

DOI: 10.13758/j.cnki.tr.2023.02.024

顾永昇, 丁建丽, 韩礼敬, 等. 基于多源环境变量的渭-库绿洲土壤颗粒含量预测研究. 土壤, 2023, 55(2): 426–432.

基于多源环境变量的渭-库绿洲土壤颗粒含量预测研究^①

顾永昇^{1,2}, 丁建丽^{1,2*}, 韩礼敬^{1,2}, 李科^{1,2}, 周倩^{1,2}

(1 新疆大学地理与遥感科学学院智慧城市与环境建模自治区普通高校重点实验室, 乌鲁木齐 830046; 2 新疆大学绿洲生态重点实验室, 乌鲁木齐 830046)

摘要: 本文以渭干河-库车河绿洲(简称渭-库绿洲)土壤颗粒为研究对象, 采集了绿洲内 50 个典型表层(0~10 cm)土壤样本, 通过相关软件, 提取到遥感指数变量、地形和气候等环境变量, 经过相关性分析确定环境变量和预测目标间的关系, 使用 R 语言构建了预测土壤颗粒含量的随机森林(random forest, RF)模型和极端梯度提升(extreme gradient boosting, XGBoost)模型。研究结果表明: XGBoost 模型的预测结果整体好于 RF 模型, 其中相关系数介于 0.39~0.78; 土壤 pH、高程及衍生变量、光谱变换变量均是两个模型预测土壤颗粒含量的重要因子; 将模型预测结果、实测数据和世界土壤数据库(HWSD)中的 3 种土壤颗粒数据作对比分析, 结果表现出模型预测数据的误差小于 HWSD 与实测数据的误差。综上所述, 通过筛选环境变量建立的 XGBoost 模型, 是预测渭-库绿洲土壤颗粒含量的有效方法。

关键词: 土壤颗粒; 高光谱; 环境变量; 机器学习

中图分类号: S152.3 **文献标志码:** A

Prediction of Soil Particle Content in Wei-Ku Oasis Based on Multi-source Environmental Variables

GU Yongsheng^{1,2}, DING Jianli^{1,2*}, HAN Lijing^{1,2}, LI Ke^{1,2}, ZHOU Qian^{1,2}

(1 Key Laboratory of Smart City and Environment Modelling of Higher Education Institute, College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi 830046, China; 2 Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi 830046, China)

Abstract: In this paper, soil particles in the Weigan River-Kuche River Oasis (referred to as the Wei-Ku oasis) were used as the research object, fifty typical surface (0–10 cm) soil samples were collected from the oasis, and environmental variables such as remote sensing index variables, topography and climate were extracted through relevant software. After correlation analysis to determine the relationship between environmental variables and prediction targets, a random forest (RF) model and an extreme gradient boosting (XGBoost) model for predicting soil particle contents were constructed using R language. The results show that the prediction results of the XGBoost model are better than those of the RF model, with the correlation coefficients ranging from 0.39 to 0.78. Soil pH, elevation and derivative variables, and spectral transformation variables are all important factors in the prediction of soil particle contents in both models. The errors of model prediction data are smaller than those of HWSD and measured data. In conclusion, the XGBoost model established by screening environmental variables is an effective method for predicting soil particle content in the Wei-Ku oasis.

Key words: Soil particles; Hyperspectra; Environmental variables; Machine learning

土壤颗粒大小是划分土壤质地的主要依据和重要特征。土壤质地受诸多因素影响, 同时它也影响着溶质和养分等物质在土壤中的运移和分布, 对提升土壤肥力和农业生产有重要的意义^[1]。渭-库绿洲是新

疆主要农业生产区, 实现该区域土壤颗粒的精准预测, 对当地农业生产和土壤质量评价具有现实意义。

目前机器学习已成为预测土壤质地常用的方法, 它具有可以控制模型过拟合、输出变量重要性

①基金项目: 新疆维吾尔自治区自然科学基金重点项目(2021D01D06)和国家自然科学基金项目(41961059)资助。

* 通讯作者(watarid@xju.edu.cn)

