

考虑空间分异性和环境变量交互作用的分区回归克里格法在土壤有机碳制图中的应用

王晓密¹, 刘久洪¹, 卜岱昀^{2*}

(1. 湖南师范大学地理科学学院, 湖南长沙 410081; 2. 中国地质调查局海口海洋地质调查中心, 海南海口 571127)

摘要: 【目的】绘制高精度土壤有机碳 (SOC) 空间分布图是精准施肥的基础, 对农业可持续发展具有重要意义。【方法】本文以法国北部为研究区, 提出一种考虑空间分异性和环境变量交互作用的分区回归克里格 (ZRK) 方法, 用于绘制高精度 SOC 含量空间分布图。全局普通克里格 (OK)、协同克里格 (COK)、回归克里格 (RK)、随机森林 (RF) 及基于单一环境变量和地理探测器分区的 OK、COK 和 RK 作为对比模型。【结果】①研究区 SOC 含量受多种环境变量的交互驱动作用, 且不同空间匀质子区域所受驱动作用不同。②空间分层聚类方法能有效挖掘 SOC 含量的空间匀质子区域, 基于空间分层聚类分区的 ZOK、ZCOK 和 ZRK 方法相较于全局 OK、COK 和 RK 模型, 其 R^2 提升了 32%~36%。③基于单一环境变量及地理探测器分区的克里格方法在本研究中表现一般, 未有效地提升 SOC 预测精度。④RF 在本研究中表现不佳, 具有较低的 R^2 和较高的均方根差和平均绝对误差; ZRK 在所有模型中表现最优, 具有最高的 R^2 及最低的均方根误差和平均绝对误差。【结论】考虑空间分异性和环境变量交互作用的 ZRK 法能有效挖掘 SOC 含量的空间匀质子区域, 提高 SOC 制图精度, 可为空间分异区域数字土壤制图提供新的模型参考。

关键词: 土壤有机碳; 空间分层聚类; 环境变量交互作用; 分区回归克里格; 数字土壤制图

中图分类号: S159.9 **文献标识码:** A **文章编号:** 0564-3945(2025)01-0047-12

DOI: 10.19336/j.cnki.trtb.2023120903

王晓密, 刘久洪, 卜岱昀. 考虑空间分异性和环境变量交互作用的分区回归克里格法在土壤有机碳制图中的应用 [J]. 土壤通报, 2025, 56(1): 47-58

WANG Xiao-mi, LIU Jiu-hong, BO Dai-yun. Application of Zoning Regression Kriging in Soil Organic Carbon Mapping Considered Spatial Heterogeneity and Interactive Environmental Variables[J]. Chinese Journal of Soil Science, 2025, 56(1): 47-58

【研究意义】土壤有机碳 (Soil organic carbon, SOC) 是极其宝贵的自然资源, 是评价土壤质量的重要指标^[1]。SOC 含量的空间分布对精准施肥和可持续土地管理具有重要意义^[2-4]。然而, SOC 含量的空间分布受多种因素的综合影响, 不可避免地存在强烈的空间异质性, 阻碍了高精度 SOC 制图的发展。因此, 考虑空间分异性对 SOC 含量空间分布的影响, 对绘制空间分异区域高精度 SOC 空间分布图具有重要意义。

【前人研究进展】目前, 基于土壤—景观模型的数字土壤制图 (Digital soil mapping, DSM) 方法主要通过引入环境变量的方式以提升预测模型对 SOC 空间变异的解释能力, 如加入气候^[5-6]、地形^[7]、

植被^[8]、土壤及土地利用^[9-10] 等因子。大量学者的研究表明, 加入有效的环境变量可获得更高的预测精度^[5, 8, 11-12], 如结合环境变量的回归克里格法 (Regression kriging, RK) 可获得相较于普通克里格 (Ordinary kriging, OK) 更高的预测精度^[13-14]。然而, 在环境复杂多样的空间分异区域, SOC 空间分布存在强烈的空间异质性, 使用全局模型难以准确预测 SOC 含量的空间分布。因此, 一些学者基于环境变量将研究区分为几个子区域, 并在子区域单独构建模型以提升预测精度^[9-10]。现有基于环境变量进行分区的方法可大致分为三类: ①基于单一环境变量进行分区, 如基于土地利用类型^[10]、地形因子^[15] 等; ②基于多个环境变量的叠置区域, 如土壤和坡度图

收稿日期: 2023-12-21; 修订日期: 2024-05-10

基金项目: 国家自然科学基金 (42101474)

作者简介: 王晓密 (1990-), 女, 湖南邵阳人, 博士, 副教授, 主要从事时空大数据挖掘、数字土壤制图、滑坡预测与风险评估等方面研究。E-mail: xiaomiw@hunnu.edu.cn

*通讯作者: E-mail: 17264499286@163.com

的叠置^[16]等；③基于地理探测器^[17]进行分区，如土地利用类型和土壤类型的交互模式^[9]及多种环境变量交互的空间异质模式^[18]等。

【本研究切入点】现有研究主要通过建立 SOC 与环境变量的线性关系构建预测模型，而忽略了环境变量交互效应对 SOC 含量空间分布的影响。且影响 SOC 含量空间分布的环境条件通常较为复杂，依赖环境变量进行分区以挖掘 SOC 空间异质性的方法存在一定局限性。**【拟解决的问题】**因此，本文以法国北部为研究区，提出考虑空间分异性和环境变量交互作用的分区回归克里格（Zoning regression kriging, ZRK）方法，拟利用空间分层聚类方法区域自适应地挖掘 SOC 含量的空间匀质子区域，进而于各子区域内探究环境变量交互作用对 SOC 含量空间分布的影响，并绘制高精度 SOC 含量空间分布图。

1 材料与方法

1.1 研究区概况

研究区位于法国北部，东邻阿尔卑斯山脉，西邻大西洋，中西部以平坦的平原，盆地和低矮起伏的丘陵为主，东南部具有海拔较高的山地，整体海拔 0~1492 m。所处地区受温带海洋性气候影响，全年温和潮湿，年平均气温约为 11.5℃，年平均降水量为 586~1429 mm，不同局部区域的地形地貌和气候降水具有显著差异。研究区具有丰富的土壤类

型，按照世界土壤数据库分类参考标准，该区的土壤主要包括棕壤、冲积土、细粒土、白粒土及灰化土等。而土地利用类型主要由耕地、草地、林地、建设用地和其他用地（如沼泽、盐泽、荒地等）组成。研究区大部分区域地势相对平坦，土地肥沃，农业发达，具有集中连片的高标准农场，盛产小麦、大豆、玉米及各种水果蔬菜。开展研究区的 SOC 空间分布特征研究，揭示其空间分异模式对保障农业生产和调节气候变化具有重要意义。研究区及土壤样本点的空间分布图如图 1 所示。

1.2 数据来源

本文的土壤数据来源于 2009 年欧洲土地利用和覆盖面积框架调查（Land Use and Coverage Area Frame Survey, LUCAS）数据集^[19-21]，其中位于法国北部的共有 1645 个土壤样本（0~20 cm）。

本文的环境变量数据包括土地利用类型、土壤类型、数字高程模型（Digital elevation model, DEM）、坡度（Slope）、年均温度（Mean annual temperature, MAT）、年均降雨（Mean annual precipitation, MAP）、归一化植被指数（Normalized difference vegetation index, NDVI）和归一化湿度指数（Normalized difference moisture index, NDMI）。其中土地利用及土壤类型数据来源于欧洲土壤数据中心（<https://esdac.jrc.ec.europa.eu/>）；DEM 数据来源于美国国家航空航天局（NASA）的陆地过程分布

