

基于高光谱技术的退耕还林地年限判别

邓永鹏, 朱洪芬, 丁皓希, 孙瑞鹏, 毕如田*

(山西农业大学资源环境学院, 太原 030000)

摘 要: 自 2000 年以来, 黄河中游坡度较大的不同区域、同一区域的不同部位在不同年度实施了退耕还林工程, 促进了黄河中游土壤质量及生态环境的改善。为了研究退耕工程对土壤及环境的影响机制, 需要快速获取退耕年限及土壤特征。该研究以黄河中游大宁县不同年限退耕还林土壤为研究对象, 获取不同年限退耕还林土壤理化性质, 同时测定不同退耕年限土壤光谱特征曲线, 以表征不同退耕年限的土壤属性及光谱特征; 以土壤原始光谱反射率数据 (Reflectance, R) 为基础, 采用 Savitzky-Golay 平滑 (Savitzky-Golay smooth, SG)、倒数的对数 (Reciprocal of Logarithm, RL)、一阶微分 (First Order Differential, FD)、去包络线 (Continuum Removal, CR)、主成分分析 (Principal Component Analysis, PCA) 以及光谱特征参数 (Spectral Characteristic Parameter, SCP) 等光谱预处理, 以原始反射率主成分 (R-PCA)、倒数的对数主成分 (RL-PCA)、一阶微分主成分 (FD-PCA)、去包络线主成分 (CR-PCA)、SCP 为输入因子, 采用 K 均值聚类 (K-means Clustering Algorithm, K-means)、支持向量机 (Support Vector Machine, SVM) 和线性判别分析 (Linear Discriminant Analysis, LDA) 构建退耕年限的分类模型并选取最优模型。结果表明: 1) 退耕年限的增加会导致土壤理化性质的显著变化, 土壤有机碳 (Soil Organic Carbon, SOC) 含量、土壤含水率逐渐增加, 土壤黏粒含量逐渐减少, 土壤粉粒、砂粒及饱和导水率呈现先增加后减少的趋势; 2) 不同退耕年限土壤光谱曲线差异细微, 预处理 CR 可显著提升光谱曲线的吸收特征, 在 480、900、1 100、1 400、1 900、2 200 和 2 350 nm 处出现明显的吸收特征; 3) 3 种分类模型取得了较为理想的分类精度, 其中 LDA 模型最优, Kappa 系数最大为 0.83; 5 种输入因子分类效果差异显著, 其中 CR-PCA 分类效果最好, 不同模型分类精度均达到 75% 以上。该研究通过土壤光谱曲线探索不同年限退耕还林土壤的光谱特征及分类方法, 可实现退耕年限的快速区分, 为退耕还林工程对土壤属性及环境影响的进一步研究提供参考。

关键词: 土壤; 有机碳; 模型; 退耕还林; 退耕年限; 高光谱; 黄河中游

doi: 10.11975/j.issn.1002-6819.2022.03.008

中图分类号: S127; S157.9

文献标志码: A

文章编号: 1002-6819(2022)-03-0066-09

邓永鹏, 朱洪芬, 丁皓希, 等. 基于高光谱技术的退耕还林地年限判别[J]. 农业工程学报, 2022, 38(3): 66-74.

doi: 10.11975/j.issn.1002-6819.2022.03.008 <http://www.tcsae.org>

Deng Yongpeng, Zhu Hongfen, Ding Haoxi, et al. Identification of the years of returning farmland to forest land using hyperspectral technology[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(3): 66-74. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.03.008 <http://www.tcsae.org>

0 引 言

黄河中游的大部分支流均处于黄土高原区, 面积约为 $34.4 \times 10^4 \text{ km}^2$, 是中国分布面积较大的生态脆弱区^[1]。为改善黄河中游的生态环境问题, 中国于 1999 年试行, 并于 2000 年正式开始实施退耕还林(草)工程^[2]。随着退耕还林工程长久实施, 研究不同年限退耕还林土壤的综合性质对评估退耕还林工程的效果具有重要意义^[3-5]。由于部分地区无法提供有效的退耕资料, 因此快速、准确地识别退耕还林地的年限对评估退耕还林工程的生态效益具有重要价值。

近年来, 光谱分析技术被广泛应用到中药材种植年限识别^[6-7]、储藏年限识别^[8-9], 野生食用菌储藏年限识别^[10]、产地识别^[11], 农作物种类区分^[12]、茶树种类识

别^[13]、土壤类型分类^[14]等不同领域内。陈泽炎等^[15]对目前中药材生长年限鉴定方法进行了总结, 发现光谱分析技术可以高效、快速地识别不同种植年限的中药材。卜海博等^[16]采用近红外光谱法对不同生长年限的林下山参和园参进行种类和年限识别, 结果表明近红外光谱法准确可靠、快速无损。杨天伟等^[11]采用傅里叶变换红外光谱技术, 准确区分了不同产地、不同年份美味牛肝菌样品, 为野生食用菌的鉴别分类提供了技术参考。虞佳维等^[17]以黄淮海地区玉米、小麦和杨树三种主要植被为研究对象, 采用高光谱技术进行区分, 结果表明光谱技术可有效用于不同植被的区分。翁海勇等^[18]采用近红外光谱分析技术实现了茶树叶片表没食子儿茶素没食子酸酯 (Epigallocatechin Gallate, EGCG) 含量的快速、准确识别, 从而为高 EGCG 含量茶树的育种提供技术支持。史舟等^[19]基于中国西藏、新疆、黑龙江、海南等地的 16 种土类的 1 581 个土壤样本, 以光谱反射率一阶微分主成分数据为自变量进行土壤光谱分类, 并根据分类结果建立了中国不同土壤类型的有机质预测模型。赵小敏等^[20]以江西红壤地区的 7 种红壤亚类土壤为研究

收稿日期: 2021-09-28 修订日期: 2021-12-10

基金项目: 国家重点研发计划 (2021YFD1600301)

作者简介: 邓永鹏, 研究方向为土地利用与信息技术。

Email: dengyp0411@163.com

*通信作者: 毕如田, 教授, 博士生导师, 研究方向为资源环境信息技术。

Email: brrt@sxau.edu.cn

对象,对不同亚类及土属的土壤高光谱特征进行分析,选取特征变量进行 Fastclus 聚类分析,结果表明,土壤亚类分类的准确度为 86.23%,土属的准确率仅为 66.37%。目前,尚未发现开展退耕还林地不同年限识别的研究,应用光谱分析技术可以为不同年限退耕还林地的分类提供技术参考。

本研究以黄河中游大宁县不同年限退耕还林地土壤为研究对象,以不同退耕年限刺槐林地土壤光谱曲线和多种土壤理化数据为基础,分析退耕年限对土壤光谱及其综合属性的影响。利用倒数的对数(Reciprocal of Logarithm, RL)、一阶微分(First Order Differential, FD)等光谱处理方法,结合主成分分析(Principal Component Analysis, PCA)、K 均值聚类(K-means Clustering Algorithm, K-means)、支持向量机(Support Vector Machine, SVM)和线性判别分析(Liner Discriminant Analysis, LDA)等方法探索不同退耕年限土壤的分类结果,实现基于土壤光谱的退耕年限快速区分,以期快速准确识别退耕年限提供技术参考。

