

# 不同利用方式下土壤有机质和全磷的 可见近红外高光谱反演\*

薛利红<sup>1†</sup> 周鼎浩<sup>2</sup> 李颖<sup>2</sup> 杨林章<sup>1</sup>

(1 江苏省农业科学院, 南京 210014)

(2 中国科学院南京土壤研究所, 南京 210008)

**摘要** 以太湖流域直湖港小流域稻田、桃园和菜地的土壤样本为研究对象, 研究了不同光谱建模方法和土地利用方式对土壤有机质和全磷高光谱反演的影响。结果表明: (1) 偏最小二乘回归分析 (Partial least squares regression, PLSR) 模型的建模和预测精度较高且稳定; 神经网络中广义回归神经网络 (Generalized regression neural network, GRNN) 网络预测精度较高但易出现过拟合现象, 反向传播神经网络 (Back propagation neural network, BPNN) 网络比较稳健但精度略低; 偏最小二乘与神经网络相结合则可综合两者优点, 改善复杂样本下的预测精度。(2) 土壤有机质的光谱反演结果优于全磷。3 种土地利用方式中, 稻田的预测效果总体优于桃园和菜地。在当前研究区域内土地利用方式对土壤有机质光谱反演影响不大, 但对全磷反演影响较大。今后利用光谱对土壤全磷反演时需分土地利用方式对模型进行校准。

**关键词** 高光谱; 偏最小二乘回归; 神经网络; 土壤有机质; 土壤全磷; 土地利用方式

**中图分类号** S153.2 **文献标识码** A

获取土壤养分的空间分布信息是控制农田面源污染、精确施肥的关键。传统的化学分析方法需要破坏性采集样品, 劳动强度大, 费时费力, 成本相对较高。而光谱分析是一种快速、低廉且无损的分析方法。自从 1965 年光谱技术首次被用来分析土壤质地和水分含量以来, 很多学者利用土壤的可见-近红外光谱对土壤特性如水分、有机质、碳、氮、铁、阳离子交换量等进行了调查研究, 结果均表明利用可见-近红外光谱技术尤其是光谱分辨率更高的高光谱技术能够成功地预测这些土壤特性<sup>[1-4]</sup>。

由于不同土壤组分对光谱的吸收叠加导致了土壤高光谱的非特异性, 再加上土壤结构等引起的散射效应, 使得土壤光谱呈现出十分复杂的吸收模式, 需要对这些光谱进行数据提取并与土壤属性进行回归建模。目前应用较多的有多元线性回归、主成分分析、偏最小二乘回归分析 (Partial least square regression, PLSR) 以及神经网络 (Artificial neu-

ral network, ANN)<sup>[5-8]</sup>。其中, 偏最小二乘回归法在处理自变量存在多重共线性和自相关问题时具有优势<sup>[9]</sup>, 在处理线性回归问题时具有较高的精度, 应用最为广泛<sup>[10]</sup>。神经网络不需要精确的数学模型, 没有任何对变量的假设 (如正态性、独立性等) 要求, 可以弥补传统统计方法的不足, 已成为光谱分析数据处理方法中的热点之一。研究表明, 神经网络在处理非线性问题时较一般方法要好, 能获得较经典最小二乘法更高的灵敏度<sup>[11-12]</sup>。

除了科学有效的光谱建模方法外, 土壤本身属性如土壤类型、潮湿度、粗糙度、养分变异等也显著影响着土壤养分光谱反演的精度及其可靠性<sup>[9, 13-15]</sup>。研究表明建模和验证时土壤样本数据集的选择是关系到模型精度及其稳定性的重要因素, 这就要求样本数据集的大小和分布要具有代表性, 不仅能代表土壤本身属性的变异情况, 还要能代表光谱的变异情况<sup>[16]</sup>。土地利用方式的不同造成土

\* 国家自然科学基金项目 (40901104, 41171235)、中国科学院知识创新工程重要方向项目 (KZCX2 - YW - QN406) 和江苏省自主创新项目 (CX(13)3014) 资助

† 通讯作者

作者简介: 薛利红 (1977—), 女, 河南修武人, 研究员, 硕士生导师, 主要从事光谱在农业生态中的应用及农业面源污染控制方面的研究工作。E-mail: lhxue@issas.ac.cn

收稿日期: 2013-03-28; 收到修改稿日期: 2013-07-29

壤养分等属性发生变化,相应的光谱特征也发生变化。然而目前关于土地利用方式对土壤属性光谱反演的影响研究相对较少。而利用人工神经网络及其与偏最小二乘结合的方法对不同土地利用方式下土壤养分含量的光谱研究更为有限。

为此,本文比较了偏最小二乘回归法和人工神经网络两种方法在土壤养分反演建模中的优劣,并尝试用二者相结合的方法来进一步提高反演精度。在此基础上,比较稻田、桃园和菜地 3 种不同土地利用方式对土壤有机质和全磷含量光谱反演的影响,探索是否能用一个通用的模型对不同土地利用方式的土壤进行预测,以期为今后土壤养分的光谱反演提供最佳的建模方法以及为模型应用时的校准提供相关参考。

## 1 材料与方法

### 1.1 样本采集与制备

稻田、桃园和菜地的土壤均采自太湖流域直湖港小流域。该流域土壤类型主要为爽水型水稻土(乌底黄泥土),稻麦轮作、菜地和水蜜桃园为目前主要的土地利用方式,其中桃园和菜地多由稻田改种而来,种植年限不一。共采集样点数为 178 个,其中稻田样品 45 个,桃园样品 67 个,菜地样品 66 个。采集表层 0~20 cm 土壤,每点呈梅花状采集 5 钻,混合均匀后作为该样点的样品。样品带回室内,自然风干,挑拣出根系等其他杂质,研磨过 100 目筛处理。磨碎后的样品分为两份,一份用作养分含量的化学分析,一份用作室内光谱分析。有机质测定采用重铬酸钾容量一外加加热法,全磷测定采用 NaOH 碱熔钼锑抗比色法<sup>[17]</sup>。

