

doi: 10.11720/wtyht.2022.0038

陈超群,戴慧敏,冯雨林,等.基于 Sentinel-2A 的孙吴地区土壤有机质反演研究[J].物探与化探,2022,46(5):1141-1148. <http://doi.org/10.11720/wtyht.2022.0038>

Chen C Q, Dai H M, Feng Y L, et al. Sentinel-2A based inversion of the organic matter content of soil in the Sunwu area[J]. Geophysical and Geochemical Exploration, 2022, 46(5): 1141-1148. <http://doi.org/10.11720/wtyht.2022.0038>

基于 Sentinel-2A 的孙吴地区土壤有机质反演研究

陈超群^{1,2,3}, 戴慧敏^{1,2,3}, 冯雨林¹, 杨泽^{1,2,3}, 杨佳佳¹

(1. 中国地质调查局 沈阳地质调查中心, 辽宁 沈阳 110034; 2. 自然资源部 黑土地演化与生态效应重点实验室, 辽宁 沈阳 110034; 3. 辽宁省黑土地演化与生态效应重点实验室, 辽宁 沈阳 110034)

摘要: 利用 Sentinel-2A 多光谱遥感影像, 结合实测土壤信息, 对黑龙江省孙吴县黑土区土壤有机质含量进行反演研究。对影像进行预处理后, 通过相关分析和随机森林(RF)选取特征波段, 采用偏最小二乘法和 BP 神经网络构建土壤有机质含量多光谱模型反演红旗林场土壤有机质含量。研究表明: 相关性选取的倒数对数一阶微分反射率波段和 RF 选择的组合波段能够有效提高土壤反演精度, 组合波段的 RF-BP 神经网络模型反演效果最佳, $R^2 = 0.7245$, RMSE = 1.3127%。本次研究可为实现土壤有机质动态监测提供技术支持和参考。

关键词: 黑土; 有机质; Sentinel-2A; 随机森林; BP 神经网络

中图分类号: P632

文献标识码: A

文章编号: 1000-8918(2022)05-1141-08

0 引言

土壤有机质是土壤质量的重要参数, 可为农作物提供各类养分, 同时对元素表生地球化学特征有重要影响。黑土作为珍贵的土壤资源, 其有机质含量是反映土壤质量的重要指标参数^[1-2]。近年来, 随着黑土逐渐退化, 土壤中有有机质成分明显减少, 估算黑土有机质含量, 扭转含量下降趋势, 是黑土地保护的重要举措^[3]。传统的土壤有机质监测主要是通过监控区进行大量野外土壤样品采集和室内化学实验分析进行反演, 这种方法周期较长, 费时费力, 精度受样品密度控制, 难以满足现代农业快速发展的需求^[4]。随着遥感技术日益成熟, 通过有机质含量的光谱差异来测定土壤有机质的含量已成为一种有效手段。

土壤有机质含量遥感反演主要包含两个研究方向: 光谱信息的处理与选择和反演模型的构建。常

采用的光谱处理方法有倒数、对数、去包络线变换等, 但选取的有机质特征波段因影像数据源不同而有所差异。屈冉等^[5]选取 Landsat TM 影像反演广西壮族自治区富川县的有机质含量, 认为土壤有机质含量与 Landsat TM 波段 5 和波段 7 的 DN 值相关性最高。陈德宝等^[6]借助 Landsat 8 遥感影像对农安县黑土区有机质进行建模反演, 表明短波红外 B6 波段反射率所建模型拟合效果最好。陈思明等^[7]对 Landsat 7 土壤光谱进行线性波谱分离重建, 认为重建光谱能显著增强与土壤有机质含量的相关性, 提高土壤有机质反演精度。在以往研究中, 线性回归和偏最小二乘回归模型(PLSR)常被用于土壤有机质含量反演。Dhawale 等^[8]结合土壤样品有机质含量和相应的土壤反射率, 选用 PLSR 建模, 均方根误差不超过 2.24%。马驰^[9]对比 Sentinel-2A 遥感影像不同波段组合的多元回归模型, R^2 均大于 0.7。目前针对有机质敏感波段的选择主要采用 Pearson 相关分析法, 反演模型也多选择线性拟合。本次研

收稿日期: 2022-01-25; 修回日期: 2022-06-27

基金项目: 中国地质调查局项目“东北黑土地 1:25 万土地质量地球化学调查”(121201007000161312)、“兴凯湖平原及松辽平原西部土地质量地球化学调查”(DD20190520)

第一作者: 陈超群(1996-), 女, 硕士研究生, 主要研究方向为生态环境遥感与地理信息系统。Email: 522110156@qq.com

通讯作者: 杨佳佳(1984-), 男, 博士, 高级工程师, 主要从事遥感应用研究工作。Email: haixianxiaomei@163.com

究借助 Sentinel-2A 遥感影像,结合黑河市孙吴县实测土壤有机质含量,通过 Pearson 相关分析和随机森林(RF)选择不同特征波段作为模型输入量,采用 PLSR 和 BP 神经网络建模,以期研究土壤表层有机质含量与遥感影像关系,并实现地面黑土区红旗林场的土壤有机质高精度快速反演。