作者简介: 顾永昇(1995—), 男, 甘肃武威人, 硕士研究生, 主要从事干旱区绿洲水盐运移研究。E-mail: 1774600807@qq.com

等优点^[2]。刘亚东等^[3]通过 RF 方法在青藏高原地区分析黏粒含量剖面分布的影响因素，其研究表明，气候和地形是影响黏粒含量剖面分布的决定性因素。Liu 等^[4]利用 MODIS 数据的衍生变量，通过构建 RF 模型对江苏省土壤颗粒、有机质、土壤 pH 等属性进行了预测。Ließ 等^[5]在厄瓜多尔山区，用 56 个采样点比较了回归树(RT)和随机森林模型(RF)预测土壤颗粒的结果。Forkuor 等^[6]基于 Landset 遥感数据，建立了多元线性回归(MLR)、随机森林回归(RFR)、支持向量机(SVM)模型预测土壤颗粒，研究结果表明机器学习预测性能优于 MLR。da Silva Chagas 等^[7]在用 RF 模型和 MLR 方法预测巴西半干旱区土壤质地空间分布时，RF 模型取得更高的预测精度。高光谱数据和光谱变换数据是预测土壤属性时常被选用的变量，通过光谱数据建立的模型能取得较高的预测精度^[8]。乔天等^[9]用筛选出的特征波段，建立土壤质地预测模型，研究结果比全波段建模预测结果更加精确。黄明祥等^[10]对海涂砂粒光谱预处理后，构建了预测砂粒的线性和非线性模型，结果表明线性模型更加稳定可靠。

前人对土壤颗粒预测时，选择的环境辅助变量多为高光谱数据、气候数据和地形数据，结合环境变量和高光谱数据预测土壤颗粒含量的研究少有报道。本文将环境变量结合实测高光谱数据作为模型输入变量，以室内实验获得的土壤砂粒、粉粒和黏粒含量为预测目标，建立 RF 和 XGBoost 预测模型。研究结果

有望为该地区的土壤监测及管理提供数据基础。

1 材料与方法

1.1 研究区概况

渭-库绿洲位于塔里木盆地中北部，其北靠天山山脉，东临塔克拉玛干沙漠，地理位置(80°37'E ~ 83°59'E, 41°06'N ~ 42°40'N)，绿洲内地势呈西北高东南低。成土母质以碳酸钙岩和盐岩为主，在风化、剥蚀等外力作用下其产物向平原区汇集。根据世界土壤数据库(HWSD)，按照 FAO-90 土壤分类系统，研究区内主要土壤类型有盐漠泥砂土(属于盐化棕漠土亚类)、火黑土(属于石灰性灰褐土亚类)、灰淤土(属灌淤土亚类)。

1.2 土样数据采集

土壤样本采集在 2017 年 7 月 2 日至 7 月 6 日完成。根据以往采样经验和绿洲内土壤质地类别，在采样区内(30 m × 30 m)用五点采样法，共采集 62 个(0 ~ 10 cm)土壤样品(图 1)。将采集的样品混合均匀后装入密封袋，经室内实验，剔除异常值和误差后，获得 50 个有效土壤样本。

1.2.1 土壤实验及数据处理 根据土水比 1 : 5 ($m : V$)配成土壤溶液，经沉淀过滤后测定土样 pH 和土壤含盐量(SSC)。土壤含水量(SMC)采用烘箱烘干后用称重法测定。采用激光粒度仪(Mircotrace S3500)测定土壤粒径，将测量数据按美国制分为：黏粒(< 0.002 mm)、粉粒(0.002 ~ 0.05 mm)、砂粒(0.05 ~ 2 mm)^[11]。