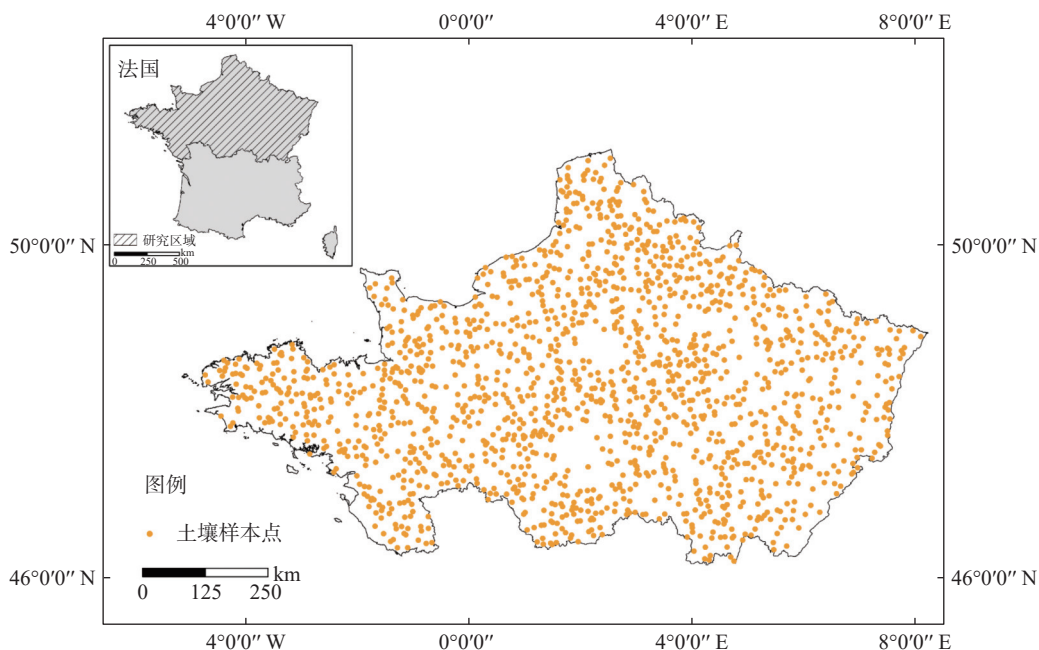


图 1 研究区及土壤样本点的空间分布

Fig.1 The spatial distribution of the study area and soil sample points

式数据档案中心 (<https://lpdaac.usgs.gov/data/>), 坡度由 DEM 数据处理获得; 年平均气温和年降水量由 WorldClim 数据库 (<https://worldclim.org/>) 获得; NDVI 和 NDMI 由 MODIS 数据的 MYD13Q1 产品 (<https://www.earthdata.nasa.gov/>) 获得, 时间为 2009 年 5 月, 与土壤样本点采集时间一致。所有环境变量均重采样至 250 m, 以确保相同的栅格大小。

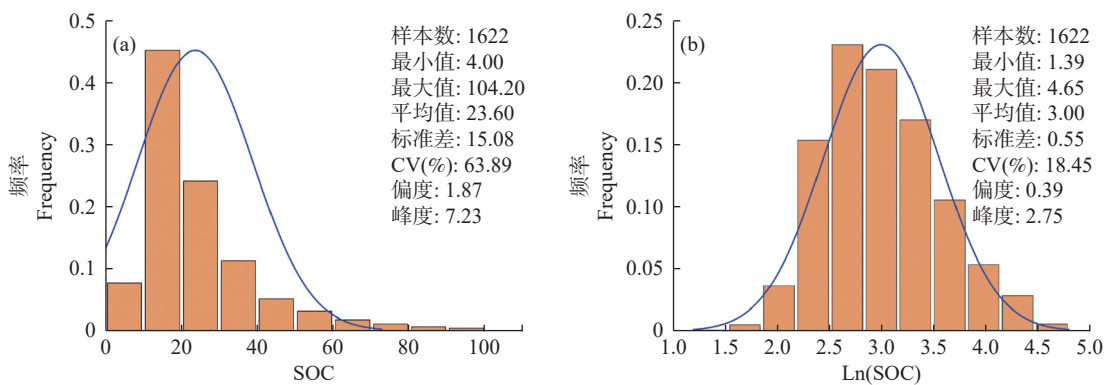
1.3 数据预处理

本文使用自然对数转换法将原始数据进行转换使其更符合正态分布, 并利用三倍标准差法去除异

常值^[22], 最终保留 1622 个土壤样本点用于后续分析 (图 2)。在构建预测模型时, 80% 的土壤样本点被随机选择作为校正集, 剩余 20% 的样本点作为验证集。本文使用 SPSS 21 用于数据转换及异常值处理, 使用 ArcGIS 10.4 中的 Subset 工具用于随机选择校正集和验证集。

1.4 研究方法

ZRK 方法是本文预测研究区 SOC 含量空间分布的主要研究方法, OK、COK、RK 及 RF 作为对比模型。



注: CV: Coefficient of variation, 变异系数。

图 2 土壤样本数据的频率分布直方图及统计数据(a)原始数据集, (b)对数转换后的数据集

Fig.2 Frequency distribution histogram of soil sample data and statistics (a) original dataset, (b) dataset after logarithmic transformation

1.4.1 OK 模型 普通克里格 (OK) 是一种广泛使用的地统计学方法, 该方法使用半变异函数描述样本间的空间变异性, 并假设其符合二阶平稳性特征, 可实现样本的最优无偏估计^[8]。通过计算理论半变异函数, 对预测点附近的观测值进行加权求和即可得到未知点的预测值。^[23]

1.4.2 COK 模型 协同克里格 (Cokriging, COK) 是在协同区域化变量理论上对 OK 方法进行的衍生, 可同时分析多个土壤属性之间的相互依赖性和地域性^[24]。COK 考虑了不同区域化变量之间的相关性, 可将预测目标从单一属性拓展到两个及以上的共区域化属性。

1.4.3 RK 模型 回归克里格 (RK) 是一种混合方法, 其在多元线性回归模型的基础上, 结合有效环境变量得到土壤属性的趋势项, 并将去除趋势项后的残差项用于 OK 插值^[25]。因此, 未知采样点 x_0 处的预测结果可以看作是回归趋势项和预测残差项之和^[9]。

1.4.4 RF 模型 随机森林 (Random Forest, RF) 是一种基于集成学习策略, 由多颗决策树组成的机器学习模型^[26]。因其在处理多元非线性数据方面的优

势, 一些研究表明 RF 在 DSM 中可获得较高的精度^[27-28]。本研究中, RF 的建模和预测在 R 语言中使用 “randomForest” 包实现。

1.4.5 ZRK 方法 分区回归克里格 (ZRK) 是本文提出的考虑空间分异性和环境变量交互作用的 SOC 含量制图方法。具体来说, ZRK 可分为两步: 首先, 利用空间分层聚类方法挖掘研究区 $Ln(SOC)$ 空间匀质子区域; 其次, 在每个匀质子区域内, 筛选显著的单一及交互环境变量, 并利用逐步回归方法建立多元线性回归方程, 进而构建 RK 模型, 绘制 SOC 含量空间分布图并进行模型验证。详细的步骤如下:

步骤 1: 挖掘 $Ln(SOC)$ 的空间匀质子区域

采用区域动态约束聚类 (Regionalization with dynamically constrained agglomerative clustering and partitioning, REDCAP) 方法^[29] 挖掘研究区 $Ln(SOC)$ 的空间匀质子区域。REDCAP 方法综合考虑了目标变量的属性相似性和空间邻接性, 可在不依赖环境变量的情况下, 将研究区划分成几个空间相邻的匀质子区域^[29-31]。空间匀质子区域挖掘大致可分为三步: 首先, 构建空间邻接矩阵; 其次, 设置 REDCAP 参

数, 构建最优参数评价指标; 最后, 执行 REDCAP, 根据评价指标, 获取最优 REDCAP 分区结果。详细的操作步骤见步骤 1.1 至 1.3。

步骤 1.1: 构建空间邻接矩阵

空间邻接矩阵是构建 REDCAP 方法的基本要素。目前主要存在两类空间邻接矩阵, 即基于距离的邻接矩阵和基于连续性的邻接矩阵。由于基于距离的邻接矩阵难以确定距离阈值且不能保证样本的空间连续性, 因此本文使用 Geoda 1.2 软件构建基于 Queen 一阶连续性的邻接矩阵并进行优化。

步骤 1.2: REDCAP 参数设置

REDCAP 方法的参数设置包括聚类方法、聚类数目、最小范围、距离函数及转换方法。其中聚类方法主要包括 5 种, 分别为“FirstOrder-SingleLinkage”, “FullOrder-WardLinkage”, “FullOrder-AverageLinkage”, “FullOrder-CompleteLinkage”和“FullOrder-SingleLinkage”。聚类数目取值范围设定为 1~7。在本研究中, 最优聚类方法和最佳聚类数分别由公式 1-5 中的指标 S 和 K 进行评估; 聚类数的最小范围设置为 10%, 以防止形成过小的簇; 距离函数采用曼哈顿距离; 并使用标准化 (Z) 法进行数据转换。

公式 1 中 S 值的取值范围为 0~1。S 值是衡量 SOC 含量空间分异性挖掘能力的指标, 其值越大, 表示簇内方差越小, 簇间方差越大, 可解释的 SOC

$$r = \{m, n | S_{(m,n)} = \max S_{(i,f)}, [i, f] \in R\}$$

$$R = \left\{ \begin{array}{l} (i, f) | K_{(i,f)} > 10, i = 1, 2, \dots, 7; f \in ['FirstOrder-SingleLinkage', \\ 'FullOrder-WardLinkage', 'FullOrder-AverageLinkage', \\ 'FullOrder-CompleteLinkage' \text{ and } 'FullOrder-SingleLinkage'] \end{array} \right\} \quad (5)$$

式中, $K_{(i,f)}$ 和 $S_{(i,f)}$ 分别表示当聚类数为 i , 聚类方法为 f 时的 K 和 S 值; 特别地, 当聚类数为 1 时, K 和 S 值等于 0。

步骤 1.3: 划分空间匀质子区域集

根据步骤 1.2 中的 REDCAP 参数设置, 执行 REDCAP, 得到研究区土壤样本点的最优聚类数和聚类方法, 利用第 n 种聚类方法将土壤样本点划分为 m 簇。同时, 将样本点转换为 Voronoi 图, 合并同一簇内样本点对应的 Voronoi 图, 最终将研究区分成 m 个空间连续, 且属性相近的空间匀质子区域。

步骤 2: 构建分区回归模型并获取 SOC 含量空间分布图

为研究环境变量对 SOC 含量的交互驱动作用,

含量的空间分异性程度越高。S 值与聚类数呈正相关, 即聚类数越多, S 值越大。S 增长率表示聚类数从 $n-1$ 增加到 n 时 S 值的增幅, 其值越大, 表示聚类效果越好。但随着聚类数的不断增加, S 值达到一定程度后, 其增长率将不再显著。通过设置 K 值 (公式 4) 可量化 S 值的增长率。K 值的取值范围为 $0 \sim 90^\circ$, 如果聚类数从 $n-1$ 增加到 n 时, K 值小于 10° , 则认为聚类数从 $n-1$ 增加到 n 时 S 值的增长不再显著。因此, 如果一个 REDCAP 聚类结果的参数满足公式 5 中的规则 r , 那么公式 5 中的聚类数 m 和聚类方法 n 将被选为聚类数和聚类方法的最优参数。

$$S = \frac{BSS}{BSS + WSS} \quad (1)$$

$$BSS = \sum_{i=1}^n q_{C_i} (\bar{x}_{C_i} - \bar{x}_e)^2 \quad (2)$$

$$WSS = \sum_{i=1}^n \sum_{j=1}^{q_{C_i}} (x_j - \bar{x}_{C_i})^2 \quad (3)$$

式中, BSS 表示簇间方差, WSS 表示簇内方差, n 是聚类数, q_{C_i} 是聚类 C_i 中土壤样本的数量, \bar{x}_{C_i} 是聚类 C_i 的中心, \bar{x}_e 是所有样本点的中心, x_j 是聚类 C_i 中土壤样本点的土壤属性值。

$$K_{(i,f)} = \arctan \frac{(S_{(i,f)} - S_{(i-1,f)}) / \max S_{(i,f)}}{(i - (i-1)) / (n-1)} \quad (4)$$

于各匀质子区域内, 分别构建 $Ln(SOC)$ 与单一及交互环境变量的多元线性关系。在每个子区域选择合理的校正集后, 使用多重共线性检验去除具有共线性的变量, 并利用逐步回归的方式建立多元线性回归方程, 构建对应的 RK 模型, 计算 RK 模型精度。如在某一子区域内未保留环境变量, 则说明该区域的环境变量未能很好的解释 SOC 含量的空间分异性, 因此, 在该子区域不建立回归方程, 并使用 OK 模型进行土壤制图。在每个匀质子区域内重复此操作并将获得的子区域 SOC 预测图进行镶嵌以获得最终的 SOC 分布图。由于采用分区插值的策略, 不同的子区域间易存在边缘突变问题。因此, 本研究在各子区域边界建立 10 km 缓冲区, 设定为边缘区域,

并采用均值平滑滤波 (40×40 窗口) 法对边缘区域进行平滑处理, 以缓解边缘突变问题。

1.5 精度评价

本文使用验证集的实测数据和预测数据计算模型的预测精度, 使用均方根误差 (RMSE)、平均绝对误差 (MAE) 及模型决定系数 (R^2) 进行模型的精度评价 (公式 6~8)。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (8)$$

式中, n 表示样本数量, O_i 表示样本 i 的实测值, P_i 表示样本 i 的预测值, \bar{O} 表示样本数据实测值的

均值。

2 结果与分析

2.1 空间匀质子区域

图 3 是一系列 REDCAP 聚类方法及对应的评价结果。图 3 (a) 表明所有聚类方法的 S 值总是随着聚类数的增加而增加, 当聚类数大于等于 4 时, “FullOrder-WardLinkage” 方法表现最优, 其 S 值高于其他方法。图 3 (b) 表明不同聚类方法随着聚类数的增加, K 值整体上呈现先升高再降低的趋势, 其中当聚类数大于等于 5 时, 聚类方法 “FullOrder-WardLinkage” 的 K 值总是低于 10° 。根据章节 1.4.5 公式 5 中 S 值和 K 值的判断标准, 本文选择 “FullOrder-WardLinkage” 作为最优的聚类方法, 并选择最优的聚类数为 4, 对应的最优聚类结果及匀质子区域如图 4 所示。