### 1.2 光谱测定

采用 ASD 公司产的 FieldSpec 3 Hi-Res (350~2 500 nm) 对水稻土风干土样进行了室内高光谱测定,采样间隔分别为 1.4 nm (350~1 050 nm) 和 2 nm (1 000~2 500 nm)。光谱分辨率为 3 nm (350~700 nm)、8.5 nm (700~1 400 nm) 和 6.5 nm (1 400~2 100 nm),测定范围为 350~2 500 nm,探头视场角为 25°,室内照明光源采用与光谱仪配套的 2 个 14.5W 卤素灯。将制备好的土壤样品均匀平铺于直径约 6.7 cm、深度约 0.8 cm 的铝盒中。探头距土壤样品表面约 13 cm。探头入射角 90°,室内光源入射角为 45°。每次光谱测定时均进行白板校正,每个样本扫描 5 次,取其平均值作为最终光谱反

射率数据。

### 1.3 光谱模型建立以及验证

**1.3.1 光谱数据处理方法** 用光谱仪自带软件 Viewspec 中的 splice correction 功能对光谱曲线进行断点拼接,并去除 350~380 nm 和 2 400~2 500 nm 噪声波段。利用 Unscrambler 软件对光谱曲线进行平滑去噪处理,平滑方法为 moving average。最后对光谱曲线进行 10 nm 重采样。

**1.3.2 光谱建模方法** 本研究分别采用偏最小二乘回归法(PLSR)、BP 神经网络(Back Propagation Neural Network, BPNN)、广义回归神经网络(Generalized Regression Neural Network, GRNN)以及 PLSR 与人工神经网络相结合的方法(PLSR + BPNN 和 PLSR + GRNN)分别建模。PLSR 采用降维的思想,在不丢失数据信息的同时精简自变量(光谱信息),去除噪声及压缩数据,并考虑了因变量即土壤养分含量信息,能更完整地表达土壤养分和光谱数据间的相关关系,建立的模型具有更好的鲁棒性。本研究在建模过程中采用了留一交叉验证的方法。BPNN 是目前使用较为广泛的一种多层前馈神经网络<sup>[18]</sup>,网络具有一定推广和概括功能<sup>[19]</sup>。其中隐藏层节点数的选择是 BP 神经网络构建的关键,节点数太少,预测误差会较大,而节点数过大,网络学习时间较长,易出现过拟合现象<sup>[10]</sup>。本研究经过反复尝试对比,最终选择单隐藏层,隐藏层节点为 7,训练函数为 Levenberg-Marquardt 算法的 trainlm,隐藏层传递函数为 tansig,输出层传递函数为 purelin。但 BPNN 在函数逼近时存在收敛速度慢和局部极小等问题,在解决样本量少且噪声较多的问题时效果并不理想,而 GRNN 在逼近能力、分类能力和学习速度等方面具有更强的优势,并且在数据缺乏时效果也较好<sup>[20-21]</sup>。本研究同时用 GRNN 来建立模型。其中,平滑参数(SPREAD)的取值是影响 GRNN 模型的关键。本研究经过反复演算对比,最终选择 SPREAD 值为 1。由于神经网络每次运算结果有一定波动,因此取 10 次运算结果的平均值作为最终结果。

为了解决人工神经网络运用时由于输入变量过多容易出现的过拟合和运行时间较长<sup>[10]</sup>等问题,本研究尝试以 PLSR 模型中得出的最优主成分作为 BPNN 和 GRNN 中的输入变量,分别建立了光谱反演土壤养分含量的 PLSR + BPNN 模型以及 PLSR + GRNN 模型。

**1.3.3 模型的建立与评价** 随机选取稻田、桃园、菜地样本的 2/3 作为建模样本集,剩余 1/3 样本作为预测样本集,分别对不同土地利用方式下的土

壤进行了建模(简称单一模型)。为了进一步比较土地利用方式对光谱反演模型精度及稳定性的影响,将3种土地利用方式下的建模样本集混合,建立综合土地利用方式反演模型(简称综合模型),再利用综合模型对不同土地利用方式下的预测样本集进行预测,并和单一模型的预测结果进行比较。建模过程中采用留一法交叉验证。模型的评价指标采用建模验证决定系数  $R_{val}^2$ 、建模验证均方根误差  $RMSE_{val}$ 、建模验证  $RPD_{val}$ 、预测决定系数  $R_{pre}^2$ 、预测均方根误差  $RMSE_{pre}$  以及预测  $RPD_{pre}$  [22-23]。当  $RPD > 3$  时,模型非常成功,  $2.25 < RPD < 3$  时,模型较为成功,  $1.75 < RPD < 2.25$  时,模型能较好地预测 [24]。为了比较综合模型和单一模型的反演效果,计算了单一模型和综合模型的预测  $R_{pre}^2$ 、 $RMSE_{pre}$  和  $RPD_{pre}$  的相对差异(%),即相对差异 = (单一模型 - 综合模型)/单一模型。对于  $R_{pre}^2$  和  $RPD_{pre}$  而言,若相对差异为正,代表单一模型优于综合模型,对于  $RMSE_{pre}$ ,则相反。差异越大,表明土地利用方式对光谱反演模型稳定性的影响也越大。

