

# 黄土高原不同地貌区农田土壤有机质预测方法研究

张万涛<sup>1</sup>, 吉静怡<sup>2,3</sup>, 李彬彬<sup>2</sup>, 王菊<sup>1</sup>, 许明祥<sup>2\*</sup>

(1 西北农林科技大学水土保持研究所, 陕西杨凌 712100; 2 中国科学院水利部水土保持研究所, 陕西杨凌 712100;  
3 西北农林科技大学资源环境学院, 陕西杨凌 712100)

**摘要:**【目的】开展黄土高原不同地貌区农田土壤有机质(SOM)预测方法研究,探讨不同预测方法在不同区域的适用性及不确定性,以便更准确地估算农田SOM空间分布特征,对土壤资源高效利用和农田精细化管理具有重要意义。【方法】在黄土高原3种典型地貌区进行试验,包括丘陵沟壑区(庄浪县)、高塬区(宁县)和平原区(武功县),分别布设样点3788、4048和3860个,分析农田土壤0—20 cm SOM含量。运用地统计学理论,分析各典型区SOM空间分布特征。提取原始样本75%为建模点,其余25%为验证点,利用普通克里格(OK)、随机森林(RF)和随机森林+普通克里格(RF+OK)等方法,结合土壤类型、地形、气候、植被、人类活动等多元影响因子,对SOM分布进行空间预测,并对预测结果进行误差分析和空间结构检验,明确各方法的不确定性。【结果】丘陵沟壑区、高塬区、平原区SOM平均含量分别为14.29、13.15、14.48 g/kg,均属于较低水平;变异系数分别为18.96%、19.54%、26.71%,呈中等变异;块金效应分别为8.60%、17.41%和10.01%,受随机性和结构性因子共同作用,且受后者影响更大;丘陵沟壑区和平原区SOM含量的Moran's *I*分别为0.26和0.14, *Z*[*I*]分别为26.56和13.51,存在显著空间自相关性,而高塬区SOM含量Moran's *I*为0.02, *Z*[*I*]为1.55,不存在空间自相关性。丘陵沟壑区、高塬区、平原区SOM含量空间分布分别受温度、海拔、降水影响最大。在平原区,RF+OK法较RF法和OK法,MSE、RMSE、MAE等误差均最小,实测值与预测值的相关系数(*r*)最高,预测值的空间结构与实测值更接近。高塬区SOM空间分布无规律,OK法在该区域不适用,RF法和RF+OK法的各项误差无明显差异,但RF法的*r*更高,且预测值的空间结构更符合宁县实际特征。在平原区,OK法预测结果的不确定性较大,RF和RF+OK方法各项误差和*r*均无明显差异,但RF方法预测值的空间结构与实测值更接近,且较其它两个地区,其SOM变异性及建模点和验证点的各项误差均最大。【结论】在不同地貌区,环境要素、空间结构不同,同一预测方法的预测精度存在差异,平原区较丘陵沟壑区和高塬区,其空间预测结果的不确定性更大。在同一地貌区,3种预测方法的预测结果存在差异,丘陵沟壑区使用RF+OK法预测SOM空间分布效果较好,而高塬区和平原区则用RF法较好。当区域SOM存在显著空间相关性,且半方差函数的拟合度较高、残差较小时,采用RF+OK方法可显著提高模型预测精度。

**关键词:** 黄土高原; 土壤有机质; 普通克里格; 随机森林; 不确定性; 空间预测

## Spatial prediction of soil organic matter of farmlands under different landforms in the Loess Plateau, China

ZHANG Wan-tao<sup>1</sup>, JI Jing-yi<sup>2,3</sup>, LI Bin-bin<sup>2</sup>, WANG Ju<sup>1</sup>, XU Ming-xiang<sup>2\*</sup>

(1 College of Forestry, Northwest A&F University, Yangling, Shaanxi 712100, China; 2 Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China;  
3 College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** 【Objectives】 This study employs different methods to predict farmland soil organic matter (SOM) in the typical geomorphic areas of the Loess Plateau. We examined the applicability and uncertainty of the prediction methods in different regions for the estimation of spatial distribution of SOM more accurately, which was of great significance for the efficient use and refined management of soil resources. 【Methods】 This study was

收稿日期: 2020-09-15 接受日期: 2020-11-16

基金项目: 国家自然科学基金项目(41771318); 国家重点研发计划项目(2017YFC0506503)。

联系方式: 张万涛 E-mail: wtzhang2018@163.com; \*通信作者 许明祥 E-mail: xumx@nwsuaf.edu.cn

conducted in the three geomorphological regions of the Loess Plateau -the hill and gully area (HGA, in Zhuanglang County), high plateau area (Ning County), and the plain area (Wugong County). We collected 3788, 4048, and 3860 soil samples, respectively, from the study areas to determine SOM content. The spatial distribution characteristics of SOM in the study areas were analyzed using geostatistics theory. 75% of the original data was extracted for modeling, and the remaining 25% were used for validation using ordinary Kriging (OK), random forest (RF), and random forest + ordinary Kriging (RF+OK) methods. The modeling techniques considered soil multi-source influencing factors such as soil type, terrain, climate, vegetation, human activities, etc. We clarified the uncertainty of each prediction method through error analysis and spatial structure inspection.

**【 Results 】** The average SOM content in the hill and gully area, high plateau area, and the plain area were 14.29, 13.15, and 14.48 g/kg. The study areas' SOM content fell into a low level, and the coefficients of variation were 18.96%, 19.54%, and 26.71%, showing medium variation. Nugget effects were 8.60%, 17.41%, and 10.01% as affected by the combination of randomness and structural factors, with the latter having a higher significant effect. The SOM content in the hilly and gully area and plain area were 0.26 and 0.14, while  $Z[I]$  were 26.56 and 13.51, showing a significant spatial autocorrelation. In the high plateau area, Moran's  $I$  of SOM content was 0.02, and  $Z[I]$  was 1.55, indicating a lack of spatial autocorrelation. The spatial distribution of SOM content in the hilly and gully areas, high plateau area, and the plain area was most affected by temperature, altitude, and precipitation, respectively. The RF+OK method had the smallest error (MSE, RMSE, MAE, etc) in the plain area compared with the RF and OK method. The correlation coefficient ( $r$ ) between the observed and predicted values was the highest, and the spatial structure of the predicted value was closer to the observed value in plain area. The spatial distribution of SOM in the high plateau area was irregular, and the OK method was not applicable in this area. There was no significant difference between the errors of the RF and RF+OK method. Still, the  $r$ -value of the RF method was higher, and the predicted value's spatial structure was close to the actual characteristics of the high plateau area. In the hill and gully area, the uncertainty of the OK method's prediction results was relatively large. There was no significant difference between the errors and  $r$  of the RF and RF+OK methods, but the spatial structure of the RF method's predicted values was closer to the observed values. Compared with the other two regions, the SOM variability and modeling and validation errors in the plain area were the largest.

**【 Conclusions 】** In different geomorphic areas, environmental factors and spatial structures are different, and the prediction accuracy of different methods vary. Compared with the hill and gully area and high plateau area, the spatial prediction results' uncertainty in the plain area is higher. We found differences in the results of the three prediction methods within the same geomorphic area. The RF+OK method in the hilly and gully area is better at predicting the spatial distribution of SOM, while the RF method is better in the high plateau and plain areas. When regional SOM has a significant spatial correlation, a high fit of the semi-variance function, and a small residual, the RF+OK method can significantly improve the model's prediction accuracy.

