

基于高光谱的土壤不同颗粒含量预测分析

张雅梅, 施梦月, 王德彩, 郭芳*

(河南农业大学林学院, 河南 郑州 450002)

摘要: 以典型黄河下游冲积平原区的土壤为研究对象, 分析土壤高光谱特征, 探讨土壤质地不同粒级颗粒含量的统一估测途径, 为土壤质地快速监测评价提供技术支持。选择原始光谱, 及其倒数、对数、标准正交变换、多元散射变化、一阶微分、二阶微分共 7 种光谱变换形式, 首先主成分降维, 然后分别建立土壤黏粒、粉粒和砂粒含量的支持向量机预测模型, 采用决定系数、均值绝对误差、均方根误差 3 种精度指标来衡量模型的预测能力。结果表明: 原始光谱的对数为最佳光谱变换形式, 具有最佳的土壤不同颗粒含量估测能力, 决定系数 $R^2 \geq 0.6853$, 均值绝对误差 $MAE \leq 0.1193$, 均方根误差 $RMSE \leq 0.1683$; 黏粒含量的变化范围相对集中, 预测能力整体表现的相对较强, $R^2 = 0.8127$, $MAE = 0.0820$, $RMSE = 0.1248$ 。通过筛选最佳光谱变换处理, 主成分降维, 支持向量机预测, 建立了土壤中黏粒、粉粒和砂粒含量的统一估测途径, 实现了简单快捷的高光谱估测。

关键词: 土壤; 颗粒含量; 高光谱; 光谱变换; 主成分分析; 支持向量机

中图分类号: S147.2 **文献标识码:** A **文章编号:** 0564-3945(2021)04-0777-08

DOI: [10.19336/j.cnki.trtb.2020022702](https://doi.org/10.19336/j.cnki.trtb.2020022702)

张雅梅, 施梦月, 王德彩, 郭芳. 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 Ya-mei, SHI Meng-yue, WANG De-cai, GUO Fang. Different Soil Particle Contents Prediction Based on Hyperspectral Data[J]. Chinese Journal of Soil Science, 2021, 52(4): 777-784

土壤不同颗粒含量是土壤最重要的特征之一, 在很大程度上决定了土壤的物理和化学性质。长期以来, 土壤测定主要依赖于土壤物理分析方法确定土壤颗粒的粒径大小, 其操作步骤复杂、成本高、速度慢、效率低。作为一种新兴技术, 高光谱技术在土壤学科中的应用不断深入, 为土壤评价与理化性质的快速监测提供了新的途径。由于土壤高光谱估测通常涉及光谱数据处理和建模等多个环节, 如何实现准确估测目前尚无统一的认识和技术方法, 仍然是一个需要研究的问题。

土壤光谱是土壤属性信息的综合反映, 能够用于高效精确的土壤属性预测建模^[1]。土壤属性建模时主要存在 3 方面的特征: 土壤光谱反映信息的综合性、光谱波段数据量较大、采样样本数量相对较少。前人基于高光谱技术预测了土壤质地中的黏粒、粉粒和砂粒不同颗粒含量^[2-8], 多见于某一种^[2-3]或某两种颗粒^[4-6]的预测, 对三种颗粒^[7-8]同时预测的相对较少。王德彩等^[7]直接利用原始光谱对平原区土壤质地的黏粒、粉粒和砂粒三种颗粒含量进行偏最小二

乘法预测。不同的光谱变换处理, 可以增强土壤在某些波段的吸收或反射特征, 从而提高土壤属性光谱建模的精度^[1], 常采用反射率倒数、对数、微分等不同的光谱变换形式^[2, 8], 但并不是所有的光谱变换都可以增强土壤光谱所反映的信息, 所以需要筛选最佳的光谱变换形式。白燕英等^[8]对河套灌区土壤光谱进行了 4 种变换, 针对三种颗粒分别选取敏感波段、最佳变换形式进行预测, 但没有获取土壤不同颗粒含量的统一预测模式。建模方法常用的有基于回归分析、偏最小二乘法、神经网络、遗传算法、支持向量机等不同的预测模型^[2-8], 不同的模型影响预测精度, 要依据模型特征挑选适宜的建模方式。典型的黄河下游冲积平原区, 土壤光谱具有独特的区域性, 则土壤质地预测方式需要探寻自身适宜的, 且基于黏粒、粉粒和砂粒三种颗粒的最佳光谱变换形式及统一预测模式鲜见报道。

本文分析典型的黄河下游冲积平原区土壤不同颗粒含量的光谱特征, 采用不同的光谱变换处理来提高数据质量并筛选最佳的光谱变换形式, 选择常

收稿日期: 2020-02-27; 修订日期: 2021-04-25

基金项目: 国家自然科学基金青年基金项目(41201210)、国家林草局生物安全与遗传资源项目(KJZXSA2019042)资助

作者简介: 张雅梅(1977-), 女, 山东菏泽人, 在读博士, 讲师, 从事森林培育及林业 3S 技术应用研究, E-mail: zym@henau.edu.cn