#### 1.4 数据处理

PLSR 分析在 Unscrambler9.7 下进行,人工神经

网络分析在 Matlab2010a 中进行。利用 Excel 和 Sigmaplot10.0 进行作图。

## 2 结果

### 2.1 不同土地利用方式下土壤养分含量及其光谱变异情况

稻田、桃园和菜地中有机质和全磷含量均具有较好的梯度分布(表1)。其中桃园有机质含量最高,最大值可达  $64.67 \text{ g kg}^{-1}$ ,这与桃园施入大量的有机肥有关,稻田和菜地有机质水平相当。不同土地利用方式下土壤全磷含量差异较大,菜地最高,平均值为  $1.13 \text{ g kg}^{-1}$ ,其次为桃园,稻田则最低,仅有  $0.38 \text{ g kg}^{-1}$ 。偏度和峰度分析表明,稻田、桃园和菜地的有机质和全磷含量均符合正态分布。图1分别展示了  $380 \sim 2400 \text{ nm}$  下稻田、桃园和菜地土壤样本的光谱平均值及其变化范围。由图可见,3种土地利用方式下光谱曲线形态无较大变化,均表现为随着波长的增加反射率逐渐上升,在  $1400 \text{ nm}$ 、 $1900 \text{ nm}$  和  $2200 \text{ nm}$  左右出现了3个明显的吸收谷。稻田和菜地的光谱反射率略高于桃园。

表1 土壤样品基本信息  
Table 1 Basic information of soil samples

土地利用方式 Land use	土壤养分 Soil nutrient	样本数 Sample size	最大值 Maximum ( $\text{g kg}^{-1}$ )	最小值 Minimum ( $\text{g kg}^{-1}$ )	平均值 Mean ( $\text{g kg}^{-1}$ )	标准差 Standard deviation ( $\text{g kg}^{-1}$ )	偏度 Skewness	峰度 Kurtosis
稻田 Paddy field	有机质 OM 全磷 TP	45	48.20 0.74	4.41 0.02	31.64 0.38	11.90 0.22	-0.794 -0.015	-0.312 -1.379
桃园 Peach orchard	有机质 OM 全磷 TP	67	64.67 1.24	18.95 0.09	36.79 0.58	9.74 0.33	0.649 0.263	0.655 -0.999
菜地 Vegetable garden	有机质 OM 全磷 TP	66	47.04 2.43	8.39 0.16	29.86 1.13	9.03 0.62	-0.660 0.494	-0.024 -1.014

### 2.2 建模方法对土壤有机质和全磷光谱反演的影响

由表2可见,对于单一模型,无论采用何种建模方法,稻田的有机质光谱反演结果均优于桃园和菜地,菜地相对最差。不同建模方法中,PLSR 和 PLSR + BPNN 效果较好。稻田的  $R_{val}^2$  达 0.92 以上,  $RMSE_{val}$  分别为  $3.44$  和  $3.23 \text{ g kg}^{-1}$ ,  $RPD_{val}$  分别为 3.5 和 3.7,  $R_{pre}^2$  分别为 0.82 和 0.75,  $RMSE_{pre}$  为  $5.03$  和  $6.26 \text{ g kg}^{-1}$ ,说明本研究建立的稻田有机质光谱反演模型非常成功。在 PLSR 下,桃园和菜地的  $R_{pre}^2$  均在 0.75 以上,表明桃园和菜地的 PLSR 模型也可以

较好地预测有机质。

对比 BPNN 和 GRNN 两种神经网络,发现 GRNN 在建模时表现略优于 BPNN,但预测时 GRNN 对桃园的预测决定系数  $R_{pre}^2$  仅为 0.19,  $RMSE_{pre}$  达  $9.08 \text{ g kg}^{-1}$ ,表明出现了过拟合现象。以 PLSR 的最佳主成分作为 GRNN 的输入变量后,过拟合问题得到了有效的改善,PLSR + GRNN 模型下桃园的  $R_{pre}^2$  提高至 0.78,  $RMSE_{pre}$  降低至  $7.74 \text{ g kg}^{-1}$ 。BP 网络的预测效果较为稳定,PLSR + BPNN 模型的预测精度提高不大。

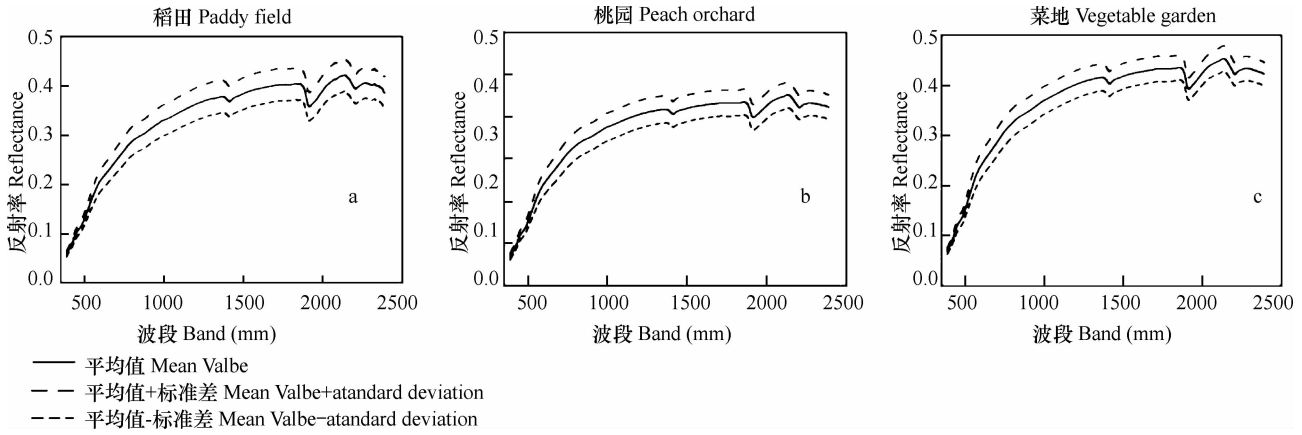


图 1 不同土地利用方式下土壤样本光谱平均值与标准差(a 为稻田,b 为桃园,c 为菜地)

Fig. 1 Mean value and standard deviation of soil spectrum relative to land use (a: Paddy field; b: Peach orchard; c: Vegetable garden)

