神经元与输出层神经元的联结权值, φ_i 为RBF高斯函数第 i 个隐含层神经元的输出, c_i 为隐含层第 i 个神经元的基函数中心值, δ 为高斯函数的方差^[25-27]。

本文的RBF组合模型选取两个常用的偏最小二乘回归(PLSR)线性预测模型和径向基函数(RBF)神经网络非线性预测模型分别对建模样本进行预测, 利用单一模型的预测结果进行算术平均、预测误差平方和倒数及熵值^[28]三种赋权方法组合后, 进行RBF模型预测, 研究选择MATLAB2014a软件工具箱中的newrb函数作为核函数进行建模, 对比验证精度, 验证精度通过均方根误差(RMSE)、预测偏差比(RPD)和决定系数(R^2)进行评价。

1.4 组合模型权重的确定

组合模型是综合各单个预测模型的预测结果, 利用不同的权系数进行加权平均后得到的预测模型, 权系数是组合预测模型最重要的一个部分^[29]。本文分别运用算术平均法、预测误差平方和倒数法和熵值法三种常用的权系数方法确定单个预测模型的权重, 确定最优的权系数组合预测模型。

算术平均法特点是单个预测模型的加权系数完全相等, 即将单个预测模型同等看待, 系数确定公式为:

$$l_k = \frac{1}{m} (k = 1, 2, \dots, m),$$

式中, l_k 为某预测模型的组合权系数, m 为单个预测模型个数。

预测误差平方和倒数法特点是单个预测模型精度越低, 加权系数则越大, 即通过预测误差平方和倒数确定权系数来降低预测精度低的单个模型在组合预测模型中的重要性, 系数确定公式为:

$$l_k = \frac{E_{kk}^{-1}}{\sum_{k=1}^m E_{kk}^{-1}} (k = 1, 2, \dots, m), E_{kk} = \sum_{i=1}^n (x_i - x_{ki})^2, \sum_{k=1}^m l_k = 1 (l_k \geq 0, k = 1, 2, \dots, m)$$

式中， l_k 为某预测模型的组合权系数， m 为单个预测模型个数， E_{kk} 为第 k 个单项预测模型的预测误差平方和， x_{ki} 为第 k 个单个预测模型在第 i 个样本中的预测值， x_i 为同一预测样本的实际值， n 为建模样本个数。

熵值法的特点是熵值越大，单个预测模型预测误差变异程度就越小，则在组合预测模型中的权系数就越大，即通过熵值体现单个预测模型在组合模型中的重要程度^[28]。

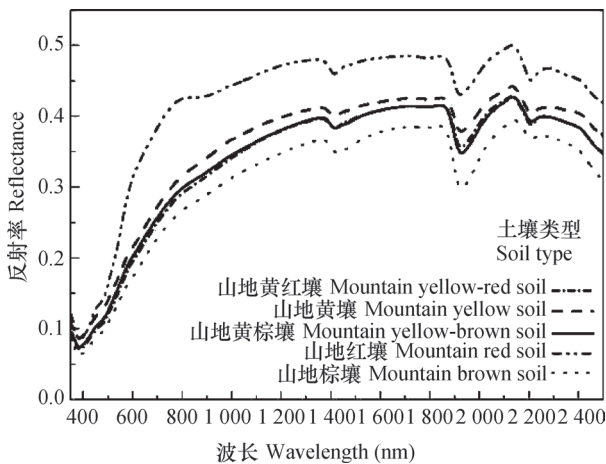


图1 不同类型土壤的反射率均值曲线

Fig.1 Reflectance spectra of soils different in soil type

2 结果

2.1 不同类型土壤的高光谱曲线特征

本文采集的数据来自于庐山森林土壤，土壤类型包括山地红壤、山地黄红壤、山地黄壤、山地黄棕壤和山地棕壤五种类型，同一类型的土壤光谱求取平均后得到其相对应的反射光谱曲线，从图1可以看出，不同土壤类型的土壤反射光谱曲线存在差异，但变化趋势上又存在一定的规律性。例如在可见光350~500 nm波段范围五种土壤类型光谱曲线均很相似，区分并不明显，而山地黄红壤、山地黄棕壤、山地黄壤和山地棕壤的光谱反射强度、曲线变化的斜率和吸收谷的强度等特征总体均较为相似，呈现的均是1 400 nm、1 900 nm、2 200 nm红外波段附近的吸收谷特征，而山地红壤由于 Fe^{3+} 含量较高，在900 nm波段附近有一个比较明显的 Fe^{3+} 的吸收谷^[21]。与表2土壤有机质含量对比，可以明显看出，土壤类型光谱曲线与有机质含量基本呈现的是随不同土壤类型有机质含量均值的增加，平均反射率降低，但由于黄棕壤、黄红壤的有机质均值含量相差不大，因此平均反射率曲线相接近；由于棕壤样本数偏少，有机质含量对于光谱曲线的影响不是很明显。