1 材料与方法

1.1 研究区概况

本研究以山西省大宁县境内退耕还林地作为研究区($110^{\circ}27'55''\sim 111^{\circ}0'40''E$, $36^{\circ}16'40''\sim 36^{\circ}36'25''N$)。大宁县地处黄河中游,地势南北高、中间低,属典型的黄土残垣沟壑区,是黄河泥沙的主要输入区之一^[21]。2000 年大宁县被定为黄河中上游退耕还林(草)生态工程试点县之一,配置有刺槐、柠条等树种^[21]。查询国家土壤信息服务平台(<http://www.soilinfo.cn/map/>)中国 1:400 万土壤类型图可知,研究区内主要土壤类型为黄绵土和褐土。

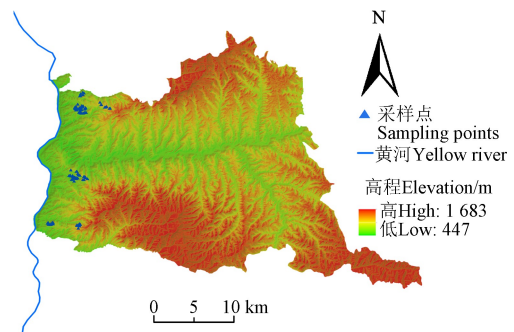
1.2 土壤样品采集与处理

2020 年 11 月在研究区内分别选择退耕 3 a (RF₃)、8 a (RF₈)、14 a (RF₁₄) 和 18 a (RF₁₈) 的刺槐林地各 32 处,共 128 个样点(图 1a)。由于黄绵土土质疏松,易受侵蚀,水土流失严重,是实施退耕还林工程的主要土壤类型^[22],因此样点土壤类型均为黄绵土。土壤样品采集选用“S”形布点法,采集 0~20 cm 表层土壤,每个采样点取 5 个点的土壤混合后作为该采样点土样。将土样分为两份,一份土样在室内自然风干、研磨并过 2 mm 筛,用于光谱测试,土壤有机碳 SOC 含量测试;一份原状土样过 2 mm 筛用于土壤质地测定。同时用环刀进行取样,测定饱和导水率、土壤含水率。其中 SOC 含量测定采用重铬酸钾氧化-外加加热法测定,土壤含水率采用经典烘干法测定,饱和导水率利用恒定水头法测定^[23]。土壤质地数据测定采用激光粒度仪 Mastersizer 3000 进行测定,测定粒径范围为 0.02~2000 μm ^[24],测定所得土壤粒径数据根据美国农业部制粒级分级划分为:黏粒(<0.002 mm),粉粒(0.002~0.05 mm),砂粒(0.05~2 mm)。

1.3 光谱测试及数据处理

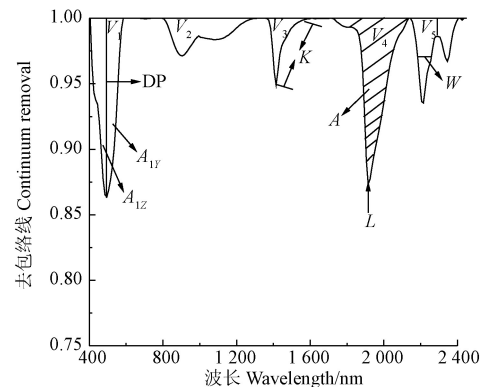
采用美国 ASD (Analytica Spectra devices, Inc) 公司的 FiledSpec 4 便携式光谱仪测定土壤反射光谱曲线,其

波谱范围为 350~2 500 nm。土壤光谱测量在暗室中进行,每次测量开始前将光谱仪预热 2 h。将土壤样本放于直径为 8 cm,深度为 2 cm 的培养皿中,用直尺轻轻刮平表面,减小土壤表面粗糙度。光源由一个 50 W 的卤素灯提供,光纤探头视场角为 10°,距离土样表面 15 cm。每次测试前采用标准白板校正,每个土样测定 10 条光谱曲线,取平均值作为该土样的实际光谱曲线。



a. 大宁县 DEM 及采样点分布

a. DEM and distribution of sampling points in Daning County



b. 土壤光谱特征参数

b. Soil spectral characteristic parameters

注: $V_1\sim V_5$ 分别为第 1、2、3、4、5 个吸收谷; DP 为吸收谷深度; L 为吸收谷最低点对应的波段位置, nm; K 为波段间斜率; A 为吸收谷面积, nm^2 ; W 为吸收谷宽度, nm; A_{1Y} 为第 1 个吸收谷右半部分面积, nm^2 ; A_{1Z} 为第 1 个吸收谷左半部分面积, nm^2 。

Note: $V_1\sim V_5$ are the 1st, 2nd, 3rd, 4th and 5th absorption valleys; DP is the depth of absorption valley; L is the band position corresponding to the lowest point of absorption valley, nm; K is the slope between bands; A is the area of absorption valley, nm^2 ; W is the width of absorption valley, nm; A_{1Y} is the right half area of the first absorption valley, nm^2 ; A_{1Z} is the left half area of the first absorption valley, nm^2 .

图 1 研究区概况及土壤光谱特征参数

Fig.1 General situation of the study area and soil spectral characteristic parameters

去除噪声较为强烈的 350~399、2 451~2 500 nm 波段。同时,为避免数据冗余,对原始光谱反射率数据(Reflectance, R)做 10 nm 重采样,并对其进行 Savitzky-Golay 平滑(Savitzky-Golay smooth, SG),对平滑后的光谱数据分别做一阶微分、倒数的对数、去包络线(Continuum Removal, CR)和光谱特征参数(Spectral Characteristic Parameter, SCP)处理,最后对 R、FD、RL 等数据进行主成分分析。其中,CR 又称为连续统去除法,可以将光谱反射率归一化到 0~1 之间,有效突出光谱特征。SCP^[25]包括:吸收谷(V)、面积(A)、深度(DP)、

吸收波段波长位置 (L)、斜率 (K)、宽度 (W)、对称度 (S) 等 (图 1b), 对称度 $S=A_{12}/A_1$ 。相关数据处理在 The Unscrambler X 10.4、ENVI 5.3、MATLAB R2019b 以及 SPSS 26.0 中完成。

1.4 分类指标选取

分析不同年限退耕还林土壤光谱特征并进行试验, 最终以原始反射率主成分 (R-PCA)、倒数的对数主成分 (RL-PCA)、一阶微分主成分 (FD-PCA)、去包络线主成分 (CR-PCA)、SCP 作为分类指标, 其中 SCP 包含: 第 1、2、3 个吸收谷的深度 (DP_1 、 DP_2 、 DP_3), 第 1、2、3 个吸收谷的对称度 (S_1 、 S_2 、 S_3), 第一个吸收谷的面积及前两个吸收谷面积之和 (A_1 、 A_1+A_2), 495~595、895~995、1 450~1 600 nm 这 3 个波段去包络线的斜率 ($K_{495-595}$ 、 $K_{895-995}$ 、 $K_{1450-1600}$)。