**Key words:** Loess Plateau; soil organic matter; ordinary Kriging; random forest; uncertainty; spatial prediction

土壤有机质 (soil organic matter, SOM) 是土壤肥力的基础, 在空间上是高度变异的时空连续体, 与环境要素关系复杂, 了解其特征和分布状况是农户、科学家和政府部门普遍关注的问题<sup>[1-2]</sup>。耕地地力监测、田间管理、固碳潜力评估等<sup>[3]</sup>均依赖于详细的土壤空间分布信息。获得详细、准确的土壤属性空间分布信息, 是土壤质量评价的基础, 对土地资源利用、数字土壤制图等具有重要意义<sup>[4]</sup>。

国内外关于土壤属性的空间预测方法已有大量

研究。过去对土壤有机质的空间变异以定性描述为主, 认为土壤属性在空间上是独立存在的, 只能概括土壤特性变化的全貌, 无法反映局部变异特征<sup>[5]</sup>。近年来, 地统计学方法逐渐被运用于空间预测中, 被证明是分析土壤要素空间分布特征的有效方法之一。但地统计学方法偏向于考虑土壤属性的结构性和空间相关性, 缺乏对空间分布过程性因素的刻画, 导致预测结果存在不确定性<sup>[6]</sup>。随着地统计学的广泛应用, 许多学者开始将地统计学方法和地

理信息系统相结合进行土壤属性空间预测, 将 GIS 和地统计学方法相结合, 发现引入定量辅助因子(气候、地形)能提高土壤养分空间分布的预测精度<sup>[6-9]</sup>, 且预测精度较基于样点数据进行空间内插的方法有明显提高<sup>[10-12]</sup>。李启权等<sup>[13]</sup>发现基于定量和定性辅助变量(土地利用、土壤类型)可以进一步提高土壤性质的空间预测。但土壤属性与环境因子间的关系并不完全是线性关系, 往往存在一些复杂的非线性关系<sup>[14]</sup>。随着人工智能技术的不断发展, 许多学者开始采用机器学习方法进行土壤属性空间分布预测, 其中随机森林以其强大的非线性拟合能力被广泛运用于土壤属性空间预测<sup>[15-19]</sup>。但土壤属性不仅与环境因子存在关系, 其自身也存在一定空间相关性, 片面的对土壤属性进行空间预测, 缺乏对环境因子、空间结构及空间相关性的认识, 会导致空间预测结果存在不确定性。同时, 不同地貌区域外部环境条件不同, 其空间变异特征及影响因素存在差异, 导致各种方法的预测结果存在差异。因此, 综合考虑土壤属性的空间结构特征及其与环境因子间的关系, 了解不同预测方法在不同区域的适用性和不确定性, 对选择适当方法进行预测至关重要。

以黄土高原的丘陵沟壑区(庄浪县)、高塬区(宁县)、平原区(武功县)为研究区, 采用耕地质量数据, 选取土壤类型、地形、气候、植被、人类活动等矢量环境因子, 利用地统计学、机器学习以及二者相结合的方法, 对黄土高原不同地貌区典型县农田 SOM 进行空间分布预测及不确定性分析, 明确不同预测方法的优势与不足、适用性及预测结果的不确定性, 为精准农业和数字化土壤制图提供依据。

## 1 材料与方法

### 1.1 研究区概况

庄浪县(E 105°46'~106°23', N 35°03'~35°28')位于甘肃东部、六盘山西麓, 是甘肃与宁夏两省交汇处, 属黄土丘陵沟壑区第三副区。地形西低东高, 境内基岩山地和丘陵沟壑占总面积的 93.5%, 海拔在 1405~2857 m, 年均降水量 510 mm, 年平均气温 8.1℃, 属于大陆性季风区温带半湿润气候。总面积 1553 km<sup>2</sup>, 耕地 11.3 万 hm<sup>2</sup>, 梯田化程度达 91.74%, 主要种植作物为玉米、小麦和马铃薯。

宁县(E 107°41'~108°34', N 35°15'~35°52')位于甘肃省庆阳市, 地处关山—六盘山褶皱带移动的鄂尔多斯盆地东南部; 地形为东北高、西南低, 海

拔在 860~1760 m; 地处大陆内部, 年均温 8.7℃, 年均降水量 565.9 mm, 属暖温带大陆性季风气候, 总面积 2653 km<sup>2</sup>, 其中耕地面积 9.5 万 hm<sup>2</sup>, 种植作物以小麦、玉米为主。

武功县(E 108°01'~108°19', N 34°12'~34°26')位于陕西省咸阳市, 地处关中平原西部, 属渭河地貌; 地形西北高、东南低, 从北向南成阶梯状, 分别为一、二、三级阶地, 海拔 411~600 m, 海拔高差 189 m; 全年平均气温 12.9℃, 年均降水量 633.7 mm, 属温带半湿润性气候区。总面积 397.8 km<sup>2</sup>, 其中耕地面积 3.4 万 hm<sup>2</sup>, 种植作物以玉米、小麦为主。

### 1.2 样品采集与处理

采样过程中, 按照“随机”、“等量”和“多点混合”等原则, 用“S”形布点, 采集 0—20 cm 耕层土壤, 尽量克服耕作、施肥等所造成的误差。用四分法取混合土样 1 kg 左右, 记录采样点位置及周边环境等相关信息。于 2007 年完成样品采集工作, 采样点基本覆盖庄浪县、宁县和武功县所有农田(图 1), 样本量分别为 3788、4048 和 3860 个。将土壤样品风干后过筛供分析测定, 采用重铬酸钾—外加热法测定 SOM 含量。

### 1.3 矢量影响因子获取

参考国内外相关研究<sup>[6, 13, 15]</sup>, 选取土壤类型、地形、气候、植被、人为干扰等影响因子构建土壤-景观模型。土壤类型数据由地方农业技术中心提供<sup>[20]</sup>; 利用 DEM(30 m) 数字高程模型提取地形因子, 其中海拔、坡度、坡向、坡度变化率、平面曲率、剖面曲率反映宏观地形因子特征, 地表起伏度、地表粗糙度反映微观地形因子特征, 地形湿润指数综合反映土壤水分分布特性, 河流能动指数、坡长坡度因子分别反映地表径流的侵蚀力和侵蚀过程; 从 www.resdc.cn 获取与采样时间相一致的温度和降水矢量栅格数据(分辨率 1 km), 并重采样至 30 m; 植被因子通过 Landsat-TM 卫星遥感影像(2007 年), 选取 3 波段和 4 波段数据计算植被归一化指数(NDVI); 根据采样点距道路和村庄距离来反映人为活动对 SOM 的扰动。

### 1.4 数据处理

采用 3 $\sigma$  准则处理 SOM 异常值。使用 R 语言对数据进行统计分析、正态检验、相关分析、随机森林回归分析、重要性排序; 在 GS\*9.0 软件中进行半方差函数分析及模型拟合, 在 GeoDa 软件中利用 K-Nearest Neighbor 空间权重矩阵进行空间自相关分