1 研究区概况

研究区孙吴县地处黑龙江省黑河市中部,位于东经 126°39'35"~128°1'6",北纬 48°59'00"~49°41'55"(图 1)。东部紧靠逊克县,西边为嫩江县,南侧

与五大连池市相挨,北方为黑河市爱辉区,总面积 4 318.9 km²。孙吴县海拔 110~755 m,属于低山丘陵区,地势总体呈西南高东北低趋势。地貌分界清晰,从西到东分别为低山沟谷区、丘陵河谷地区和沿江平原。土壤类型以暗棕壤和草甸黑土为主^[10]。气候属于寒温带大陆性季风气候,年均气温 -0.6℃,年均降雨约 550 mm,冻结期较长,无霜期短^[11]。本文选择孙吴县红旗林场地区进行遥感反演,红旗林场位于孙吴县西北方向,范围为东经 126°41'25"~127°14'34",北纬 49°16'32"~49°30'58",界内发育孙吴县最高山峰松木山。

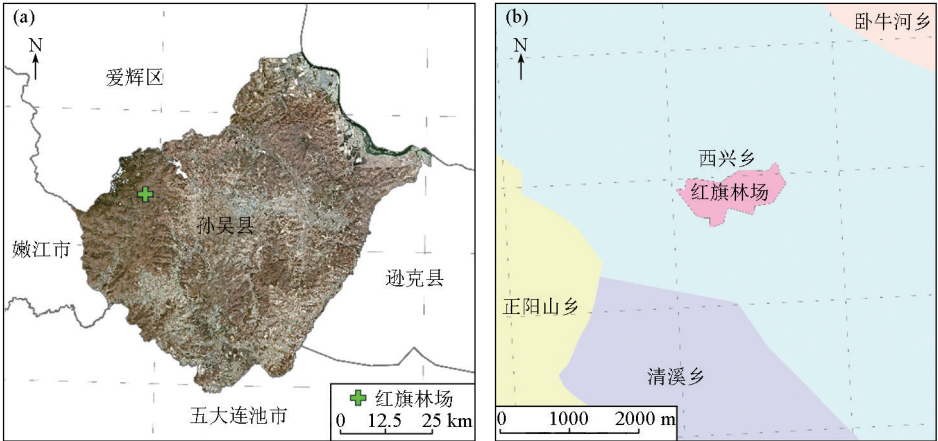


图 1 孙吴县遥感影像(a)及红旗林场位置(b)
Fig. 1 Remote sensing image of Sunwu County (a) and the location of Hongqi Forest Farm (b)

2 数据采集与处理

2.1 土壤采集与有机质测定

按照《土地质量地球化学评价规范》(DZ/T 0295—2016)采样要求,在孙吴县采集土壤时去除表面枯枝落叶等杂物,用刻槽法垂直采集地表至 20 cm 深土样,保证上下均匀采集,并去除动、植物残留体、砾石、肥料团块等。土壤有机质含量采用硫酸—重铬酸钾法测定。共计采集 806 个样品,其中 564 个土样作为建模集,242 个样品为测试集,统计信息如表 1 所示。

表 1 土壤样品中有机质含量统计信息

Table 1 Statistical information of organic matter content in soil samples					
	个数	最小值/%	最大值/%	均值/%	标准差/%
建模集	564	0.8620	11.8266	5.7226	2.0316
测试集	242	1.1896	11.9128	5.8393	2.1074

2.2 遥感数据获取与处理

选取研究区内 2018 年 11 月 7 日裸土无雪时期

的 Sentinel-2A 影像,云覆盖 0%。影像的预处理包括几何校正、大气校正、图像镶嵌及图像剪裁等操作。所选影像为 Level-1C 上层大气反射产品,已经过系统几何精校正处理,其精度在一个像元内,满足研究需求。借助 SNAP 软件中 Sen2cor280 工具箱实现大气校正,校正后丢失卷云波段 B10。为提高土壤有机质与光谱反射率(R)相关性,对遥感影像进行倒数($1/R$)、对数($\lg R$)、幂函数(R^a)、一阶微分(FDR)、二阶微分(SDR)及倒数对数一阶微分(FDLR)处理。

3 算法原理

3.1 Pearson 相关分析

为获取土壤有机质光谱响应波段,研究中采用 Pearson 相关判断反射率与有机质含量之间的线性相关性。公式为:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(X_i - \bar{X})^2} \sqrt{(Y_i - \bar{Y})^2}}, \quad (1)$$

式中: r 表示相关系数; X_i 为样品 i 的波段的反射率; Y_i 为样品 i 的有机质含量; \bar{X} 为反射率均值; \bar{Y} 为有机质含量均值。为防止出现虚假相关,引入显著水平 sig 进行判断。将通过显著性检验的波段及其变换作为土壤有机质的特征波段。

3.2 随机森林

2001 年 Breiman 提出随机森林(random forest, RF)算法,主要优势体现在处理多维数据集时无需降维,最大程度的保留数据集原始信息^[12-13]。本次研究中以 X. IncMSE 指标为重要性选择依据,X. IncMSE 值越大,表明该波段信息越能反应有机质含量。