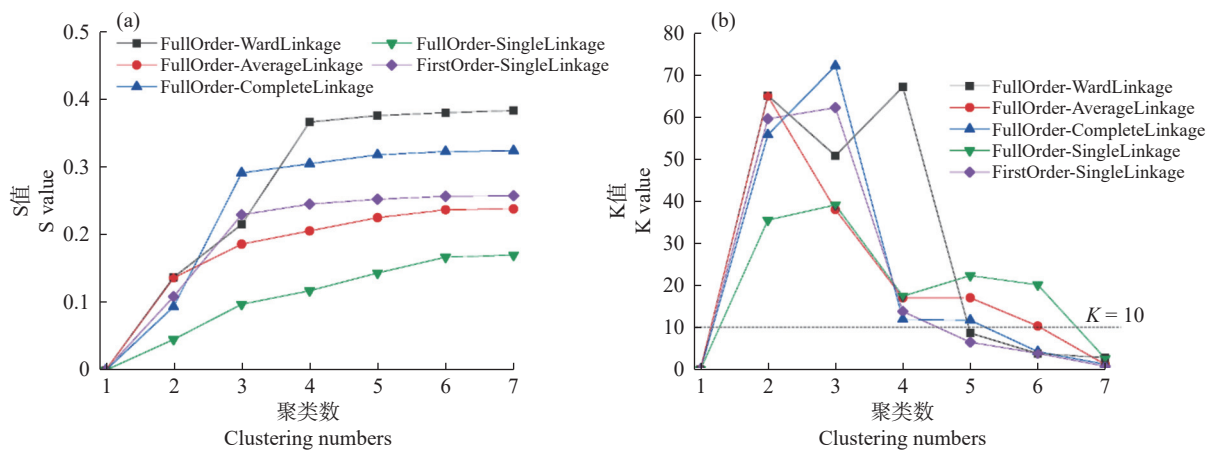


图 3 不同聚类方法在不同聚类数下的 S 和 K 值

Fig.3 The S and K values of different clustering methods under different clustering numbers.

表 1 是研究区及各空间匀质子区域内土壤样本点 $Ln(SOC)$ 的描述性统计数据。全局土壤样本的 $Ln(SOC)$ 均值为 2.998 g kg^{-1} , 标准差为 0.553, 变异系数 (Coefficient of variation, CV) 为 18.5%。通过 REDCAP 分区后, 所有子区域中土壤样本点 $Ln(SOC)$ 的变异系数均小于全局的变异系数, 说明基于 REDCAP 分区能在一定程度上降低土壤样本点的空间变异性。

2.2 ZRK 方法的预测结果

基于图 4 的空间匀质子区域集, 采用 ZRK 于各子区域内构建 $Ln(SOC)$ 与环境变量的驱动关系模型 (表 2)。结果表明环境变量对 SOC 含量存在交互驱动作用, 且不同的匀质子区域内, 环境变量对

SOC 含量的交互驱动作用不同。区域 1 主要受 NDMI 和高程的交互作用, 及温度和降雨的交互作用共同影响; 区域 2 主要受降雨和 NDMI 的交互影响; 区域 3 主要受降雨、温度及 NDVI 和 NDMI 的交互作用共同影响; 区域 4 则主要受 NDVI 和 DEM 的交互驱动作用。

基于表 2 的回归方程, 使用 ZRK 绘制研究区的 SOC 含量空间分布图 (图 5), 并验证其精度 (表 3)。由表 4 可知, ZRK 的 RMSE 为 0.406, R^2 为 0.488。由图 5 可知, 在同一空间匀质子区域内, SOC 含量具有相似性, 在不同的空间匀质子区域间, SOC 含量的空间分布具有差异性。其中区域 4 位于研究区东南部, 具有研究区最高的 SOC 含量; 区域 2 位于

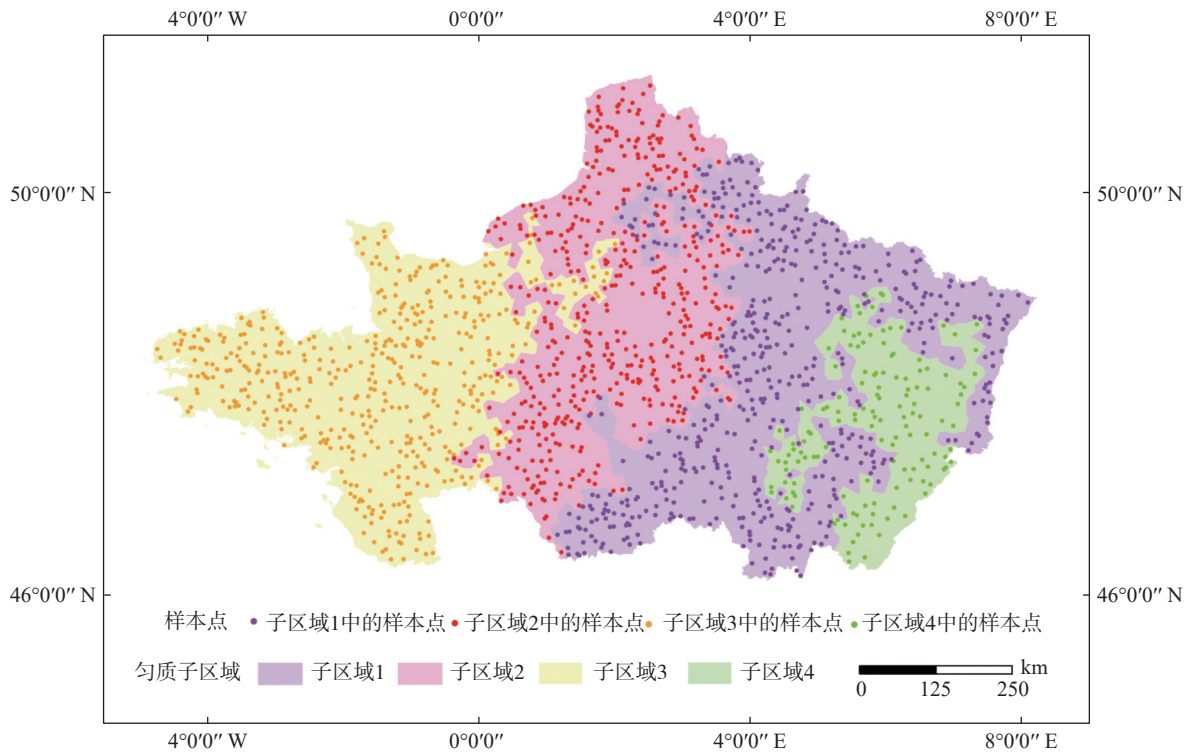


图 4 研究区 Ln(SOC) 含量的空间匀质子区域集

Fig.4 The spatial homogeneous zones of Ln(SOC) content in the study area

表 1 研究区及各空间匀质子区域内土壤样本点 Ln(SOC) 的描述性统计数据

Table 1 Descriptive statistics of Ln(SOC) in the study area and each spatial homogeneous zone.