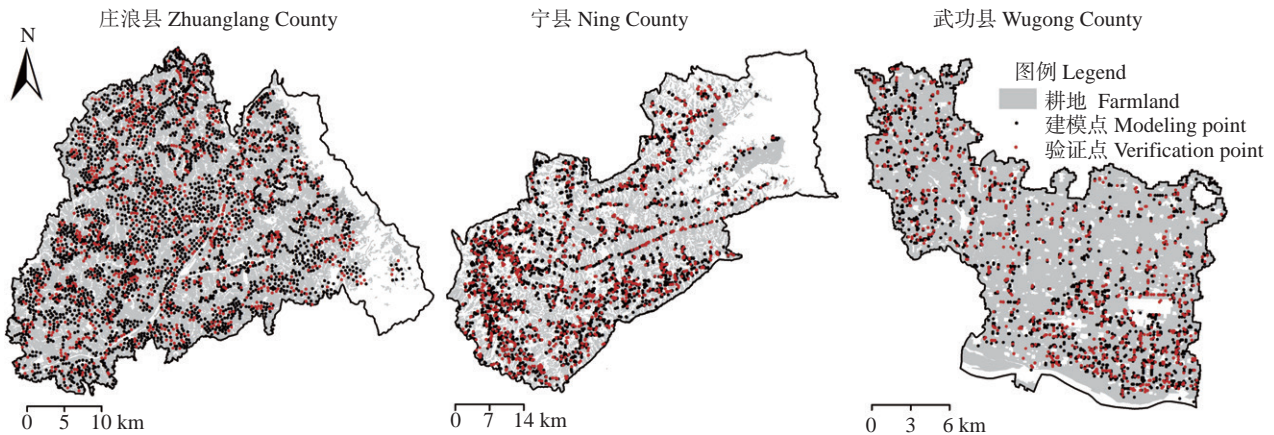


图 1 研究区样点分布图

Fig. 1 Distribution of sample points in study area

析; 使用 ArcGIS10.2 进行空间分析、普通克里格插值、独立数据集验证以及数字土壤制图。

### 1.5 研究方法

随机森林 (random forest, RF)<sup>[21]</sup> 是一种集成机器学习方法, 它利用随机重采样技术 bootstrap 和节点随机分裂技术构建多棵决策树, 通过投票得到最终分类结果。RF 算法可以有效地避免共线性问题, 支持高维度特征, 同时其受噪声和离群样点数据影响小, 学习速度快, 模型稳定性较好。计算公式如下:

$$Z_{RF}(x_i) = \sum_{i=1}^n a_i x_i \quad (1)$$

式中:  $Z_{RF}(x_i)$  为样点土壤有机质含量值;  $x_i (i = 1, 2, 3 \dots n)$  为各样点的环境因子数 (点要素);  $a_i (i = 1, 2, 3 \dots n)$  为 RF 回归系数。

地统计分析是基于区域化变量理论基础的一种空间分析方法, 半方差函数表示一定范围内变量属性的空间依赖性, 其公式为<sup>[22]</sup>:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^N [Z(x_i) - Z(x_{i+h})]^2 \quad (2)$$

式中:  $\gamma(h)$  为半方差函数;  $h$  为两个两样点的分割距离;  $N(h)$  为相距为  $h$  的样点对数目;  $Z(x_i)$  和  $Z(x_{i+h})$  分别为区域化变量  $Z(x)$  在位置  $x_i$  和  $x_{i+h}$  处的实测值。

若变异函数和相关分析的结果表明某一属性的空间相关性存在, 则可以利用普通克里格 (ordinary Kriging, OK) 进行插值。根据 SOM 含量实测值与 RF 模型预测值的残差值进行半方差函数分析获取最优的拟合模型及模型参数, 进而对土壤有机质含量的残差值进行最优 OK 插值, 计算公式如下:

$$R(x_i) = Z(x_i) - Z_{RF}(x_i) \quad (3)$$

$$Z_R(x_i) = \sum_{i=1}^n \lambda_i \cdot R_{RF}(x_i) \quad (4)$$

式中:  $R(x_i)$  是 RF 实测值与预测值的残差;  $Z(x_i)$  是实测值;  $Z_{RF}(x_i)$  是 RF 预测值;  $Z_R(x_i)$  是实测点  $x_i$  处残差的预测值;  $\lambda_i$  是分配给每个残差值的权重, 且  $\sum \lambda_i = 1$ ,  $n$  是样本数目。

结合式 (1)、(4), 基于 RF 模型的 SOM 含量预测值与基于 OK 法的残差估计值进行空间加运算的公式如下:

$$Z_{RF+OK}(x_i) = Z_{RF}(x_i) + Z_R(x_i) \quad (5)$$

式中:  $Z_{RF+OK}(x_i)$  是 RF 和 OK 方法相结合的预测值;  $Z_{RF}(x_i)$  是 RF 算法在实测点  $x_i$  处的预测值;  $Z_R(x_i)$  是 OK 方法在实测点  $x_i$  处残差的预测值。

### 1.6 预测精度评价

采用独立数据集验证方法, 利用 ArcGIS 中 Create Subsets 模块随机从原始样本中提取 75% 作为建模点, 其余 25% 为验证点。选择均方误差 (MSE)、均方根误差 (RMSE)、平均绝对误差 (MAE)、验证点的测定值与模拟值的相关系数 ( $r$ ) 来评价插值的精度。MSE、RMSE、MAE 越小、 $r$  值越大, 预测结果的精度越高。另外, 对预测值进行半方差函数 (公式 2) 和空间自相关分析, 进一步从空间结构和空间自相关性方面验证预测结果的可靠性。莫兰指数 (Moran's  $I$ ) 为空间相关的计算指标, 用来衡量集聚特征的一个综合性评价统计特征参数, 可用于检验某一空间要素在空间上是否存在依赖关系, 计算公式如下:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij} \cdot (X_i - \bar{X})(X_j - \bar{X})}{\left( \sum_{i=1}^n \sum_{j=1}^n W_{ij} \right) \sum_{i=1}^n (X_i - \bar{X})^2} \quad (6)$$

式中:  $n$  为空间数据的个数;  $X_i$  和  $X_j$  分别为  $i$  区、 $j$  区空间要素的属性值;  $\bar{X}$  为所有空间数据的平均值;  $W_{ij}$  为空间权重矩阵的元素, 空间权重矩阵一般为对称矩阵, 且  $W_{ij} = 0$ ,  $I$  值介于  $[-1, 1]$ 。空间自相关类型可分为高-高 (H-H)、低-低 (L-L)、高-低 (H-L)、低-高 (L-H) 4 种, 其中 H-H 和 L-L 为正相关, 表现为属性值在空间聚集现象, H-L 和 L-H 为负相关类型, 表现为空间分散或差异性现象<sup>[23]</sup>。

可以通过标准化统计量  $Z(I)$  来检验空间要素空间自相关性的显著性水平, 计算公式如下:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}; \quad E(I) = \frac{-1}{n-1} \quad (7)$$

式中:  $Var(I)$  为 Moran's  $I$  的理论方差;  $E(I)$  为

Moran's  $I$  的理论期望值。在 0.05 水平下, 当  $|Z| > 1.96$  时, 表明该空间要素存在空间自相关, 否则不存在空间自相关。

## 2 结果与分析

### 2.1 土壤有机质含量描述性统计特征

对各典型县 SOM 数据进行 K-S 检验, 偏度系数较小, 所有 SOM 近似服从正态分布<sup>[20]</sup>(表 1)。庄浪县、宁县、武功县 SOM 平均含量分别为 14.29、13.15、14.48 g/kg, 变异系数分别为 18.96%、19.54%、26.71%, 均呈中等变异。依照第二次土壤普查养分分类标准, 各典型县 SOM 含量均处于较低水平。

表 1 研究区土壤有机质含量特征

Table 1 Characteristics of SOM content in study area

地貌 Geomorphical area	研究区 Study area	样本数 No. of samples	最小值 Min. (g/kg)	最大值 Max. (g/kg)	均值 Mean (g/kg)	标准差 SD (g/kg)	CV (%)	偏度 Skewness	峰度 Kurtosis	K-S
丘陵沟壑区 Hill and gully area	庄浪县 Zhuanglang	3788	5.04	23.75	14.29	2.71	18.96	0.19	0.55	0.00
高塬区 High plateau area	宁县 Ning county	4048	5.30	21.10	13.15	2.57	19.54	0.21	0.20	0.00
平原区 Plain area	武功县 Wugong	3868	3.00	26.20	14.48	3.87	26.71	0.12	-0.06	0.00