*通讯作者: E-mail: gg19771979@126.com

用的主成分分析法进行降维, 选择适合小样本、非线性和高维数的支持向量机方法^[9], 建立黏粒、粉粒和砂粒三种颗粒含量预测模型, 以期对土壤高光谱估测的模式有新的认识, 为典型的黄河下游冲积平原区土壤质地快速监测评价提供技术支持。

1 研究区概况

封丘县位于河南省东北部, 处于 34°53′~35°14′N, 114°14′~114°46′E 之间。县境南北长 38.2 km, 东西宽 48.7 km, 面积约 1220.5 km²。地处黄淮海平原, 属于黄河故道区, 是典型的黄河下游冲击平原。属于暖温带半干旱半湿润季风气候, 年均降雨量约 618 mm。地势由西南向东北倾斜, 属于黄淮海地区典型土壤。主要分布有两大土类, 即潮土土类和风沙土土类。其中, 潮土是该地区分布最广、面积最大的耕作土壤, 风沙土在该地区多呈断续条带状分布。

2 数据来源

2.1 土壤数据

以研究区行政边界为基础, 布设 3 km × 3 km 网格的采样点, 进行规则采样, 得到 101 个样本。采集 0~20 cm 的表层土壤, 经自然风干、研磨, 过 2 mm 孔筛, 然后采用激光衍射法 (Beckman Coulter LS230, 美国, 测试粒径范围 0.04~2000 μm) 进行土壤颗粒粒径分析。按美国制进行分级: 黏粒 (<0.002 mm), 粉粒 (0.05~0.002 mm), 砂粒 (2~0.05 mm)。

2.2 光谱数据

采用便携式地物光谱仪 (FieldSpec 3, 美国), 测量土壤样品的反射光谱数据。在室内黑暗环境中, 每个土壤样品旋转 4 次, 每个角度扫描 10 次, 共 40 次, 然后取平均值作为样本的光谱反射率。测得的原始光谱波长范围是 350~2500 nm, 而光谱中 350~380 nm、2400~2500 nm 之间的数据信噪比较高, 需要进行剔除, 所以研究只选取了 380~2400 nm 波段的反射光谱。380~2400 nm 的光谱数据进行 1 nm 间隔的重采样, 共输出 2021 个波段。

3 研究方法

3.1 光谱变换

为了分析不同光谱变换形式对建模及其精度的

影响, 寻求土壤黏粒、粉粒和砂粒含量的最佳光谱变换形式, 选取了 6 种常用的光谱变换形式, 即: 原始光谱的倒数 (1/x)^[8]、对数 (lnx)^[8]、标准正交变换 (Standard Normal Variate, SNV)、多元散射变化 (Multiplicative Scatter Correction, MSC)^[2]、一阶微分 (x′)^[2]、二阶微分 (x′′)^[10], 加上原始光谱 (x), 共 7 种光谱形式。

3.2 主成分分析

原始光谱及其 6 种变换形式, 共 7 种光谱数据, 波长范围均为 380~2400 nm 波段, 共 2021 个波段, 数据量比较大, 若直接采用 7 种光谱进行预测, 将会因数据量太大, 导致预测时运算次数减少、速度变慢, 不能得到最佳的模型。由此采用经典的主成分分析法进行降维, 降低数据冗余、减少数据量。研究区有效样本 100 个 (101 个样本, 剔除 1 个方差最大的离散点, 剩余 100 个样本), 相对于 2021 个光谱波段来说比较少, 主成分分析也可以较好的限制变量多样本少的问题。

3.3 数据标准化

每个样本包含光谱的主成分变量、黏粒含量变量、粉粒含量变量和砂粒含量变量, 其量纲单位不统一, 所以必须对样本数据进行标准化处理。研究采用极差标准化 (正规化方法), 将全部变量均规范化到区间 [0, 1]。

$$\hat{x}_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}$$

式中, \hat{x}_{ij} 表示标准化后第 i 行第 j 列的值, x_{ij} 表示第 i 行第 j 列的值 ($i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n$)。

3.4 支持向量机

支持向量机 (Support Vector Machines, SVM) 是一种基于统计学习理论的机器学习算法, 其以统计学习理论为体系, 通过寻求结构风险最小化, 追求在有限信息的条件下得到最优结果, 在数据处理方面具有适用于高维度特征空间、小样本统计学习、抗噪声影响能力强等优点^[9-12]。土壤光谱受到有机质、水分等诸多因素的影响, 是一个复杂的过程, 且样本数量有限, 所以研究选择较适宜的支持向量机模型进行预测。SVM 建模的关键是选择参数 (惩罚系数 cost 和核函数半径 gamma) 和核函数。cost 和 gamma 通过交叉验证的方法来获取, 以获取的 cost