与有机质反演结果类似,对于单一模型,无论采用何种建模方法,全磷的光谱反演效果依然以稻田最好,桃园和菜地差别不明显(表 3)。不同建模方法中,PLSR、PLSR + BPNN 及 PLSR + GRNN 的效果较好。与 PLSR 相比,PLSR 与人工神经网络结合对预测效果的提高作用相对有限。PLSR 下,稻田

$R^2_{val}$  为 0.77,  $RMSE_{val}$  为  $0.07 \text{ g kg}^{-1}$ ,  $RPD_{val}$  达到 3.2,  $R^2_{pre}$  为 0.76,  $RMSE_{pre}$  为  $0.11 \text{ g kg}^{-1}$ ,  $RPD_{pre}$  为 2.0, 表明该模型能成功地对稻田土壤全磷进行反演;桃园和菜地的建模验证和预测的决定系数也均大于 0.6,  $RPD$  均大于 1.5, 表明利用 PLSR 可以对桃园和菜地的全磷进行较好地预测。

表 2 不同土地利用方式下五种建模方法对土壤有机质光谱反演的影响

Table 2 Spectral inversion of soil organic matter using 5 different modeling methods relative to land use

建模方法 Regression model	土地利用方式 Land use	建模 Calibration			预测 Prediction		
		$R^2_{val}$	$RMSE_{val} (\text{g kg}^{-1})$	$RPD_{val}$	$R^2_{pre}$	$RMSE_{pre} (\text{g kg}^{-1})$	$RPD_{pre}$
PLSR	稻田 Paddy field	0.92	3.44	3.5	0.82	5.03	2.4
	桃园 Peach orchard	0.74	4.69	2.1	0.82	4.71	2.3
	菜地 Vegetable garden	0.68	5.22	1.7	0.75	4.79	1.9
BPNN	稻田 Paddy field	0.87	3.99	3.0	0.76	5.89	2.0
	桃园 Peach orchard	0.72	5.35	1.9	0.63	6.34	1.7
	菜地 Vegetable garden	0.64	5.68	1.6	0.51	8.77	1.0
GRNN	稻田 Paddy field	0.96	3.79	3.2	0.82	6.79	1.8
	桃园 Peach orchard	0.81	5.89	1.7	0.19	9.08	1.2
	菜地 Vegetable garden	0.88	4.47	2.0	0.61	7.16	1.3
PLSR + BPNN	稻田 Paddy field	0.93	3.23	3.7	0.75	6.26	1.9
	桃园 Peach orchard	0.74	5.58	1.8	0.66	6.62	1.6
	菜地 Vegetable garden	0.70	5.39	1.7	0.57	7.47	1.2
PLSR + GRNN	稻田 Paddy field	0.94	7.57	1.6	0.71	8.97	1.3
	桃园 Peach orchard	0.82	7.18	1.4	0.78	7.74	1.4
	菜地 Vegetable garden	0.80	6.83	1.3	0.73	7.85	1.2

注: PLSR, 偏最小二乘回归法; BPNN, 反向传播神经网络; GRNN, 广义回归神经网络; 下同 Note: PLSR, Partial least square regression; BPNN, Back propagation neural network; GRNN, Generalized regression neural network. The same below

同样,GRNN 在预测桃园土壤全磷时依然出现了过拟合现象,没有 BP 网络表现稳定。将其与 PLSR 相结合则可以在明显改善这一现象,全磷的建模效果和预测效果均有所提高。这再次证明了

利用 PLSR 选取的主成分可以在有效提取主要信息的同时消除噪声干扰,其与人工神经网络相结合既能有效防止过拟合现象,又能提高模型的精度和稳定性。

表 3 不同土地利用方式下五种建模方法对土壤全磷的光谱反演结果

Table 3 Spectral inversion of soil total phosphorus using 5 different modeling methods relative to land uses

建模方法 Regression model	土地利用方式 Land use	建模 Calibration			预测 Prediction		
		$R^2_{val}$	$RMSE_{val}$ (g kg <sup>-1</sup> )	$RPD_{val}$	$R^2_{pre}$	$RMSE_{pre}$ (g kg <sup>-1</sup> )	$RPD_{pre}$
PLSR	稻田 Paddy field	0.77	0.07	3.2	0.76	0.11	2.0
	桃园 Peach orchard	0.61	0.21	1.6	0.60	0.22	1.5
	菜地 Vegetable garden	0.60	0.40	1.6	0.64	0.40	1.6
BPNN	稻田 Paddy field	0.82	0.10	2.3	0.62	0.14	1.5
	桃园 Peach orchard	0.76	0.17	2.0	0.58	0.23	1.4
	菜地 Vegetable garden	0.74	0.31	2.0	0.40	0.62	1.0
GRNN	稻田 Paddy field	0.91	0.10	2.2	0.62	0.17	1.3
	桃园 Peach orchard	0.89	0.17	2.0	0.15	0.33	1.0
	菜地 Vegetable garden	0.81	0.39	1.6	0.30	0.58	1.1
PLSR + BPNN	稻田 Paddy field	0.88	0.08	2.7	0.76	0.12	1.7
	桃园 Peach orchard	0.77	0.17	2.0	0.63	0.24	1.4
	菜地 Vegetable garden	0.72	0.37	1.7	0.44	0.56	1.1
PLSR + GRNN	稻田 Paddy field	0.91	0.12	1.8	0.80	0.15	1.4
	桃园 Peach orchard	0.80	0.24	1.4	0.77	0.27	1.2
	菜地 Vegetable garden	0.75	0.43	1.4	0.58	0.50	1.2

### 2.3 土地利用方式对土壤有机质和全磷光谱反演的影响

对于有机质而言,混合样本综合建模精度在 0.61~0.77 之间,以 BPNN 及 PLSR + BPNN 较好,

建模精度为 0.77,  $RMSE_{val}$  分别为 5.09 和 5.15 g kg<sup>-1</sup>,  $RPD_{val}$  分别为 2.1 和 2.0(表 4),表明该综合模型能较成功地预测土壤有机质。