表2 不同类型土壤有机质含量统计特征

Table 2 Statistics of organic matter contents in different types of soils

土壤类型 Soil types	样本数 Number of samples	均值 Mean (g kg ⁻¹)	最小值 Minimum (g kg ⁻¹)	最大值 Maximum (g kg ⁻¹)	标准差 SD	变异系数 CV (%)
山地棕壤 Mountain brown soil	3	63.87	57.88	73.56	8.47	13.26
山地黄棕壤 Mountain yellow-brown soil	34	66.89	12.35	150.94	30.27	45.25
山地黄壤 Mountain yellow soil	2	49.75	35.28	64.23	20.47	41.14
山地黄红壤 Mountain yellow-red soil	8	57.41	28.90	86.94	19.24	33.52
山地红壤 Mountain red soil	143	22.43	4.80	100.54	14.15	63.10

2.2 土壤有机质的反射光谱特征

考虑到不同土壤类型有机质含量所反映出来的光谱特征并不明显，为更准确地探析有机质的光谱特征，选择山地红壤这一特定土壤类型的土样来分析，将有机质划分为低于 10 g kg^{-1} 、 $10 \sim 20 \text{ g kg}^{-1}$ 、 $20 \sim 30 \text{ g kg}^{-1}$ 、 $30 \sim 40 \text{ g kg}^{-1}$ 和高于 40 g kg^{-1} 五个等别，每等取光谱及连续统平均曲线（图2）进行对比。

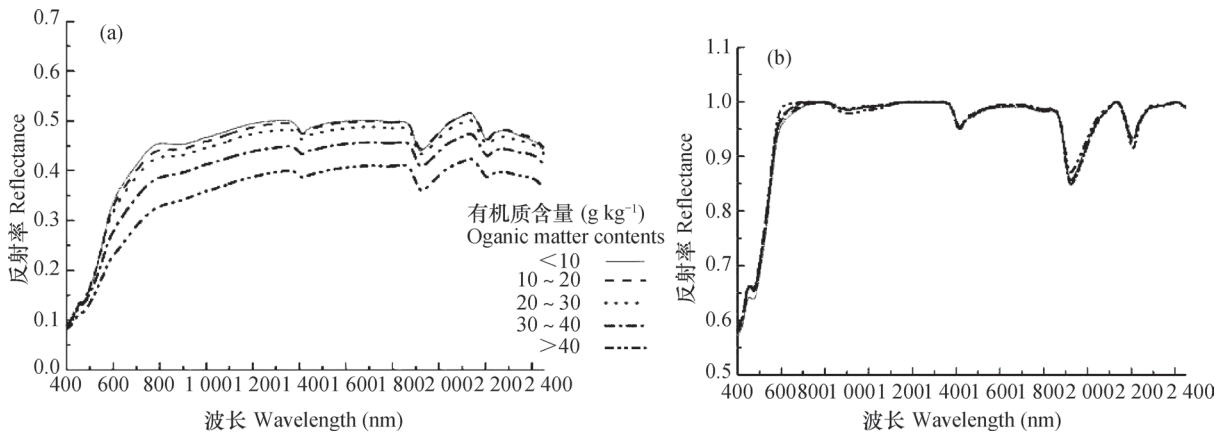


图2 山地红壤有机质含量的反射率 (a) 与连续统去除 (b) 曲线

Fig. 2 Reflectance (a) and continuum removal spectral curves (b) of mountain red soil in soil organic matter content

的光谱反射曲线在 1400 nm 、 1900 nm 和 2200 nm 附近有明显的水分吸收特征，但不同有机质含量等别的土样之间差异不明显，说明这些波段与有机质含量的相关性不大；从反射率曲线可以看出在 $400 \sim 600 \text{ nm}$ 波段存在一个小的吸收谷，通过连续统曲线（图2b）可以看出，在 420 nm 附近有一个小的反射峰， 480 nm 附近有一个小的吸收谷， 600 nm 附近又出现一个较高的反射峰，同时 800 nm 附近有较明显的反射峰，在 900 nm 附近有较明显的吸收谷，并且该反射峰与吸收谷的特征与有机质含量等级有关，有机质含量越低，特征越明显，含量越高，特征越不明显；而在连续统去除曲线中， $420 \sim 480 \text{ nm}$ 附近的波峰及吸收谷特征除小于 10 g kg^{-1} 含量的有机质光谱曲线之外，其他含量的光谱曲线差别不明显， $600 \sim 800 \text{ nm}$ 附近随有机质含量的增加，波峰越明显，而在 $900 \sim 2450 \text{ nm}$ 波段范围，曲线与有机质含量的特征并不是很明显，说明在 $900 \sim 2450 \text{ nm}$ 波段附近的有机质光谱特征并不明显。因此，庐山林地山地红壤有机质含量与反射率之间的关系表现主要为 $400 \sim 600 \text{ nm}$ 及

从反射率曲线（图2a）可以看出，山地红壤的光谱曲线随有机质含量均值的升高，光谱反射率降低，但低于 10 g kg^{-1} 与 $10 \sim 20 \text{ g kg}^{-1}$ 范围的趋势并不是很明显，有机质含量大于 20 g kg^{-1} 的土样趋势更明显。并且在 $400 \sim 600 \text{ nm}$ 波段范围这种差异性不明显，但在 $600 \sim 1400 \text{ nm}$ 波段范围光谱曲线的差异性更明显。