3.3 偏最小二乘回归模型

偏最小二乘回归分析(partial least square regression, PLSR)常用于遥感光谱反演建模,优势在于建模过程中集中了主成分分析、典型相关分析和线性回归分析方法的优点^[14]。其建模思路为:设自变量 $\mathbf{X} = [x_1, x_2, \dots, x_p]_{n \times p}$, 因变量为 $\mathbf{Y} = [y]_{n \times 1}$, 其中 \mathbf{X} 为波谱曲线反射率及其变换形式, \mathbf{Y} 为土壤成分含量, n 为采集样本数, p 为特征波段数。从自变量信息中提取最大变异信息成分 μ_1 , 且与因变量呈最大相关性。在提取第一主成分 μ_1 后,建立 \mathbf{Y} 与 μ_1 的回归方程,若模型精度满意,则算法停止;否则继续利用 \mathbf{X} 和 \mathbf{Y} 被成分解释后的残余信息进行多次迭代提取,直到回归方程达到满意精度。

为了检验反演模型的精度及稳定性,借助决定系数 R^2 和均方根误差 $RMSE$ 作为模型评价指标,公式如下:

$$R^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^n (y_i - \bar{y})^2, \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (3)$$

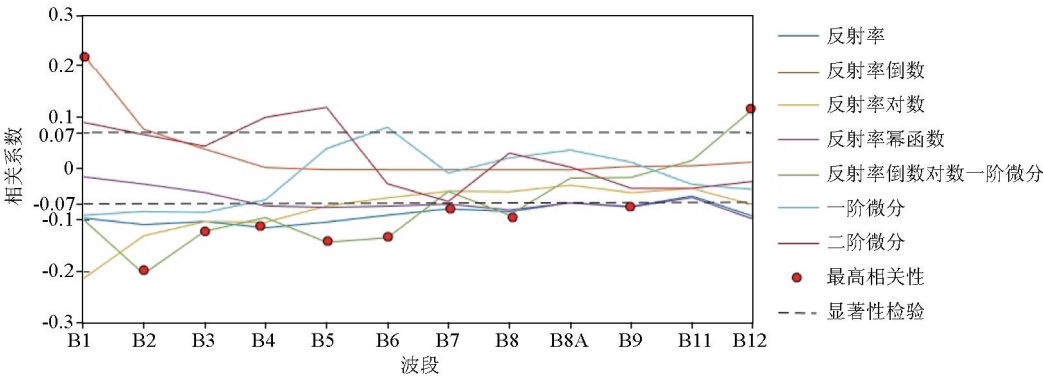


图2 波段反射率及其变换与土壤有机质含量相关性

Fig. 2 Correlation between band reflectivity and transformations and soil organic matter content

式中: n 为样本个数; y_i 为土壤有机质含量真实值; \bar{y} 为有机质含量平均值; \hat{y}_i 为有机质含量预测值。 R^2 越大,表明模型越稳定,拟合效果越好。 $RMSE$ 值越小,表明模型的估测能力越稳定^[15]。

3.4 BP 神经网络

BP 神经网络(误差反向传播)是人工神经网络的一种,由输入层—隐含层—输出层组成^[16]。学习过程包含信号的正向传播与误差的反向传播,正向传播时,信号从输入层传入,经隐含层逐层处理后,最后达到输出层。若输出层的实际输出与期望输出不符,则转向误差的反向传播阶段^[17]。BP 神经网络具有较强的非线性处理能力和自适应特点,能够较好地拟合光谱反射率与土壤有机质含量的关系。通过经验公式(4)确定隐含层的节点数取值范围,结合训练结果的精度选择最佳隐含层的节点数:

$$N = \sqrt{n + m} + k, \quad (4)$$

式中: n 为输入层节点数; m 为输出层节点数; k 为 1~10 之间的常数; N 为隐藏层节点数。

4 黑土区土壤有机质反演

4.1 特征波段选择

4.1.1 相关性波段选择

在 SPSS 26 平台下计算土壤有机质含量与 Sentinel-2A 遥感影像反射率及其变换间的相关性。如图 2 所示,有机质含量与光谱反射率呈现负相关,但相关性不高,各波段不同数学变换的相关系数最高值绝大多数出现在 FDLR 变换,表明该预处理方法能有效提高 Sentinel-2A 反射率与土壤有机质的相关性。将不同数学变换中通过显著性检验的波段作为相应反演的特征波段,同时组合各波段中最高相关系数的变换形式作为一种响应波段参考,其中 B8 和 B11 无通过显著性检验的变换,故不予讨论(表 2)。

表 2 相关性选取特征波段

Table 2 Feature bands selected by correlation analysis

数学变换	波段
R	B1、B2、B3、B4、B5、B6、B7、B8、B9、B12
$1/R$	B1、B2
$\lg R$	B1、B2、B3、B4、B5、B12
R^a	B4、B5、B6、B7、B8、B9、B12
FDR	B1、B2、B3、B4、B5、B6
SDR	B1、B4、B5
FDLR	B1、B2、B3、B4、B5、B6、B8、B12
组合波段	$1/B1$ 、 $FDL(B2)$ 、 $\lg(B3)$ 、 $B4$ 、 $FDL(B5)$ 、 $FDL(B6)$ 、 $B7$ 、 $FDL(B8)$ 、 $(B9)^a$ 、 $FDL(B12)$

