

doi: 10. 6046/gtzyyg. 2019. 01. 15

引用格式: 王永敏,李西灿,田林亚,等. 土壤有机质含量地面高光谱估测模型对比分析[J]. 国土资源遥感,2019,31(1):110 – 116. (Wang Y M,Li X C,Tian L Y,et al. Comparison and analysis of estimation models of soil organic matter content established by hyperspectral on ground[J]. Remote Sensing for Land and Resources,2019,31(1):110 – 116.)

土壤有机质含量地面高光谱估测模型对比分析

王永敏¹, 李西灿², 田林亚¹, 贾 斌³, 杨 惠⁴

(1. 河海大学地球科学与工程学院,南京 211100; 2. 山东农业大学信息科学与工程学院,泰安 271018; 3. 中建四局第三建筑工程有限公司,遵义 563000; 4. 东南大学交通学院,南京 210000)

摘要:采用高光谱技术获得的数据进行土壤有机质含量的反演和估测是近年来的研究热点。为确定有效的估测建模方法,利用地面实测的土壤高光谱反射率及有机质含量等数据,采用小波分析方法实现去噪,包络线去除法实现建模参数提取和数据量压缩,结合多种不同的数据变换方法,利用 BP 神经网络法、多元线性回归法及最小二乘回归法建立不同的估测模型。对比发现,BP 神经网络模型的估测效果优于回归模型,其中结合对数的平方变换和神经网络所建立的模型为最优估测模型,模型的决定系数达到 0.933,检验样本的均方根误差达到 0.069。实验证明,BP 神经网络 + 对数的平方变换模型的学习机制适用于土壤有机质含量地面高光谱估测且效果好。通过在建模因子层面上进行数据变换建立了较好的估测模型,其研究方法、模型和结论,对土壤有机质含量地面高光谱估测具有一定的参考意义。

关键词: 地面高光谱; 土壤有机质; 数据变换; 估测模型; 对比分析

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001 – 070X(2019)01 – 0110 – 07

0 引言

土壤有机质含量是农作物生长的必要物质之一,对其进行快速、准确测定和估计对发展精细农业和提高农田管理的科学性具有重要意义^[1]。传统的土壤有机质含量测定方式手段繁琐,耗时耗力,且有一定的局限性。近年来,高光谱技术的发展,使得快速实现土壤有机质含量估测成为可能,因而研究合理有效的土壤有机质含量高光谱估测方法非常重要。邱华旭等^[2]通过对光谱反射率做常规模型变换,发现适当的模型变换可以使土壤性质指标与反射率数据的相关性得到有效提高;郭超凡等^[3]和 Kokaly 等^[4]对光谱吸收特性采用连续移除法和多元逐步线性回归法处理,可以提高高光谱估测模型的精度;陈增文等^[5]发现土壤有机质含量与光谱反射率呈负相关,为利用土壤高光谱反射率数据进行有机质含量估测提供了理论支持和方法引导。但是,诸多研究在光谱敏感波段方面的结论并不一致,彭杰等^[6]研究土壤理化特征发现,350 ~ 2 500 nm

波段可以反映土壤特性的细微差异,可用于反演土壤理化参数;廖钦洪等^[7]和 Liu 等^[8]采用一阶微分归一化技术处理黑土的反射率数据,研究发现光谱波段范围为 620 ~ 810 nm 时对有机质敏感;Karnieli 等^[9]发现在 1 720 nm,2 180 nm 和 2 309 nm 处光谱曲线存在吸收特性。此外,采用不同的数据处理方法,所建模型精度也存在差异,沈润平等^[10]通过人工神经网络方法建立土壤有机质含量估测模型,总均方根误差达到了 1.31;于雷等^[11]通过连续小波变换和支持向量机回归建立的估测模型决定系数达到了 0.83;彭小婷等^[12]采用包络线去除法,结合偏最小二乘回归法,建立模型的决定系数达到了 0.763。但是,目前综合比较各种土壤有机质含量高光谱估测方法的研究较少。

为探究不同方法用于土壤有机质含量高光谱估测的精度,确定有效的估测建模方法,本文通过小波分析和包络线去除法实现去噪和参数提取,采用 9 种数据变换方法,对提取的建模因子进行变换并结合神经网络法、多元线性回归法和最小二乘回归法共建立了 27 种地面高光谱土壤有机质含量估测模