### 2.2 土壤有机质空间结构特征

趋势分析结果 (图 2) 表明, 丘陵沟壑区 SOM 从西到东或从南到北呈逐渐增加趋势, 平原区从西到东或从北到南呈逐渐增加趋势, 高塬区在各个方向上无明显趋势特征。半方差函数分析结果 (表 2) 表明, 丘陵沟壑区和平原区 SOM 含量最优理论模型为指数模型, 高塬区 SOM 含量最优模型为高斯函数, 块金效应表现为高塬区 (17.41%) > 平原区 (10.01%) > 丘陵沟壑区 (8.60%)。各典型地貌区 SOM 空间变异受结构性因素和随机性因素共同作用, 且受结构性因素影响更大。各建模点的空间结构与全部样点无明显差异, 具有良好的空间代表性。

空间自相关分析结果 (图 3) 表明, 丘陵沟壑区、高塬区、平原区 SOM 含量 Moran's  $I$  分别为 0.26、0.02 和 0.14。丘陵沟壑区和平原区 SOM 的  $Z[I]$  均大于 1.96, 存在显著空间自相关性; 高塬区 SOM 的  $Z[I]$  小于 1.96, 无显著空间自相关性。通过 Moran's  $I$  散点图 (图 3) 发现, 丘陵沟壑区和平原区 SOM 在空间上主要分布在一、三象限, 表现为高值与低值集聚并存, 而高塬区则表现为在 4 个象限离散分布。

### 2.3 土壤有机质空间分布模型预测

对 17 个环境因子采用随机森林回归模型进行建模分析, 以环境变量的 1/3 为 metry 节点值, 丘陵沟壑区、高塬区、平原区的 ntree 高于 400、500 和 240 时预测误差趋于稳定。利用 importance 函数对环境变量进行重要性排序 (图 4), %MSE 越大, 表示环境因子越重要。不同环境因子对研究区 SOM 空间预测的重要性存在差异。在丘陵沟壑区, 温度是 SOM 空间预测的最佳环境辅助因子; 在高塬区, 地形因子对 SOM 影响较大, 其中海拔对其影响最大; 在平原地区, 降雨对 SOM 空间分布影响最大。从不同方法的空间预测结果 (图 5) 来看, 研究区 SOM 含量空间分布的总体趋势一致。但从局部分布特征来看, OK 方法预测结果在空间上具有较好的连续性, 表现为高值区和低值区的斑块状分布; 在高塬区和平原区, RF 和 RF + OK 方法预测结果在空间上无明显差异, 能较好地刻画细节信息; 在丘陵沟壑区, RF + OK 较 RF 方法, 即刻画了局部细节信息, 又反映了 SOM 在空间上的集聚特征。

### 2.4 土壤有机质空间预测不确定性分析

对建模点和验证点的预测结果进行精度检验



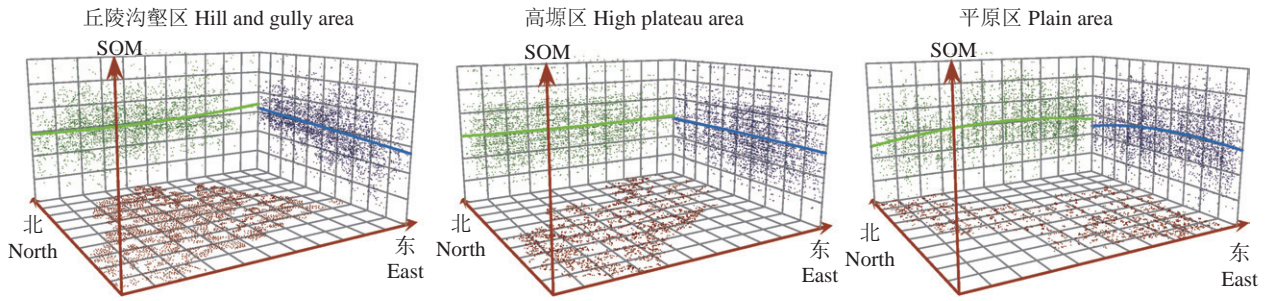


图 2 典型地貌区有机质含量趋势分析图

Fig. 2 Trend analysis chart of SOM contents in typical geomorphic areas

表 2 研究区土壤有机质属性半方差函数模型结构及其参数

Table 2 Semi-variance structures of SOM and their parameters in study area

地貌 Geomorphic area	样本 Samples	数量 No. of samples	模型 Model	块金值 Nugget	基台值 Partial	块金效应 (%) Nugget effect	变程 (m) Range	R <sup>2</sup>	残差 Residual
丘陵沟壑区 Hill and gully area	全部样点 All samples	3788	Exponential	0.61	7.10	8.60	2190	0.61	0.89
	建模点 Modeling point	2841	Exponential	0.65	7.20	9.10	2100	0.58	0.88
高塬区 High plateau area	全部样点 All samples	4048	Gaussian	1.08	6.20	17.41	1212	0.01	0.04
	建模点 Modeling point	3036	Gaussian	1.10	6.31	17.43	1195	0.00	0.06
平原区 Plain area	全部样点 All samples	3868	Exponential	1.19	11.88	10.01	960	0.41	1.31
	建模点 Modeling point	2895	Exponential	1.37	12.35	11.07	990	0.39	1.78

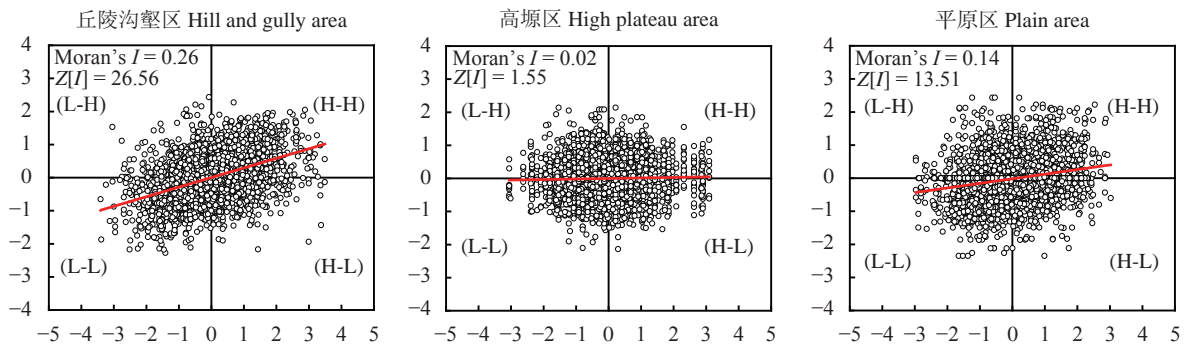


图 3 典型地貌区 Moran's I 散点图

Fig. 3 Spatial autocorrelation of farmland SOM contents in typical geomorphic areas

(表 3、表 4)，在丘陵沟壑区，建模点的各项误差表现为  $RF + OK < RF < OK$ ，验证点的各项误差表现为  $RF + OK < OK < RF$ ，实测值与预测值的相关系数 ( $r$ ) 表现为  $RF + OK > RF > OK$ 。在高塬区和平原区，OK 方法的各项误差均最大， $r$  值较低，而 RF 和 RF + OK 法的各项误差无明显差异，但前者在高塬区  $r$  值更高。通过对比不同方法预测值与实测