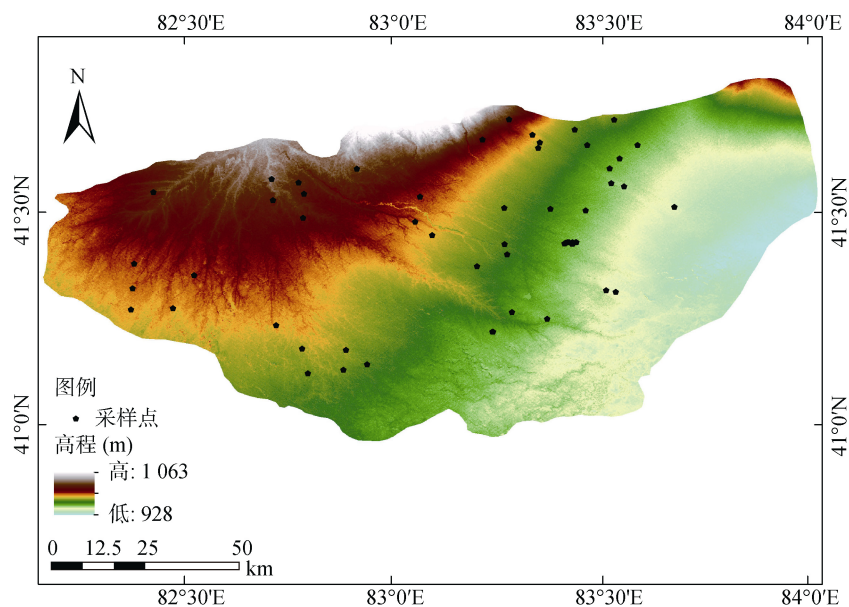


图 1 采样点示意图
Fig. 1 Sampling sites

1.2.2 光谱测量及数据处理 在暗室环境下采用 FieldSpec3 型光谱仪, 对每个样本测量 10 次后取均值, 即为该样本的光谱数据。去除边缘噪声较大的 350 ~ 400 nm 和 2 401 ~ 2 500 nm 的光谱曲线, 用一阶微分(FD)和 Savitzky-Golay (SG)平滑方法对其余波段进行预处理。通过 SPSS 软件对 3 种光谱数据进行主成分(PCA)分析, 原始光谱选择前 3 个主成分(YPC1、YPC2、YPC3), 一阶微分选择前 5 个主成分(FDPC1、FDPC2、FDPC3、FDPC4、FDPC5), SG 平滑选择前 2 个主成分(SGPC1、SGPC2), 作为模型输入变量。

1.3 遥感影像获取及预处理

在 Google Earth Engine(GEE)平台, 获取 2017 年 7 月 4 日 L1T 级的 Landsat8 OLI 影像, 其空间分辨率为 30 m, 波段运算后得到归一化植被指数(NDVI)和增强型植被指数(EVI)。土壤容重(BD)数据来源于

HWSD, 空间分辨率为 1 km, 下载地址 <http://data.tpdc.ac.cn>; 土壤有机碳(SOC)、阳离子交换量(CEC)数据来源于 <https://soilgrids.org/>, 空间分辨率 250 m, 通过 ArcMap 处理后得到研究区土壤 BD、SOC 和 CEC 数据。

1.4 地形和气候变量

在土壤的形成和发育过程中, 受母质、时间、人类活动诸多环境因素影响^[12]。因时间和人类活动没有定量数据表达, 选择地形和气候作为环境变量。地形(DEM)变量及衍生变量用 SAGA GIS 软件计算, 数据来源 <https://www.gscloud.cn>, 空间分辨率 30 m。下载 2017 年 CRU TS 气候数据集作为气候数据, 数据来源 <https://crudata.uea.ac.uk/cru/data/hrg/>, 分辨率为覆盖陆地表面 0.5°。经 ArcMap 重采样(30 m 空间分辨率), 得到 7 月的月均温(TEM)和月均降水量(PRE)。以上环境变量见表 1。

表 1 环境变量信息
Table 1 Environmental variable information

环境变量	变量名	数据处理	环境变量	变量名	数据处理
土壤因子	土壤含水量(SMC)	室内实验	气候因子	月均温(TEM)	ArcMap 处理
	土壤 pH			月均降水量(PRE)	
	土壤含盐量(SSC)	ArcMap 处理	植被因子	归一化植被指数(NDVI)	GEE 计算
	土壤容重(BD)			增强型植被指数(EVI)	
	土壤有机碳(SOC)		SAGA GIS 提取	光谱因子	一阶微分主成分 1(FDPC1)
阳离子交换量(CEC)	一阶微分主成分 2(FDPC2)				
地形因子	数字高程模型(DEM)	一阶微分主成分 3(FDPC3)			
	山体阴影(AH)	一阶微分主成分 4(FDPC4)			
	坡向(AS)	一阶微分主成分 5(FDPC5)			
	闭合洼地(CD)	SG 平滑主成分 1(SGPC1)			
	收敛指数(CI)	SG 平滑主成分 2(SGPC2)			
	河网准基面(CNBL)	原始光谱主成分 1(YPC1)			
	河网距离(CND)	原始光谱主成分 2(YPC2)			
	坡长变量(LSF)	原始光谱主成分 3 (YPC3)			
	谷深(VD)				
	剖面曲率(PrC)				
	相对坡度位置(RSP)				
	坡度(S)				
	总汇水面积(TCA)				
	地形湿度指数(TWI)				