型,比较不同方法所建模型精度,最终确定有效的估测建模方法,以期对相关研究提供一定的参考。

1 数据与方法

1.1 数据采集

本文所用数据为陕西省横山县土壤样本的地面实测高光谱反射率和有机质含量数据。陕西省横山县(E108°56′~110°02′,N37°22′~38°14′)位于陕西省北部,其地势西南高,东北低,中间高,南部低,呈倾斜趋势,海拔约 890~1 534 m;由于处在鄂尔多斯草原与黄土高原的过渡地带,水土流失严重,“三沙七丘”地貌景观独特,主要分布有黄绵土、绵沙土、黑垆土和风沙土等土壤类型。

土样采集利用梅花桩取样法,保证采样点尽可能均匀分布在研究区内,顾及不同土壤类型,在每个测区内随机选择 4~5 个具有代表性的测点(白界乡、岔龙湾、高镇和石湾镇 4 个采样区测点数为 5,其余为 4),选择天气晴朗、光照稳定、土壤干燥时除掉测点上方的腐殖落叶层并进行 5 次光谱测定取平均值作为样本点的光谱反射率值,然后取 15 cm 左右深度的表层土,共采集 84 个样本,样本采集区分布情况如图 1 所示。



图 1 土壤样本采集区域分布示意图

Fig.1 Area of collecting soil samples

光谱反射率采集使用 ASD FieldSpec FR 野外光谱仪,采用 3°视场角,面对太阳方向,探头距离目标 20 cm 进行垂直观测,观测范围直径近 2 cm。光谱反射率波段范围为 350~2 500 nm,其中 350~1 000 nm 光谱分辨率为 3 nm,1 000~2 500 nm 光谱分辨率为 10 nm。对土壤样本的有机质含量和水含量在室内进行分析测定。土壤样本的理化参数统计结果如表 1 所示。在数据处理之前,通过比较分析,剔除了数据噪声较严重的 4 个样本,剩余 80 个样本数据用于本研究。经多次实验,随机选择 50 个样本用于建模,剩余 30 个样本用于检验。

表 1 土壤样本有机质含量和水含量统计

Tab.1 Statistical characteristics of soil organic matter content and water content (%)

样本集	有机质最小值	有机质最大值	有机质平均值	有机质标准差	水含量最小值	水含量最大值	水含量平均值	水含量标准差
总体样本	0.124	1.289	0.600	0.250	0.46	38.65	9.87	7.65
建模样本	0.124	1.289	0.608	0.238	4.83	38.65	9.65	7.62
检验样本	0.176	1.289	0.575	0.241	0.46	35.32	9.97	7.69

1.2 数据处理

1.2.1 小波分析去噪

在数据获取过程中,由于测定方法、外界环境和操作等原因,获得的光谱反射率数据不可避免地存在噪声,因此对原始数据进行了去噪处理。小波分析是常用且效果较好的去噪方法之一,本文通过 Matlab 编程实现光谱反射率的小波去噪。选用 Bior1.3 小波基函数进行 3 层小波分解,当小波系数小于阈值时,既不采用硬阈值法置零,也不采用软阈值法按照固定向量向零附近收缩,而是通过归一化相关系数设置默认阈值^[13],归一化相关系数公式为

$$W_{J,K} = \sqrt{PW_J/PCW_J} \quad , \quad (1)$$

$$PW_J = \sum_K w_{J,K}^2 \quad , \quad (2)$$

$$PCW_J = \sum_K cw_{J,K}^2 \quad , \quad (3)$$

$$cw_{J,K} = w_{J,K} \times w_{J+1,K} \quad , \quad (4)$$

式中: $cw_{J,K}$ 为尺度 J 上 K 点处的相关系数; $w_{J,K}$ 和 $w_{J+1,K}$ 分别为小波系数,由信号小波系数和噪声小波系数 2 部分组成。通过默认阈值对信号进行重构,实现原始光谱反射率数据的去噪。