值的空间结构及其模型参数发现 (表 5)，在丘陵沟壑区，RF + OK 方法的预测值与实测值的空间结构相似，而 RF 和 OK 方法的预测值和实测值差异较大。在高塬区，不同方法的预测值与实测值间差异较大，且 RF 方法的预测值空间结构更接近实测值。在平原区，3 种预测方法的块金值差异较小，均能较好的表现出 SOM 的空间结构，且 RF 方法预测值的空

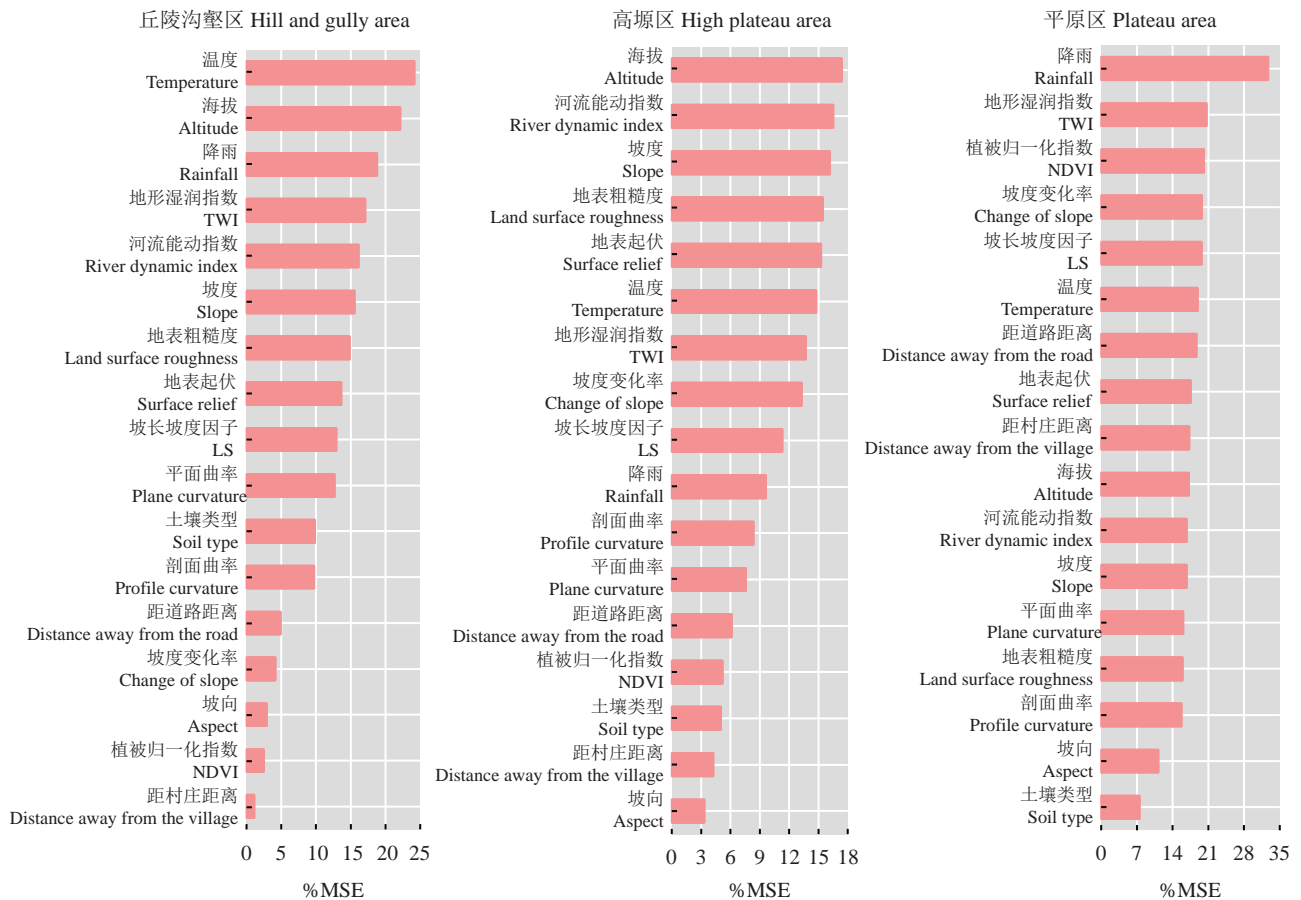


图4 环境因子重要性排序

Fig. 4 Ranking of importance of environmental factors

[注 (Note): TWI—Topographic wetness index; LS—Slope length and slope factor; NDVI—Normalized difference vegetation index.]

间结构与实测值更为接近。空间自相关分析显示(表6), 在丘陵沟壑区, RF + OK 较 RF 和 OK 方法, 其预测值的 Moran's  $I$  和  $Z[I]$  更接近实测值; 在高原区, RF 较 RF + OK 和 OK 方法, 实测值的 Moran's  $I$  和  $Z[I]$  更接近实测值; 在平原区, 3 种方法预测值与实测值差异较大, 且 RF 方法的预测值与实测值的 Moran's  $I$  和  $Z[I]$  更接近。

### 3 讨论

土壤 SOM 与环境变量关系复杂, 且在空间上具有一定异质性<sup>[24]</sup>, 导致其在空间预测中存在较大不确定性。有研究表明通过半方差函数或土壤与环境关系对土壤属性进行空间预测, 可以提高预测精度<sup>[14, 25]</sup>, 但缺乏对土壤属性影响过程和空间结构的综合考虑。同时研究区地貌单一, 未考虑不同预测方法在不同地貌区的适用性。因此综合考虑土壤属性与环境因子之间的关系以及土壤属性本身的空间结构和相关性, 探讨不同预测方法在不同地貌区土壤属性空间预测的不确定性和适用性, 对提高 SOM 的

空间预测精度具有重要意义。

地形因子支配着地表和土壤中水热资源的重新分配, 影响土壤生态系统的物质循环过程和强度, 而降水和温度是影响陆地有机碳储量空间分布的主要气候变量<sup>[26]</sup>, 被广泛地运用于土壤属性的空间预测中。土壤类型和植被不同, 其 SOM 含量存在差异<sup>[27]</sup>, 利用土壤类型和植被归一化指数作为辅助变量可以有效提高预测精度<sup>[28]</sup>。同时, 农田土壤作为人类赖以生存的物质基础, 受人类活动(耕作方式、施肥等)影响显著<sup>[29]</sup>。本研究表明, 在构建土壤景观模型的过程中, 由于地貌类型的差异, 不同环境因子参与模型的相对重要性不同。在丘陵沟壑区, 温度对 SOM 空间分布影响最大, 可能是因为丘陵沟壑区地势差异较大, 导致温度发生变化, 进而影响了土壤养分的输入、分解、积累<sup>[30]</sup>。高原区受海拔影响更大, 这与张志霞<sup>[20]</sup>发现坡向是高原区最重要因子不同。原因可能与高原区独特的地貌特征有关, 高原区农田呈断续分布, 面积较小, 塬面总体平坦, 而内部塬面有起伏, 削弱了坡向对 SOM 空间分布的影