表 4 综合土地利用方式下土壤有机质和全磷的光谱建模精度比较

Table 4 Precisions of the five spectral modeling methods for prediction of soil SOM and TP of fields under comprehensive land uses ( $n = 119$ )

土壤参数 Soil parameter	统计指标 Statistic index	PLSR	BPNN	GRNN	PLSR + BPNN	PLSR + GRNN
有机质 OM	$R^2_{val}$	0.69	0.77	0.61	0.77	0.73
	$RMSE_{val}$ (g kg <sup>-1</sup> )	5.86	5.09	8.84	5.15	9.54
	$RPD_{val}$	2.6	2.1	1.2	2.0	1.1
全磷 TP	$R^2_{val}$	0.35	0.67	0.50	0.62	0.50
	$RMSE_{val}$ (g kg <sup>-1</sup> )	0.43	0.32	0.48	0.34	0.52
	$RPD_{val}$	1.2	1.7	1.1	1.6	1.0

对五种建模方法下单一模型和综合模型的  $R^2_{pre}$ 、 $RMSE_{pre}$  和  $RPD_{pre}$  指标的相对差异比较发现(图 2),单一模型仅在 GRNN 及 PLSR + GRNN 建模方法下精度

高于综合模型,其中 GRNN 下桃园由于出现过拟合现象才导致相对差异较大。而在 PLSR + BPNN 建模方法下,单一模型的预测效果明显低于综合模型,  $R^2_{pre}$  下

降了 18%~21%,  $RMSE_{pre}$  增加了 24%~43%, 菜地的  $RPD_{pre}$  下降了 75%。在 PLSR 下, 单一模型仅稻田的预测精度略微增加, 增幅小于 15%。

选择 PLSR 下的综合模型和单一模型分别作有机质实测值与预测值的散点图(图 3), 发现综合模

型对稻田、菜地和桃园的预测效果与单一模型持平, 实测值与预测值均匀分布于 1:1 线附近,  $R^2$  分别为 0.82 和 0.83。表明土地利用方式对土壤有机质光谱反演模型的稳定性影响并不大, 今后进行有机质光谱反演时不必考虑土地利用方式的影响。

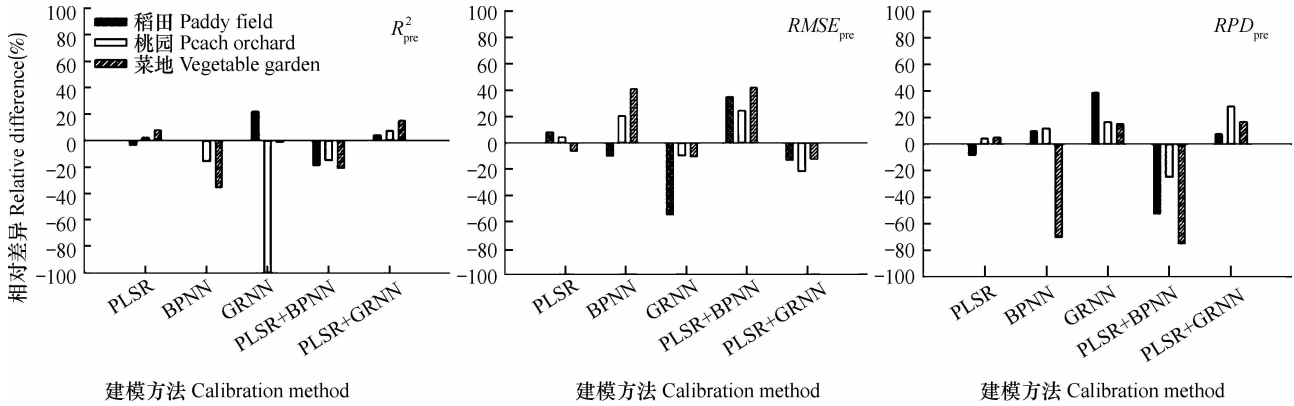


图 2 单一模型和综合模型对土壤有机质预测效果的比较

Fig. 2 Comparison between simple and integrated models in prediction of soil organic matter

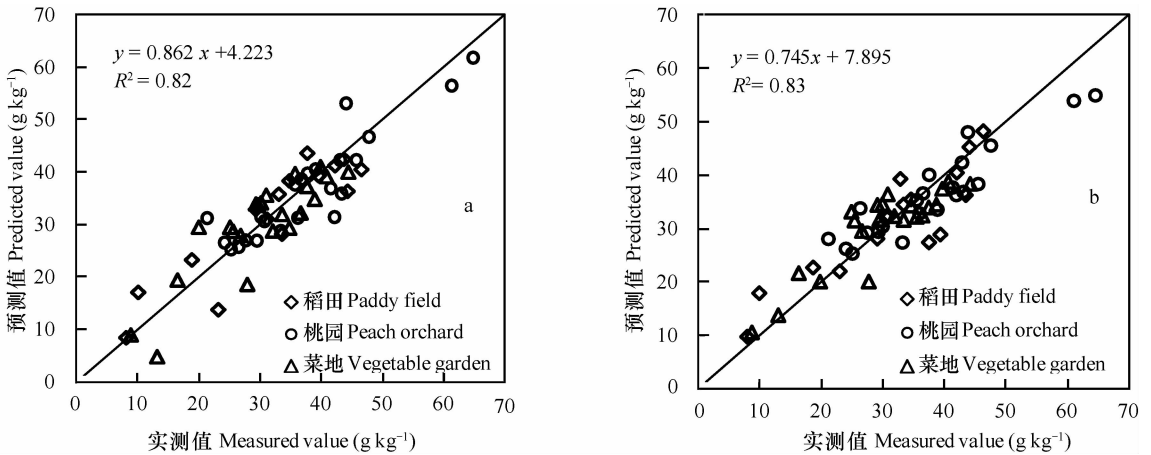


图 3 PLSR 下单一模型(a)与综合模型(b)对土壤有机质的预测效果

Fig. 3 Scatter plot of measured and predicted soil organic matter using simple PLSR model and integrated model, separately ( $n = 59$ )