1.2.2 包络线去除法提取参数

由于原始的光谱反射率数据量大,本文采用包络线去除法压缩数据量并提取参数。利用 VB 编程并结合 Excel 表格,选择与有机质含量相关性高的连续波段为敏感波段,并在选取的敏感波段上进行包络线去除,共提取了高度面积、深度面积、高差面积、最大高度、最大深度、最大高差、对称度和位置等参数。包络线提取参数的运算公式分别为

$$C = R_i/[R_{\text{start}} + T \times (\lambda_i - \lambda_{\text{start}})] \quad , \quad (5)$$

$$K = (R_{\text{end}} - R_{\text{start}})/(\lambda_{\text{end}} - \lambda_{\text{start}}) \quad , \quad (6)$$

$$H = (1/C) - 1 \quad , \quad (7)$$

$$P = 1 - C \quad , \quad (8)$$

$$D = (1/C) - C \quad , \quad (9)$$

$$AH = \sum_s^e H \quad , \quad (10)$$

$$AP = \sum_s^e P \quad , \quad (11)$$

$$AD = \sum_s^e D \quad , \quad (12)$$

$$MH = \text{Max} \{ H \}_s^e \quad , \quad (13)$$

$$MP = \text{Max} \{ P \}_s^e \quad , \quad (14)$$

$$MD = \text{Max} \{ D \}_s^e \quad , \quad (15)$$

$$DS = K / (AP - K) \quad , \quad (16)$$

$$LM = \text{Index}(\text{Max} \{ H \}_s^e) \quad , \quad (17)$$

式中： C 为包络线结果值； R_i 为 i 处的原始光谱反射率数据； R_{start} 和 R_{end} 分别为起止点处的光谱反射率； λ_i 为 i 处的波长； λ_{start} 和 λ_{end} 分别为起止点处的波长； T 为待处理的光谱反射率曲线起止点间的斜率； K 为最大深度处深度值； H 、 P 和 D 分别为高度、深度和高差； AH 为高度面积； AP 为深度面积； AD 为高差面积； MH 为最大高度； MP 为最大深度； MD 为最大高差； DS 为对称度； LM 为位置； s 和 e 分别为起点和终点位置。按照相关性由高到低的顺序，选取其中相关性较高的 6 个特征作为建模因子。

1.2.3 数据变换

为提高建模因子与土壤有机质含量的相关性，选择 9 种数据变换方法进行变换处理，包括平方变换 (square transform, SQ)、倒数变换 (reciprocal transform, RE)、指数变换 (exponential transform, EXP)、对数变换 (log transform, LOG)、指数的平方变换 (exponential square transform, ES)、对数的平方变换 (logarithm square transform, LS)、一阶微分变换 (first order differential transform, DE1)、对数一阶微分变换 (logarithm first order differential transform, LGD1)、指数一阶微分变换 (exponential first order differential

transform, EXD1)，其中一阶微分用差分表示。由于位置因子的指数变换数值过大，先将位置因子缩小 100 倍后进行指数变换。

1.3 建模及精度评定方法

采用 BP(back propagation) 神经网络法、多元线性回归法 (multivariable linear regression, MLR) 和最小二乘回归法 (least squares regression, LSR)，结合 9 种数据变换方法，共建立了 27 种土壤有机质含量地面高光谱估测模型。

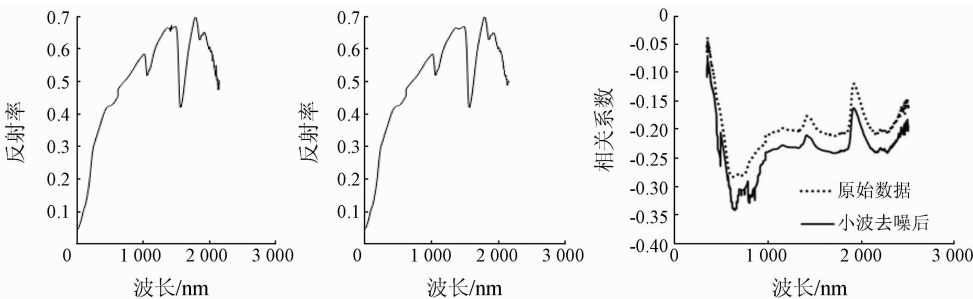
网络参数设置对神经网络模型的结果质量有很大影响，因此选择合适的参数 (隐含层数、学习速率、平滑因子、学习误差和分级迭代级数) 对模型的精度很重要。由于目前没有很成熟的参数设置模型可供使用，本文通过大量实验，反复对比，最终确定本实验的参数设置为：输入层数为 6，对应提取的 6 个建模因子所对应的数据；输出层数为 1，对应样本有机质含量，人为确定前 50 个样本数据作为学习样本，后 30 个样本数据作为检验样本；隐含层数为 15，学习速率为 1.5 h，平滑因子为 0.7，学习误差为 0.03，分级迭代级数为 14。通过编程实现 MLR 和 LSR 建模。