和 γ 为基础, 再选择广泛应用的径向基 (RBF) 函数作为核函数。

研究区 100 个样本, 按照黏粒含量与砂粒含量的比值排序, 每隔 5 个选取一个样本, 共 20 个样本, 用于模型检验, 剩余的 80 个样本用于模型建立。主成分分析获取的主成分, 作为建模的变量, 分别获取 7 种光谱变换形式中黏粒、粉粒、砂粒含量的 SVM 模型。选择决定系数 (Determination Coefficient, R^2)、均值绝对误差 (Mean Absolute Error, MAE)、均方根误差 (Root Mean Squared Error, RMSE) 3 个精度指标, 进行模型的最终评价。

本文采用 SPSS 24.0 软件进行土壤数据描述性分析和主成分提取, The Unscrambler 9.7 软件进行光谱变换, WEKA 3.7 和 Libsvm 1.0.6 工具箱实现建模和预测, SigmaPlot 10.0 作图。

4 结果与分析

4.1 土壤不同颗粒含量统计特征分析

由表 1 可见, 研究区整体样本的黏粒、粉粒和砂粒含量分别位于 4.1% ~ 21.3%、4.0% ~ 65.1% 和 13.6% ~ 89.8% 之间, 表明黏粒含量分布最为集中, 粉粒含量居中, 砂粒含量的分布最为离散。整体样本的黏粒、粉粒和砂粒含量的均值分别为 10.95%、35.55% 和 53.51%, 表明样本土壤以砂粒含量为主, 土壤质地类型主要为砂壤质。不同样本类型间, 黏粒、粉粒和砂粒含量的变异系数差别特别小, 均在 0.29 ~ 0.42 范围内, 通常认为, 变异系数 (Coefficient of Variation, Cv) $Cv < 0.1$ 为弱变异性, $0.1 \leq Cv \leq 1.0$ 为中等变异性, $Cv > 1.0$ 为高度变异性^[13]。由此可知, 各颗粒组成均呈中等变异性。

表 1 土壤黏粒、粉粒和砂粒含量的统计特征值
Table 1 Statistical characteristics of clay, silt and sand contents

颗粒粒级 Particle grade	样本类型 Sample type	样本数量 Sample number	最小值 Minimum (%)	最大值 Maximum (%)	平均值 Mean (%)	标准差 Standard deviation	变异系数 Coefficient of variation	偏度 Skewness	峰度 Kurtosis
黏粒	整体样本	100	4.1	21.3	10.95	3.47	0.32	0.85	0.38
	建模集	80	4.1	21.3	11.03	3.57	0.32	0.87	0.41
	验证集	20	6.2	17.1	10.63	3.11	0.29	0.60	-0.20
粉粒	整体样本	100	4.0	65.1	35.55	14.11	0.40	-0.26	0.18
	建模集	80	4.2	65.1	35.53	13.99	0.39	-0.24	0.27
	验证集	20	4.0	63.4	35.61	14.94	0.42	-0.35	0.22
砂粒	整体样本	100	13.6	89.8	53.51	16.97	0.32	-0.07	-0.01
	建模集	80	13.6	88.8	53.44	16.93	0.32	-0.10	0.06
	验证集	20	19.5	89.8	53.77	17.55	0.33	0.08	-0.02

4.2 光谱特征分析

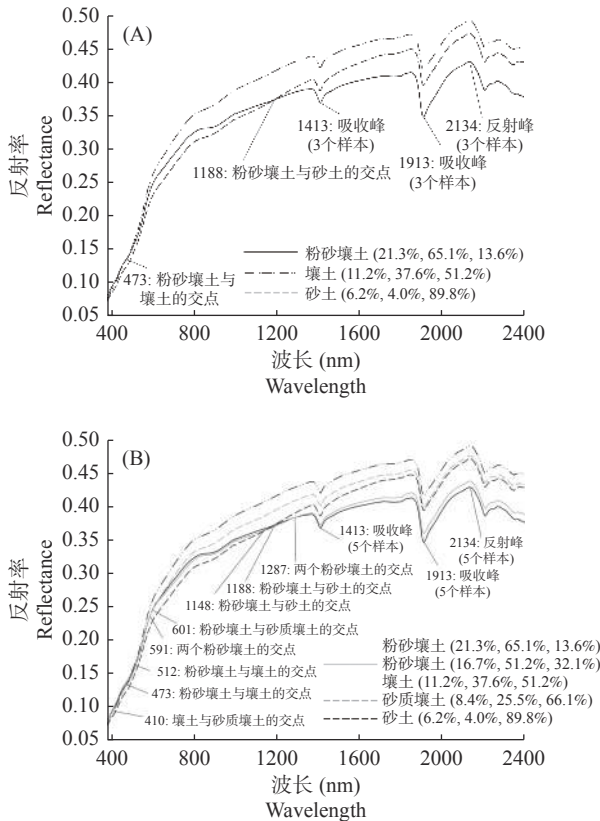
按照黏粒、粉粒含量的降低, 砂粒含量的升高, 选择了 5 个样本, 分析不同土壤光谱曲线。粉砂壤土 (21.3%, 65.1%, 13.6%), 其黏粒含量和粉粒含量最大, 砂粒含量最小; 粉砂壤土 (16.7%, 51.2%, 32.1%), 其黏粒含量和粉粒含量降低, 砂粒含量升高; 壤土 (11.2%, 37.6%, 51.2%), 其黏粒含量、粉粒含量和砂粒含量最接近平均值 10.95%、35.55%、53.51%; 砂质壤土 (8.4%, 25.5%, 66.1%), 其黏粒含量和粉粒含量继续降低, 砂粒含量继续升高; 砂土 (6.2%, 4.0%, 89.8%), 其黏粒含量接近最小值, 粉粒含量最小, 砂粒含量最大。