1.5 模型构建和评价

随机森林(RF)是多棵决策树的组合, 其中树彼此间相互独立, 在多棵树中完成对样本的训练和预测^[2]。RF 不同于线性回归要假设目标预测变量的概率分布, 并能够防止过拟合问题^[6]。在 R 语言中用 caret 包把样本数据 60% 划为训练集, 40% 划为测

试集, 可取得较好的预测效果。模型参数 ntree 为 500 和 1 000, mtry 为 2、3 和 5。

极端梯度提升(XGBoost)算法具有正则化、并行处理运算、内置交叉验证和高度的算法灵活性等优势^[13]。其模型结构相对简单, 避免过拟合且准确率较高。模型参数 eta=0.1, gamma 默认, max-depth=6,

nrounds=500。

R^2 用来表示模型预测精度，RMSE 和 MAE 用于计算模型预测数据的误差。其计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (\bar{x} - y_i)^2} \quad (1)$$

$$BMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (3)$$

式中： x_i 表示样本的观测值， y_i 是模型预测值， n 为总样本量， \bar{x} 是观测数据的均值。

2 结果与分析

2.1 土壤粒径描述性统计特征

50 个采样点的土壤粒径统计结果如表 2 所示。研究区内土壤颗粒含量砂粒最多，粉粒其次，黏粒最少。3 种土壤颗粒的变异系数随着颗粒粒径的减小而升高，说明研究区内土壤颗粒含量的异质性较强。

表 2 土壤粒径描述性统计(%)
Table 2 Descriptive statistics of soil particle sizes

土壤颗粒	最小值	最大值	平均值	标准差	变异系数(%)
砂粒	7.28	95.24	58.95	28.35	48.09
粉粒	3.88	89.09	42.45	29.32	69.08
黏粒	0.19	6.45	1.98	1.30	73.33

采用 SigmaPlot 绘制土壤质地三重图(图 2)。根据美国制土壤质地分类标准，渭-库绿洲的土壤质地主要为砂壤质。

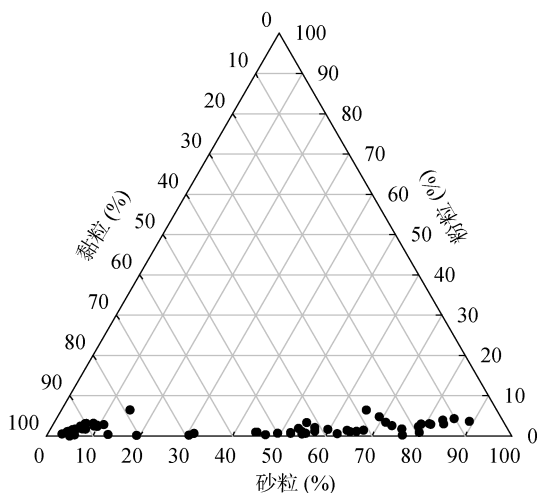


图 2 土壤质地三重图

Fig. 2 Ternary map for soil texture

2.2 土壤理化属性及环境条件

研究区内土壤采样点的基本理化属性和环境条件如表 3 所示。可以看出采样点土壤呈微碱性；SOC 含量和 NDVI 值因绿洲内土壤类别和植被覆盖度的差异而变化较大；CEC 是土壤保肥指标，绿洲内的土壤保肥能力处于中等水平。由于绿洲内降水少蒸发量大，加之肥力较好的成土母质，形成了发达的绿洲滴灌农业。