1.5 模型建立与精度评价

本研究采用 K-means、SVM、LDA 三种方法构建不同年限退耕还林地土壤光谱分类模型, 对比分析三种模型的分​​类精度。K-means 优点是处理大样本时精度较高, 本研究先对指标标准化以消除指标量级对结果的影响。SVM 对解决中小样本、非线性和高维数据具有独特的优势, 本研究选用径向基核函数, 惩罚参数 (C) 和 Gamma 参数通过交叉验证方法进行选择, C 为 3, Gamma 参数为 0.33; LDA 常用于降维和分类, 核心思想是投影后类内方差最小, 类间方差最大^[26], 本文 LDA 变换的时候主要是对自变量数据即光谱数据进行了变换, 重新投影到一个超平面上, 使得样本有更大的类间距离, 更小的类内距离。分类模型构建分别在 SPSS 26.0、MATLAB R2019b 和 The Unscrambler X 10.4 中完成。

将不同退耕年限土壤样本分别按 3:1 划分训练集和验证集, 各退耕年限得到 24 个建模样本和 8 个验证样本, 故本研究建模集为 96 个样本, 验证集为 32 个样本。使用混淆矩阵验证模型的分​​类精度, 并且使用总精度和 Kappa 系数对模型精度进行评价, 计算公式如下:

总精度 = $\frac{\sum_{a=1}^j P_{aa}}{M} \times 100\%$ (1)

Kappa = $\frac{M \sum_{a=1}^j P_{aa} - \sum_{a=1}^j (P_{a+} \cdot P_{+a})}{M^2 - \sum_{a=1}^j (P_{a+} \cdot P_{+a})}$ (2)

式中 j 为混淆矩阵中列的数量; P_{aa} 为混淆矩阵中第 a 行第 a 列的像元数, 表示正确分类的个数; P_{a+} 和 P_{+a} 分别为第 a 行和第 a 列总样本个数; M 代表验证样本的总个数。

2 结果与分析

2.1 不同年限退耕土壤的基本理化性质

从表 1 可知, 研究区总样本土壤有机碳介于 2.12~30.11 g/kg 之间, 变异系数为 60.52%, 属于中等变异程度, 不同退耕年限土壤有机碳含量依次为 $RF_{18}>RF_{14}>RF_8>RF_3$ 。含水率最大值为 25.39%, 最小值为 3.77%。饱和导水率介于 0.01~10.98 mm/min 之间, 变异系数为

219.61%, 属于强变异程度。黏粒质量百分数介于 4.03%~17.81% 之间, 变异系数为 26.46%, 属于中等变异程度。砂粒质量百分数最大值为 55.95%, 最小值为 4.70%, 不同退耕年限砂粒质量百分数依次为 $RF_8>RF_3>RF_{14}>RF_{18}$ 。

表 1 土壤样品基本理化性质
Table 1 Basic physical and chemical properties of soil samples

样本 Sample	指标 Index	含水率 Water content/ %	有机碳 Organic carbon/ (g·kg ⁻¹)	饱和导水率 Saturated water conductivity/ (mm·min ⁻¹)	黏粒 Clay/%	粉粒 Silt/%	砂粒 Sand/%
整体 Total	最大值	25.39	30.11	10.98	17.81	81.74	55.95
	最小值	3.77	2.12	0.01	4.03	40.02	4.70
	均值	9.71	9.22	0.51	7.86	72.09	20.05
	标准差	4.32	5.58	1.12	2.08	5.39	6.40
	变异系数/%	44.49	60.52	219.61	26.46	7.48	31.92
退耕 3a (RF ₃) Returning Farmland 3 years	最大值	15.52	13.13	0.63	17.81	80.45	32.93
	最小值	5.06	2.39	0.02	5.05	58.09	9.65
	均值	9.51	6.16	0.23	8.09	71.70	20.20
	标准差	2.66	2.99	0.16	2.45	4.56	5.51
	变异系数/%	27.97	48.54	69.57	30.28	6.36	27.28
退耕 8a (RF ₈) Returning farmland 8 years	最大值	24.81	26.78	10.98	17.78	81.74	55.95
	最小值	4.12	2.12	0.01	4.03	40.02	4.70
	均值	9.74	8.96	0.94	7.47	71.53	21.00
	标准差	4.92	6.33	2.12	2.47	7.95	9.37
	变异系数/%	50.51	70.65	225.53	33.07	11.11	44.62
退耕 14a (RF ₁₄) Returning farmland 14 years	最大值	24.96	24.96	2.67	14.16	81.25	34.44
	最小值	3.77	2.52	0.03	5.47	59.09	8.27
	均值	9.50	10.06	0.33	7.98	72.06	19.96
	标准差	4.50	5.30	0.45	1.88	4.98	5.68
	变异系数/%	47.37	52.68	136.36	23.56	6.91	28.46
退耕 18a (RF ₁₈) Returning farmland 18 years	最大值	25.39	30.11	1.48	12.46	77.74	28.84
	最小值	4.06	2.79	0.08	5.74	65.11	9.80
	均值	10.07	11.68	0.56	7.89	73.03	19.09
	标准差	4.96	5.83	0.28	1.37	3.15	4.13
	变异系数/%	49.26	49.91	50.00	17.36	4.31	21.63

2.2 不同年限退耕土壤的光谱特征

为了分析不同退耕年限土壤光谱曲线的特征, 对不同退耕年限土壤光谱曲线求平均值, 得到不同退耕年限土壤光谱反射率平均值曲线 (图 2a)。从图 2a 可以看出, 不同退耕年限的土壤光谱曲线形状基本相似, 总体上呈现递增的趋势。在可见光波段范围内, 随着波长的增加, 光谱反射率迅速增大, 在近红外波段范围内, 随着波长的增加, 光谱反射率增加的速度逐渐减缓。总体上反射率依次为 $RF_3>RF_8>RF_{14}>RF_{18}$, 与 SOC 含量呈负相关 ($r\leq -0.65$)。在可见光 (400~580 nm) 波段则表现为 $RF_8>RF_3>RF_{14}>RF_{18}$, 砂粒含量越高, 光谱反射率越大。在 610~900 nm 波段范围内, 反射率依次为 $RF_3>RF_8>RF_{18}>RF_{14}$, 这可能与土壤中氧化铁的含量有关, 有研究指出, 在可见光波段内氧化铁含量增加会导致反射率明显下降^[27]。本研究对不同退耕年限土壤光谱反射率平均值曲线做去包络线处理, 从而深入分析不同退耕年限土壤的光谱特征 (图 2b)。