从图 1 可知, 不同粒径颗粒含量的土壤光谱曲线存在交点, 交点集中于可见光和长波近红外波段, 集中于可见光波段的交点相对较多, 集中于长波近

红外波段的交点相对较少。从图 1 (A) 和 (B) 比较可知, 随着样本的增加, 黏粒和粉粒含量的逐渐降低, 以及砂粒含量的逐渐升高, 样本于 380 ~ 1369 nm 波段范围的反射率变化越来越复杂, 不同样本间的交点数量从 2 个激增为 8 个, 其中可见光范围增加了 4 个, 占总增加量的 66.7%, 表明土壤光谱在可见光波段范围内比较敏感、稳定性差; 1369 ~ 2400 nm 波段范围, 不存在不同样本间的交点, 于 1413 nm 出现第一个明显的共同吸收峰, 1913 nm 出现最大的共同吸收峰, 2134 nm 出现共同的最大反射峰, 这 3 个典型的共同峰值不受样本数量的影响, 其他峰值的影响也比较小, 这都表明土壤光谱在此范围内的反射率比较稳定。

从光谱曲线的整体上看, 5 个样本的光谱曲线形态和趋势大致相同, 吸收峰和反射峰位置几乎一致。

砂粒含量最高的砂土（6.2%，4.0%，89.8%）于 380~1188 nm 波段范围内反射率最低的，黏粒和粉粒含量最大的样本粉砂壤土（21.3%，65.1%，13.6%）于 1369~2400 nm 波段范围内反射率最低。



注:图例括号内数据依次是土壤黏粒、粉粒和砂粒含量值

图 1 不同土壤光谱反射率曲线

Fig.1 Spectral reflectance curves of different soil

4.3 主成分分析

将 100 个样本的 2021 个光谱变量值，通过主成分分析降维转换为少数不相关的成分。7 种光谱形式，按累积方差贡献率不少于 95%、96%、97%、98%、99%、99.9% 的条件选取主成分，结果见表 2。为了

避免光谱过度降维和模型过度拟合，最终选择按累积方差贡献率不少于 98% 的条件，选取各自的主成分。 x' 选取了 71 个主成分， x'' 选取了 87 个主成分，两者的主成分数量比较多，可见降维效果不太好。而其他的变换形式，降维的效果相对较好。

4.4 预测分析

x' 光谱变换形式，黏粒、粉粒、砂粒含量的 SVM 模型， $R^2 \leq 0.25$ ， $RMSE \geq 0.3$ ，模型精度太差，与其他光谱变换形式没有可比性，所以直接舍弃此建模。由表 3 可知， x'' 光谱变换形式，黏粒、粉粒、砂粒含量的 SVM 模型， $R^2 < 0.45$ ，远远小于其他模型，但是 MAE 和 RMSE 相对精度还可以。6 种变换形式，18 个 SVM 模型， R^2 的最大值是 0.8877，最小是 0.3945；MAE 最小是 0.0820，最大值是 0.3046；RMSE 最小值是 0.1248，最大值是 0.3549；3 个指标的最大值最小值之间的差距较大，说明建立模型的精度差距也比较大。

由图 2 可知，黏粒含量的 R^2 : 标准正交变换 > 多元散射 > 原始光谱 > 原始光谱的对数 > 原始光谱的倒数；粉粒含量的 R^2 : 标准正交变换 > 原始光谱的对数 > 多元散射 > 原始光谱 > 原始光谱的倒数；砂粒含量的 R^2 : 标准正交变换 > 多元散射 > 原始光谱的对数 > 原始光谱 > 原始光谱的倒数。依据 R^2 ，土壤 3 种颗粒 SVM 模型的综合预测能力表现为：标准正交变换 > 多元散射 > 原始光谱的对数 > 原始光谱 > 原始光谱的倒数。