表 3 土壤属性和环境描述

Table 3 Description properties of soils and environment				
指标	均值	标准差	偏度(%)	变异系数(%)
pH	7.8	0.43	-0.20	5.51
BD(g/cm ³)	1.41	0.08	0.50	5.64
SOC(g/kg)	12.65	4.25	1.54	33.54
CEC(cmol/kg)	17.91	2.07	-0.07	11.53
TEM(°C)	27.31	0.96	-0.52	3.5
PRE(mm)	16.51	2.11	-0.2	12.77
NDVI	0.22	0.15	0.42	67.34
母质	碳酸钙岩和盐岩的坡积物、洪积物和灌淤积物			
植被	怪柳、盐爪爪、云杉、圆柏、苔草等			

2.3 土壤颗粒的影响因素分析

将预处理的环境变量和实测土壤颗粒数据进行 Pearson 分析。由表 4 可知，pH 与砂粒呈负相关，由于砂粒孔隙度大，在强烈的蒸发下，土壤水和致酸离子解离后，导致土壤呈酸性。pH 与粉粒呈正相关，随着土壤粒径的减小，土壤颗粒的保水性能会提升，土壤颗粒间的 OH⁻ 和 H⁺ 彼此交换中改变土壤酸碱性。土壤光谱反射率受颗粒粒径影响，粒径大的颗粒之间能保持更多的空气和水，使得光谱吸收率增加；粒径小的土壤颗粒，因孔隙度的变小使颗粒间结合更为紧密，光谱反射率变大^[14]。相关研究表明，光谱数据在通过微分变换后，土壤光谱反射率与土壤粒径呈负相关^[15]。由于绿洲内高程起伏较小，减小了地形对土壤颗粒再次分配的影响，因此地形变量和土壤颗粒有较强的相关性^[16]。