对于全磷而言, 混合样本综合建模效果不甚理想(表 4),  $R^2_{val}$  在 0.35~0.67 之间,  $RPD_{val}$  在 1.0~1.7 之间。不同建模方法中, 以 PLSR+BPNN 和 BPNN 相对较好。与土壤有机质不同, PLSR 方法的效果要差于人工神经网络, 这可能是由于 PLSR 对于噪声和误差的抗干扰能力不如神经网络。由于全磷没有特征的光谱吸收峰, 其对整个光谱信息的贡献量可能也较小, 当不同土地利用方式的样本进行混合时, 无疑增加了样本的噪声, 从而使得建模精度下降。

对五种建模方法下单一模型和综合模型的  $R^2_{pre}$ 、 $RMSE_{pre}$  和  $RPD_{pre}$  指标的相对差异比较发现(图

4), 稻田的预测效果明显不如桃园和菜地, 与有机质完全不同。除了 BPNN 和 PLSR+BPNN 综合模型对菜地的预测效果略优于单一模型外, 其余  $R^2_{pre}$  和  $RPD_{pre}$  的相对差异均为正值,  $RMSE_{pre}$  均为负值, 表明综合模型效果总体不如单一模型。综合模型与单一模型的相对差异大都在 10% 以上, 尤其是稻田, 相对差异最大可达 100%。

同样选择 PLSR 下的综合模型和单一模型分别作预测样本集土壤全磷实测值与预测值的散点图(图 5), 发现综合模型预测结果明显差于单一模型, 其低估了菜地土壤样本, 却高估了桃园土壤样本, 从而导致实测值与预测值的总体  $R^2$  仅为 0.57, 远低

于单一模型的 0.75。说明土地利用方式对土壤全磷光谱反演模型的稳定性和精度影响较大,今后在

研究全磷光谱反演时必须根据土地利用方式进行分别建模校准。

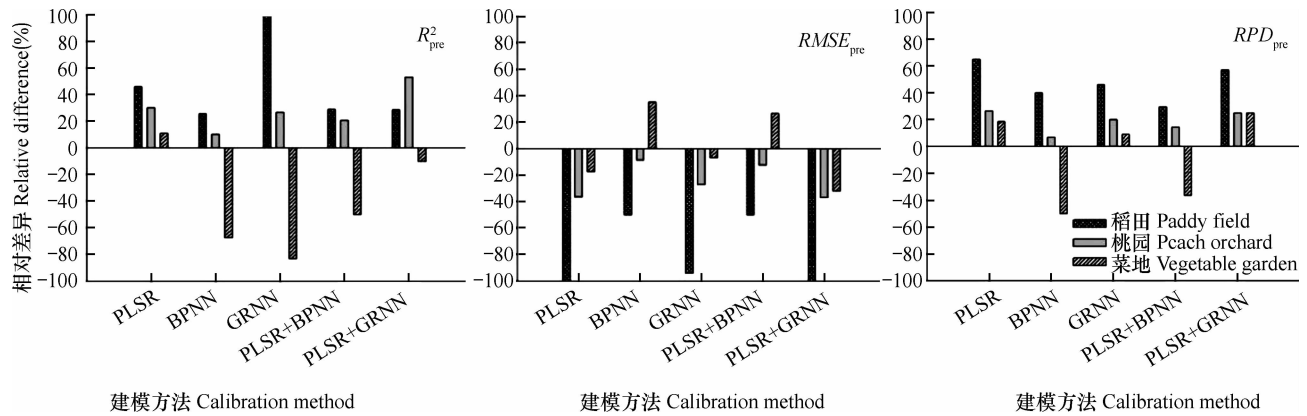


图 4 单一模型和综合模型对土壤全磷预测结果的比较

Fig. 4 Comparison between simple and integrated models in prediction of soil total phosphorus

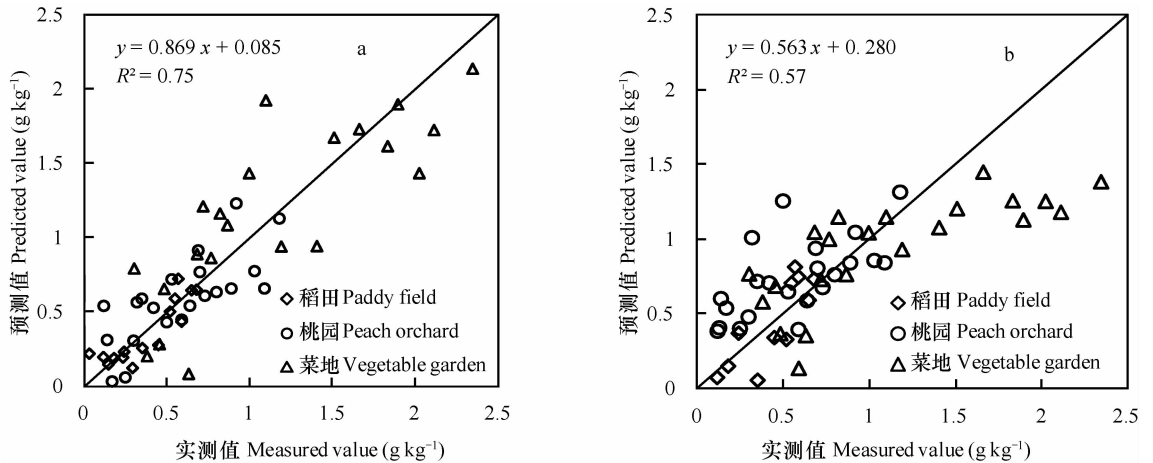


图 5 PLSR 下单一模型(a)与综合模型(b)对土壤全磷的预测结果

Fig. 5 Scatter plot of measured and predicted soil total phosphorus using single PLSR model and integrated model, separately ( $n = 59$ )