由图 3 可知，黏粒含量的 MAE: 原始光谱的对数 < 原始光谱 < 原始光谱的倒数 < 原始光谱的二阶微分 < 标准正交变换 < 多元散射；粉粒含量的 MAE: 原始光谱的倒数 < 原始光谱的对数 < 原始光谱 < 原始光谱的二阶微分 < 标准正交变换 < 多元散射；砂

表 2 主成分分析的结果

Table 2 Results of principal component analysis

光谱变换形式 Spectral transformation	主成分数量 Principal component number					
	累计贡献率95% Cumulative contribution rate 95%	累计贡献率96% Cumulative contribution rate 96%	累计贡献率97% Cumulative contribution rate 97%	累计贡献率98% Cumulative contribution rate 98%	累计贡献率99% Cumulative contribution rate 99%	累计贡献率99.9% Cumulative contribution rate 99.9%
x	2	2	2	3	3	5
$1/x$	2	2	2	3	3	5
$\ln x$	2	2	2	3	3	5
MSC	3	3	4	4	5	9
SNV	3	3	4	4	5	10
x'	44	52	60	71	83	97
x''	72	76	81	87	93	99

表 3 不同模型的最优参数及精度指标
Table 3 Optimal parameters and precision indices of different models

光谱变换形式 Spectral transformation	SVM模型 SVM model	最优模型参数 Optimal parameter		精度指标 Accuracy index		
		Cost	Gamma	R^2	MAE	RMSE
x	黏粒含量	9.1927	0.1109	0.8188	0.0898	0.1254
	粉粒含量	4.9500	0.1020	0.6323	0.1419	0.1820
	砂粒含量	3.1382	0.1109	0.6960	0.1540	0.1826
1/x	黏粒含量	3.2391	0.4136	0.7937	0.1163	0.1542
	粉粒含量	0.8173	0.5145	0.6216	0.1096	0.1727
	砂粒含量	0.5145	0.8173	0.6558	0.1070	0.1638
ln x	黏粒含量	8.3855	0.1109	0.8127	0.0820	0.1248
	粉粒含量	3.0373	0.2118	0.6853	0.1173	0.1683
	砂粒含量	0.8173	0.6155	0.6972	0.1193	0.1578
MSC	黏粒含量	5.5600	0.2118	0.8532	0.2512	0.2822
	粉粒含量	9.1927	0.1109	0.6718	0.3046	0.3549
	砂粒含量	9.8991	0.1109	0.7414	0.2900	0.3363
SNV	黏粒含量	7.6791	0.2118	0.8877	0.2156	0.2473
	粉粒含量	4.5509	0.1109	0.6924	0.2537	0.3025
	砂粒含量	6.1655	0.1109	0.7616	0.2515	0.2988
x''	黏粒含量	1.3218	0.1109	0.4480	0.2115	0.2758
	粉粒含量	7.2755	0.0010	0.4255	0.1864	0.2343
	砂粒含量	6.5691	0.0010	0.3945	0.1868	0.2353

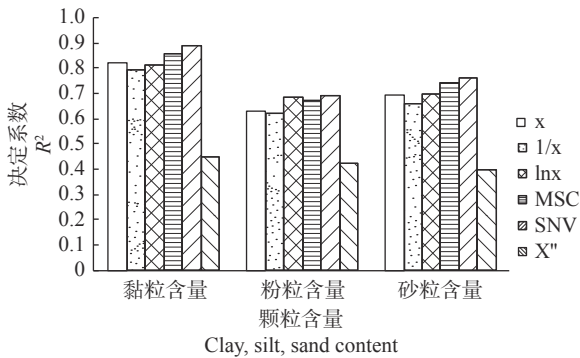


图 2 不同光谱变换形式土壤黏粒、粉粒和砂粒含量预测模型的 R^2
Fig.2 R^2 of clay, silt and sand contents prediction model in soil with different spectral transformation forms

粒含量的 MAE: 原始光谱的倒数 < 原始光谱的对数 < 原始光谱 < 原始光谱的二阶微分 < 标准正交变换 < 多元散射。依据 MAE, 土壤 3 种颗粒 SVM 模型的综合预测能力表现为: 原始光谱的倒数 > 原始光谱的对数 > 原始光谱。

由图 4 可知, 黏粒含量的 RMSE: 原始光谱的对数 < 原始光谱 < 原始光谱的倒数 < 标准正交变换 < 原始光谱的二阶微分 < 多元散射; 粉粒含量的 RMSE: 原始光谱的对数 < 原始光谱的倒数 < 原始光谱 < 原始光谱的二阶微分 < 标准正交变换 < 多元散射; 砂粒含量的 RMSE: 原始光谱的对数 < 原始光谱的倒数 < 原始光谱 < 原始光谱的二阶微分 < 标准正交变换 <