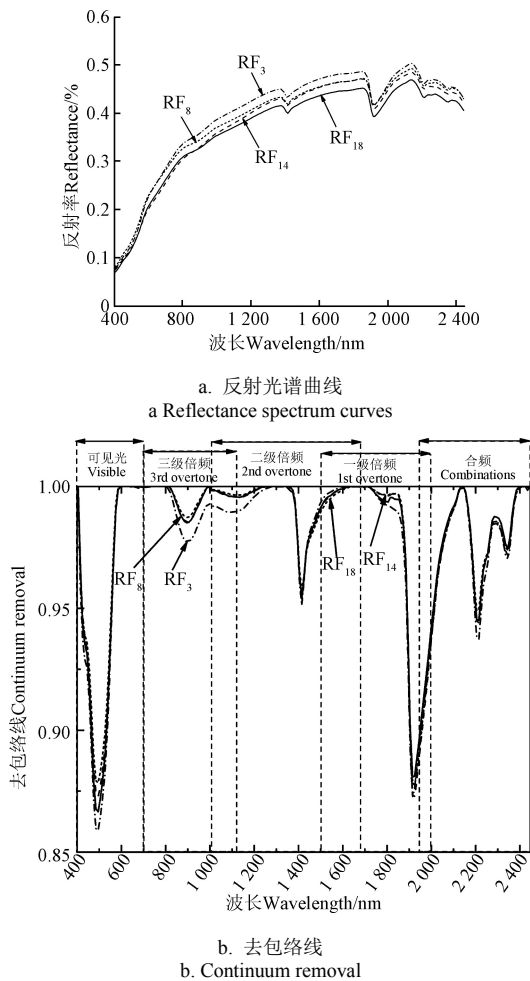


图2 不同年限退耕土壤反射光谱曲线及去包络线
Fig.2 Reflectance spectrum curves and continuum removal of soil with different years of conversion

从图2b中可以看出,在480、900、1 100、1 400、1 900、2 200、2 350 nm处去包络线都出现了明显的吸收特征。在400~600 nm波段范围内,去包络线值依次为 $RF_{18} > RF_{14} > RF_8 > RF_3$,与SOC含量呈正相关,并且表现为吸收谷面积随着去包络线值的增大而减小。在800~1 000 nm和1 000~1 200 nm波段范围内,去包络线值表现出的规律相同,均为 $RF_{18} > RF_{14} > RF_8 > RF_3$,造成 RF_8 和 RF_{14} 去包络线近乎相同的原因可能是 RF_8 和 RF_{14} 土壤中氧化铁含量相近,常认为400~1 100 nm处吸收谷的出现与氧化铁的存在有着一定的关系。在1 400 nm处的吸收谷是受到土壤表面吸附水、黏土矿物O-H结构晶体等吸收引起的^[28],1 900 nm处的吸收谷主要是由土壤中水分子的O-H官能基在1 900 nm处发生一级倍频伸缩震动和转角震动造成的,土壤含水率增加,光谱反射率降低,同时土壤含水率的增加也与黏粒的增加有着一定的关系。对比不同年限退耕土壤饱和和导水率可以发现,饱和和导水率依次为 $RF_8 > RF_{18} > RF_{14} > RF_3$,这与1 900 nm处吸收谷最低点去包络线值的规律相同。在2 200~2 300 nm附近存在着Al-OH黏土矿物(高岭石)的吸收带,从而导致了光谱曲线明显的吸收谷^[29]。已有研究表明^[30],不同粒径土壤的光谱差异随波长的增加而增大,因此

2 000~2 500 nm处的光谱差异很可能由土壤黏粒、粉粒的差异引起。在2 350 nm处吸收谷的最低点去包络线值依次为 $RF_8 > RF_{18} > RF_{14} > RF_3$,与土壤黏粒含量呈负相关。对土壤理化性质和光谱曲线进行综合分析,发现不同年限退耕土壤光谱曲线的特征有较大差异,这些差异不仅仅是单一土壤属性造成的,因此不同年限退耕还林土壤的光谱分类不可以仅依据单一土壤属性进行。

2.3 不同年限退耕土壤的分类

2.3.1 基于K-means的分类

从聚类结果(表2)可以看出,不同输入因子构建的聚类模型,聚类精度有很大差距。以CR-PCA作为输入因子时,模型精度最高,总精度达到81.25%,Kappa系数为0.75。其中 RF_{14} 有1个样本错分为 RF_{18} , RF_{18} 有1个样本错分为 RF_{14} , RF_8 有3个样本错分为 RF_3 ,仅有1个样本错分为 RF_{18} ,并且除 RF_8 的聚类精度仅有50.00%外,其余不同年限退耕土壤聚类精度都达到85%以上,说明以CR-PCA为输入因子,K-means为聚类模型时对不同年限退耕土壤的整体聚类效果较好,对 RF_8 的聚类效果则较差。以R-PCA作为输入因子时,模型精度最差,总精度为46.88%,Kappa系数仅为0.29,仅有 RF_{18} 的聚类精度达到60%以上,其中 RF_3 有2个样本错分为 RF_8 , RF_8 有3个样本错分为 RF_3 , RF_{14} 有4个样本错分为 RF_3 , RF_{18} 有2个样本错分为 RF_{14} 。以RL-PCA、FD-PCA、SCP作为输入因子时,模型聚类精度较为均衡,从整体上来看,FD-PCA和CR-PCA是K-means模型中相对较好的分类指标, RF_3 、 RF_{14} 以及 RF_{18} 均以CR-PCA的聚类精度为最高,而 RF_8 的聚类精度则以FD-PCA为最高。

2.3.2 基于SVM的分类

从表2中可以看出,以R-PCA、RL-PCA为输入因子时,模型精度较低,总精度和Kappa系数分别为50.00%、46.88%和0.33、0.29,其中 RF_{14} 的分类精度均只有25.00%,为本研究最低的分类精度,并且有3个样本错分为 RF_3 ,2个样本错分为 RF_8 ,1个样本错分为 RF_{18} ,说明以SVM作为分类模型,R-PCA和RL-PCA作为输入因子时, RF_{14} 的分类效果较差。以FD-PCA作为输入因子时,模型精度最高,总精度为84.38%,Kappa系数为0.79,不同年限退耕土壤的分类精度都达到了75.00%及以上, RF_8 的分类精度更是达到100%,其中 RF_3 有1个样本错分为 RF_{14} ,有1个样本错分为 RF_{18} , RF_{14} 有1个样本错分为 RF_3 ,有1个样本错分为 RF_{18} , RF_{18} 有1个样本错分为 RF_{14} ,说明FD-PCA可以显著地提高不同年限退耕土壤的分类精度。同K-means分类模型相似,以CR-PCA、SCP为输入因子时,模型分类精度较为均衡。总体上来看,以SVM作为分类模型时,FD-PCA会显著提高不同年限退耕土壤的分类精度,CR-PCA、SCP次之。

2.3.3 基于LDA的分类

由表2可知,分类结果总精度依次为 $CR-PCA > FD-PCA = SCP > R-PCA > RL-PCA$ 。对比以R-PCA和RL-PCA为输入因子时不同退耕年限的分类精度可以

发现, 除 RF_3 的分类精度达到 80%以上, 其余退耕年限土壤的分类精度均小于 65%, RF_{14} 的分类精度只有 37.5%, 并且 RF_{14} 错分的样本大多为 RF_3 和 RF_8 。以 CR-PCA 作为输入因子时, 除 RF_{14} 的分类精度较低外, 其余退耕年限土壤的分类精度均达到 85%以上, 具有较好的分类效果, 其中 RF_8 有 1 个样本错分为 RF_{18} , RF_{14} 有 1 个样本错分为 RF_3 , 1 个样本错分为 RF_8 , RF_{18} 有 1