从曲线形态方面可以看出，不同有机质含量

$600 \sim 800 \text{ nm}$ 波段范围，并且在 $600 \sim 800 \text{ nm}$ 波段反射率的影响特征要明显大于 $400 \sim 600 \text{ nm}$ 波段；光谱特征主要反映在 $400 \sim 900 \text{ nm}$ 波段，例如反射率曲线在 800 nm 附近存在反射峰、 900 nm 附近的吸收谷及连续统去除曲线中 420 nm 附近的发射峰、 480 nm 附近的吸收谷和 600 nm 附近的反射峰。这与彭杰^[29]所得出的 $570 \sim 630 \text{ nm}$ 波段是有机质最敏感的波段一致。

2.3 基于土壤有机质间接反演有效铁

从有效铁含量与光谱的相关曲线与有机质含量的光谱相关曲线（图3）的对比可以看出，相关曲线高度相似，尤其在 $600 \sim 1850 \text{ nm}$ 波段范围，趋势接近完全一致，但在 1900 nm 波段附近，有效铁与有机质的相关曲线略有不同，有机质的相关性要明显高于有效铁，且幅度有所增加，说明土壤有机质对土壤有效铁的高光谱特征具有很强的掩盖作用，因此，通过土壤光谱对土壤有效铁含量进行直接反演比较困难。

可见近红外光谱分析技术已逐渐成为土壤有机质、全氮等多种理化参数分析常用的方法，土壤有

表3 不同土壤有机质含量统计特征

土壤有机质含量分等 Soil organic matter content classification (g kg ⁻¹)	样本数 Number of samples	均值 Mean (g kg ⁻¹)	最小值 Minimum (g kg ⁻¹)	最大值 Maximum (g kg ⁻¹)	标准差 SD	变异系数 CV (%)
< 10	21	7.61	4.80	9.88	1.44	18.93
10 ~ 20	54	15.67	10.02	19.74	2.94	18.78
20 ~ 30	42	24.89	20.37	29.40	2.55	10.24
30 ~ 40	16	34.02	30.55	39.71	3.26	9.57
> 40	10	61.10	40.09	100.54	18.45	30.19

质是影响土壤光谱在可见近红外区的主要成分, 而由于土壤中Fe等其他成分受土壤性质的影响, 含量的变化较大, 对土壤光谱的影响是不稳定的, 并且有机质对铁氧化物的反演具有明显的影响^[30-33], 因此, 可通过有机质含量与有效铁含量之间的相关性分析探讨土壤有效铁含量与土壤光谱间的间接关系。

从表1中可以发现, 受地理位置与差异影响, 庐山林地土壤样品中有机质含量与有效铁含量的离散程度均比较高, 且离散程度很接近, 从图3中土壤有机质含量与有效铁含量与土壤光谱的相关曲线可以看出, 相关曲线趋势趋于一致, 可见, 庐山林地土壤样品的有机质含量与有效铁含量之间存在一定的相关性。并且土壤有机质含量的光谱特征明

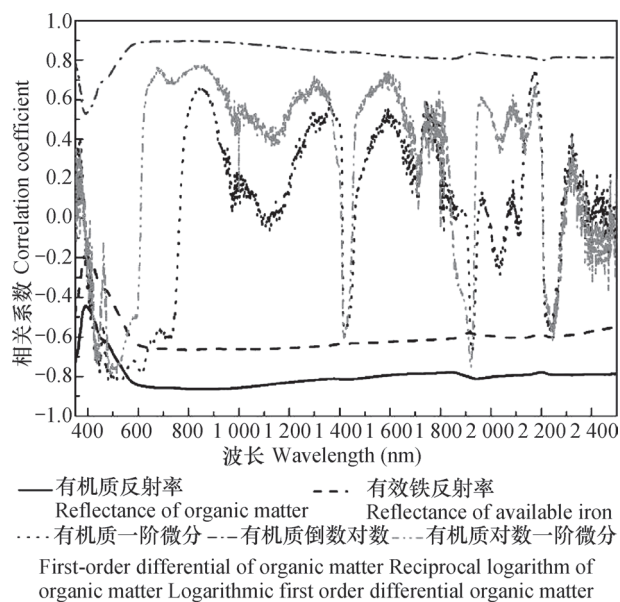


图3 土壤有效铁、土壤有机质含量与反射率及其变化光谱的相关性

Fig. 3 Correlation of reflectance and its variation spectra with soil organic matter and soil available iron content

显, 可通过土壤光谱对土壤有机质含量进行直接反演。由于有机质含量与有效铁含量之间存在一定的相关性, 因此, 为通过构建土壤有机质含量与土壤有效铁含量之间的相关模型, 进而对土壤有效铁含量进行土壤光谱的间接反演提供了可能。