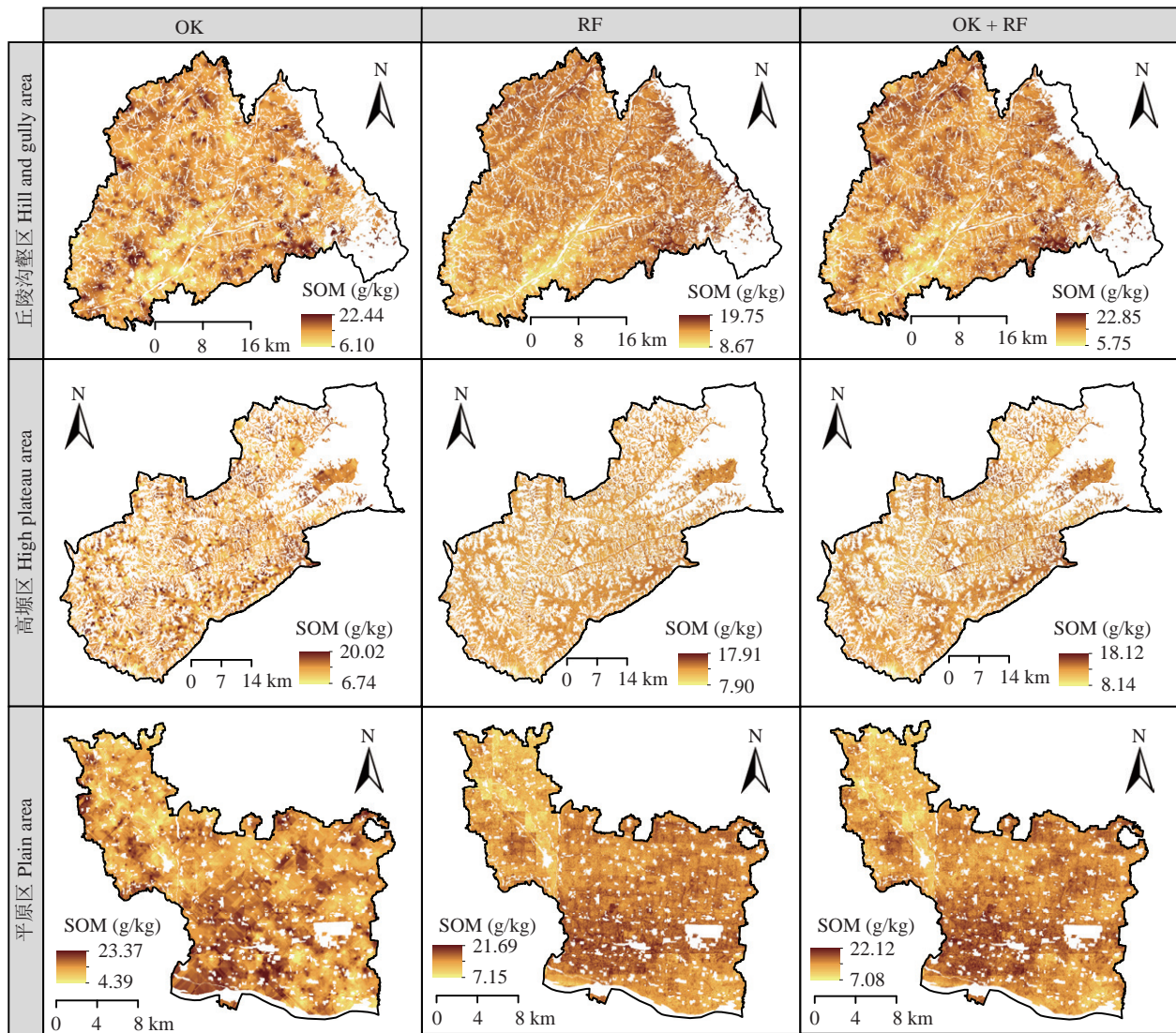


图 5 基于不同预测方法的土壤有机质含量空间分布

Fig. 5 Spatial distribution of soil organic matter content with different prediction methods

[注 (Note): OK—普通克里格方法 Ordinary Kriging method; RF—随机森林方法 Random forest method; RF + OK—随机森林与普通克里格相结合方法 Random forest and ordinary Kriging method.]

响。平原地区受降水影响最大，受地形因素影响较小，通过年均降水量可以有效预测平原地区 SOM 空间分布。土壤类型和植被对研究区 SOM 空间分布影响较小。一方面可能与土壤分类级别有关，土壤分类越低，对 SOM 空间变异的解释能力越大<sup>[28]</sup>。另一方面可能与植被多样性有关，研究区农田种植的作物种类相对单一，主要以玉米和小麦为主，削弱了 NDVI 对其影响。平原区较丘陵沟壑区和高原区，SOM 空间分布受人类活动影响更大。说明在平原区利用距村庄和道路距离可以有效预测 SOM 的空间分布。

本研究发现，在不同地貌区，不同方法预测精度存在差异。在丘陵沟壑区，OK 法 SOM 存在显著

空间自相关性，可以有效预测 SOM 空间分布，但较 RF 和 RF + OK 法的预测结果不确定性较大，可能是因为 OK 法侧重于描述土壤属性的空间相关性，未考虑土壤与环境间的关系。RF 方法充分考虑了环境因子与土壤有机质含量的复杂非线性关系，预测精度较 OK 方法有所提升，这与 Zhang 等<sup>[18]</sup>和 César 等<sup>[19]</sup>研究结果一致。但 RF 方法对非采样点的预测精度较差。可能是因为土壤属性在空间上是连续的，具有一定的空间相关性，RF 算法未考虑到样点间的空间结构相关性，导致预测结果存在较大不确定性。RF + OK 方法不仅考虑了土壤属性与环境因子之间的复杂关系，而且考虑到各土壤采样点间的空间结构和相关性，因此该方法的误差较小，其



表 3 不同地貌区建模点与验证点各项误差

Table 3 Various errors of modeling points and verification points in different geomorphic areas

地貌 Geomorphic area	方法 Method	均方误差 MSE		均方根误差 RMSE		平均绝对误差 MAE	
		建模点	验证点	建模点	验证点	建模点	验证点
		Modeling point	Verification point	Modeling point	Verification point	Modeling point	Verification point
丘陵沟壑区 Hill and gully area	OK	6.24	6.38	2.48	2.53	1.84	1.91
	RF	1.25	6.50	1.12	2.55	0.83	1.96
	RF + OK	0.07	5.99	0.26	2.45	0.12	1.88
高原区 High plateau area	OK	9.21	8.45	3.04	2.91	2.4	2.32
	RF	1.52	7.02	1.23	2.65	0.94	2.11
	RF + OK	1.49	7.03	1.22	2.65	0.92	2.11
平原区 Plateau area	OK	16.32	15.32	1.04	3.90	3.10	3.03
	RF	5.99	14.67	2.45	3.83	1.79	3.00
	RF + OK	5.93	14.74	2.43	3.84	1.78	3.00

注 (Note): OK—普通克里格方法 Ordinary Kriging method; RF—随机森林方法 Random forest method; RF + OK—随机森林与普通克里格相结合方法 Random forest and ordinary Kriging method.

表 4 不同地貌区农田有机质实测点与预测点相关系数 ( $r$ )

Table 4 Correlation coefficients between measured and predicted points of farmland SOM in different geomorphic areas

地貌区 Geomorphic area	OK	RF	RF + OK
丘陵沟壑区 Hill and gully area	0.436	0.978	0.996
高原区 High plateau area	0.031	0.944	0.934
平原区 Plateau area	0.271	0.797	0.794

注 (Note): OK—普通克里格方法 Ordinary Kriging method; RF—随机森林方法 Random forest method; RF + OK—随机森林与普通克里格相结合方法 Random forest and ordinary Kriging method.