采用模型的决定系数 R^2 、检验样本的均方根误差 (root mean square, RMSE) 和估测模型的差异显著性 sig 值作为估测模型的精度和可靠性评定因子。各模型的 sig 值由 SPSS 软件获得，用以检验模型的显著性；根据 R^2 和 $RMSE$ ，比较不同模型的精度和可靠性，最终选出最佳模型。其中， $\text{sig} < 0.05$ 则说明模型是显著的， R^2 越大，模型的可靠性越强、模型越稳定； $RMSE$ 越小，模型精度越高。

2 结果与分析

2.1 小波分析去噪结果

通过 Bior1.3 小波基函数对原始光谱反射率数据进行 3 层小波分解，实现去噪处理 (图 2)。



(a) 小波去噪前光谱反射率 (b) 小波去噪后光谱反射率 (c) 去噪前后反射率与有机质含量相关系数

图 2 小波去噪前后光谱反射率与有机质含量相关系数对比

Fig.2 Comparison of correlation coefficient between spectral reflectance of original and wavelet denoising and organic matter content

图 2(a)和(b)中可以看出(以 77 号样本为例,其他样本处理效果与此类似),小波处理使得原始数据中的噪声得到去除,在 1 400 nm 附近和 2 000 ~ 2 500 nm 表现明显。由图 2(c)可以发现原始光谱反射率数据与有机质含量呈负相关且相关性不高,小波分析去噪提高了光谱反射率与有机质含量的相关性。因此,采用小波去噪后的光谱反射率数据代替原始数据建立估测模型。

2.2 参数提取

由图 2 的数据得知,在波段 590 ~ 770 nm 之间,光谱反射率与有机质含量相关系数较高,是光谱敏感波段,为减小数据处理量,选择波段 600 ~ 700 nm 之间的光谱反射率数据,编程实现包络线去除法,并对参数进行提取。以 66 号土壤样本为例,图 3 为包络线去除前后光谱反射率对比,其他样本处理后的曲线趋势与此类似。

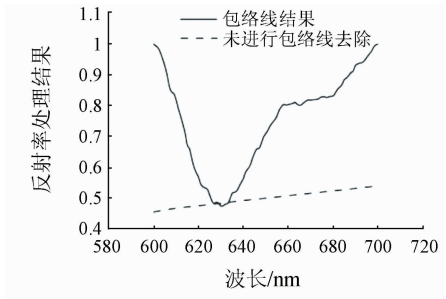


图 3 包络线去除结果对比
Fig. 3 Continuum removal results

实验共提取高度面积、深度面积、高差面积、最大高度、最大深度、最大高差、对称度和位置等多项指标因子。按照相关性由高到低的顺序,选取位置、对称度、深度面积、最大深度、最大高差和最大高度等 6 个参数作为建模因子,但是各个因子与有机质含量相关性不高,不能直接用于建模,故采用不同数据变换方法对各因子进行变换处理(表 2)。

表 2 相关系数对比
Tab. 2 Comparison of correlation coefficient

变换方法	AP	DS	MP	MH	MD	LM
原始	-0.205	-0.217	-0.196	-0.151	-0.171	0.230
SQ	-0.288	-0.380	-0.286	-0.172	-0.365	0.296
RE	0.296	0.357	0.356	0.256	0.386	-0.295
EXP	0.482	-0.534	-0.518	-0.472	-0.484	0.305
LOG	-0.390	-0.362	-0.211	-0.209	-0.318	0.295
ES	0.431	-0.437	-0.368	-0.428	-0.499	0.401
LS	-0.539	-0.596	-0.432	-0.631	-0.628	0.521
DE1	-0.282	-0.329	-0.287	-0.273	-0.286	0.224
LGD1	-0.462	-0.309	0.508	0.414	0.511	0.424
EXD1	0.337	-0.321	0.409	0.458	0.406	-0.245