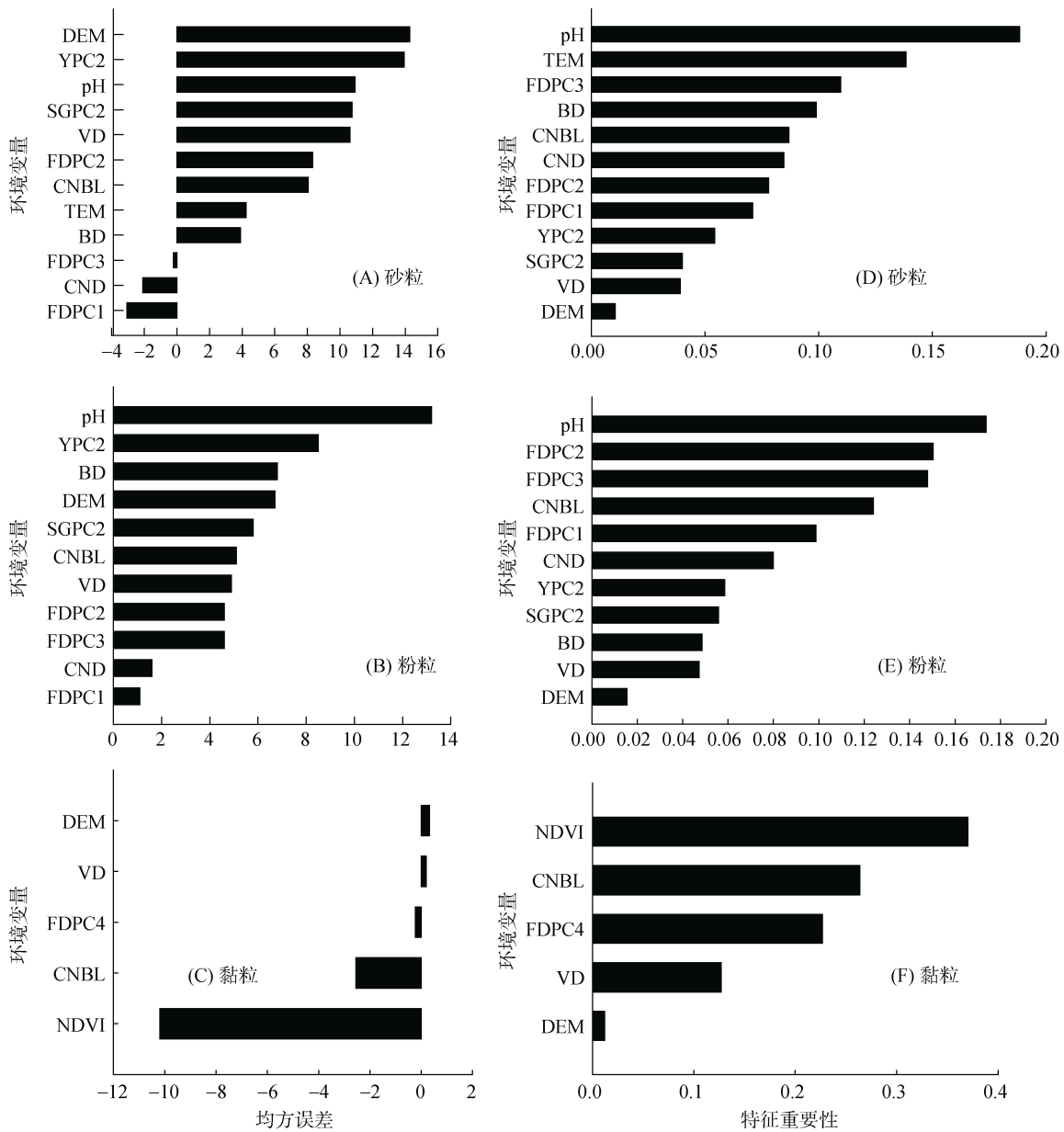
2.4 环境变量重要性分析

在 R 语言中对环境变量进行重要性排序(图 3)。两个预测模型中，pH、光谱变换变量和地形变量是预测砂粒和粉粒含量的重要因子。pH 受生物、气候及人类作用等因素影响，土壤中游离的酸碱离子在土壤溶液交换过程中改变土壤酸碱性^[17]。相关研究表明，变换后的高光谱数据，在参与建模时综合预测能力好于原始光谱^[18]。李爱迪^[19]的研究结果表明：Elevation、TWI 等地形因子是预测土壤质地的重要变

表 4 土壤颗粒与环境变量的相关性
Table 4 Correlation of soil particle contents with environmental variables

环境变量	砂粒	粉粒	黏粒	环境变量	砂粒	粉粒	黏粒
pH	-0.568**	0.582**	-0.172	DEM	-0.368**	0.386**	-0.313*
BD	0.259*	-0.257*	-0.091	FDPC1	0.308*	-0.313*	0.042
NDVI	-0.031	0.043	-0.262*	FDPC2	-0.617**	0.627**	-0.071
TEM	0.253*	-0.249	-0.145	FDPC3	-0.353**	0.361**	-0.096
CNBL	-0.361**	0.380**	-0.327**	FDPC4	0.102	-0.085	-0.377**
CND	-0.280*	0.287*	-0.081	SGPC2	-0.401**	0.414**	-0.180
VD	0.383**	-0.400**	0.269*	YPC2	-0.371**	0.383**	-0.160

注：*、**表示相关性达 $P<0.05$ 和 $P<0.01$ 显著水平(双尾)。



(A、B、C 为 RF 模型中变量重要性排序；D、E、F 为 XGBoost 模型中变量重要性排序)

图 3 环境变量重要性

Fig. 3 Importance of environmental variables

量。在预测黏粒含量的变量重要性排序中，两个模型的排序出现较大差异，是因为 RF 模型中是用均方误差作为变量重要性的评价指标，XGBoost 模型是以变量划分后对样本的覆盖度为变量重要性衡量指标。

2.5 预测结果分析和对比

模型预测结果如表 5 所示。RF 模型对粉粒的预测效果最好，砂粒次之，黏粒的预测效果最差。对比 RF 模型，XGBoost 模型的预测效果，砂粒最好，粉粒次之，黏粒有所提升。预测单个土壤颗粒时，XGBoost 模型对砂粒的预测效果最好；RF 模型对粉粒预测有优势，误差也相应减小；XGBoost 对黏粒的预测结果好于 RF 模型。从整体预测结果来看，XGBoost 模型好于 RF 模型。

表 5 土壤颗粒含量预测精度验证
Table 5 Prediction accuracy validation of soil particle contents