预测结果更符合研究区 SOM 空间实际分布特征。在高原区, SOM 不存在空间自相关性, 且半方差函数模型拟合度为零, 不满足地统计学插值条件, 其无偏估计的可靠性较低<sup>[31]</sup>。RF + OK 法和 RF 法的各项误差无明显差异, 但 RF 对非采样点的预测效果更好, 预测结果的空间结构和空间相关性更符合研究区实际特征。因此, 在高原区, 采用 RF 方法对 SOM 空间预测效果更好。在平原区, OK 方法预测结果的不确定性较大, RF 与 RF + OK 法的各项误差和  $r$  值无明显差异, 但从空间结构和空间相关性来看, RF 方法较其他两种方法, 其预测结果的空间结构和空间相关性更接近实测值, 说明 RF 方法优于其他两种方法。平原区较其他两个地区, SOM 含量异

质性较大, 3 种方法预测结果各项误差及空间结构差异较大, 因此平原地区 SOM 空间预测结果的不确定性要高于丘陵沟壑区和高原区。这一方面可能与人为干扰(耕作管理、耕作方式)有关<sup>[32-33]</sup>, 平原区人均耕地面积较小, 农户管理方式多样<sup>[33]</sup>; 另一方面可能是平原区受地形因素限制较小, 农业生产和机械化水平高, 农田投入大, 导致该区 SOM 空间变异较丘陵区和高原区大<sup>[34]</sup>。

综上所述, 本研究针对 SOM 在空间预测中存在的问题, 利用 OK、RF、OK + RF 3 种空间预测方法, 结合多源环境影响因子, 从地形地貌的角度研究了不同空间预测方法在区域上的适用性和不确定性。但在黄土高原地区, 土壤侵蚀和人为因素对 SOM 的影响可能更大, 从侵蚀和人为干扰强度角度探讨区域 SOM 空间分布具有重要意义。其次, 布点方法和数量也是影响土壤属性空间分布的重要因素, 显著影响预测结果, 未来需加强布点方法和空间预测方法关系研究, 进而实现对土壤属性的精准预测。另外, 土壤属性与环境因子关系复杂, 如何综合考虑二者间的复杂关系、土壤属性本身的空间相关性及土壤属性间的交互关系来更精准的对土壤属性进行空间预测, 未来仍需深入研究。

## 4 结论

1) 黄土高原不同地貌区 SOM 变异性表现为平原区 > 高原区 > 丘陵沟壑区, 均呈中等程度变异。研

表 5 不同地貌区农田土壤有机质预测值半方差模型及参数

Table 5 Semi-variance model and parameters of the predicted value of farmland SOM in different geomorphic areas

地貌 Geomorphic area	方法 Method	模型 Model	块金值 Nugget	基台值 Partial	块金效应 Nugget effect (%)	变程 (m) Range	$R^2$	残差 Residual
丘陵沟壑区 Hill and gully area	OK	Gaussian	0.87	2.27	38.04	5144	0.91	0.15
	RF	Exponential	0.57	2.78	20.50	3180	0.61	0.26
	RF + OK	Exponential	0.59	6.59	8.95	2250	0.62	0.78
高原区 High plateau area	OK	Spherical	0.16	2.70	5.93	1940	0.14	0.04
	RF	Exponential	0.25	2.32	10.83	1230	0.07	0.01
	RF + OK	Exponential	0.27	2.51	10.55	1740	0.36	0.03
平原区 Plain area	OK	Exponential	0.62	4.91	0.13	1620	0.80	0.26
	RF	Exponential	0.71	5.72	12.41	1230	0.43	0.75
	RF + OK	Exponential	0.80	6.40	0.12	1380	0.61	0.68

注 (Note): OK—普通克里格方法 Ordinary Kriging method; RF—随机森林方法 Random forest method; RF + OK—随机森林与普通克里格相结合方法 Random forest and ordinary Kriging method.

表 6 不同地貌区农田有机质空间自相关分析

Table 6 Spatial autocorrelation analysis of farmland SOM in different geomorphic areas

指标 Index	丘陵沟壑区 Hill and gully area				高原区 High plateau area				平原区 Plain area			
	实测值 Measured value	预测值 Predictive value			实测值 Measured value	预测值 Predictive value			实测值 Measured value	预测值 Predictive value		
		OK	RF	RF + OK		OK	RF	RF + OK		OK	RF	RF + OK
Moran's $I$	0.28	0.77	0.35	0.31	0.00	0.22	0.02	0.09	0.14	0.34	0.26	0.31
$Z[I]$	22.44	62.89	28.16	24.99	0.30	18.52	1.39	7.52	11.80	28.04	22.01	26.86

注 (Note): OK—普通克里格方法 Ordinary Kriging method; RF—随机森林方法 Random forest method; RF + OK—随机森林与普通克里格相结合方法 Random forest and ordinary Kriging method.

究区 SOM 受随机性和结构性因子共同作用, 且受结构性因素影响更大。丘陵区沟壑区和平原区 SOM 存在显著空间自相关性, 而高原区 SOM 不存在空间自相关性。

2) 不同地貌区, 参与土壤景观模型构建的环境因子的相对重要性不同。丘陵沟壑区、高原区、平原区影响 SOM 空间分布的主控因素分别为温度、海拔和降水量。

3) 不同地貌区, 环境要素、空间变异和空间相关性不同, 同一预测方法的预测精度存在差异, 平原区较丘陵沟壑区和高原区, 其空间预测结果的不确定性更大。在同一地貌区, 不同预测方法的预测精度存在差异, 丘陵沟壑区采用 RF + OK 法预测 SOM 空间分布效果较好, 高原区和平原区则使用 RF 效果较好。当区域 SOM 存在显著空间相关性, 且半方差函数的拟合度较高, 残差较小时, 采用 RF + OK 方法可以显著地提高模型的预测精度。

#### 参 考 文 献:

- [1] 孙波, 解宪丽. 全球变化下土壤功能演变的响应和反馈[J]. 地球科学进展, 2005, 20(8): 903-909.  
Sun B, Xie X L. Response and feedback of soil function evolvement to global change[J]. Advances in Earth Science, 2005, 20(8): 903-909.
- [2] Jones C, Mcconnel L C, Coleman K, et al. Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in soil[J]. Global Change Biology, 2004, 10(1): 154-166.
- [3] 赵明松, 张甘霖, 王德彩, 等. 徐淮黄泛平原土壤有机质空间变异特征及主控因素分析[J]. 土壤学报, 2013, 50(1): 1-11.  
Zhao M S, Zhang G L, Wang D C, et al. Spatial variability of soil organic matter and its dominating factors in Xu-Huai Alluvial plain[J]. Acta Pedologica Sinica, 2013, 50(1): 1-11.
- [4] 黄思华, 濮励杰, 解雪峰, 等. 面向数字土壤制图的土壤采样设计研究进展与展望[J]. 土壤学报, 2020, 57(2): 259-272.  
Huang S H, Pu L J, Xie X F, et al. Review and outlook of designing of soil sampling for digital soil mapping[J]. Acta Pedologica Sinica, 2020, 57(2): 259-272.