将土壤有机质含量与土壤有效铁含量进行 Pearson 相关分析, 得出在 0.01 水平 (双侧) 上的相关系数为 0.726, 呈显著相关。进而可得到有效铁含量随有机质含量变化的散点图和拟合线性方程。由图4可以看出, 随着有机质含量的升高, 有效铁含量呈线性上升, 因此可得出有机质含量与有效铁含量之间的二项式拟合方程。

而通过土壤有机质含量与变化光谱的相关曲线 (图3) 可以看出, 在 403 ~ 525 nm 可见光区域附近, 经过一阶微分和对数一阶微分变换处理后的相关性要较原始光谱反射率显著提高, 但是

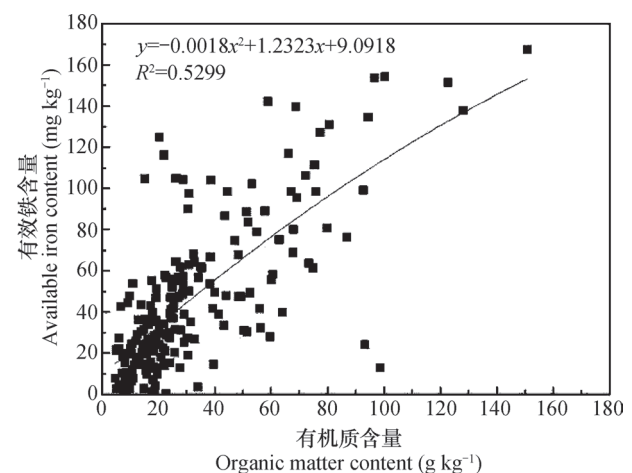


图4 有机质含量与有效铁含量的散点图及拟合方程

Fig. 4 Scatter plot and binomial equation of organic matter contents and available iron contents

549nm ~ 2 400nm可见光—近红外波段范围内，土壤有机质含量与原始光谱反射率呈显著的负相关，并且相关性大于一阶微分和对数一阶微分处理后的光谱反射率，而经过倒数对数变化处理后的相关性与原始光谱反射率是呈对称的关系。因此，从土壤反射光谱与有机质含量的相关曲线可以看出，土壤反射光谱与有机质含量在可见光—近红外波段呈连续的相关性，且对应的波段范围较大，因此，本文在构建土壤有机质含量的光谱模型中选用原始光谱数据进行建模。

2.4 单一模型的选择与验证结果

单一预测模型的选择与相应权重系数的确定是组合模型预测精度的两个重要影响因素，单一模型的精度对组合模型的预测结果至关重要，模型间的预测误差越小组合模型的预测误差就越小，同时模型数量并不是越多越好 [24]。

考虑偏最小二乘回归 (PLSR) 是光谱分析中最为经典和常用的线性建模方法，非线性处理方法中人工神经网络具有较强的处理能力，而径向基函数 (RBF) 神经网络模型较应用最广的BP神经网络模型在算法方面要快很多，更适合预测结构未知对象的模型，函数逼近能力和精度方面均要优于BP神经网络模型 [21, 25, 34]。本文选择PLSR和RBF神经网络模型作为单一预测模型。PLSR对于模型线性部分有很好的分析效果，RBF神经网络模型则对模型的非线性部分有很好的逼近分析效果。

分别建立土壤有机质含量PLSR和RBF光谱预测模型，将得到的预测值通过有机质含量与有效铁含量的二项式获取有效铁含量的间接预测值。图5a和图5b分别为实测土壤有机质含量和PLSR、RBF单个模型验证结果，图5c和图5d为实测土壤有效铁含量和单个模型间接验证结果。对比发现，对