多元散射。依据 RMSE, 土壤 3 种颗粒 SVM 模型的综合预测能力表现为: 原始光谱的对数 > 原始光谱的倒数 > 原始光谱。

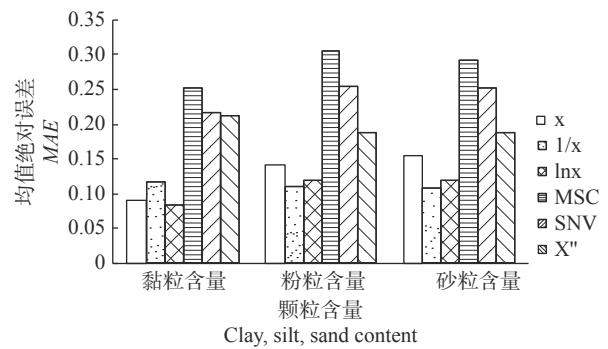


图 3 不同光谱变换形式土壤黏粒、粉粒和砂粒含量预测模型的 MAE
Fig.3 MAE of clay, silt and sand contents prediction model in soil with different spectral transformation forms

从模型预测能力角度看, 依据不同的精度指标, 不同光谱变换形式的 SVM 预测能力不同。按照 R^2 , 预测能力为: 标准正交变换 > 多元散射 > 原始光谱的对数; 按照 MAE, 则为: 原始光谱倒数 > 原始光谱对数; 按照 RMSE, 则为: 原始光谱的对数 > 原始光谱的倒数。从选择不同光谱变换形式最佳的预测能力着手, 需要综合评价 3 个精度指标。原始光谱倒数和原始光谱对数两者之间的 RMSE 差距远远大于 MAE 的差距, 所以依据 RMSE 和 MAE 两个精度

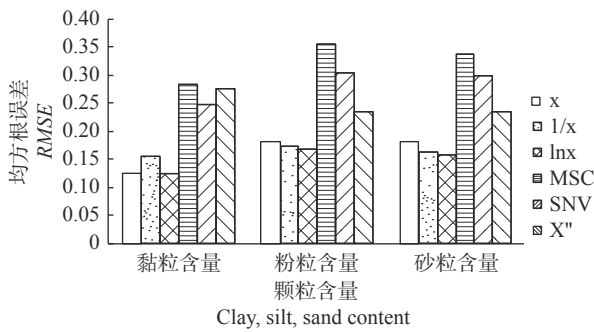


图 4 不同光谱变换形式土壤黏粒、粉粒和砂粒含量预测模型的 $RMSE$

Fig.4 $RMSE$ of clay, silt and sand contents prediction model in soil with different spectral transformation forms

指标, 选择原始光谱对数的变换形式为精度最佳的建模方式。再结合 R^2 , 则选择原始光谱的对数为最佳光谱变化形式比较合理。

从土壤不同颗粒含量的角度看, 无论哪种光谱变换形式, 黏粒含量的 R^2 均最大, 且 $RMSE$ (除原始光谱的二阶微分) 和 MAE (除原始光谱倒数、原始光谱的二阶微分) 两值均最小, 说明黏粒含量 SVM 模型的预测能力整体较强。粉粒含量的 R^2 值 (除原始光谱的二阶微分) 均最小, 砂粒含量的 R^2 (除原始光谱的二阶微分) 均处于黏粒含量和粉粒含量之间的范围。粉粒含量和砂粒含量的 $RMSE$ 、 MAE 精度指标表现出来的规律性不太明显。

5 讨论

选择原始光谱, 及其倒数、对数、标准正交变换、多元散射变化、一阶微分、二阶微分, 共 7 种光谱形式, 建立 SVM 模型, R^2 、 MAE 和 $RMSE$ 的最高值与最低值之间的差距都比较大, 说明建立模型的预测能力差距也比较大, 表明不同光谱变换形式对土壤不同颗粒含量的预测影响较大, 这与黄明祥等的研究结果一致^[2]。其中原始光谱的对数为最佳光谱变化形式, 具有最佳的预测能力, 而原始光谱一阶微分变换形式的 SVM 模型预测能力因很差而被直接舍弃。原始光谱的预测能力与原始光谱的对数比较接近, 所以原始光谱的预测能力也是比较好, 这与张娜等的研究结果一致^[6]。

精度指标的作用是衡量模型预测能力, 不同的精度指标其含义不同, 衡量模型建模能力的侧重点也不同。分别以 R^2 、 MAE 和 $RMSE$ 来进行模型预测能力的排序, 3 种情况不同光谱变换形式的 SVM 预测能力各不相同。所以要想获取最佳的模型建模能

力的排序, 需要不同的精度指标来综合衡量, 不同的精度指标相互制约相互关联有机统一, 才能筛选出最佳模型。

选取样本土壤不同颗粒含量的分布范围不同, 影响模型的预测。研究区整体样本中, 黏粒含量分布最为集中, SVM 模型的预测能力整体表现的都相对较强。粉粒和砂粒含量的分布比较离散, 模型的预测能力表现的相对较弱。砂粒含量的分布最为离散, 却没有表现出整体最差的预测能力。可见, 土壤不同颗粒含量的分布范围对模型预测能力的影响存在复杂的非线性关系。