### 3 讨论

前人针对外界条件因素(土壤潮湿度、土壤粗糙度、土壤类型等)对光谱分析的研究很多,但土地利用方式的影响却鲜有研究。而建模方法方面,PLSR 和 ANN 结合的方法虽在国外一些文献中有记载<sup>[25-26]</sup>,但应用于土壤有机质和全磷的光谱反演研究很少。

本文对不同建模方法的比较发现,建模和预测精度较高的两种方法是 PLSR 和 PLSR + BPNN。人工神经网络中的 GRNN 网络建模精度略优于 BPNN,但在预测验证时易出现过拟合现象;BPNN 网络相对比较稳定,但运行时间较长,因此,实际选取哪种网络应综合考虑精度要求和时间要求。

PLSR 与 BPNN 相结合,综合了偏最小二乘不易过拟合、可以精简提取信息的优点以及神经网络鲁棒性强、适用范围广等优点,程序运行时间远低于 BPNN 和 GRNN,预测效果较好,值得以后借鉴。Mouazen 等<sup>[10]</sup>比较了主成分回归(PCR)、PLS 与 NN 对土壤属性的光谱预测效果,也发现 PLSR-NN 模型较 PLSR 和 PCR 模型提高了土壤属性光谱预测的精度。

本文研究表明,土壤有机质的光谱反演精度要优于土壤全磷,与前人研究一致<sup>[21]</sup>。这主要是因为土壤有机质是土壤的主要组成成分,具有明显的吸收峰,而全磷在土壤中含量相对较低,多以无机态存在,没有相应的特征吸收峰。国内外大量关于土壤有机质光谱反演的研究均表明,土壤有机质的光谱反演是目前土壤属性中最为成功也最为稳定

的一个<sup>[27]</sup>,这也证实了本文关于土地利用方式对有机质光谱反演影响较小的结论。土地利用方式的不同造成了土壤有机质的含量和质量产生了变异,如稻田土壤有机质主要来源于秸秆和残茬的输入,其主要成分应该是降解的纤维素;桃园和菜地的土壤有机质则主要来源于商品有机肥或粪肥,其组成多为腐殖质。另外由于稻田和旱作的水肥等管理方式不同,导致土壤的颗粒组成也产生了分化,桃园和菜地等旱作土壤其颗粒更为细密,这些均对土壤光谱产生了影响<sup>[27]</sup>,从而影响到有机质和全磷反演的精度。Stenberg 等<sup>[16]</sup>认为较细的颗粒造成光的散射从而掩盖了有机质的吸收,导致了有机质的反演精度下降。这可能是稻田土壤有机质反演精度高于桃园和菜地的一个原因。另外样本来源的变异性也是影响光谱模型精度的一个重要因素,样本集的变异程度越大,其预测效果就越好<sup>[16]</sup>。本研究中稻田土壤样本有机质含量的变异系数(37.6%)要高于桃园和菜地样本,这也是稻田有机质反演精度最高的另一个原因。

土地利用方式对土壤全磷光谱反演影响较大,综合模型的预测精度明显下降,与单一模型的相对差异大于10%。这可能是因为土壤磷没有特征吸收峰,相较有机质而言,土壤类型、土地利用方式及其他养分含量等土壤背景的干扰对其预测影响更为显著,混合样本下的土壤背景较单一土地利用方式下的更为复杂,因此精度有所下降。表明在对土壤全磷进行光谱反演时,需要采用与预测样本相对比较一致的土壤样本对模型先进行校准。土壤全磷的混合反演模型对稻田的预测效果差于桃园和菜地,可能是因为样本范围的不一致问题,稻田全磷含量分布在值较小的区域内(0.02 ~ 0.74 g kg<sup>-1</sup>),混合样本集的上限为2.43 g kg<sup>-1</sup>,而土壤全磷与光谱特征之间的关系并不是单纯的线性关系。这也表明在利用光谱对土壤属性进行反演时,要注意建模数据集和预测数据集的分布范围,两者要相对比较一致时才能获得较好的预测精度。

## 4 结 论

偏最小二乘回归分析模型的建模和预测精度较高且稳定;人工神经网络中广义回归神经网络预测精度较高但易出现过拟合现象,反向传播神经网络网络比较稳健但精度略低;偏最小二乘与人工神经网络相结合则可综合两者优点,改善复杂样本下

的预测精度。土壤有机质的光谱反演结果优于全磷。3种土地利用方式中,稻田的预测效果总体优于桃园和菜地。在目前研究区域内,土地利用方式对土壤有机质光谱反演影响不大,但对全磷反演影响较大,今后利用光谱对土壤全磷反演时需分土地利用方式对模型进行校准。