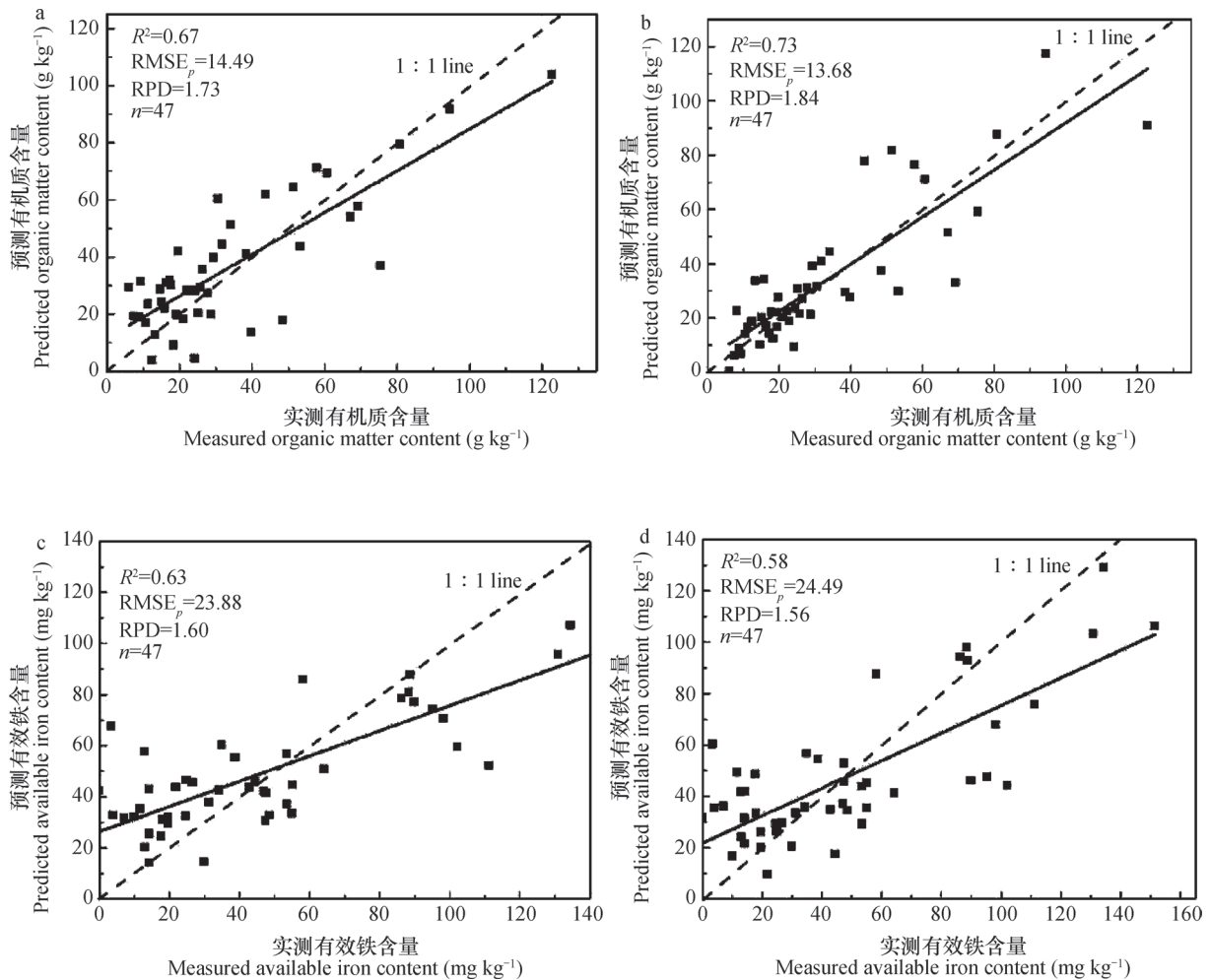


图5 土壤有机质、有效铁预测模型偏最小二乘回归 (a, c)、径向基函数 (b, d) 结果比较

Fig. 5 Comparison between models of Partial Least Squares Regression (PLSR) (a, c) and Radial Basis Function (RBF) (b, d) in prediction of soil organic matter content and available iron content

有机质含量的直接预测验证中, RBF模型的验证结果要好于PLSR模型, 其验证 R^2 由0.68提高至0.72, $RMSE_p$ 由14.49 $g\ kg^{-1}$ 降低至13.68 $g\ kg^{-1}$, RPD由1.74提高至1.83, RBF模型的拟合曲线较PLSR模型更加接近1:1线, 尤其是对于有机质含量较高的土样PLSR模型存在明显的低估现象, 说明RBF模型在有机质含量光谱直接预测验证中, 在解决高值低估的问题上较PLSR有更好的优势; 但在有效铁含量的间接预测验证中, RBF模型的验证结果与PLSR模型拟合效果相差不大, 甚至精度要略低于PLSR模型, 这主要是因为有效铁含量的预测是通过二项式线性模型进行间接预测, 而PLSR模型也

属于线性模型, 间接拟合过程中误差小, 而RBF模型属于非线性模型, 间接拟合过程中易造成预测的误差偏大。但总体比较, RBF模型在土壤属性含量的光谱直接预测的优势比较明显。

2.5 不同权重组合模型与验证结果

从验证结果(表4)可以看出, 无论在土壤有机质含量的直接预测还是在土壤有效铁含量的间接预测验证中, 组合模型与单个预测模型相比, 决定系数 R^2 更高, 均方根误差($RMSE_p$)与预测偏差比RPD效果更好, 因此综合比较, 说明用组合模型来预测土壤属性含量较单个预测模型精度更高、效果更好。

表4 不同预测模型的土壤有机质与有效铁含量验证结果

Table 4 Results of organic matter and soil available iron contents using different models

模型 Model	有机质 Organic matter			有效铁 Available iron		
	R^2	均方根误差 $RMSE_p$ ($g\ kg^{-1}$)	预测偏差比 RPD	R^2	均方根误差 $RMSE_p$ ($mg\ kg^{-1}$)	预测偏差比 RPD
偏最小二乘回归PLSR	0.68	14.49	1.74	0.63	23.88	1.6
径向基函数RBF	0.73	13.68	1.84	0.58	24.49	1.56
RBF组合模型RBF combination model						
算术平均权系数 ^①	0.81	11.87	2.12	0.68	21.86	1.75
预测平方和倒数权系数 ^②	0.79	12.32	2.04	0.62	23.46	1.63
熵值权系数 ^③	0.81	11.54	2.18	0.7	21.6	1.77