个样本错分为 RF_{14} 。以 SCP 作为输入因子时, RF_8 和 RF_{14} 分类精度差异较大, 其中 RF_8 分类精度为 100.00%, RF_{14} 分类精度为 62.50%, RF_{14} 中有 1 个样本错分为 RF_8 , 2 个样本错分为 RF_{18} 。以 FD-PCA 作为输入因子时, 模型分类效果较为均衡。从整体来看, LDA 作为分类模型, 对 RF_{14} 的分类效果较差; 以 CR-PCA 和 FD-PCA 为输入因子时, 分类精度有较大提升。

表 2 不同模型分类结果
Table 2 Classification results of different models

模型 Model	输入因子 Input factor	划分类型 Classified type	样本个数 Number of samples				聚类精度 Clustering accuracy/%	总精度 Total accuracy/%	Kappa 系数 Kappa coefficient
			RF_3	RF_8	RF_{14}	RF_{18}			
K 均值聚类 K-means Clustering Algorithm (K-Means)	R-PCA	RF_3	4	2	1	1	50.00	46.88	0.29
		RF_8	3	3	1	1	37.50		
		RF_{14}	4	1	3	0	37.50		
		RF_{18}	1	0	2	5	62.50		
	RL-PCA	RF_3	4	2	0	2	50.00	62.50	0.50
		RF_8	0	6	0	2	75.00		
		RF_{14}	1	0	5	2	62.50		
		RF_{18}	2	0	1	5	62.50		
	FD-PCA	RF_3	5	3	0	0	62.50	78.13	0.71
		RF_8	1	7	0	0	87.50		
		RF_{14}	0	1	6	1	75.00		
		RF_{18}	0	0	1	7	87.50		
	CR-PCA	RF_3	8	0	0	0	100.00	81.25	0.75
		RF_8	3	4	0	1	50.00		
		RF_{14}	0	0	7	1	87.50		
		RF_{18}	0	0	1	7	87.50		
	SCP	RF_3	5	1	0	2	62.50	56.25	0.42
		RF_8	0	4	1	3	50.00		
		RF_{14}	0	2	4	2	50.00		
		RF_{18}	0	1	2	5	62.50		
支持向量机 Support Vector Machine (SVM)	R-PCA	RF_3	5	2	1	0	62.50	50.00	0.33
		RF_8	0	5	0	3	62.50		
		RF_{14}	3	2	2	1	25.00		
		RF_{18}	0	2	2	4	50.00		
	RL-PCA	RF_3	4	1	2	1	50.00	46.88	0.29
		RF_8	2	4	1	1	50.00		
		RF_{14}	3	2	2	1	25.00		
		RF_{18}	1	1	1	5	62.50		
	FD-PCA	RF_3	6	0	1	1	75.00	84.38	0.79
		RF_8	0	8	0	0	100.00		
		RF_{14}	1	0	6	1	75.00		
		RF_{18}	0	0	1	7	87.50		
	CR-PCA	RF_3	6	0	1	1	75.00	78.13	0.71
		RF_8	0	6	1	1	75.00		
		RF_{14}	0	2	6	0	75.00		
		RF_{18}	0	1	0	7	87.50		
	SCP	RF_3	6	1	1	0	75.00	75.00	0.67
		RF_8	0	6	1	1	75.00		
		RF_{14}	0	1	7	0	87.50		
		RF_{18}	0	1	2	5	62.50		
线性判别分 析(LDA) Liner Discriminant Analysis	R-PCA	RF_3	7	1	0	0	87.50	59.38	0.46
		RF_8	1	4	1	2	50.00		
		RF_{14}	2	2	3	1	37.50		
		RF_{18}	0	2	1	5	62.50		
	RL-PCA	RF_3	7	1	0	0	87.50	56.25	0.42
		RF_8	2	4	1	1	50.00		
		RF_{14}	2	3	3	0	37.50		
		RF_{18}	1	2	1	4	50.00		
	FD-PCA	RF_3	7	0	0	1	87.50	81.25	0.75
		RF_8	1	7	0	0	87.50		
		RF_{14}	1	1	6	0	75.00		
		RF_{18}	1	0	1	6	75.00		
	CR-PCA	RF_3	8	0	0	0	100.00	87.50	0.83
		RF_8	0	7	0	1	87.50		
		RF_{14}	1	1	6	0	75.00		
		RF_{18}	0	0	1	7	87.50		
	SCP	RF_3	7	0	0	1	87.50	81.25	0.75
		RF_8	0	8	0	0	100.00		
		RF_{14}	0	1	5	2	62.50		
		RF_{18}	1	0	1	6	75.00		

注: R-PCA 为原始反射率主成分, RL-PCA 为倒数对数主成分, FD-PCA 为一阶微分主成分, CR-PCA 为去包络线主成分, SCP 为光谱特征参数。

Note: R-PCA is the principal component of original reflectance, RL-PCA is the principal component of logarithm of the reciprocal, FD-PCA is the principal component of first-order differential, CR-PCA is the principal component of continuum removal, and SCP is the spectral characteristic parameter.

2.4 分类精度比较

以 R-PCA、RL-PCA 为输入因子时, 精度最高的模型分别是 LDA 模型和 K-means 模型, 总精度和 Kappa 系数分别为 59.38%、0.46 和 62.50%、0.50。以 FD-PCA、CR-PCA、SCP 为输入因子时, 除 FD-PCA 以 SVM 模型精度最高, 其余均为 LDA 模型精度最高, 总精度分别为 84.38%、87.50%、81.25%, Kappa 系数分别为 0.79、0.83、0.75, 一定程度上提高了不同退耕年限土壤光谱分类的精度。在 K-means、SVM、LDA 模型中, 以 FD-PCA、CR-PCA 为输入量的模型要略优于以其他三种输入量的分类模型。从整体上来看, 以 CR-PCA 为输入因子, 基于 LDA 模型构建的分类模型总精度最高, 达到 87.50%; 以 FD-PCA 为输入因子, 基于 SVM 构建的分类模型精度次之, 也达到了 84.38%; 以 R-PCA 为输入因子构建的 K-means 分类模型和以 RL-PCA 为输入因子构建的 SVM 分类模型总精度最差, 仅仅只有 46.88%; 通过对比五种输入因子在三种模型中的分类总精度可以发现, 以 CR-PCA 和 FD-PCA 为输入因子的分类总精度均达到了 75% 以上, 其中 CR-PCA 总精度最大达到 87.50%, FD-PCA 总精度最大为 84.38%, 并且对比不同年限聚类精度, CR-PCA 除 K-means 模型中的 RF₈ 聚类精度较低外, 其余聚类精度均要优于 FD-PCA, 因此在本研究中 CR-PCA 是区分不同年限退耕土壤的最优输入因子。