## 参 考 文 献

- [1] Aichi H, Fouad Y, Walter C, et al. Regional predictions of soil organic carbon content from spectral reflectance measurements. *Biosystems Engineering*, 2009, 104(3): 442—446
- [2] Bowers S A, Hanks R J. Reflection of radiant energy from soil. *Soil Science*, 1965, 100(3): 130—138
- [3] Dalal R C, Henry R J. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Science Society of America Journal*, 1986, 50(1): 120—123
- [4] Volkan B A, van Es H M, Akbas F, et al. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*, 2010, 74(2): 229—238
- [5] Vasques G, Grunwald S, Sickman J. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma*, 2008, 146(1): 14—25
- [6] 王圣毫, 李智, 郑维平, 等. 人工神经网络在近红外光谱建模中的应用及研究现状. *红外*, 2012(8): 9—15. Wang S H, Li Z, Zheng W P, et al. Research and application of ANN method in near infrared spectrum modeling (In Chinese). *Infrared*, 2012(8): 9—15
- [7] 吴海云, 刘洋, 左月明. 近红外光谱数据分析方法的研究进展. *农产品加工(学刊)*, 2010(3): 76—79. Wu H Y, Liu Y, Zuo Y M. An overview of spectral data processing in NIR analytical technique (In Chinese). *Academic Periodical of Farm Products Processing*, 2010(3): 76—79
- [8] 徐广通, 袁洪福, 陆婉珍. 现代近红外光谱技术及应用进展. *光谱学与光谱分析*, 2000, 20(2): 134—142. Xu G T, Yuan H F, Lu W Z. Development of modern near infrared spectroscopic techniques and its applications (In Chinese). *Spectroscopy and Spectral Analysis*, 2000, 20(2): 134—142
- [9] Viscarra Rossel R A, Walvoort D J J, McBratney A B, et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 2006, 131(1/2): 59—75
- [10] Mouazen A M, Kuang B, De Baerdemaeker J, et al. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, 2010, 158(1/2): 23—31
- [11] 高建波, 胡东成. 小波变换和神经网络用于红外光谱定量分析. *清华大学学报: 自然科学版*, 2001, 41(3): 121—124. Gao J B, Hu D C. Quantitative analysis of the infrared spectrum



- using wavelet transforms and artificial neural networks (In Chinese). *Journal of Tsinghua University: Science and Technology*, 2001, 41(3): 121—124
- [12] 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
- [13] Fontán J M, Calvache S, López-Bellido R J, et al. Soil carbon measurement in clods and sieved samples in a mediterranean vertisol by visible and near-infrared reflectance spectroscopy. *Geoderma*, 2010, 156(3/4): 93—98
- [14] 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
- [15] 张娟娟, 田永超, 朱艳, 等. 不同类型土壤的光谱特征及其有机质含量预测. *中国农业科学*, 2009, 42(9): 3154—3163. Zhang J J, Tian Y C, Zhu Y, et al. Spectral characteristics and estimation of organic matter contents of different soil types (In Chinese). *Scientia Agricultura Sinica*, 2009, 42(9): 3154—3163
- [16] Stenberg B, Rossel R A V, Mouazen A M, et al. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 2010, 107: 163—215
- [17] 鲍士旦. 土壤农化分析. 第3版. 北京: 中国农业出版社, 1999. Bao S D. Soil and agricultural chemistry analysis (In Chinese). 3rd ed. Beijing: China Agriculture Press, 1999
- [18] 穆云峰. RBF神经网络学习算法在模式分类中的应用研究. 大连: 大连理工大学, 2006. Mu Y F. Study on the application of RBF neural network learning algorithm for pattern classification (In Chinese). Dalian: Dalian University of Technology, 2006
- [19] 孙一兵. 浅议 BP 网络的优缺点及改进. *科技创新导报*, 2009(24): 18. Sun Y B. Discuss of advantages and disadvantages of BP neural network and its improvement (In Chinese). *Science and Technology Innovation Herald*, 2009(24): 18
- [20] 谷志红, 牛东晓, 王会青. 广义回归神经网络模型在短期电力负荷预测中的应用研究. *中国电力*, 2006, 39(4): 11—14. Gu Z H, Niu D X, Wang H Q. Research on application of general regression neural network in short-term load forecasting (In Chinese). *Electric Power*, 2006, 39(4): 11—14
- [21] 史峰, 王小川, 郁磊, 等. MATLAB神经网络30个案例分析. 北京: 北京航空航天大学出版社, 2010. Shi F, Wang X C, Yu L, et al. 30 cases of Matlab neural networks (In Chinese). Beijing: Beihang University Press, 2010
- [22] Farifteh J, van der Meer F, Atzberger C, et al. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, 2007, 110(1): 59—78
- [23] Williams P C, Norris K. Near-infrared technology in agricultural and food industries. St. Paul, Minnesota: American Association of Cereal Chemists, Inc., 1987: 330
- [24] Maleki M R, van Holm L, Ramon H, et al. Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. *Biosystems Engineering*, 2006, 95(3): 425—436
- [25] Despagne F, Massart DL. Variable selection for neural networks in multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, 1998, 40(2): 145—163
- [26] Janik L J, Cozzolino D, Damberg R, et al. The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. *Analytica Chimica Acta*, 2007, 594(1): 107—118
- [27] Ben-Dor E, Banin A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, 1995, 59(2): 364—372

## PREDICTION OF SOIL ORGANIC MATTER AND TOTAL PHOSPHORUS WITH VIS-NIR HYPERSPECTRAL INVERSION RELATIVE TO LAND USE

Xue Lihong<sup>1†</sup> Zhou Dinghao<sup>2</sup> Li Ying<sup>2</sup> Yang Linzhang<sup>1</sup>

(1 *Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China*)

(2 *Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*)

**Abstract** Effects of spectral modeling methods and land use patterns were explored on hyperspectral inversion of soil organic matter (SOM) and total phosphorus (TP) in soil samples collected from paddy fields, peach orchards and vegetable gardens in the Zhihugang catchment, Taihu Lake Region. Results show that the PLSR (Partial least square regression) model was quite high and stable in modeling and prediction precision; the GRNN (General regression neural network) of ANN (Artificial neural network) was also quite high in prediction precision, but prone to overfitting; the BPNN (Back Propagation Neural Network) was relatively stable, but slightly low in precision; and the combined PLSR-ANN model improved in prediction precision by combining the advantages of the two in handling complicated samples. The spectral inversion of SOM was better than that of TP, and among the three patterns of land use, paddy fields were fitter

than the other two for use of the models in prediction of SOM and TP. In the current study zone, patterns of land use did not have much effect on spectral inversion of SOM, but did much on that of soil TP. It is, therefore, essential to calibrate the models in light of land use patterns in conducting spectral inversion of soil TP.

**Key words** Hyperspectrum; PLSR(Partial least square regression); ANN(Artificial neural network); Soil organic matter; Total phosphorus; Land use

(责任编辑:卢 萍)