4.1.2 RF 重要性选择

采用 R 语言的 randomForest 包实现土壤有机质特征波段选取,其中默认生成 500 棵决策树,并进行 5 次重复十折交叉验证,结合最简原则选择不同光谱变换下的特征波段。以对数变换为例,图 3 交叉验证曲线展示了模型误差与用于拟合的自变量数量的关系,当波段数为 6 时,误差下降幅度基本保持不变,结合简约性原则,选择重要程度值从大到小排序前 6 的波段作为有机质反演建模的输入参数,实验中 R 、 $1/R$ 、 R^a 、FDR、SDR 以及 FDLR 均需要 6 个重要变量表示土壤有机质含量。为提高变量表达精度,将所有波段的变换作为 RF 的因变量,并需要 X.IncMSE 值前 26 个波段变换作为建模输入参数集,以精准表达有机质含量信息(图 4)。RF 重要波段选取结果如表 3 所示。

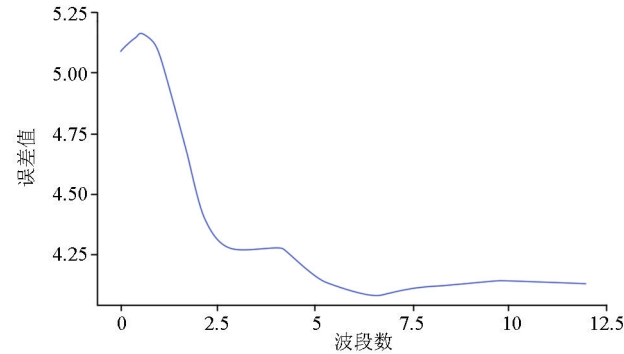


图 3 对数变化交叉验证曲线

Fig. 3 Cross validation curve of lgR

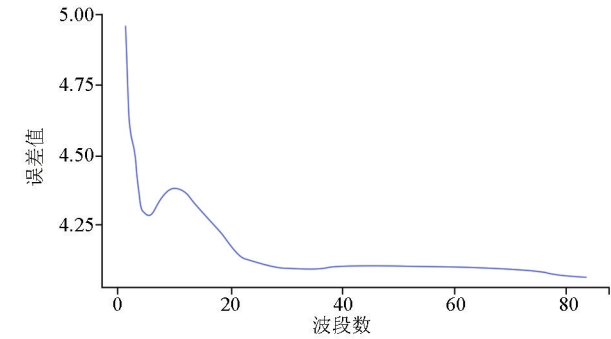


图 4 所有波段交叉验证曲线

Fig. 4 Cross validation curve of all bands

表 3 RF 重要波段

Table 3 Important bands of RF

数学变换	波段
R	B1、B2、B3、B5、B7、B8A
$1/R$	B1、B2、B3、B5、B7、B8A
$\lg R$	B1、B2、B3、B5、B6、B7
R^a	B2、B3、B5、B6、B7、B8A
FDR	B1、B2、B3、B4、B6、B12
SDR	B1、B3、B4、B5、B11、B12
FDLR	B1、B2、B3、B4、B8、B12
组合波段	B1、B7、B11、 $1/B1$ 、 $1/B2$ 、 $1/B3$ 、 $1/B11$ 、 $1/B1$ 、 $\lg(B2)$ 、 $FDL(B1)$ 、 $FDL(B2)$ 、 $FDL(B3)$ 、 $FDL(B5)$ 、 $FDL(B6)$ 、 $FDL(B8)$ 、 $FD(B1)$ 、 $FD(B5)$ 、 $FD(B6)$ 、 $FD(B12)$ 、 $SD(B1)$ 、 $SD(B2)$ 、 $SD(B3)$ 、 $SD(B5)$ 、 $SD(B8)$ 、 $SD(B11)$ 、 $SD(B12)$

4.2 土壤有机质反演

4.2.1 PLSR 模型反演

将 Pearson 相关分析(表 2)和 RF(表 3)提取的特征波段作为自变量,土壤有机质含量作为因变量,建立有机质含量 PLSR 反演模型,如表 4 所示,结果显示相关-PLSR 模型和 RF-PLSR 模型反演精度结果相近,FDLR 变换和组合波段都能有效提高模型反演精度,其中针对传统的相关-PLSR 模型,FDLR 光谱变换的拟合程度最好,RF-PLSR 模型中组合波段的效果更为显著。但 PLSR 模型下建模集和测试集的决定系数 R^2 均未超过 0.1。