3 讨论

土壤光谱曲线是土壤属性的综合体现。对不同年限退耕土壤的光谱曲线进行分析, 发现反射光谱曲线形状大致相似, 反射率依次为 $RF_3 > RF_8 > RF_{14} > RF_{18}$, 与 SOC 含量呈负相关, 与南锋等^[31]研究结果相符, 南锋等^[31]对黄土高原煤矿复垦区农田 SOC 进行预测, 发现 SOC 含量与反射率呈负相关。在 400~580 nm 波段范围内, 反射率大小与土壤砂粒含量呈正相关, 这与张雅梅等^[30]的研究结果不符, 张雅梅等^[30]对土壤质地不同粒径颗粒含量进行统一预测, 发现在 380~1 100 nm 波段范围内, 砂粒含量最高的砂土反射率最低。这可能是 SOC 含量和砂粒含量共同作用产生的结果。对光谱曲线做去包络线处理, 发现在 480 和 900 nm 处, 随着 SOC 含量的增加, 吸收谷变浅, 面积减小, 这与赵明松等^[32]的研究结果相一致, 赵明松等^[32]对江苏中部水稻土和潮土的去包络线进行分析, 发现 480、900 nm 处 SOC 含量增加导致吸收谷减小。在 800~1 200 nm 波段范围内, 土壤光谱曲线也表现出一定的差异, 这可能由于氧化铁含量的不同所造成的。去包络线在 1 400、1 900 nm 处的吸收谷也表现出显著差异, 造成这一差异的主要原因可能是由于土壤水分的不同。

目前研究多采用高光谱非成像技术、成像技术及高分影像等对土壤类型进行区分, 未能考虑诸如退耕还林地不同年限的区分。本文以地面高光谱数据为基础, 实现了退耕还林地不同年限分类研究, 可以为高分影像分类时的光谱变换形式和波段选择提供一定参考, 并且后

续研究中可以将图像与光谱结合, 探究分类精度更高的模型。本研究以 R-PCA、RL-PCA、FD-PCA、CR-PCA 和 SCP 为输入指标, 基于 K-means、SVM 和 LDA 构建退耕还林地不同年限分类模型。通过对比不同分类模型的精度, 发现以 CR-PCA 为输入指标构建的 LDA 分类模型精度最高。造成这一结果的原因可能是 LDA 属于监督学习的线性分类方法, 对线性问题有较好的处理效果。比较不同分类模型中输入指标的精度发现, CR-PCA 在 K-means、SVM 和 LDA 模型中的分类精度依次为 81.25%、78.13% 和 87.50%, 这充分说明对光谱数据进行 CR 处理可以显著增强土壤属性对光谱曲线产生的综合影响, 进而提高模型的精度, 其中主要波段范围为 405~595、805~995、1 355~1 495、1 835~2 135、2 145~2 285 nm, 后期利用高分影像数据时可以着重考虑选择这些波段范围。孟祥添等^[33]等采用高分 5 号影像数据, 对原始光谱曲线进行去包络线、主成分分析等处理, 进而区分东北典型黑土区土壤类型, 发现 CR-PCA 分类精度较原始反射率提高了 9.15%, 说明去包络线可以增强光谱差异性, 与本研究结果一致。

本研究将高光谱技术应用到退耕还林土壤研究中, 并实现了退耕还林地不同年限识别, 为研究退耕还林土壤属性变化及影响提供了一种思路和方法。然而本研究构建的模型较为简单, 未构建目前广泛应用的机器学习模型^[34-35], 会对分类模型的精度有一定的影响, 后续将会对分类模型和输入指标做进一步的研究, 为退耕还林对土壤属性影响的深入研究及退耕还林工作奠定基础。

4 结论

为了明确退耕工程对土壤属性的影响, 以及不同年限退耕土壤的光谱特征, 实现快速获取退耕年限的目的, 本研究以大宁县不同年限退耕还林地土壤为研究对象, 获取土壤理化性质及光谱曲线数据, 探讨不同年限退耕土壤理化性质的变化及对光谱曲线造成的影响, 同时构建分类模型, 选取最优分类模型及最优输入因子。主要结论:

1) 随着退耕年限的增长, 土壤有机碳含量逐渐增加, 砂粒含量先增加后减少, 土壤有机碳含量与原始反射率大小呈负相关。

2) 不同年限退耕土壤原始光谱曲线形态相似, 整体呈递增的趋势; 进行去包络线处理后显著提升光谱曲线的吸收特征, 在 480 nm 处出现了由土壤有机碳引起的吸收谷, 900 和 1 100 nm 的吸收谷可能是由于氧化铁含量不同导致的, 1 400 和 1 900 nm 主要是由于土壤水分引起的吸收特征, 2 200 和 2 350 nm 处的吸收特征可能是由于不同粒径颗粒含量的不同所导致的。

3) 以去包络线主成分为输入量, 基于线性判别分析构建的分类模型精度最高, 达到 87.50%, 为本研究最优分类模型; 以去包络线主成分为输入因子的分类模型, 总精度均达到了 75% 以上, 最大为 87.50%, 说明在本研究中去包络线主成分是区分不同年限退耕土壤的最优输入因子。

[参 考 文 献]