样本集 Sample dataset	数量 Count	均值 (g kg ⁻¹) Mean	标准差 Standard deviation	变异系数 (%) Coefficient of variation	偏度 Skewness	峰度 Kurtosis
全局土壤样本点	1622	2.998	0.553	18.5	0.393	2.746
区域1中的样本点	539	3.084	0.477	15.5	0.274	3.187
区域2中的样本点	483	2.542	0.353	13.9	1.079	7.062
区域3中的样本点	463	3.176	0.493	15.5	0.359	2.987
区域4中的样本点	137	3.672	0.385	10.5	0.081	2.885

表 2 各匀质子区域内 Ln(SOC) 与环境变量的逐步回归方程

Table 2 The stepwise regression of Ln(SOC) and environmental variables in each homogeneous zone

模型 Model	逐步回归方程 Stepwise regression
R_Z1	$\text{Ln}(\text{SOC}) = 4.099 + 0.092 \times X4 \times X5 - 0.016 \times X1 \times X2$
R_Z2	$\text{Ln}(\text{SOC}) = 2.352 + 0.054 \times X1 \times X4$
R_Z3	$\text{Ln}(\text{SOC}) = -1.835 + 1.066 \times X1 - 0.182 \times X2 + 0.837 \times X3 \times X4$
R_Z4	$\text{Ln}(\text{SOC}) = 2.613 + 1.348 \times X3 \times X5$

注: R_Z1, R_Z2, R_Z3 及 R_Z4 分别表示基于空间匀质子区域 1, 2, 3 及 4 的校正集构建的回归模型; X1, X2, X3, X4 和 X5 分别代表 ln(MAP), MAT, NDVI, NDMI 及 ln(DEM)。

研究区中部, SOC 含量最低; 位于东部的区域 1 和位于西部的区域 3, 具有中等的 SOC 含量。

2.3 不同预测模型的精度比较

表 3 为不同方法 Ln(SOC) 的理论半变异函数及相关参数, 通过对比交叉验证精度, 选择最优变异函数拟合模型。块基比反映 Ln(SOC) 变异的自相关程度^[32], 由表 3 可知, OK、COK 和 RK 的块基

比分别为 44.5%, 39.0% 和 66.2%, 表明研究区的 Ln(SOC) 具有中等的空间自相关性, 通过 REDCAP 方法分区后, 各子区域的 Ln(SOC) 具有强烈或中等的空间自相关性。

为验证 ZRK 方法的有效性, 本文将基于全局样本点的 OK、COK、RK 及 RF 模型作为对比。由表 4 可知, ZRK 优于全局克里格及 RF 模型, 其模型验

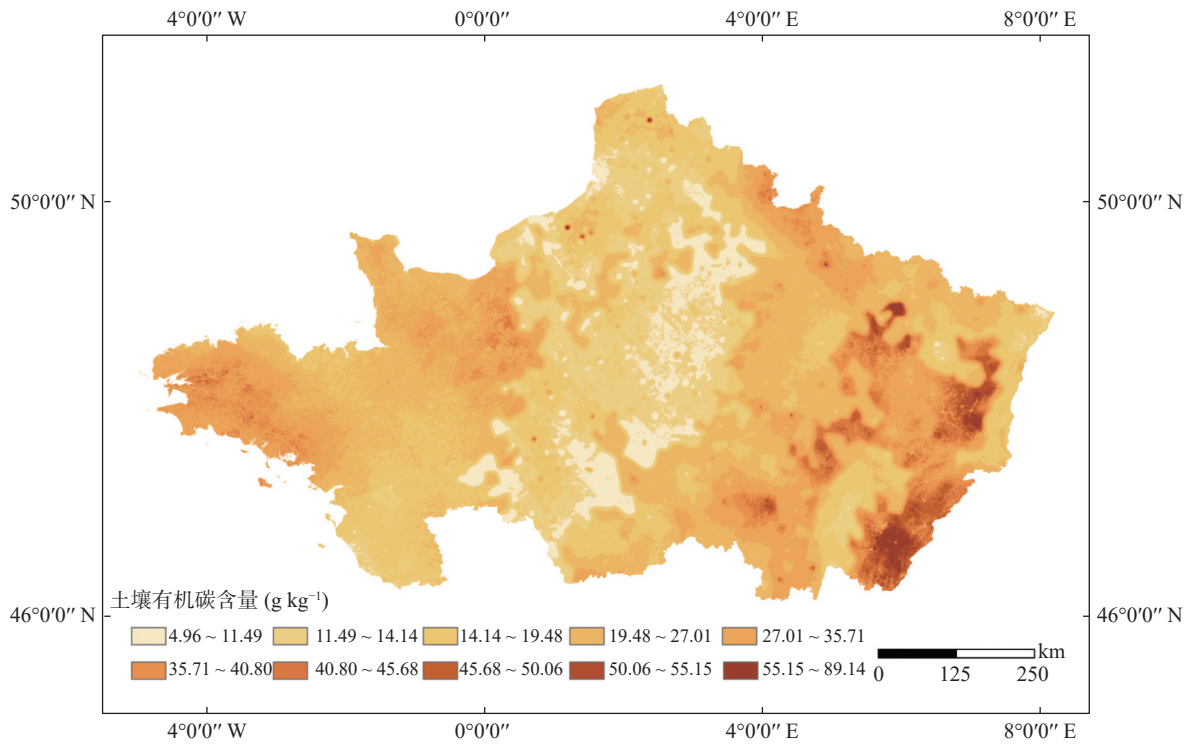


图 5 利用分区回归克里格绘制的 SOC 含量空间分布图
Fig.5 The spatial distribution map of SOC with ZRK

表 3 不同方法 $Ln(SOC)$ 的理论半变异函数及相关参数

Table 3 Theoretical semi-variogram model and related parameters of $Ln(SOC)$ of different methods

方法 Method	区域 Zone	模型 Model	块金值 Nugget (C_0)	偏基台值 Partial Sill (C)	块基比 Proportion ($C_0 / (C_0 + C)$, %)	变程 (m) Range
OK	研究区	Exponential	0.160	0.200	44.5	714232
COK	研究区	Exponential	0.174	0.272	39.0	1620004
RK	研究区	Exponential	0.155	0.079	66.2	214808
	区域1	Exponential	0	0.186	0	19902
ZOK	区域2	Exponential	0	0.100	0	12735
	区域3	Spherical	0.153	0.110	58.2	390218
	区域4	Exponential	0.101	0.063	61.6	126533
ZCOK	区域1	Exponential	0.118	0.118	50.0	160173
	区域2	Exponential	0	0.101	0	12735
	区域3	Exponential	0.156	0.147	51.5	910365
ZRK	区域4	Exponential	0.103	0.062	62.3	130700
	区域1	Stable	0	0.216	0	113137
	区域2	Exponential	0	0.101	0	12630
	区域3	Gaussian	0.149	0.055	73.1	174945
	区域4	Stable	0	0.143	0	71703

证精度 R^2 比全局克里格模型高 32% ~ 42%，比 RF 模型精度高 70%。

为验证 REDCAP 分区策略的有效性，本文基于 REDCAP 分区，进一步使用 OK 和 COK 模型（即 ZOK 和 ZCOK）进行 SOC 含量空间分布预测。由表 4 可知，结合 REDCAP 分区的 ZOK、ZCOK 和 ZRK 方法，其 R^2 比经典的 OK、COK 及 RK 模型高 32% ~ 36%，证明了 REDCAP 分区的有效性。除