①Mean weight coefficient; ②Predicted squared and reciprocal weight coefficients; ③Entropy weight coefficient

而从表4的组合模型可以看出, 不同的权系数方法验证结果也存在差异, 本文中三种权系数确定方法对比可以看出, 算术平均权系数与熵值权系数组合模型的验证结果相差不大, 优势比较明显, 而熵值权系数法的验证结果要略优于算术平均权系数法, 在土壤有机质含量直接建模的验证结果中, 熵值权系数组合模型的决定系数(R^2)将单个预测模型最低的0.68提高至0.81, 将单个模型的预测偏差比(RPD)从1.74提高至2.18, 而在土壤有效铁含量间接建模的验证结果中, 熵值权系数法将单个间接预测模型的决定系数(R^2)从0.58提高至0.70, 预测偏差比(RPD)从1.56提高至1.77。由于本文只选用了两种不同的单个预测模型, 一种是线性模型、另一种是非线性模型, 因此权系数确定方法中倒数权系数法由于只考虑两个单个模型的整体预测误差, 受单个模型数量限制及对单个模型具体影响

因素考虑不周的因素影响, 预测结果没有算术平均权系数及熵值权系数组合模型的优越性明显。

3 结论

通过土壤有效铁与有机质之间的相关性分析, 本研究建立的RBF神经网络组合模型对于庐山林地土壤有效铁含量的间接预测具有一定的可行性, 且通过组合模型在土壤有机质含量直接预测的结果可以看出, 对比传统的PLSR模型预测精度有了显著的提高。而不同的权系数法对于组合模型的预测精度会略有不同, 为组合模型在土壤光谱其他属性含量反演中的应用提供了思路。

参考文献

- [1] 苟文平, 刘世全, 张世熔, 等. 西藏土壤有效铁含量及

- 其影响因素. 山地学报, 2007, 25 (3): 359—363
Gou W P, Liu S Q, Zhang S R, et al. Status of soil available iron and its affecting factors in Tibet (In Chinese). *Journal of Mountain Science*, 2007, 25 (3): 359—363
- [2] 杨媛媛, 李占斌, 宋晓强, 等. 丹江水源区小流域土壤有效铁空间分布特征. 西安理工大学学报, 2015, 31 (2): 201—206
Yang Y Y, Li Z B, Song X Q, et al. Spatial distribution of soil available iron in a small watershed of Danjiang River water resource district (In Chinese). *Journal of Xi'an University of Technology*, 2015, 31 (2): 201—206
- [3] 范玉兰, 黄晓成, 卢映琼, 等. 赣南地区脐橙园土壤有效铁含量分布特征研究. 中国南方果树, 2015, 44 (3): 53—58
Fan Y L, Huang X C, Lu Y Q, et al. Study on the distribution characteristics of soil available iron content in Gannan navel orange orchard (In Chinese). *South China Fruits*, 2015, 44 (3): 53—58
- [4] Adiloglu A. Determination of suitable chemical extraction methods for the available iron content of brown forest soil in turkey. *Eurasian Soil Science*, 2006, 39 (9): 961—967
- [5] 田焯, 沈润平, 丁国香. 支持向量机在土壤镁含量高光谱估算中的应用. 土壤, 2015, 47 (3): 602—607
Tian Y, Shen R P, Ding G X. Application of support vector machine on soil magnesium content estimation based on hyper-spectra (In Chinese). *Soils*, 2015, 47 (3): 602—607
- [6] 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. 农业工程学报, 2015, 31 (14): 103—109
Yu L, Hong Y S, Geng L, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31 (14): 103—109
- [7] 吕杰, 徐静, 闫振国. 基于小波神经网络的矿区土壤铜含量反演研究. 矿业研究与开发, 2015, 35 (4): 68—70
Lü J, Xu J, Yan Z G. Remote sensing inversion on copper content in mine tailings based on wavelet neural network (In Chinese). *Mining Research and Development*, 2015, 35 (4): 68—70
- [8] 章文龙, 曾从盛, 高灯州, 等. 闽江河口湿地土壤全磷高光谱遥感估算. 生态学报, 2015, 35 (24): 8085—8093
Zhang W L, Zeng C S, Gao D Z, et al. Estimating the soil total phosphorus content based on hyper-spectral remote sensing data in the Min River estuarine wetland (In Chinese). *Acta Ecologica Sinica*, 2015, 35 (24): 8085—8093
- [9] Chodak M, Ludwig B, Khanna P, et al. Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands. *Journal of Plant Nutrition and Soil Science*, 2002, 165 (1): 27—33
- [10] Kemper T, Sommer S. Estimate of heavy metal contamination in soil after a mining accident using reflectance spectroscopy. *Environmental Science & Technology*, 2002, 36 (12): 2742—2747
- [11] Siebielec G, McCarty G W, Stuczynski T I, et al. Near-and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *Journal of Environmental Quality*, 2004, 33 (6): 2056—2069
- [12] Ren H Y, Zhuang D F, Singh A N, et al. Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study. *Pedosphere*, 2009, 19 (6): 719—726
- [13] 李巨宝, 田庆久, 吴均昭. 浚阳河两岸农田土壤Fe、Zn、Se元素光谱响应研究. 遥感信息, 2005 (3): 10—13
Li J B, Tian Q J, Wu J Z. The study of spectral responses of agricultural soils for Fe, Zn and Se in the area of Fuyang riversides (In Chinese). *Remote Sensing Information*, 2005 (3): 10—13
- [14] 彭杰, 李曦, 周清, 等. 氧化铁对有机质光谱特性的影响分析. 遥感学报, 2013, 17 (6): 1396—1412
Peng J, Li X, Zhou Q, et al. Influence of iron oxide on the spectral characteristics of organic matter (In Chinese). *Journal of Remote Sensing*, 2013, 17 (6): 1396—1412
- [15] 彭杰, 向红英, 周清, 等. 土壤氧化铁的高光谱相应研究. 光谱学与光谱分析, 2013, 33 (2): 502—506
Peng J, Xiang H Y, Zhou Q, et al. Influence of soil iron oxide on VNIR diffuse reflectance spectroscopy (In Chinese). *Spectroscopy and Spectral Analysis*, 2013, 33 (2): 502—506
- [16] 严晓明, 郑之. BP与RBF网络在一些非线性曲线拟合上性能的比较. 福建农林大学学报(自然科学版), 2011, 40 (6): 653—656
Yan X M, Zheng Z. Performance comparison between some nonlinear curves fitting with BP and RBF (In Chinese). *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 2011, 40 (6): 653—656
- [17] Bates J M, Granger C W J. The combination of