- [1] 陆大道, 孙东琪. 黄河流域的综合治理与可持续发展[J]. 地理学报, 2019, 74(12): 2431-2436.
Lu Dadao, Sun Dongqi. Development and management tasks of the Yellow River Basin: A preliminary understanding and suggestion[J]. Acta Geographica Sinica, 2019, 74(12): 2431-2436. (in Chinese with English abstract)
- [2] 曾鸣. 不同年限退耕还林(草)效果研究[D]. 沈阳: 沈阳农业大学, 2018.
Zeng Ming. Study on the Effect of Returning Farmland to Forestry (Grass) in Different Years[D]. Shenyang: Shenyang Agricultural University, 2018. (in Chinese with English abstract)
- [3] 张恒硕, 查同刚, 张晓霞. 晋西黄土区退耕年限对土壤物理性质的影响[J]. 北京林业大学学报, 2020, 42(6): 123-133.
Zhang Hengshuo, Zha Tonggang, Zhang Xiaoxia. Effects of converting years from farmland to forestland on soil physical properties in the loess area of western Shanxi Province, northern China[J]. Journal of Beijing Forestry University, 2020, 42(6): 123-133. (in Chinese with English abstract)
- [4] 周佳佳, 郝小玲, 吴洁, 等. 晋南典型丘陵沟壑区退耕还林 16 年后土壤物理性质变化[J]. 世界生态学, 2021, 10(1): 70-79.
Zhou Jiajia, Hao Xiaoling, Wu Jie, et al. Changes of soil physical properties after returning farmland to forest for 16 years in typical hilly and gully regions of southern Shanxi[J]. International Journal of Ecology, 2021, 10(1): 70-79. (in Chinese with English abstract)
- [5] 史君怡, 王国梁, 刘国彬. 退耕还林(草)对黄土高原不同土地利用方式下土壤剖面水分含量的影响[J]. 水土保持研究, 2018, 25(2): 81-86.
Shi Junyi, Wang Guoliang, Liu Guobin. Effect of graining for green on water content of soil profile under different land use patterns in the loess plateau[J]. Research of Soil and Water Conservation, 2018, 25(2): 81-86. (in Chinese with English abstract)
- [6] 郝安辉, 张兴国, 李坤伦, 等. 不同生长年限及不同部位川明参红外光谱研究[J]. 时珍国医国药, 2016, 27(11): 2665-2667.
Hao Anhui, Zhang Xingguo, Li Kunlun, et al. The infrared spectroscopic study of radix chuanmingshinis violacei with different growing periods and different parts[J]. Lishizhen Medicine and Materia Medica Research, 2016, 27(11): 2665-2667. (in Chinese with English abstract)
- [7] 杨会云, 刘飞, 杨春艳, 等. 红外光谱结合判别分析对滇重楼生长年限的鉴别[J]. 中国实验方剂学杂志, 2015, 21(15): 35-39.
Yang Huiyun, Liu Fei, Yang Chunyan, et al. Identification of Paradis Rhizoma with different growing years by infrared spectroscopy combined with discriminant analysis[J]. Chinese Journal of Experimental Traditional Medical Formulae, 2015, 21(15): 35-39. (in Chinese with English abstract)
- [8] 杨方, 钟莹, 周鸿昆, 等. 基于表面增强拉曼光谱技术的陈皮年份检测[J]. 激光与光电子学进展, 2017, 54(3): 302-307.
Yang Fang, Zhong Ying, Zhou Hongkun, et al. Detection of years of dried tangerine peels based on surface-enhanced raman spectroscopy[J]. Laser & Optoelectronics Progress, 2017, 54(3): 302-307. (in Chinese with English abstract)
- [9] 杨少壮, 李灿, 李辰, 等. 不同贮存年限陈皮的太赫兹光谱和成像的差异分析[J]. 现代食品科技, 2019, 35(12): 258-266.
Yang Shaozhuang, Li Can, Li Chen, et al. The differences in the dried tangerine peels stored for different years revealed by terahertz spectroscopy and imaging[J]. Modern Food Science and Technology, 2019, 35(12): 258-266. (in Chinese with English abstract)
- [10] 张钰, 李杰庆, 李涛, 等. 不同储藏年限绒柄牛肝菌紫外 & 红外光谱数据融合鉴别研究[J]. 现代食品科技, 2018, 34(2): 218-224, 245.
Zhang Yu, Li Jieqing, Li Tao, et al. Research on identification of boletus tomentipes with different storage period by UV and FT-IR combined with data fusion[J]. Modern Food Science and Technology, 2018, 34(2): 218-224, 245. (in Chinese with English abstract)
- [11] 杨天伟, 李涛, 李杰庆, 等. 不同年份和产地美味牛肝菌的红外光谱鉴别研究[J]. 光谱学与光谱分析, 2016, 36(7): 2117-2123.
Yang Tianwei, Li Tao, Li Jieqing, et al. Study on the discrimination of boletus edulis from different years and origins with FTIR[J]. Spectroscopy and Spectral Analysis, 2016, 36(7): 2117-2123. (in Chinese with English abstract)
- [12] 王崇, 吴见. 农作物种类高光谱遥感识别研究[J]. 地理与地理信息科学, 2015, 31(2): 29-33, 2.
Wang Dong, Wu Jian. Study on crop variety identification by hyperspectral remote sensing[J]. Geography and Geo-Information Science, 2015, 31(2): 29-33, 2. (in Chinese with English abstract)
- [13] 李晓丽, 魏玉震, 徐劼, 等. 基于高光谱成像的茶叶中 EGCG 分布可视化[J]. 农业工程学报, 2018, 34(7): 180-186.
Li Xiaoli, Wei Yuzhen, Xu Jie, et al. EGCG distribution visualization in tea leaves based on hyperspectral imaging technology[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(7): 180-186. (in Chinese with English abstract)
- [14] 李丹, 彭智平, 韩留生, 等. 基于土壤反射光谱特性的广东省稻田土壤快速分类[J]. 热带地理, 2015, 35(1): 29-34.
Li Dan, Peng Zhiping, Han Liusheng, et al. Rapid soil classification of paddy field in guangdong province based on visible and near infrared reflectance spectra[J]. Tropical Geography, 2015, 35(1): 29-34. (in Chinese with English abstract)
- [15] 陈泽炎, 赖长江生, 魏旭雅, 等. 中药材生长年限鉴定方法的研究进展[J]. 中国中药杂志, 2021, 46(6): 1357-1367.
Chen Zeyan, Lai Changjiangsheng, Wei Xuya, et al. Research progress on identification methods of growth years of traditional Chinese medicinal materials[J]. China Journal of Chinese Materia Medica, 2021, 46(6): 1357-1367. (in Chinese with English abstract)