4.2.2 BP 神经网络模型反演

结合式(4),根据相关性和 RF 重要程度选取获得的特征波段数,确定 BP 神经网络模型中隐含层层数。网络的训练函数为 Trainlm,输入层和输出层传递函数分别为 Tansig 和 Purelin,表 5 为相关分析和 RF 两种特征响应波段模拟结果,对比 PLSR 拟合结果(表 4),非线性拟合的多光谱遥感影像反射率与土壤有机质含量模型精度能得到显著提高。由于多光谱遥感的光谱分辨率较低,光谱包含的土壤信息较为复杂,因此无法类比高光谱土壤有机质遥感反演,线性回归拟合模型不能有效提取影像上土壤有机质含量信息^[18-20]。相关性提取波段与 RF 提取的重要波段在进行 BP 神经网络建模时,建模集和测试集的 R^2 主要集中在 0.2~0.5, RMSE 集中在 1.3%~1.4%。相关-BP 神经网络模型中 FDLR 建模拟合程度最高,建模集 R^2 为 0.623 7, RMSE 为 1.354 8%,测试集 R^2 为 0.444 6, RMSE 为 1.266 4%。RF-BP 神经网络模型中组合波段建模拟合程度最高,建模集 R^2 为 0.724 5, RMSE 为 1.312 7%,测试集 R^2 为 0.541 8, RMSE 为 1.372 2%。

表 4 基于 PLSR 模型的土壤有机质反演
Table 4 Inversion of soil organic matter by PLSR

数学变换	拟合模型	建模集		测试集	
		R^2	RMSE/%	R^2	RMSE/%
相关性	R $y = 6.3252 - 0.1679x_1 - 0.3836x_2 - 0.4512x_3 - 0.6285x_4 - 0.6827x_5 - 0.6340x_6 + 0.5943x_7 - 0.5943x_8 - 0.5199x_9 - 0.5565x_{10}$	0.0275	2.0263	0.0155	2.0925
	$1/R$ $y = 5.2818 + 0.0001x_1 + 0.000017x_2$	0.0439	1.9847	0.0657	2.0375
	$\lg R$ $y = 5.4429 - 0.5202x_1 - 0.2827x_2 + 1.1558x_3 - 1.5976x_4 - 0.5254x_5 + 3.3470x_6$	0.0525	1.9757	0.0609	2.0406
	R^a $y = 6.0894 + 48.0295x_1 - 110.9565x_2 + 21.8818x_3 + 78.5713x_4 - 33.5780x_5 - 10.4766x_6 - 5.2279x_7$	0.0159	2.0136	0.0121	2.0917
	FDR $y = 5.8376 - 23.9683x_1 + 67.1411x_2 - 63.9701x_3 + 44.2371x_4$	0.0223	2.0071	0.0409	2.0632
	SDR $y = 6.1351 + 4.19x_1 + 14.4844x_2 + 59.2323x_3$	0.0200	2.0093	0.0238	2.0800
	FDLR $y = 5.0094 + 1.7352x_1 - 6.2089x_2 + 6.6447x_3 - 5.4056x_4 + 3.6822x_5 - 11.4086x_6 - 3.5433x_7 - 3.0895x_8$	0.0524	1.9779	0.071	2.0366
	组合 $y = 8.8137 - 0.0162x_1 + 1.7147x_2 + 0.4484x_3 - 0.1764x_4 - 1.1268x_5 - 0.4075x_6 - 0.5233x_7 - 1.5618x_8 - 0.8610x_9 - 1.2728x_{10}$	0.0434	1.9852	0.0400	1.9933
随机森林	R $y = 5.77391 - 8.6669x_1 - 0.6491x_2 + 9.5163x_3 - 26.4628x_4 + 8.9020x_5 + 6.6144x_6$	0.0203	2.0318	0.0035	2.0539
	$1/R$ $y = 5.80955 + 0.0002x_1 + 0.0001x_2 + 0.0009x_3 - 0.1832x_4 - 0.0226x_5 + 0.2048x_6$	0.0234	2.2798	0.09	2.0390
	$\lg R$ $y = 4.40343 - 0.4786x_1 - 0.0567x_2 + 0.0415x_3 + 0.0548x_4 + 0.0692x_5 + 0.0786x_6$	0.0480	2.0029	0.0408	2.0124
	R^a $y = 5.95664 - 0.4164x_1 - 0.8538x_2 - 2.1607x_3 - 2.2302x_4 - 2.3068x_5 - 2.6665x_6$	0.0051	2.0475	0.0023	2.0525
	FDR $y = 5.65109 - 33.6899x_1 + 85.8885x_2 - 75.5978x_3 - 13.8874x_4 + 28.8994x_5 - 6.2782x_6$	0.0365	2.0149	0.0207	2.0356
	SDR $y = 5.55202 + 13.2022x_1 - 29.3378x_2 + 42.5214x_3 + 53.8592x_4 - 3.5595x_5 - 2.0586x_6$	0.0331	2.0496	0.0200	2.0520
	FDLR $y = 4.69522 - 0.0386x_1 - 0.7984x_2 - 0.6553x_3 - 0.1570x_4 - 0.0148x_5 + 0.0701x_6$	0.0463	2.0046	0.0418	2.0118
	组合 $y = 1.2967x_1 + 0.2947x_2 + 2.6967x_3 + 0.5528x_4 + 0.8238x_5 + 0.7684x_6 - 0.2272x_7 - 0.6389x_8 + 1.0292x_9 + 1.3686x_{10} - 1.8264x_{11} + 0.7614x_{12} - 0.1926x_{13} + 1.2489x_{14} - 0.1014x_{15} - 0.3382x_{16} + 0.9709x_{17} - 0.3017x_{18} + 1.5403x_{19} + 2.7079x_{20} + 0.0221x_{21} - 3.4156x_{22} + 3.8703x_{23} + 4.6421x_{24} - 0.5339x_{25} - 1.746x_{26} - 1.7177$	0.0760	1.9518	0.0363	2.0183