- forecasts. *Operational Research Society*, 1969, 20 (4): 451—468
- [18] 王景明, 卢志红, 吴建富, 等. 庐山土壤类型的特点与分布规律. *江西农业大学学报*, 2010, 32 (6): 1284—1290
Wang J M, Lu Z H, Wu J F, et al. The characteristics and distribution patterns of the soil types in MT. Lushan (In Chinese). *Acta Agriculturae Universitatis Jiangxiensis*, 2010, 32 (6): 1284—1290
- [19] 鲍士旦. 土壤农化分析. 北京: 中国农业出版社, 1999: 30—34
Bao S D. Soil agri-chemical analysis (In Chinese). Beijing: China Agriculture Press, 1999: 30—34
- [20] 马青杰. 土壤有效性铜、锌、铁、锰简易测定方法. *农业科技与信息*, 2010 (1): 34—34
Ma Q J. A simple method for the determination of copper, zinc, iron and manganese in soil (In Chinese). *Journal of Agricultural Science and Information*, 2010 (1): 34—34
- [21] 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
- [22] 马海蛟, 崔晨风. 基于可见-近红外光谱和神经网络的土壤类型鉴别. *江苏农业科学*, 2014, 42 (4): 284—286
Ma H J, Cui C F. Identification of soil type based on visible and near infrared spectroscopy and neural network (In Chinese). *Jiangsu Agricultural Sciences*, 2014, 42 (4): 284—286
- [23] 高宁, 张建中. MATLAB在RBF神经网络模型中的应用. *农业网络信息*, 2009 (2): 110—116
Gao N, Zhang J Z. Application of RBF neural network model based on MATLAB (In Chinese). *Agriculture Network Information*, 2009 (2): 110—116
- [24] 黄仁东, 张海彬, 杨志辉, 等. 基于多准则的组合预测模型权重研究及其应用. *中南大学学报 (自然科学版)*, 2015, 46 (5): 1778—1785
Huang R D, Zhang H B, Yang Z H, et al. Research and application of multi-criteria combination forecast mode (In Chinese). *Journal of Central South University (Science and Technology)*, 2015, 46 (5): 1778—1785
- [25] 刘旭, 于国祥, 沈西挺. 基于神经网络的预测模型比较研究. *河北省科学院学报*, 2007, 24 (4): 7—12
Liu X, Yu G X, Shen X T. Comparative research on forecasting models based on neural networks (In Chinese). *Journal of the Hebei Academy of Sciences*, 2007, 24 (4): 7—12
- [26] 尹燕平, 卢文喜, 许晓鸿, 等. 基于RBF神经网络的土壤侵蚀预测模型研究. *水土保持研究*, 2013, 20 (2): 25—28
Yin Y P, Lu W X, Xu X H, et al. Research on soil erosion prediction model based on RBF neural network (In Chinese). *Research of Soil and Water Conservation*, 2013, 20 (2): 25—28
- [27] 卫敏, 余乐安. 具有最优学习率的RBF神经网络及其应用. *管理科学学报*, 2012, 15 (4): 50—57
Wei M, Yu L A. A RBF neural network with optimum learning rates and its application (In Chinese). *Journal of Management Sciences in China*, 2012, 15 (4): 50—57
- [28] 刘玉成. 熵法确定权重的地基沉降组合预测模型. *中国科技论文在线*, 2010, 6 (11): 875—878
Liu Y C. A combined prediction model for foundation subsidence using the entropy weight method (In Chinese). *Sciencepaper Online*, 2010, 6 (11): 875—878
- [29] 付华, 许凡, 徐耀松, 等. 基于变权重RBF组合模型的煤与瓦斯突出强度预测. *中国安全科学学报*, 2013, 23 (8): 65—70
Fu H, Xu F, Xu Y S, et al. Prediction of coal and gas outburst intensity based on variable weight RBF combination model (In Chinese). *China Safety Science Journal*, 2013, 23 (8): 65—70
- [30] 彭杰, 周清, 张杨珠, 等. 有机质对土壤光谱特性的影响研究. *土壤学报*, 2013, 50 (3): 517—524
Peng J, Zhou Q, Zhang Y Z, et al. Effect of soil organic matter on spectral characteristics of soil (In Chinese). *Acta pedologica sinica*, 2013, 50 (3): 517—524
- [31] Bo S, Rossel R A V, Mouazen A M, et al. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 2010, 107: 163—214
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
- [32] 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
- [33] 刘艳芳, 卢延年, 郭龙, 等. 基于地类分层的土壤有机质光谱反演校正样本集的构建. *土壤学报*, 2016, 53 (2): 332—341