2.3 建模因子变换结果比较

由于原始因子与有机质含量的相关性不高,通过 1.2.3 节中的方法对建模因子数据进行变换处理。计算各因子与有机质含量的相关系数(表 2)结果表明,提取参数经 LS 变换后,除最大深度外,其他参数与有机质含量的相关系数均达到 0.5 以上,能有效提高提

取参数与土壤有机质含量的相关性。此外,EXP,ES,LGD1 和 EXD1 的变换效果也较好。

2.4 建模结果及对比分析

分别计算 27 个模型的 sig, R² 和 RMSE,图 4 为 3 种建模方法所建模型的 R² 和 RMSE 对比结果,具体数值如表 3 所示。

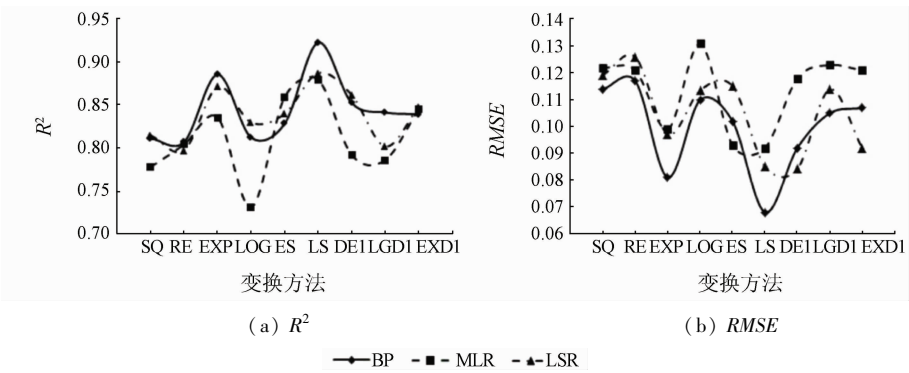


图 4 3 种模型的 R²和 RMSE 对比
Fig. 4 Comparison of three models' R²and RMSE

表 3 27 种估测模型检验样本集结果统计

Tab. 3 27 hyperspectral estimation models' results of test samples

参数	方法	SQ	RE	EXP	LOG	ES	LS	DE1	LGD1	EXD1	均值
R^2	BP	0.811	0.807	0.885	0.812	0.828	0.933	0.852	0.839	0.841	0.845
	MLR	0.778	0.804	0.835	0.731	0.859	0.879	0.792	0.845	0.786	0.812
	LSR	0.815	0.797	0.871	0.826	0.833	0.887	0.861	0.847	0.802	0.839
$RMSE$	BP	0.114	0.117	0.081	0.110	0.102	0.069	0.092	0.107	0.105	0.099
	MLR	0.122	0.121	0.099	0.131	0.093	0.092	0.118	0.121	0.123	0.113
	LSR	0.119	0.125	0.097	0.114	0.115	0.085	0.084	0.092	0.114	0.105
sig	BP	0.015	0.019	0.001	0.011	0.006	0.00	0.003	0.013	0.007	0.008
	MLR	0.027	0.031	0.033	0.042	0.002	0.001	0.013	0.006	0.035	0.021
	LSR	0.019	0.028	0.001	0.021	0.005	0.001	0.002	0.007	0.016	0.011