- [16] 卜海博, 聂黎行, 王丹, 等. 近红外光谱法无损识别林下山参及其生长年限[J]. 光谱学与光谱分析, 2012, 32(7): 1801-1805.
Bu Haibo, Nie Lixing, Wang Dan, et al. Nondestructive identification of the root of mountain cultivation ginseng and growth years by near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2012, 32(7): 1801-1805. (in Chinese with English abstract)
- [17] 虞佳维, 程志庆, 张劲松, 等. 高光谱信息的农林植被种类区分[J]. 光谱学与光谱分析, 2018, 38(12): 3890-3896.
Yu Jiawei, Cheng Zhiqing, Zhang Jinsong, et al. An approach to distinguishing between species of trees and crops based on hyperspectral information[J]. Spectroscopy and Spectral Analysis, 2018, 38(12): 3890-3896. (in Chinese with English abstract)
- [18] 翁海勇, 许金钊, 陶铸, 等. 高 EGCG 含量茶树品种光谱识别模型构建[J]. 中国农机化学报, 2021, 42(6): 111-117.
Weng Haiyong, Xu Jinchai, Tao Zhu, et al. Construction of spectral screening model for tea cultivars with high EGCG content[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(6): 111-117. (in Chinese with English abstract)
- [19] 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.
Shi Zhou, Wang Qianlong, Peng Jie, et al. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations[J]. Scientia Sinica: Terrae, 2014, 44(5): 978-988. (in Chinese with English abstract)
- [20] 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.
Zhao Xiaomin, Yang Meihua. Hyper-spectral characteristics of major types of soils in Red Soil region of Jiangxi Province, China[J]. Acta Pedologica Sinica, 2018, 55(1): 31-42. (in Chinese with English abstract)
- [21] 卫新菊. 实施退耕还林 改善生态环境[J]. 中国农村小康科技, 2006(8): 43-44, 46.
Wei Xinju. Implement returning farmland to forest to improve ecological environment[J]. Journal of Agriculture, 2006(8): 43-44, 46. (in Chinese with English abstract)
- [22] 黄自立, 张文孝. 陕北黄绵土的性质与改良利用[J]. 土壤通报, 1987(3): 108-110.
- [23] 王紫薇, 邵明安, 黄来明, 等. 青海省东部不同土地利用方式下土壤饱和导水率分布及其影响因素[J]. 水土保持学报, 2021, 35(3): 150-155.
Wang Ziwei, Shao Mingan, Huang Laiming, et al. Distribution and influencing factors of soil saturated hydraulic conductivity under different land use patterns in eastern qinghai province[J]. Journal of Soil and Water Conservation, 2021, 35(3): 150-155. (in Chinese with English abstract)
- [24] 段世航, 崔若然, 江荣风, 等. 激光衍射法测定土壤粒径分布的研究进展[J]. 土壤, 2020, 52(2): 247-253.
Duan Shihang, Cui Ruoran, Jiang Rongfeng, et al. Research advance in determining soil particle size distribution by laser diffraction method[J]. Soils, 2020, 52(2): 247-253. (in Chinese with English abstract)
- [25] 刘焕军, 张小康, 张新乐, 等. 面向土壤分类的高光谱反射特征参数模型[J]. 遥感学报, 2017, 21(1): 105-114.
Liu Huanjun, Zhang Xiaokang, Zhang Xinle, et al. Hyperspectral reflectance characteristics parameter extraction for soil classification model[J]. Journal of Remote Sensing, 2017, 21(1): 105-114. (in Chinese with English abstract)
- [26] 师芸, 马东晖, 吕杰, 等. 基于流形光谱降维和深度学习的高光谱影像分类[J]. 农业工程学报, 2020, 36(6): 151-160.
Shi Yun, Ma Donghui, Lyu Jie, et al. Hyperspectral image classification based on manifold spectral dimensionality reduction and deep learning method[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(6): 151-160. (in Chinese with English abstract)
- [27] 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 Jie, Zhang Yangzhu, Zhou Qing, 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. (in Chinese with English abstract)
- [28] Rossel R A V, Behrens T. Using data mining to model and interpret soil diffuse reflectance spectra[J]. Geoderma, 2010, 158(1): 46-54.
- [29] 国佳欣, 赵小敏, 郭熙, 等. 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, et al. Inversion of organic matter content in red soil based on PLSR-BP composite model[J]. Acta Pedologica Sinica, 2020, 57(3): 636-645. (in Chinese with English abstract)
- [30] 张雅梅, 施梦月, 王德彩, 等. 基于高光谱的土壤不同颗粒含量预测分析[J]. 土壤通报, 2021, 52(4): 777-784.
Zhang Yamei, Shi Mengyue, Wang Decai, et al. Different soil particle contents prediction based on hyperspectral data[J]. Chinese Journal of Soil Science, 2021, 52(4): 777-784. (in Chinese with English abstract)
- [31] 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.
Nan Feng, Zhu Hongfen, Bi Rutian. 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. (in Chinese with English abstract)
- [32] 赵明松, 谢毅, 陆龙妹, 等. 基于高光谱特征指数的土壤有机质含量建模[J]. 土壤学报, 2021, 58(1): 42-54.
Zhao Mingsong, Xie Yi, Lu Longmei, et al. Modeling for soil organic matter content based on hyperspectral feature indices[J]. Acta Pedologica Sinica, 2021, 58(1): 42-54. (in Chinese with English abstract)
- [33] 孟祥添, 鲍依临, 刘焕军, 等. 基于高分 5 号影像的东北典型黑土区土壤分类[J]. 农业工程学报, 2020, 36(16): 231-238.
Meng Xiangtian, Bao Yilin, Liu Huanjun, et al. Soil

- classification in typical black soil areas of Northeast China based on Gaofen-5 images[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(16): 231-238. (in Chinese with English abstract)
- [34] 唐海涛, 孟祥添, 苏循新, 等. 基于 CARS 算法的不同类型土壤有机质高光谱预测[J]. 农业工程学报, 2021, 37(2): 105-113.
Tang Haitao, Meng Xiangtian, Su Xunxin, et al. Hyperspectral prediction on soil organic matter of different types using CARS algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(2): 105-113. (in Chinese with English abstract)
- [35] 白婷, 丁建丽, 王敬哲. 基于机器学习算法的土壤有机质质量比估算[J]. 排灌机械工程学报, 2020, 38(8): 829-834.
Bai Ting, Ding Jianli, Wang Jingzhe. Estimation of soil organic matter content based on machine learning[J]. Journal of Drainage and Irrigation Machinery Engineering(JDIME), 2020, 38(8): 829-834. (in Chinese with English abstract)

Identification of the years of returning farmland to forest land using hyperspectral technology

Deng Yongpeng, Zhu Hongfen, Ding Haoxi, Sun Ruipeng, Bi Rutian^{*}

(College of Resource and Environment, Shanxi Agricultural University, Taiyuan 030000, China)

Abstract: Soil quality and ecological environment have been improved in the middle reaches of the Yellow River in China. This improvement can attribute to the national project of returning farmland to forest in different years since 2000, particularly on the great slopes. Therefore, it is necessary to rapidly acquire the years of returning farmland and soil characteristics, in order to evaluate the ecological benefits of the project. Taking the middle reaches of the Yellow River as the study area, this study aims to obtain the soil's physical and chemical properties, as well as the soil's spectral curves for the returning cropland to forest in different years using hyperspectral imaging technology. Some spectral preprocessing were utilized, including the Savitzky-Golay Smoothing (SG), Reciprocal Logarithm (RL), the First-Order Differential (FD), Continuum Removal (CR), Principal Component Analysis (PCA), and Spectral Characteristic Parameters (SCP). The classification models were constructed for the years of returning cropland to forest using the K-means clustering (K-means), support vector machine (SVM), and Linear Discriminant Analysis (LDA). Among them, the input factors were set as The Principal Component Of Original Reflectance (R-PCA), Principal Component Of Logarithm Of The Reciprocal (RL-PCA), Principal Component Of First-Order Differential (FD-PCA), Principal Component Of Continuum Removal (CR-PCA), and SCP. The results showed that: 1) The content of Soil Organic Carbon (SOC) increased gradually, and the content of sand particles increased first and then decreased, with the increase of the years of returning cropland. The content of SOC was negatively correlated with the soil original reflectance. 2) There was a similar shape of soil original spectral curve in the different years of returning farmland, indicating the overall increasing trend. The CR preprocessing was significantly improved the absorption of the spectral curve, with the outstanding absorption characteristics at 480, 900, 1 100, 1 400, 1 900, 2 200, and 2 350 nm. 3) The highest accuracy (87.50%) was achieved in the classification model of LDA with the CR-PCA as input factor, which was the optimal classification model. The second highest accuracy (84.38%) was found in the classification model of SVM with the FD-PCA as the input factor. All the classification models with the CR-PCA as input factor shared the highest accuracy of more than 75%, with the maximum of 87.50 %, indicating that the CR-PCA was the optimal input factor to distinguish the different years of returning farmland in this case, followed by the FD-PCA. As such, the rapid distinction of the years was fully realized for the returning farmland to forest, according to the spectral characteristics and classification through the soil spectral curves. The finding can provide a strong reference for the soil properties and environmental impacts of the returning farmland to forest.

Keywords: soils; organic carbon; models; returning farmland to forest; years of returning farmland; hyperspectrum; middle reaches of the Yellow River