- [5] 杨奇勇, 杨劲松, 刘广明. 土壤速效养分空间变异的尺度效应[J]. 应用生态学报, 2011, 22(2): 431-436.  
Yang Q Y, Yang J S, Liu G M. Scale-dependency of spatial variability of soil available nutrients[J]. Chinese Journal of Applied Ecology, 2011, 22(2): 431-436.
- [6] 张素梅, 王宗明, 张柏, 等. 利用地形和遥感数据预测土壤养分空间分布[J]. 农业工程学报, 2010, 26(5): 188-194.  
Zhang S M, Wang Z M, Zhang B, *et al.* Prediction of spatial distribution of soil nutrients using terrain attributes and remote sensing data[J]. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(5): 188-194.
- [7] 赵明松, 刘斌寅, 卢宏亮, 等. 基于地理加权回归的地形平缓区土壤有机质空间建模[J]. 农业工程学报, 2019, 35(20): 102-110.  
Zhao M S, Liu B Y, Lu H L, *et al.* Spatial modeling of soil organic matter over low relief areas based on geographically weighted regression[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(20): 102-110.
- [8] Song X D, Brus D J, Liu F, *et al.* Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China[J]. Geoderma, 2016, 261(1): 11-22.
- [9] Yang S H, Liu F, Song X D, *et al.* Mapping topsoil electrical conductivity by a mixed geographically weighted regression kriging: A case study in the Heihe River Basin, northwest China[J]. Ecological Indicators, 2019, 102: 252-264.
- [10] Li Q Q, Yue T X, Wang C Q, *et al.* Spatially distributed modeling of soil organic matter across China: An application of artificial neural network approach[J]. Catena, 2013, 104(1): 210-218.
- [11] Phachomphon K, Dlamini P, Chaplot V. Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables[J]. Geoderma, 2009, 155(3-4): 372-380.
- [12] Shi W J, Liu J Y, Du Z P, *et al.* Surface modelling of soil properties based on land use information[J]. Geoderma, 2011, 162(3-4): 347-357.
- [13] 李启权, 王昌全, 岳天祥, 等. 基于定性和定量辅助变量的土壤有机质空间分布预测: 以四川三台县为例[J]. 地理科学进展, 2014, 33(2): 259-269.  
Li Q Q, Wang C Q, Yue T X, *et al.* Prediction of distribution of soil organic matter based on qualitative and quantitative auxiliary variables: A case study in Santai County in Sichuan Province[J]. Progress in Geography, 2014, 33(2): 259-269.
- [14] Mcbratney A B, Mendonça S M L, Minasny B. On digital soil mapping[J]. Geoderma, 2003, 117(1): 53-52.
- [15] 王茵茵, 齐雁冰, 陈洋, 等. 基于多分辨率遥感数据与随机森林算法的土壤有机质预测研究[J]. 土壤学报, 2016, 53(2): 342-354.  
Wang Y Y, Qi Y B, Chen Y, *et al.* Prediction of soil organic matter based on multi-resolution remote sensing data and random forest algorithm[J]. Acta Pedologica Sinica, 2016, 53(2): 342-354.
- [16] Wadoux J C, Brus D J, Heuvelink G B M. Sampling design optimization for soil mapping with random forest[J]. Geoderma, 2019, 355(1): 113-123.
- [17] Fabian E G, Jasmin B, Volkmar M, *et al.* From geological to soil parent material maps-A random forest-supported analysis of geological map units and topography to support soil survey in South Tyrol[J]. Geoderma, 2019, 354(15): 1-20.
- [18] Zhang H, Wu P B, Yin A J, *et al.* Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model[J]. Science of the Total Environment, 2017, 592(15): 704-713.
- [19] César D S C, Waldir D C J, Silvio B 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.
- [20] 张志霞. 黄土高原农田土壤有机碳空间变异性[D]. 杨凌: 中国科学院大学硕士学位论文, 2014.  
Zhang Z X. Spatial variation of soil organic carbon of croplands on the Loess Plateau[D]. Yangling: MS Thesis of University of Chinese Academy of Sciences, 2014.
- [21] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [22] 张志霞, 许明祥, 刘京, 等. 黄土高原不同地貌区土壤有机碳空间变异与合理取样数研究[J]. 自然资源学报, 2014, 29(12): 2103-2113.  
Zhang Z X, Xu M X, Liu J, *et al.* Spatial variation and reasonable sampling number of soil organic carbon under different geomorphic types on the Loess Plateau[J]. Journal of Natural Resources, 2014, 29(12): 2103-2113.
- [23] 熊昌盛, 栾乔林, 韦仕川. 基于空间自相关的耕地质量分布格局[J]. 地域研究与开发, 2016, 35(5): 128-132, 148.  
Xiong C S, Luan Q L, Wei S C. Spatial pattern disparity of arable land quality based on spatial autocorrelation analysis[J]. Areal Research and Development, 2016, 35(5): 128-132, 148.
- [24] 张志霞, 许明祥, 吴永斌, 等. 黄土丘陵区县域农田土壤有机碳空间变异性: 以甘肃庄浪县为例[J]. 植物营养与肥料学报, 2013, 19(4): 955-963.  
Zhang Z X, Xu M X, Wu Y B, *et al.* Spatial variation of soil organic carbon of croplands in hilly Loess Plateau region of China: A case study in Zhuanglang County, Gansu Province[J]. Journal of Plant Nutrition and Fertilizers, 2013, 19(4): 955-963.
- [25] Meersmans J, De R F, Canters F, *et al.* A multiple regression approach to assess the spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium)[J]. Geoderma, 2008, 143(1-2): 1-13.
- [26] Gustavo S, Michael I B, Tomas D, *et al.* Variation in soil carbon stocks and their determinants across a precipitation gradient in West Africa[J]. Global Change Biology, 2012, 18(5): 1670-1683.
- [27] 王云强, 张兴昌, 李顺姬, 等. 小流域土壤矿质氮与地形因子的关系及其空间变异性研究[J]. 环境科学, 2007, 28(7): 1567-1572.  
Wang Y Q, Zhang X C, Li S J, *et al.* Spatial variability and the relationships of soil mineral N and topographic factors in a small watershed[J]. Environmental Science, 2007, 28(7): 1567-1572.
- [28] 顾成军, 史学正, 于东升, 等. 省域土壤有机碳空间分布的主控因子: 土壤类型与土地利用比较[J]. 土壤学报, 2013, 50(3): 425-432.  
Gu C J, Shi X Z, Yu D S, *et al.* Main factor controlling SOC spatial distribution at the province scale as affected by soil type and land use[J]. Acta Pedologica Sinica, 2013, 50(3): 425-432.
- [29] Lemke R L, Vandenbygaart A J, Campbell C A, *et al.* Crop residue removal and fertilizer N: Effects on soil organic carbon in a long-term crop rotation experiment on a Udic Boroll[J]. Agriculture, Ecosystems Environment, 2010, 135(1): 42-51.

- [30] Chevallier T, Cournac L, Hamdi S, *et al.* Temperature dependence of CO<sub>2</sub> emissions rates and isotopic signature from a calcareous soil[J]. *Journal of Arid Environments*, 2016, 135: 132–139.
- [31] 陈锦赋. 基于地统计方法的二氧化碳浓度空间插值研究[J]. *现代计算机 (专业版)*, 2012, (16): 22–25.  
Chen J F. Research on the spatial interpolation of concentrations of carbon dioxide based on the geostatistics[J]. *Modern Computer*, 2012, (16): 22–25.
- [32] 张晓伟, 许明祥, 师晨迪, 等. 半干旱区县域农田土壤有机碳固存速率及其影响因素: 以甘肃庄浪县为例[J]. *植物营养与肥料学报*, 2012, 18(5): 1086–1095.  
Zhang X W, Xu M X, Shi C D, *et al.* Soil organic carbon sequestration rate and its influencing factors in farmlands of semi-arid regions: A case study in Zhuanglang County, Gansu Province[J]. *Journal of Plant Nutrition and Fertilizers*, 2012, 18(5): 1086–1095.
- [33] 张圣民, 许明祥, 张志霞, 等. 黄土高原不同地貌类型区农田土壤有机碳采样布点方法研究[J]. *自然资源学报*, 2010, 25(3): 443–453.  
Zhang S M, Xu M X, Zhang Z X, *et al.* Methods of sampling soil organic carbon in farmlands with different landform types on the Loess Plateau[J]. *Journal of Natural Resources*, 2010, 25(3): 443–453.
- [34] 闫豫疆. 县域级平原与丘陵农田土壤养分空间差异性综合研究[D]. 武汉: 华中农业大学硕士学位论文, 2012.  
Yan Y J. The synthetic study on spatial variation of agricultural soil nutrient on Plain and Hill of country level[D]. Wuhan: MS Thesis of Huazhong Agricultural University, 2012.