表 4 不同模型的预测精度

Table 4 Accuracy of different models

方法 Method	均方根误差 RMSE	决定系数 R^2	平均绝对误差 MAE
OK	0.437	0.348	0.336
COK	0.438	0.345	0.338
RK	0.429	0.371	0.330
ZOK	0.415	0.464	0.318
ZCOK	0.413	0.470	0.316
ZRK	0.406	0.488	0.311
RF	0.464	0.287	0.365

此之外, 基于单一环境变量及地理探测器的分区策略同样作为对比。在本实验中, 基于连续环境变量的分区方法使用自然断点法将连续的环境变量进行分段, 并使用 R 语言中的 NbClust 包确定最佳的分段数。基于类型变量的分区分别按照不同的土地利用类型、土壤类型进行区域划分, 样本点过少的类统一为其他。基于地理探测器的方法可参考王劲峰^[17, 33]、Liu^[9] 等的相关研究, 并使用 R 语言中的 GD 包构建地理探测器。为与 ZRK 方法进行对比, 于上述基于环境变量的各子区域内, 同样采用经典

的克里格模型开展 SOC 制图, 模型预测精度如表 5 所示。结果表明, 基于单一环境变量及地理探测器分区的克里格模型未能有效提升模型精度, 而基于 REDCAP 分区的 ZOK、ZCOK 和 ZRK 方法相较于基于环境变量分区的方法拥有更高的预测精度, 其 R^2 提升了 20%~191%, 进一步证明了基于 REDCAP 分区的优势及有效性。在所有模型中, ZRK 的 R^2 均高于其余模型, RMSE、MAE 均低于其余模型, 精度最高。

表 5 基于单一环境变量及地理探测器分区的克里格模型精度

Table 5 Accuracy of kriging models based on single environmental variable and geographical detector partition

分区策略 Zoning strategy	子区域数 Number of zones	模型 Model	均方根误差 RMSE	决定系数 R^2	平均绝对误差 MAE
基于温度分区	3	OK	0.432	0.331	0.327
		COK	0.431	0.336	0.330
		RK	0.423	0.359	0.325
基于降雨分区	2	OK	0.383	0.376	0.302
		COK	0.382	0.381	0.298
		RK	0.371	0.415	0.292
基于高程分区	3	OK	0.455	0.332	0.352
		COK	0.450	0.346	0.355
		RK	0.445	0.362	0.348
基于坡度分区	3	OK	0.455	0.317	0.354
		COK	0.458	0.306	0.354
		RK	0.443	0.353	0.350
基于NDVI分区	4	OK	0.458	0.330	0.358
		COK	0.461	0.323	0.359
		RK	0.452	0.350	0.354
基于NDMI分区	4	OK	0.464	0.332	0.357
		COK	0.461	0.339	0.357
		RK	0.460	0.342	0.354
基于土地利用类型分区	4	OK	0.469	0.302	0.352
		COK	0.457	0.355	0.337
		RK	0.459	0.331	0.352
基于土壤类型分区	7	OK	0.454	0.290	0.345
		COK	0.459	0.276	0.354
		RK	0.491	0.171	0.369
基于地理探测器分区	4	OK	0.450	0.337	0.338
		COK	0.456	0.319	0.340
		RK	0.432	0.387	0.326

3 讨论

3.1 环境变量对 SOC 含量空间分布的影响

地形、气候及植被等环境变量是解释 SOC 含量空间分异的重要驱动因子^[34-36]。表 2 和图 5 表明研究区 SOC 含量受环境变量的交互驱动作用, 且不同的空间匀质子区域所受环境变量影响不同。区域 1

位于研究区东部, 为山地丘陵地貌。该区域的 SOC 含量与环境变量呈非线性关系。高程及 NDMI 的交互作用对区域 1 的 SOC 含量具有正向的影响, 而温度和降水的交互作用则对 SOC 含量呈现负向影响。较高的温度及充沛的降水将增强土壤微生物的活性, 加快 SOC 的分解^[37]。区域 2 位于研究区中部, 地势平坦, SOC 含量与 NDMI 及降雨的交互作用成正相

关。充沛的降水及较高的 NDMI 可防止有机土壤的氧化, 保存更多的 SOC。区域 3 位于研究区西部, 西邻大西洋, 处于温带海洋性气候, 其 SOC 含量受降水、温度、NDVI 及 NDMI 的共同影响。其中, 降水对 SOC 含量具有正向驱动作用; 而温度对 SOC 含量则具有负向的驱动作用; NDMI 和 NDVI 的交互作用与该区域的 SOC 含量呈正相关。区域 4 位于研究区东南部, 属于林地山区, 海拔较高, 在四个匀质子区域中其 SOC 含量最高, 主要受 NDVI 和 DEM 的正向交互驱动影响。总体而言, 研究区 SOC 含量的空间分布呈现显著的空间异质性, 受多种环境变量的交互影响, 且不同的匀质子区域的驱动作用不同。

3.2 不同预测模型的结果分析

图 6 是基于 OK、COK、RK、ZOK、ZCOK 及 ZRK 六种模型绘制的研究区 SOC 含量空间分布图。从研究区的空间结构和变化趋势来看, 六幅分布图

在全局上表现出相似的 SOC 含量空间分布特征, 如研究区的 SOC 含量整体表现为东南部和西部边缘较高, 而中部较低, 但在局部区域也表现出一定的差异性。在经典克里格模型中, RK 考虑了显著环境变量的影响, 其描绘的 SOC 含量空间分布图比 OK 及 COK 的分布图更详细。但 RK 仅使用一个回归模型考虑全局环境变量对 SOC 含量的影响, 忽略了环境变量对 SOC 含量影响的空间分异性。ZOK 和 ZCOK 两种方法考虑了空间异质性的影响, 但在考虑环境变量方面还存在不足。ZRK 方法同时考虑了 SOC 含量的空间异质性及环境变量交互作用的影响, 相较于其他模型, 其具有更详细的 SOC 含量空间分布图。基于多元交互环境变量的 RF 模型在本研究中并未获得理想的精度, 其 R^2 仅为 0.287, 低于全局 OK 精度。可能的原因在于本研究区范围较大, SOC 含量空间分布存在显著的空间异质性, 全局环境变量能解释的 SOC 空间变异性有限等。