结合表 4 中对比 R 、 $1/R$ 等波段变换的不同提取方法, Pearson 相关分析中选择的 FDLR 反射率变换在线性回归和非线性回归中都展现较高的拟合效果。因为 Pearson 相关性分析获取的是有机质含量与反射率间简单直线性相关的方向和密切程度, 因此光谱处理的程度直接决定了与有机质含量的相关性, 进而影响了模型反演精度。FDLR 变换中光谱倒数对数计算可以有效放大相似光谱间的差别, 再经过一阶微分处理后消除部分线性的背景, 同时降低噪声光谱对目标光谱的影响程度。但相关性选择的组合波段包含了反射率变量处理的不同级别, 当

加入相关性低的波段变换, 很有可能引入了土壤其他成分的特征信息, 导致有机质反演精度降低。而在 RF 重要程度选择中存在误差验证, 结合每次选择的特征集计算袋外误差率, 最后选择袋外误差率最低的特征集作为回归模型的输入集。筛选出的波段可能与有机质含量相关性低, 但叠加其他波段光谱特征反而提高了有机质估测精度。因此 RF 波段选取时, 组合所有光谱反射率及变换信息后筛选出的特征响应波段更能充分反映有机质含量信息, 建模精度显著提升。

表 5 基于 BP 神经网络模型的土壤有机质反演
Table 5 Inversion of soil organic matter by BP neural network

	数学变换	隐藏层个数	建模集		测试集	
			R^2	RMSE/%	R^2	RMSE/%
相关性	R	5	0.3635	1.3998	0.2711	1.4472
	$1/R$	11	0.2816	1.4057	0.2291	1.4283
	$\lg R$	6	0.3392	1.3967	0.2818	1.4207
	R^a	11	0.2726	1.4041	0.2388	1.4555
	FDR	10	0.2074	1.4149	0.1697	1.4368
	SDR	9	0.2005	1.4106	0.1977	1.4388
	FDLR	6	0.6237	1.3548	0.4446	1.2664
	组合	16	0.5637	1.3548	0.4305	1.3659
随机森林	R	7	0.2906	1.4000	0.2601	1.4362
	$1/R$	7	0.2603	1.4068	0.2241	1.4262
	$\lg R$	14	0.2883	1.4068	0.2664	1.4231
	R^a	14	0.1663	1.4152	0.2499	1.4200
	FDR	7	0.3751	1.3980	0.2860	1.4090
	SDR	11	0.2783	1.4107	0.2057	1.4520
	FDLR	13	0.4544	1.3750	0.3653	1.2420
	组合	16	0.7245	1.3127	0.5418	1.3722

4.3 红旗林场有机质空间分布

对比 8 种影像变换的相关-PLSR、相关-BP 神经网络、RF-PLSR、RF-BP 神经网络建模情况,选择模型拟合度最高、稳定性最好的 RF-BP 神经网络模型作为多光谱遥感数据的土壤有机质反演模型,并预测孙吴县红旗林场的有机质含量分布(图 5)。红旗林场的有机质范围大致在 0.1%~18.8%,主要集中在 3.742 6%~12.455 2%,平均含量为 7.939 9%,呈

现出中间高、四周低的分布趋势。该地区土壤有机质地球化学分析结果显示,红旗林场有机质平均含量为 8.51%,分布趋势为中西部含量偏高,向北侧逐渐降低,该结果与遥感反演程度几乎吻合。但由于在林场附近采集土壤样品较少,因此地球化学实测结果分布较为粗糙,而遥感反演获取的土壤有机质含量更为细致。

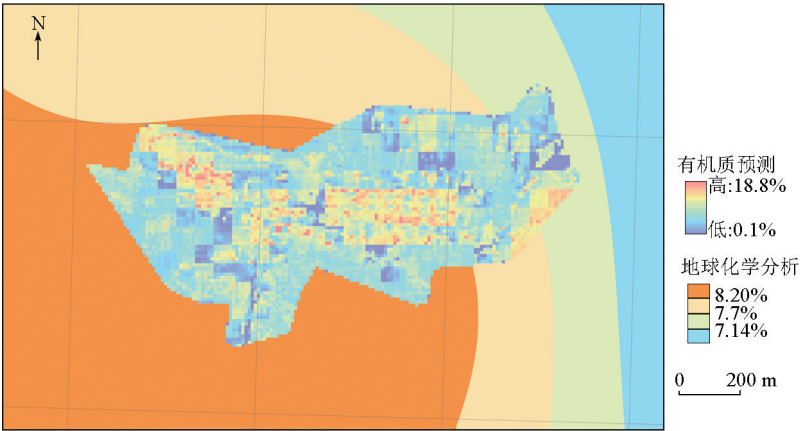


图 5 红旗林场土壤有机质遥感反演和地球化学对比
Fig.5 Distribution of soil organic matter in Hongqi Forest Farm