由图 4(a)可以发现,BP 与 LSR 模型的 R^2 均较高,除几种情况(RE 和 LGD1 变换对 2 种建模方法效果相近;ES 变换结合 MLR 模型效果优于 BP 和 LSR 模型)外,均明显优于 MLR 模型,说明建模因子与土壤有机质含量之间不是简单的线性关系,MLR 模型可靠性不高。同时,比较 BP 模型和 LSR 模型,发现对应多种数据变换方法,二者所建模型的 R^2 差别不大,但是 LS 变换对 BP 模型更敏感,所建模型的 R^2 明显优于其他模型。由图 4(b)可以看出,BP 模型的 $RMSE$ 较小,MLR 模型的 $RMSE$ 较大,说明 BP 模型的估测精度高于 MLR 模型。综合发现,BP 模型结合 LS 变换所建模型的 R^2 最大, $RMSE$ 最小,因此本实验确定的最佳估测模型为 BP + LS 模型。表 3 中数据也显示,所有模型的 sig 值均小于 0.05,说明所建估测模型均是显著的,结果可靠,所做比较具有意义。比较各模型 R^2 和 $RMSE$ 的均值结果可以发现,从整体而言,BP 模型和 LSR 模型均优于 MLR 模型,且 BP 模型更优于 LSR 模型。

图 5 为 BP + LS 模型实测值与估测值的对比结果,横坐标表示有机质含量的实测值,纵坐标为 BP + LS 模型得到的有机质含量的估测值,由表 3 中数据可知,BP + LS 模型的 R^2 可达 0.933,检验样本的 $RMSE$ 为 0.069。

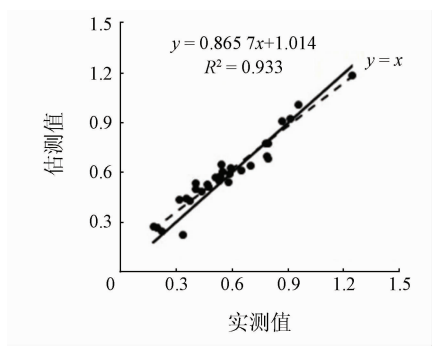


图 5 估测值与实测值对比结果

Fig. 5 Comparison of predicted and measured values

图 5 中虚线为估测点的趋势线,直线为 $y = x$ 指标线,在直线下方的点表明有机质含量的估测值低

于实测值,而直线上方的点表明有机质含量的估测值高于实测值。从图 6 可以看出,当有机质含量 < 0.6% 时,估测值普遍高于实测值;当有机质含量在 0.6% ~ 0.8% 之间时,估测值大多低于实测值;当有机质含量 > 0.8% 时,估测值高于实测值。当有机质含量偏小或者偏大时,BP + LS 模型得到的有机质含量的估测值多高于实测值,但总体而言,差别不大,模型精度较高。LS 变换有效提高了建模因子与有机质含量的相关性,同时,BP 神经网络具有较强的非线性映射能力,能较好地模拟土壤有机质含量和地面实测高光谱反射率之间的关系,此外,BP 神经网络在建模时,能通过学习自动实现输出数据的合理规则,因此 BP + LS 估测模型具有较高的精度和可靠性。

3 结论

1)数据质量对建模效果影响显著。通过比较原始光谱数据和小波去噪后光谱反射率数据与有机质含量的相关系数,发现经小波分析处理后,在 590 ~ 770 nm 之间,光谱反射率与有机质含量的相关性显著提高。研究提高数据质量及数据与目标因子相关性的理论和方法具有重要意义。

2)比较不同数据变换的结果,发现数据变换可以提高数据和土壤有机质含量的相关性,但是建模因子对不同数据变换方法响应程度不同,不恰当的变换可能使参数的相关系数不增反降,综合比较,对数的平方变换是一种有效的变换方法,对 3 种建模方法效果均较好,在同类模型中精度和可靠性最高,因为对数的平方运算能较好地消除数据异方差问题,缩小差距。

3)针对建模因子与有机质含量相关性不高的问题,改变了以往对光谱反射率数据进行变换的思路,提出了对提取的建模因子直接进行数据变换,然后利用变换后的数据建立高光谱土壤有机质含量估测模型的方法。本文利用土壤地面实测高光谱数据

(350 ~ 2 500 nm 和有机质含量数据,通过小波分析和包络线去除法提取特征参数,选取与有机质含量相关系数较高的参数作为建模因子,经过 9 种数据变换方法处理后,结合 BP 神经网络法、多元线性回归法和最小二乘回归法共建立 27 个土壤有机质含量估测模型,对比各模型的评价指标,最终确定 BP 神经网络 + 对数的平方变换模型为最佳估测模型。