模型	土壤颗粒	训练集			验证集		
		R ²	RMSE	MAE	R ²	RMSE	MAE
RF	砂粒	0.86	8.40	6.10	0.52	18.54	14.92
	粉粒	0.89	8.33	6.20	0.75	15.18	12.17
	黏粒	0.83	0.78	0.53	0.03	1.61	1.20
XGBoost	砂粒	0.91	1.49	1.38	0.78	14.60	10.26
	粉粒	0.88	3.52	1.86	0.54	18.25	13.21
	黏粒	0.81	0.89	0.55	0.39	0.99	0.84

通过 ArcMap 提取 HWSD 中研究区内的土壤颗粒数据。对模型预测数据、实测土壤颗粒数据和 HWSD 中的 3 种土壤颗粒数据的误差(RMSE、MAE)进行对比分析。从图 4 中可以看出，本研究中两个模型的预测误差整体上均小于 HWSD 和实测数据的误差。

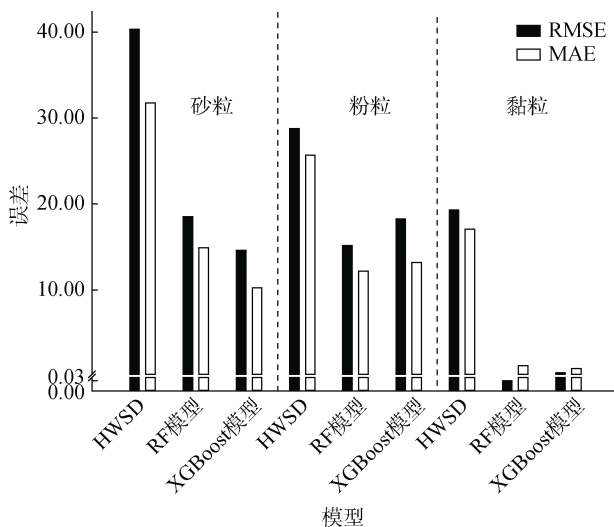


图 4 数据误差对比

Fig. 4 Comparison of data error

3 讨论

选择土壤属性变量、环境变量和光谱变量等，构建了 RF 和 XGBoost 预测土壤颗粒含量模型。从预测结果来看(表 5 和图 4)，本文两种模型的预测结果比马重阳等^[20]预测土壤属性的结果有所提升；与 da Silva Chagas 等^[7]预测干旱区土壤颗粒的研究结果相似。模型输入变量对预测结果也有较大影响，在相关研究中，通过高光谱数据建立的预测模型，相较于只有土壤属性变量、环境变量和地形变量建立的预测模型，能取得更高的预测精度^[21-22]。同时，本研究与前人研究也存在差异之处，魏宇宸等^[23]和其他学者^[5, 7]在预测土壤颗粒含量时，RF 模型预测效果最好。本研究中，RF 模型在预测黏粒时精度较低，可能是因为 RF 模型将 FDPC4、CNBL、NDVI 环境变量重要性计算为负数。

徐佳等^[24]利用机器学习方法从土壤属性角度出发，推测关键成土的环境要素研究中发现，各土壤属性中 pH 对地表温度、年降水量和年均温环境变量的贡献性较高。在本文中，pH 也是环境变量中重要的土壤属性因子，其对砂粒和粉粒的预测结果影响较大。DEM 及相关衍生变量是影响土壤颗粒组成的重要因素，在以往的研究中常被选为预测土壤颗粒含量的关键因子^[25]。在本研究中 DEM、CNBL、CND 等地形因子，在模型预测的环境变量中均占据较高的重要性。光谱信息是反映土壤属性的有效数据，用光谱数据建立机器学习评估粒径含量和分布模型，达到较高的预测精度^[26]。在本研究中，预测砂粒和粉粒含量时，有较多的光谱变量参与建模，预测精度也较高；预测黏粒含量时，只有较少的光谱变量参与建模，是导致模型预测精度较低的一部分原因。

实测土壤颗粒含量数据的离散程度，会对模型预测结果产生不确定性的影响^[18]。研究区实测土壤颗粒数据中砂粒和粉粒的分布比较集中，两种模型预测精度整体较高；而黏粒数据的分布较离散，使 RF 模型没有发挥本有的预测性能。由于建模样本量过小，致使本文中出现了验证集对比建模集精度下降的问题。在以后的研究中，应采用更加科学合理的采样方法以及增加样本数量，利用更优的环境变量筛选方法和多种变量组合方案，以降低模型预测的不确定性，提高预测精度。