5 结论

对 Sentinel-2A 多光谱遥感影像反射率进行 1/ R 、 $\lg R$ 、 R^a 、FDR、SDR 及 FDLR 变换,结合不同模型实现了土壤有机质的反演,取得很好应用成效。主要得出以下结论:

1)通过相关性分析法建模时反射率的 FDLR 变

换模型拟合程度最好,而采用 RF 算法筛选的组合波段在反演时能有效提高土壤有机质含量建模精度。

2)多光谱遥感影像光谱分辨率较低,因此线性拟合模型无法准确估测土壤有机质含量,需要非线性模型以实现光谱信息与有机质含量的有效拟合。

3)对比不同遥感影像预处理操作下的相关-

PLSR、相关-BP 神经网络、RF-PLSR、RF-BP 神经网络建模情况,RF-BP 神经网络模型反演土壤有机质含量拟合程度最高,建模集 R^2 为 0.724 5, $RMSE$ 为 1.312 7%,测试集 R^2 为 0.541 8, $RMSE$ 为 1.372 2%。

参考文献 (References):

[1] 孔牧, 杨少平. 森林沼泽景观区有机质对元素表生地球化学特征的影响机制[J]. 物探与化探, 2008, 32(1): 31-32, 74.
Kong M, Yang S P. Preliminary research into the disturbed principle of organic material to character of supergene-geochemistry in forest marsh landscape andscape area [J]. Geophysical and Geochemical Exploration, 2008, 32(1): 31-32, 74.

[2] Rasmussen C, Heckman K, Wieder W R, et al. Beyond clay: Towards an improved set of variables for predicting soil organic matter content[J]. Biogeochemistry, 2018, 137(5): 297-306.

[3] 戴慧敏, 刘凯, 宋运红, 等. 东北地区黑土退化地球化学指示与退化强度[J]. 地质与资源, 2020, 29(6): 510-517. DOI: 10.13686/j.cnki.dzyz. 2020. 06. 002.
Dai H M, Liu K, Song Y H, et al. Black soil degradation and intensity in northeast China: Geochemical indication [J]. Geology and Resources, 2020, 29(6): 510-517. DOI: 10.13686/j.cnki.dzyz. 2020. 06. 002.

[4] 刘焕军, 张美薇, 杨昊轩, 等. 多光谱遥感结合随机森林算法反演耕作土壤有机质含量[J]. 农业工程学报, 2020, 36(10): 134-140.
Liu H J, Zhang M W, Yang H X, et al. Inversion of cultivated soil organic matter content combining multi-spectral remote sensing and random forest algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(10): 134-140.

[5] 屈冉, 张雅琼, 聂忆黄, 等. 基于多光谱遥感影像的富川县表层土壤有机质含量反演[J]. 环境与可持续发展, 2019, 44(1): 154-157.
Qu R, Zhang Y Q, Nie Y H, et al. Inversion of surface soil organic matter content in Fuchuan county based on multi spectral remote sensing image [J]. Environment and Sustainable Development, 2019, 44(1): 154-157.

[6] 陈德宝, 陈桂芬. 基于 Landsat8 遥感图像的黑土区土壤有机质含量反演研究[J]. 中国农机化学报, 2020, 41(6): 194-198.
Chen D B, Chen G F. Inversion of soil organic matter content in black soil region based on landsat8 remote sensing image [J]. Journal of Chinese Agricultural Mechanization, 2020, 41(6): 194-198.

[7] 陈思明, 邹双全, 毛艳玲, 等. 土壤光谱重建的湿地土壤有机质含量多光谱反演[J]. 光谱学与光谱分析, 2018, 38(3): 912-917.
Chen S M, Zou S Q, Mao Y L, et al. Inversion of soil organic matter content in wetland using multispectral data based on soil spectral reconstruction [J]. Spectroscopy and Spectral Analysis, 2018, 38(3): 912-917.

[8] Dhawale N M, Adamchuk V I, Prasher S O, et al. Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer[J]. European Journal of Soil Science,

2015, 66(4): 661-669.

[9] 马驰. 基于 Sentinel-2A 遥感影像土壤有机质含量的反演研究[J]. 北方园艺, 2020(2): 94-100.
Ma C. Inversion of soil organic matter content based on sentinel-2A remote sensing image [J]. Northern Horticulture, 2020(2): 94-100.

[10] 刘鹏. 孙吴县耕地质量评价[D]. 哈尔滨: 东北农业大学, 2020.
Liu P. Evaluation of cultivated land quality in Sunwu County[D]. Harbin: Northeast Agricultural University, 2020.

[11] 李丹丹. 黑河市耕地地力评价与土壤改良对策研究[D]. 哈尔滨: 东北农业大学, 2018.
Li D D. Investigation and evaluation on cultivated land fertility of Heihe City [D]. Harbin: Northeast Agricultural University, 2018.

[12] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.