本文探究了多种数据变换方法和建模方法的组合在高光谱土壤有机质含量估测中的效果和精度,并最终确定了相对最优的建模方法和估测模型,对该领域高光谱技术的应用推广具有一定的参考意义。但由于本文没有顾及土壤水含量对有机质含量估测的影响,仍存在一定的局限性,未来将继续深入研究。

参考文献 (References) :

[1] 王 晓. 土壤含水量高光谱特性与估测模型研究 [D]. 泰安:山东农业大学,2012.
Wang X. Studies on Soil Moisture Hyper - spectrum Characteristics and Estimating Model [D]. Taian:Shandong Agricultural University, 2012.

[2] 邱华旭,黄张裕,李希灿. 土壤性质指标光谱反演数据变换模型研究[J]. 测绘科学,2013,38(3):131 - 133.
Qiu H X, Huang Z Y, Li X C. Data transformation model analysis on spectral retrieved deduction of soil properties index [J]. Science of Surveying and Mapping, 2013, 38(3): 131 - 133.

[3] 郭超凡,郭道宇. 基于可见光波段包络线去除的湿地植物叶片叶绿素估算[J]. 生态学报,2016,36(20):6538 - 6546.
Guo C F, Guo X Y. Estimation of wetland plant leaf chlorophyll content based on continuum removal in the visible domain [J]. Acta Ecologica Sinica, 2016, 36(20): 6538 - 6546.

[4] Kokaly R F, Clark R N. Spectroscopic determination of leaf biochemistry using band - depth analysis of absorption features and stepwise multiple linear regression [J]. Remote Sensing of Environment, 1999, 67(3): 267 - 287.

[5] 陈增文,陈光水,钟菱芳,等. 基于高光谱遥感的土壤有机碳含量估算研究进展[J]. 亚热带资源与环境学报,2009,4(1):78 - 87.
Chen Z W, Chen G S, Zhong X F, et al. Review on estimation of soil organic carbon content based on hyperspectral measurements [J]. Journal of Subtropical Resources and Environment, 2009, 4(1): 78 - 87.

[6] 彭 杰,张杨珠,周 清,等. 土壤理化特性与土壤光谱特征关系的研究进展[J]. 土壤通报,2009,40(5):1204 - 1208.
Peng J, Zhang Y Z, Zhou Q, et al. The progress on the relationship physics - chemistry properties with spectrum characteristic of the soil [J]. Chinese Journal of Soil Science, 2009, 40(5): 1204 - 1208.

[7] 廖钦洪,顾晓鹤,李存军,等. 基于连续小波变换的潮土有机质含量高光谱估算[J]. 农业工程学报,2012,28(23):132 - 139.
Liao Q H, Gu X H, Li C J, et al. Estimation of fluvo - aquic soil organic matter content from hyperspectral reflectance based on con-

tinuous wavelet transformation [J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, 28(23): 132 - 139.

[8] Liu H J, Zhang Y Z, Zhang B, et al. Novel hyperspectral reflectance models for estimating black - soil organic matter in Northeast China [J]. Environmental Monitoring and Assessment, 2009, 154(1 - 4): 147 - 154.

[9] Karnieli A, Verchovsky I, Hall J K, et al. Geographic information system for semi - detailed mapping of soils in semi - arid region [J]. Geocarto International, 1998, 13(3): 29 - 42.

[10] 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[J]. 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 [J]. Acta Pedologica Sinica, 2009, 46(3): 391 - 397.

[11] 于 雷,洪永胜,周 勇,等. 连续小波变换高光谱数据的土壤有机质含量反演模型构建[J]. 光谱学与光谱分析,2016,36(5):1428 - 1433.
Yu L, Hong Y S, Zhou Y, et al. Inversion of soil organic matter content using hyperspectral data based on continuous wavelet transformation [J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1428 - 1433.

[12] 彭小婷,高文秀,王俊杰. 基于包络线去除和偏最小二乘的土壤参数光谱反演[J]. 武汉大学学报(信息科学版),2014,39(7):862 - 866.
Peng X T, Gao W X, Wang J J. Inversion of soil parameters from hyperspectra based on continuum removal and partial least squares regression [J]. Geomatics and Information Science of Wuhan University, 2014, 39(7): 862 - 866.