土壤颗粒粒径最大的砂土 (6.2% 4.0% 89.8%), 在 380 ~ 1188 nm 波段范围内, 是 5 个样本中反射率最低的, 这表明随着土壤粒径的增加, 土壤光谱反射率呈现一定程度的降低, 这与前人研究结果一致^[14-15]。这是由于随着土壤颗粒粒径的增加, 颗粒间的空隙增大, 总面积降低, 从而降低了光的反射作用。黏粒和粉粒含量最大的粉砂壤土 (21.3% 65.1% 13.6%) 于 1369 ~ 2400 nm 波段范围内反射率最低。近红外是水分比较敏感的波段, 土壤光谱反射率会随着土壤含水量的增加而降低^[16], 并在 1440、1990 nm 附近出现两个较强水分光谱吸收峰^[16-17], 这与本研究土壤光谱的两个峰谷 (1413、1913 nm) 相一致。为了最大消除水分的影响, 研究基于风干磨碎土, 但是由于土壤中黏粒对水分的保水性比较强, 在水分敏感波段, 还是受到一定的影响, 使得光谱反射率降为最低。一般认为, 土壤有机质敏感波段在 600 ~ 800 nm 之间^[18] 或者 620 ~ 810 nm 之间^[19], 去除有机质后, 土壤光谱反射率在全波段均有所增加, 尤其在可见光波段变化明显^[20-21], 这表明土壤光谱受到有机质的影响, 且主要影响于可见光范围。由此推测, 不同粒级颗粒含量的土壤光谱在可见光波段范围形成的交点, 有机质可能是主导因素, 尚需进一步的研究。

研究基于风干磨碎土, 土壤光谱信息没有完全消除水分的影响, 还受到有机质等多种因素的影响^[22-23], 但是从筛选出来的原始光谱对数的模型预测能力来看, 结果比较理想, 实现了简单快捷的预测。从模型的精度来看, 虽然干土建立的模型比较理想, 但若用于精准预测, 精确度还不够高。所以干土建立的模型, 可以较好地运用于估测, 或普通精度的预测, 精准预测还需要通过大量的工作去除水分^[24]、

有机质^[25]等因素对光谱反射率的影响。

6 结论

(1) 不同光谱变换形式对土壤不同颗粒含量的预测影响较大, 筛选出原始光谱的对数为最佳光谱变换形式, 具有最佳的预测能力。其建立的黏粒、粉粒和砂粒三种颗粒 SVM 模型的 R^2 可达 0.6853 ~ 0.8127, MAE 低至 0.0820 ~ 0.1193, $RMSE$ 低至 0.1248 ~ 0.1683。

(2) 土壤不同颗粒含量的分布范围影响模型精度, 黏粒含量的变化范围比较集中, 其预测能力整体表现的都相对较强。原始光谱、原始光谱的对数、原始光谱的标准正交变换、原始光谱的多元散射变化 4 种变换处理建立的黏粒含量 SVM 模型, R^2 均为 3 种颗粒的最大值, MAE 和 $RMSE$ 均为 3 种颗粒的最小值。

(3) 土壤光谱, 通过筛选最佳光谱变换处理, 主成分降维, 然后支持向量机建模的预测模式, 黏粒、粉粒和砂粒 3 种颗粒含量预测精度相对较高, 实现了土壤中不同颗粒的统一估测途径, 为土壤质地快速监测评价提供了技术支持。

参考文献:

- [1] 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*, 2020
- [2] 黄明祥, 程街亮, 王珂, 等. 海涂土壤高光谱特性及其砂粒含量预测研究[J]. *土壤学报*, 2009, 46(5): 932 - 937.
- [3] 李春蕾, 许端阳, 陈蜀江. 基于高光谱遥感的新疆北疆地区土壤砂粒含量反演研究[J]. *干旱区地理*, 2012, 35(3): 473 - 478.
- [4] 乔天, 吕成文, 肖文凭, 等. 基于遗传算法的土壤质地高光谱预测模型研究[J]. *土壤通报*, 2018, 49(4): 773 - 778.
- [5] 王德彩, 蔚霖, 张俊辉, 等. 基于正交信号校正的 Vis-NIR 光谱土壤质地预测[J]. *河南农业大学学报*, 2017, 51(3): 408 - 413.
- [6] 张娜, 张栋良, 李立新, 等. 基于高光谱的区域土壤质地预测模型建立与评价——以河套灌区解放闸灌域为例[J]. *干旱区资源与环境*, 2014, 28(5): 67 - 72.
- [7] 王德彩, 郭登巍, 赵明松, 等. 平原区土壤质地的反射光谱预测与地统计制图[J]. *土壤通报*, 2012, 43(2): 257 - 262.
- [8] 白燕英, 魏占民, 刘全明, 等. 基于高光谱的河套灌区农田表层土壤质地反演研究[J]. *地理与地理信息科学*, 2013, 29(5): 68 - 71.
- [9] Kashif G M, Asefa T, Kembowski M W, et al. Soil moisture prediction using support vector machine[J]. *Journal of the American Water Resources Association*, 2006, vol. 42(4): 1033 - 1046.
- [10] 沈强, 张世文, 夏沙沙, 等. 基于支持向量机的土壤有机质高光谱反演[J]. *安徽理工大学学报(自然科学版)*, 2019, 39(4): 39 - 45.
- [11] 张强, 黄生志, 陈晓宏. 基于支持向量机的土壤湿度模拟及预测研究[J]. *土壤学报*, 2013, 50(1): 59 - 67.
- [12] 乔红波, 焦红涛, 吴旭, 等. 基于支持向量机模型的冬小麦全蚀病为害等级遥感监测[J]. *麦类作物学报*, 2014, 34(12): 1694 - 1698.
- [13] 张世文, 王胜涛, 刘娜, 等. 土壤质地空间预测方法比较[J]. *农业工程学报*, 2011, 27(1): 332 - 339.
- [14] 马创, 申广荣, 王紫君, 等. 不同粒径土壤的光谱特征差异分析[J]. *土壤通报*, 2015, 46(2): 292 - 298.
- [15] 乔星星, 冯美臣, 杨武德, 等. 粒径对土壤光谱特性的影响[J]. *山西农业科学*, 2016, 44(9): 1325 - 1327, 1359.
- [16] 李晨, 张国伟, 周治国, 等. 滨海盐土土壤水分的高光谱参数及估测模型[J]. *应用生态学报*, 2016, 27(2): 525 - 531.
- [17] Stenberg B, Rossel R A V, Mouazen A M, et al. Chapter five: Visible and near infrared spectroscopy in soil science[J]. *Advances in Agronomy*, 2010, 107: 163 - 215.
- [18] 吴裕, 申广荣, 刘璐, 等. 土壤有机质高光谱特征及其反演研究[J]. *上海交通大学学报(农业科学版)*, 2019, 37(4): 37 - 44.
- [19] 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.
- [20] 朱亚星, 于雷, 洪永胜, 等. 土壤有机质高光谱特征与波长变量优选方法[J]. *中国农业科学*, 2017, 50(22): 4325 - 4337.
- [21] 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.
- [22] Macros R N, Jose Alexander M D. Spectral reflectance methodology in comparison to traditional soil analysis[J]. *Soil Science Society American Journal*, 2006, 70: 393 - 407.
- [23] Kooistra L, Wanders J, Epema G F, et al. The potential of field spectroscopy for the assessment of sediment properties in river floodplains[J]. *Analytica Chimica Acta*, 2003, 484: 189 - 200.
- [24] 于雷, 洪永胜, 朱亚星, 等. 去除土壤水分对高光谱估算土壤有机质含量的影响[J]. *光谱学与光谱分析*, 2017, 37(7): 2143 - 2151.
- [25] 彭杰, 周清, 张杨珠, 等. 有机质对土壤光谱特性的影响研究[J]. *土壤学报*, 2013, 50(3): 517 - 524.

Different Soil Particle Contents Prediction Based on Hyperspectral Data

ZHANG Ya-mei, SHI Meng-yue, WANG De-cai, GUO Fang*
(College of Forestry, Henan Agricultural University, Zhengzhou 450002)

Abstract: The hyperspectral characteristics of different particle contents of soil in the typical alluvial plain of the lower Yellow River were investigated in order to provide technical support for rapid monitoring and evaluation of soil texture. Seven spectral transformation forms were selected, including the original spectrum and its reciprocal, logarithm, standard orthogonal transformation, multivariate scattering change, the first derivative and the second derivative. First, principal component analysis was applied to reduce dimension. Then, predictive models of the contents of clay, silt and sand were established with support vector machines. Three accuracy indices were selected including determination coefficient, mean absolute error and root mean squared error. The results showed that the logarithm of the original spectrum was the best spectral transformation form due to the best prediction with the $R^2 \geq 0.6853$, the $MAE \leq 0.1193$ and the $RMSE \leq 0.1683$. The variation range of clay content was relatively concentrated, showing the best prediction by the R^2 of 0.8127, MAE of 0.0820 and $RMSE$ of 0.1248. The soil spectrum was reduced dimension with principal component by selecting the best spectral transformation. Whereafter, the support vector machine modeling was used to predict the contents of clay, silt and sand in soil, which was realized a simple and fast hyperspectral estimation on soil texture.

Key words: Soil; Particle content; Hyperspectral; Spectral transformation; Principal component analysis; Support vector machines

[责任编辑: 韩春兰]