4 结论

1)通过 Pearson 相关性分析得出的环境变量，构

建了 RF 和 XGBoost 模型预测土壤砂粒、粉粒、黏粒含量,并取得较好的建模效果。XGBoost 模型的预测精度整体较高,尤其是预测黏粒含量。

2)数字高程模型、原始光谱主成分 2、土壤 pH 和月均温是预测砂粒含量的重要环境变量;土壤 pH、一阶微分主成分 2、土壤容重和数字高程模型等是预测粉粒含量的重要环境变量;归一化植被指数,河网基准面,一阶微分主成分 4 和谷深是预测黏粒含量的重要环境变量。

3)对模型得到的预测数据、实测数据和世界土壤数据库(HWSD)中的土壤颗粒数据进行对比分析,模型预测数据比 HWSD 中土壤颗粒数据更接近实测数据的范围。

参考文献:

- [1] 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.
- [2] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [3] 刘亚东,李旺平,赵林,等. 青藏高原温泉地区土壤黏粒含量剖面分布模式及其影响因素[J]. *土壤*, 2021, 53(3): 637-645.
- [4] Liu F, Rossiter D G, Song X D, et al. An approach for broad-scale predictive soil properties mapping in low-relief areas based on responses to solar radiation[J]. *Soil Science Society of America Journal*, 2020, 84(1): 144-162.
- [5] Mareike Ließ, Bruno Glaser, Bernd Huwe. Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models[J]. *Geoderma*, 2012, 170: 70-79.
- [6] Forkuor G, Hounkpatin O K L, Welp G, et al. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models[J]. *PLoS One*, 2017, 12(1): e0170478.
- [7] da Silva Chagas C, de Carvalho W Jr, Bhering S B, et al. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions[J]. *CATENA*, 2016, 139: 232-240.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 吴克宁,赵瑞. 土壤质地分类及其在我国应用探讨[J]. *土壤学报*, 2019, 56(1): 227-241.
- [12] 丁建丽,王飞. 干旱区大尺度土壤盐度信息环境建模——以新疆天山南北中低海拔冲积平原为例[J]. *地理学报*, 2017, 72(1): 64-78.
- [13] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system[C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA. New York: ACM, 2016: 785-794.
- [14] 杨雪红. 土壤粒径对土壤光谱特征的影响[J]. *科技信息*, 2010(25): 390-391, 154.
- [15] 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.
- [16] 张世文,黄元仿,苑小勇,等. 县域尺度表层土壤质地空间变异与因素分析[J]. *中国农业科学*, 2011, 44(6): 1154-1164.
- [17] 耿增超,戴伟. *土壤学*[M]. 北京: 科学出版社, 2011.
- [18] 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.
- [19] 李爱迪. 地形因素影响下重庆市主要土壤的质地类型空间分布预测研究[D]. 重庆: 西南大学, 2019.
- [20] 马重阳,孙越琦,巫振富,等. 基于不同模型的区域尺度耕地表层土壤有机质空间分布预测[J]. *土壤通报*, 2021, 52(6): 1261-1272.
- [21] 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.
- [22] 卢宏亮,赵明松,刘斌寅,等. 基于随机森林模型的安徽省土壤属性空间分布预测[J]. *土壤*, 2019, 51(3): 602-608.
- [23] 魏宇宸,赵美芳,朱昌达,等. 基于景观及微地形特征的丘陵区土壤属性预测[J]. *应用生态学报*, 2022, 33(2): 467-476.
- [24] 徐佳,刘峰,吴华勇,等. 基于人工神经网络和随机森林学习模型从土壤属性推测关键成土环境要素的研究[J]. *土壤通报*, 2021, 52(2): 269-278.
- [25] Laborczi A, Szatmári G, Takács K, et al. Mapping of topsoil texture in Hungary using classification trees[J]. *Journal of Maps*, 2016, 12(5): 999-1009.
- [26] Parent E J, Parent S É, Parent L E. Determining soil particle-size distribution from infrared spectra using machine learning predictions: Methodology and modeling[J]. *PLoS One*, 2021, 16(7): e0233242.