[13] 孔玲军. MATLAB 小波分析超级学习手册 [M]. 北京:人民邮电出版社,2014:100 - 160.
Kong L J. Super Learning Manual of Wavelet Analysis Based on MATLAB [M]. Beijing: Posts and Telecom Press, 2014: 100 - 160.

[14] Ben - Dor E, Banin A. Visible and near infrared (0.4 - 1.1 μm) analysis of arid and semiarid soil [J]. Remote Sensing of Environment, 1994, 48(3): 261 - 274.

[15] 李希灿,王 静,李玉环,等. 基于模糊集分析的土壤质量指标高光谱反演[J]. 地理与地理信息科学,2008,24(4):25 - 28.
Li X C, Wang J, Li Y H, et al. High spectral retrieved deduction of soil quality index based on fuzzy sets analysis [J]. Geography and Geo - Information Science, 2008, 24(4): 25 - 28.

[16] 沈士喆,张小龙,衡 伟. 一种自适应阈值的预筛选 Harris 角点检测方法[J]. 数据采集与处理,2011,26(2):207 - 213.
Shen S Z, Zhang X L, Heng W. Improved Harris corner detection algorithm based on auto - adaptive threshold and pre - selection [J]. Journal of Data Acquisition and Processing, 2011, 26(2): 207 - 213.

[17] 邓小炼,王长耀,亢 庆,等. 一种基于角点提取的遥感影像地面控制点自适应匹配算法[J]. 宇航学报,2006,27(1):45 - 50,88.
Deng X L, Wang C Y, Kang Q, et al. An adaptive matching algorithm for ground control point of remote sensing image based on corner extraction [J]. Journal of Astronautics, 2006, 27(1): 45 - 50, 88.

[18] 何 挺,王 静,林宗坚,等. 土壤有机质光谱特征研究[J]. 武汉大学学报(信息科学版),2006,31(11):975 - 979.

He T,Wang J,Lin Z J,et al. Spectral features of soil organic matter [J]. Geomatics and Information Science of Wuhan University, 2006,31(11):975-979.

[19] 郭斗斗,黄绍敏,张水清,等. 多种潮土有机质高光谱预测模型的对比分析[J]. 农业工程学报,2014,30(21):192-200.

Guo D D,Huang S M,Zhang S Q,et al. Comparative analysis of various hyperspectral prediction models of fluvo-aquic soil organic matter[J]. Transactions of the Chinese Society of Agricultural Engineering,2014,30(21):192-200.

Comparison and analysis of estimation models of soil organic matter content established by hyperspectral on ground

WANG Yongmin¹, LI Xican², TIAN Linya¹, JIA Bin³, YANG Hui⁴

(1. School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China; 2. College of Information Science and Engineering, Shandong Agricultural University, Taian 271018, China; 3. CCFED the Third Construction Engineering Co., Zunyi 563000, China; 4. School of Transportation, Southeast University, Nanjing 210000, China)

Abstract: Using the data obtained by hyperspectral techniques to estimate the content of soil organic matter is a hotspot in recent years. For the purpose of determining the effective estimation modeling method, specific data such as reflectance obtained by hyperspectral on ground and organic matter content were used in this paper. Wavelet analysis was used to remove the noise, and continuum removal was used to extract the parameters and compress the data. Combining a variety of different data transformation methods and utilizing BP neural networks, multiple linear regression (MLR) and least squares regression (LSR), many different estimation models of soil were established. It is found that the neural network method is superior to the regression model among various data transformation methods after comparing different estimation models established by the three modeling methods. The optimal estimation model is the model established by the combination of logarithmic square transformation and neural network. The R^2 of the model is 0.933 and the $RMSE$ is 0.069. The authors creatively carried out the data transformation at the modeling factor level and established the good estimation model. It is shown that the learning mechanism of BP + LS model is suitable for hyperspectral estimation of soil organic matter and works well. The methods, models and conclusions of this paper have some reference significance for the hyperspectral estimation of soil organic matter.

Keywords: hyperspectra on ground; soil organic matter; data conversion; estimation model; comparative analysis
(责任编辑: 陈 理)