- Liu Y F, Lu Y N, Guo L, et al. Construction of calibration set based on the land use types in visible and near-infrared (VIS-NIR) model for soil organic matter estimation (In Chinese). *Acta Pedologica Sinica*, 2016, 53 (2): 332—341
- [34] 沈润平, 丁国香, 魏国栓, 等. 基于人工神经网络的土壤有机质含量高光谱反演. *土壤学报*, 2009, 46 (3): 391—397
- Shen R P, Ding G X, Wei G S, et al. Retrieval of soil organic matter content from hyper-spectrum based on ANN (In Chinese). *Acta Pedologica Sinica*, 2009, 46 (3): 391—397

Composite-Model-Based Indirect Reversion of Soil Available Iron Spectrum of Forest Soil in Lushan

XIE Wen ZHAO Xiaomin[†] GUO Xi YE Yingcong LI Weifeng WANG Xiaoyan ZHANG Jiajia
(Key Laboratory of Poyang Lake Watershed Agricultural Resources and Ecology of Jiangxi Province, Jiangxi Agricultural University/College of Forestry, Nanchang 330045, China)

Abstract 【Objective】As iron is one of the nutrient elements essential to plant growth, the content of soil available plays an important role in evolution of forest environment. The technology of hyper-spectral remote-sensing (RS) provides a new means for determination of soil physical and chemical components in laboratory. 【Method】In this study, the relationship between soil organic matter and available iron was used to predict the content of available iron in soil indirectly. Besides the traditional single factor prediction model has its own limitation. In order to solve the problem of errors with the single-factor model, this study brought forth a composite model to improve accuracy of the prediction of soil organic matter contents in forest soils at a regional scale with the Vis-NIR spectrum technique. A total of 190 soil samples were collected from the 0~20 cm soil layers of the forests typical of Lushan region in Jiangxi Province. An ASD FieldSpec3 spectrograph diameter equipped with a high intensity contact probe was used to measure original spectral reflectance of the samples in line with standard procedure of the laboratory conditions, and mean while, the soil samples were analyzed for physical and chemical properties. Out of the 190 soil samples, 143 were picked out as samples for modeling and the remaining 47 verification ones. 【Result】The results showed that a significant positive correlation was found between the contents of soil organic matter and soil available iron, and then the binomial model can be built. Based on the results of spectral inversion of soil organic matter content, the contents of soil available iron were retrievable indirectly. Among the spectral inversion models, based on the full band (400~2450 nm) of soil spectra in this study, PLSR (Partial Least Square Regression) of the optimal linear fitting model and RBF (Radial Basis Function) neural network of the nonlinear fitting model were selected to form a combination to figure out arithmetic mean weight coefficients and to project an optimal combination model based on squared, reciprocal and entropy weight coefficients. Accuracies of the predictions of soil available iron content were evaluated by root mean squared error ($RMSE_p$), ratio of partial deviation (RPD) and determination coefficients (R^2). Results show that the combination model is superior to the two separate models in prediction accuracy. Among the combination models, the entropy weight coefficient combination model is the best, with determination coefficient (R^2) in verification model, root mean squared error ($RMSE_p$) and ratio of standard deviation of determination to standard deviation (RPD) of the soil organic matter prediction being 0.81, 11.54 g kg⁻¹ and 2.18, the soil available iron indirect prediction being 0.70, 21.60 mg kg⁻¹ and 1.77, respectively. The combination model is able to make use to a maximum margin of various information of the samples for prediction, reduce effectively the impacts of random factors

in using single prediction models, enhance prediction stability and raise prediction capability of the models.

【Conclusion】 All the findings of the study demonstrate that it is feasible to in directly predict soil available iron contents in forest soils by making use of hyper-spectral RS data. In the end, it can be concluded that the combination model can play a pretty good role in predicting soil organic matter content and indirect predicting soil available iron content.

Key words Soil spectrum; Prediction of soil available iron; RBF neural network; Partial least squares regression (PLSR) ; Combination model

(责任编辑: 檀满枝)