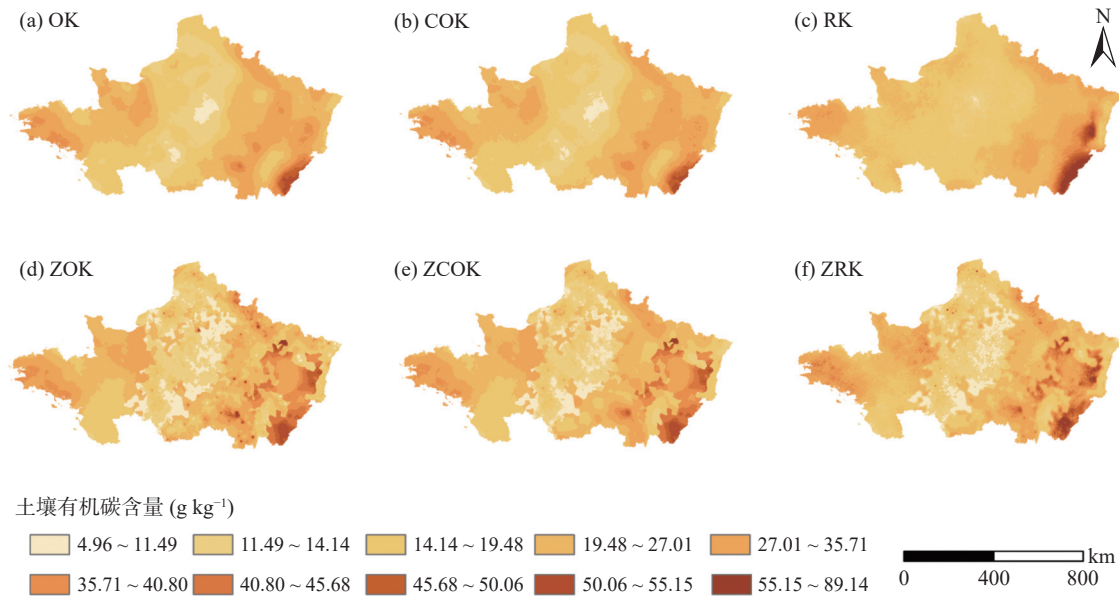


图 6 不同模型预测的 SOC 分布图

Fig.6 SOC distribution maps predicted by different models

相关研究表明 SOC 在不同的土地利用类型或土壤类型中具有显著差异^[9-10]。因此, 本文基于 REDCAP、土地利用类型及土壤类型进行单因素方差分析, 以分析不同分区方法对 SOC 空间变异的解释程度。由表 6 可知, 三种分区方法获取的子区域内校正集的 $\ln(\text{SOC})$ 含量均存在显著差异 ($P < 0.05$), 但相较于土地利用和土壤类型分区, REDCAP 可解释的 $\ln(\text{SOC})$ 空间变异程度更高, 达

36.7%, 而土地利用类型对 $\ln(\text{SOC})$ 空间变异的解释度为 4.8%, 土壤类型为 11.3%。由表 5 可知, 基于土地利用及土壤类型分区, 进行 OK、COK 和 RK 模型的构建, 其预测精度并不理想, 低于全局克里格精度。而基于其他单一环境变量的分区方法与经典克里格相结合, 同样未有效的提升预测精度, 相反, 如选择了不合理的环境变量进行分区还将导致 SOC 含量的预测精度降低, 主要原因在于这些环境

变量对 SOC 含量的驱动作用有限, 未能有效挖掘 SOC 含量的空间匀质区域。仅有基于降雨分区的经典克里格模型相较于全局克里格模型, 精度具有一定提升, 原因在于本研究区域范围较大, 相较于其他环境变量, 气候因子对 SOC 含量的驱动作用更显著。有学者表明基于地理探测器进行分区的克里格模型可获得较高的精度^[9]。本文研究结果表明, 基于

地理探测器的回归克里格模型精度略优于全局 RK 模型, 但精度提升有限, 地理探测器方法虽能获得显著环境变量的交互区域, 但在空间分异强烈的区域, 环境复杂多样, 基于地理探测器分区的方法过度依赖环境变量, 其挖掘的空间子区域可解释的 SOC 含量空间异质性同样有限。

表 6 基于 REDCAP、土地利用类型及土壤类型分区的 $\ln(SOC)$ 单因素方差分析

Table 6 One-way ANOVA of $\ln(SOC)$ based on different zones from REDCAP, land use types and soil types

分区方法 Zoning method	项目 Item	平方和 Sum of squares	自由度 Degree of freedom	均方误差 Mean square	S (%)	F	P
REDCAP	组间 Between groups	143.833	3	47.944	36.7	250.219	0.000**
	组内 Within groups	247.751	1293	0.192			
	总数 Sum	391.584	1296				
土地利用	组间 Between groups	18.751	3	6.250	4.8	21.569	0.000**
	组内 Within groups	374.970	1294	0.290			
	总数 Sum	393.720	1297				
土壤类型	组间 Between groups	45.437	6	7.573	11.3	27.409	0.000**
	组内 Within groups	356.695	1291	0.276			
	总数 Sum	402.132	1297				

注: S 表示分区方法组间方差与总方差的比值, 其值越大表示分区方法可解释的 SOC 空间变异程度越高; F 统计量用于衡量组间差异与组内差异之间的相对大小; **表示相关性达到 $P < 0.01$ 显著水平 (双尾)。

4 结论

本文提出了一种考虑空间分异性和环境变量交互作用的分区回归克里格 (ZRK) 方法, 并成功应用于绘制法国北部 SOC 含量空间分布图。全局克里格 (OK、COK、RK)、RF 和基于单一环境变量及地理探测器分区的克里格作为 ZRK 的对比模型。结果表明: ①研究区受多种环境变量的交互影响, 且不同空间匀质子区域所受的驱动作用不同。②REDCAP 可有效挖掘 SOC 含量的空间匀质子区域。基于 REDCAP 分区的 ZOK、ZCOK 和 ZRK 方法相较于全局克里格模型, 其 R^2 提升了 32% ~ 36%, 证明了基于 REDCAP 分区绘制 SOC 空间分布图的有效性。③基于单一环境变量及地理探测器分区的克里格方法在本研究中表现较差, 表明在空间分异区域, 基于环境变量进行分区难以准确挖掘 SOC 含量的空间匀质子区域。④RF 在本研究中表现不佳, 具有较低的 R^2 和较高的均方根差和平均绝对误差; ZRK 在所有模型中表现最优, 具有最高的 R^2 及最低的均方根差和平均绝对误差, 其 R^2 相较于全局克里格模型提升了 32% ~ 42%, 相较于 RF 提升了 70%。

参考文献:

- [1] Lal R. Soil carbon sequestration impacts on global climate change and food security[J]. *Science (New York, NY)*, 2004, 304(5677): 1623 – 1627.
- [2] Yang L, He X, Shen F, et al. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data[J]. *Soil and Tillage Research*, 2020, 196: 104465.
- [3] He X, Yang L, Li A, et al. Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images[J]. *Catena*, 2021, 205: 105442.
- [4] Karunaratne S B, Bishop T F A, Baldock J A, et al. Catchment scale mapping of measureable soil organic carbon fractions[J]. *Geoderma*, 2014, 219-220: 14 – 23.
- [5] 罗梅, 郭龙, 张海涛, 等. 基于环境变量的中国土壤有机碳空间分布特征[J]. *土壤学报*, 2020, 57(1): 48 – 59.
- [6] 庞龙辉, 刘峰, 赵霞, 等. 青海省表层土壤属性数字制图[J]. *土壤通报*, 2019, 50(03): 505 – 513.
- [7] 杨顺华, 张海涛, 郭龙, 等. 基于回归和地理加权回归 Kriging 的土壤有机质空间插值[J]. *应用生态学报*, 2015, 26(6): 1649 – 1656.
- [8] Wu Z, Liu Y, Han Y, et al. Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data[J]. *Science of the Total Environment*, 2021, 754: 142120.