[13] 彭刘亚, 解惠婷, 冯伟栋. 基于随机森林算法的砂土液化预测方法[J]. 物探与化探, 2020, 44(6): 1429-1434.
Peng L Y, Xie H T, Feng W D. The method of predict sand liquefaction based on random forest algorithm [J]. Geophysical and Geochemical Exploration, 2020, 44(6): 1429-1434.

[14] 王琨, 肖克炎, 丛源. 对数比变换和偏最小二乘法在地球化学组合异常提取中的应用——以湘西北铅锌矿为例[J]. 物探与化探, 2015, 39(1): 141-148.
Wang K, Xiao K Y, Cong Y. Log-ratio transformation and PLS methods for identifying integrated geochemical anomalies: A case study of lead-zinc mineralization in northwestern Hunan [J]. Geophysical and Geochemical Exploration, 2015, 39(1): 141-148.

[15] 陈昊宇, 杨光, 韩雪莹, 等. 基于连续小波变换的土壤有机质含量高光谱反演[J]. 中国农业科技导报, 2021, 23(5): 132-142. DOI: 10.13304/j.nykjdb. 2020. 0742.
Chen H Y, Yang G, Han X Y, et al. Hyperspectral inversion of soil organic matter content based on continuous wavelet transform journal of agricultural science and technology [J]. Journal of Agricultural Science and Technology, 2021, 23(5): 132-142. DOI: 10.13304/j.nykjdb. 2020. 0742.

[16] 叶勤, 姜雪芹, 李西灿, 等. 基于高光谱数据的土壤有机质含量反演模型比较[J]. 农业机械学报, 2017, 48(3): 164-172.
Ye Q, Jiang X Q, Li X C, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data [J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 164-172.

[17] 王启元, 赵艳玲, 房铄东, 等. 基于多光谱遥感的裸土土壤含水量反演研究[J]. 矿业科学学报, 2020, 5(6): 608-615.
Wang Q Y, Zhao Y L, Fang S D, et al. Inversion of soil moisture in bare soil based on multi-spectral remote sensing [J]. Journal of Mining Science and Technology, 2020, 5(6): 608-615.

[18] 汤超. 淮北矿区有机质含量反演[J]. 农业与技术, 2021, 41(13): 123-128. DOI: 10.19754/j.nyyjs. 20210715035.
Tang C. Inversion of organic matter content in Huaibei mining area [J]. Agriculture and Technology, 2021, 41(13): 123-128. DOI: 10.19754/j.nyyjs. 20210715035.

[19] 谢树刚. 基于高光谱的黄河三角洲土壤有机质含量估测模型研究[D]. 泰安: 山东农业大学, 2021. DOI: 10.27277/d.cnki.

gsdnu. 2021. 000631.
Xie S G. Research on estimation model of soil organic matter content in Yellow River Delta based on hyperspectral [D]. Tai'an: Shandong Agricultural University, 2021. DOI: 10. 27277/d. cnki. gsdnu. 2021. 000631.

模型研究[J]. 地质与资源, 2020, 29(1): 68-75, 84. DOI: 10. 13686/j. cnki. dzyz. 2020. 01. 006.
Tao P F, Wang J H, Li Z Z, et al. Research of soil nutrient content inversion model based on hyperspectral data [J]. Geology and Resources, 2020, 29(1): 68-75, 84. DOI: 10. 13686/j. cnki. dzyz. 2020. 01. 006.

[20] 陶培峰, 王建华, 李志忠, 等. 基于高光谱的土壤养分含量反演

Sentinel-2A based inversion of the organic matter content of soil in the Sunwu area

CHEN Chao-Qun^{1,2,3}, DAI Hui-Min^{1,2,3}, FENG Yu-Lin¹, YANG Ze^{1,2,3}, YANG Jia-Jia¹

(1. Shenyang Center of China Geological Survey, Shenyang 110034, China; 2. Key Laboratory of Black Soil Evolution and Ecological Effect, Ministry of Natural Resources, Shenyang 110034, China; 3. Key Laboratory of Black Soil Evolution and Ecological Effect, Liaoning Province, Shenyang 110034, China)

Abstract: This study conducted the inversion of the organic matter content in the soil of the black soil area in Sunwu County, Heilongjiang Province using the Sentinel-2A multispectral remote sensing images and the surveyed soil data. After preprocessing the images, the characteristic bands were selected through correlation analysis and using the random forest (RF) method. Subsequently, a multispectral inversion model for the organic matter content of the soil was built using the partial least square method and the BP neural network, and the inversion of the organic matter content of the soil in the Hongqi Forest Farm was conducted. According to the obtained results, the bands selected based on the reciprocal of the logarithm of the first-order differential of reflectance through the correlation analysis and the combined bands selected using the RF method can effectively improve the inversion precision of the organic matter content in the soil, and the RF-BP neural network model for the combined bands yielded the optimal inversion performance ($R^2=0.7245$ and $RMSE=1.3127\%$). The results of this study will provide technical support and reference for the dynamic monitoring of the organic matter content in soils.

Key words: black soil; organic matter; Sentinel-2A; random forest; BP neural network

(本文编辑:蒋实)