- [9] Liu Y, Chen Y, Wu Z, et al. Geographical detector-based stratified regression kriging strategy for mapping soil organic carbon with high spatial heterogeneity[J]. *Catena*, 2021, 196: 104953.
- [10] 吴子豪, 刘艳芳, 陈奕云, 等. 综合土地利用及空间异质性的土壤有机碳空间插值模型 [J]. *应用生态学报*, 2018, 29(1): 238 – 246.
- [11] Wu Z, Wang B, Huang J, et al. Estimating soil organic carbon density in plains using landscape metric-based regression Kriging model[J]. *Soil and Tillage Research*, 2019, 195: 104381.
- [12] 刘焕军, 谢雅慧, 潘越, 等. 黑土区田块土壤有机质空间分异及分布研究 [J]. *土壤通报*, 2018, 49(06): 1445 – 1451.
- [13] García-Tomillo A, Mirás-Avalos J M, Dafonte-Dafonte J, et al. Estimating soil organic matter using interpolation methods with a electromagnetic induction sensor and topographic parameters: a case study in a humid region[J]. *Precision Agriculture*, 2016, 18(5): 882 – 897.
- [14] Shen Q, Wang Y, Wang X, et al. Comparing interpolation methods to predict soil total phosphorus in the Mollisol area of Northeast China[J]. *Catena*, 2019, 174: 59 – 72.
- [15] 张婧, 毕如田, 丁皓希, 等. 基于地形分类的黄土丘陵区土壤有机质插值方法 [J]. *江苏农业科学*, 2021, 49(8): 182 – 189.
- [16] Baldwin D, Naithani K J, Lin H. Combined soil-terrain stratification for characterizing catchment-scale soil moisture variation[J]. *Geoderma*, 2017, 285: 260 – 269.
- [17] 王劲峰, 徐成东. 地理探测器: 原理与展望 [J]. *地理学报*, 2017, 72(1): 116 – 134.
- [18] Du Z, Gao B, Ou C, et al. A Quantitative Analysis of Factors Influencing Organic Matter Concentration in the Topsoil of Black Soil in Northeast China Based on Spatial Heterogeneous Patterns[J]. *ISPRS International Journal of Geo-Information*, 2021, 10(5): 348.
- [19] Orgiazzi A, Ballabio C, Panagos P, et al. LUCAS Soil, the largest expandable soil dataset for Europe: a review[J]. *European Journal of Soil Science*, 2017, 69(1): 140 – 153.
- [20] Toth G, Jones A, Montanarella L. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union[J]. *Environ Monit Assess*, 2013, 185(9): 7409 – 7425.
- [21] Panagos P, Van Liedekerke M, Borrelli P, et al. European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies[J]. *European Journal of Soil Science*, 2022, 73(6).
- [22] Chen Y, Feng X, Fu B. An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018[J]. *Earth System Science Data*, 2021, 13(1): 1 – 31.
- [23] Zhang S, Xiao Z, Zhang H, et al. Key factors determining soil organic carbon changes after freeze-thaw cycles in a watershed located in northeast China[J]. *Science of the Total Environment*, 2022, 828: 154525.
- [24] 郭龙, 张海涛, 陈家赢, 等. 基于协同克里格插值和地理加权回归模型的土壤属性空间预测比较 [J]. *土壤学报*, 2012, 49(5): 6.
- [25] Hengl T, Heuvelink G B M, Stein A. A generic framework for spatial prediction of soil variables based on regression-kriging[J]. *Geoderma*, 2004, 120(1-2): 75 – 93.
- [26] Williams G. Random Forests [M]/Williams G. Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. New York, NY; Springer New York. 2011: 245-268.
- [27] 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.
- [28] 袁玉琦, 陈瀚阅, 张黎明, 等. 基于多变量与 RF 算法的耕地土壤有机碳空间预测研究——以福建亚热带复杂地貌区为例 [J]. *土壤学报*, 2021, 58(04): 887 – 899.
- [29] Guo D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP)[J]. *International Journal of Geographical Information Science*, 2008, 22(7): 801 – 823.
- [30] Guo D. Greedy Optimization for Contiguity-Constrained Hierarchical Clustering [Z]. 2009: 591-596.10.1109/icdmw.2009.75
- [31] Guo D, Wang H. Automatic Region Building for Spatial Analysis[J]. *Transactions in GIS*, 2011, 15: 29 – 45.
- [32] Cambardella C A, Moorman T B, Novak J M, et al. Field-Scale Variability of Soil Properties in Central Iowa Soils[J]. *Soil Science Society of America Journal*, 1994, 58(5).
- [33] Wang J-F, Zhang T-L, Fu B-J. A measure of spatial stratified heterogeneity[J]. *Ecological Indicators*, 2016, 67: 250 – 256.
- [34] Ma Y-J, Xie T, Li X-Y. Spatial variation of soil organic carbon in the Qinghai Lake watershed, northeast Qinghai-Tibet Plateau[J]. *Catena*, 2022, 213: 106187.
- [35] Zeraatpisheh M, Garosi Y, Reza Owliaie H, et al. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates[J]. *Catena*, 2022, 208: 105723.
- [36] Guo L, Zhang H, Chen Y, et al. Combining Environmental Factors and Lab VNIR Spectral Data to Predict SOM by Geospatial Techniques[J]. *Chinese Geographical Science*, 2019, 29(2): 258 – 269.
- [37] Qi R, Li J, Lin Z, et al. Temperature effects on soil organic carbon, soil labile organic carbon fractions, and soil enzyme activities under long-term fertilization regimes[J]. *Applied Soil Ecology*, 2016, 102: 36 – 45.

Application of Zoning Regression Kriging in Soil Organic Carbon Mapping Considered Spatial Heterogeneity and Interactive Environmental Variables

WANG Xiao-mi¹, LIU Jiu-hong¹, BO Dai-yun^{2*}

(1. School of Geographic Sciences, Hunan Normal University, Changsha 410081, China;

2. Haikou Marine Geological Survey Center of China Geological Survey, Haikou 571127, China)

Abstract: [Objective] The aims were to precisely map the spatial distribution of soil organic carbon (SOC), in order to widely carry out the precision fertilization for sustainable agricultural development. [Method] A zoning regression kriging (ZRK) method considered spatial heterogeneity and interactive environmental variables was proposed to draw a high-precision spatial distribution map of SOC content in northern France. The global ordinary kriging (OK), co-kriging (COK), regression kriging (RK), random forest (RF), and OK, COK, and RK based on one single environmental variable and geographical detector were used for comparison. [Result] ① The SOC content in the study area was driven by the interaction of various environmental variables, and the driving effects of different spatial homogeneous zones were different. ② The spatial hierarchical clustering method could effectively use the spatial homogeneous zones of SOC content. Compared with the global OK, COK and RK models, the R^2 values of the ZOK, ZCOK and ZRK methods based on spatial hierarchical clustering partitions were improved by 32% ~ 36 %. ③ The kriging methods based on single environmental variable and geographical detector partition performed generally in this study, and did not effectively improve the SOC prediction accuracy. ④ RF performed poorly in the study, with lower R^2 and higher RMSE and MAE, while ZRK performed best in all models, with the highest R^2 and the lowest RMSE and MAE. [Conclusion] The ZRK method, which takes into account spatial heterogeneity and interactive environmental variables, can effectively use the spatial homogeneous zones of SOC content and draw the spatial distribution map of SOC content considered the interaction of environmental variables, which provides a new model reference for digital soil mapping in spatial heterogeneous regions.

Key words: Soil organic carbon; Spatial hierarchical clustering; Interaction of environmental variables; Zoning regression kriging; Digital soil mapping

[责任编辑: 裴久渤]