

基于时间窗口与极端气候信息的禹城市农田 SOM 空间分布预测^①

夏迎新¹, 李道诚¹, 肖二龙¹, 宁立新^{1,2*}, 梁星宇¹, 颜 君¹

(1 山东农业大学信息科学与工程学院, 山东泰安 271018; 2 山东农业大学农业大数据研究中心, 山东泰安 271018)

摘 要: 土壤有机质(SOM)对土壤质量极其重要。传统 SOM 空间分布预测研究依赖固定时间段遥感图像均值建模, 一定程度上忽视了季节性和植被动态变化的影响。选择最佳时间窗口对 SOM 空间分布预测研究极为关键。华北平原由于裸土期极短、云雾遮挡等原因, 遥感数据不确定性大, 且相关研究较少。本研究以山东省禹城市为对象, 根据植被和耕地条件划分遥感图像时间窗口, 利用 2017 年 Landsat-8 遥感影像数据, 基于随机森林模型, 评估了基于不同时间窗口遥感图像的 SOM 制图精度及其差异, 最后引入极端气候因子, 探究了其对 SOM 空间分布的影响。结果表明: 仅使用遥感变量时, 基于各时间窗口的 SOM 预测精度为返青播种期>旺盛生长期>休耕养护期>成熟收获期, R^2 范围为 0.257~0.330; 引入环境协变量后, 基于各时间窗口的 SOM 预测精度显著提升, R^2 范围为 0.413~0.477, 且基于返青播种期的预测精度最高, 即最佳时间窗口为 2—5 月; 基于最佳时间窗口并引入极端气候变量, SOM 预测精度 R^2 达到 0.501, 暖夜日数和年月最小日最低气温是影响 SOM 空间分布的重要变量。本研究基于华北平原的独特环境, 深刻阐释了时间窗口和极端气候因子的作用, 为类似区域研究提供了新思路和新方法。

关键词: 时间窗口; 土壤有机质; 数字土壤制图; 极端气候信息

中图分类号: S159.9 **文献标志码:** A

Prediction of Spatial Distribution of Soil Organic Matter in Agricultural Fields in Yucheng City Based on Time Window and Extreme Weather Information

XIA Yingxin¹, LI Daocheng¹, XIAO Erlong¹, NING Lixin^{1,2*}, LIANG Xingyu¹, YAN Jun¹

(1 College of Information Science and Engineering, Shandong Agricultural University, Taian, Shandong 271018, China; 2 Agricultural Big-Data Research Center, Shandong Agricultural University, Taian, Shandong 271018, China)

Abstract: Soil organic matter (SOM) is crucial to soil quality. Traditional studies on SOM spatial distribution prediction establish models using the mean values of remote sensing images collected over fixed periods, which neglects the influence of seasonal variations and vegetation dynamics to a certain degree. Therefore, selecting an optimal time window is important for SOM spatial distribution prediction research. The North China Plain suffers from high uncertainty in remote sensing data due to its short bare soil period and frequent cloud and haze occlusion, and relevant studies on this region are still limited. Taking Yucheng City in Shandong Province as the study area, this research classified time windows for remote sensing images based on vegetation and cultivated land conditions. Using 2017 Landsat-8 imagery and the random forest model, we evaluated the accuracy and differences of SOM mapping based on remote sensing data from different time windows. Finally, extreme climate factors were introduced to explore their impacts on the spatial distribution of SOM. The results indicated that when only remote sensing variables were used, the SOM prediction accuracy of different time windows ranked as: reviving and sowing period > vigorous growth period > fallow and maintenance period > maturity and harvesting period, with the coefficient of determination (R^2) ranging from 0.257 to 0.330. After incorporating environmental covariates, the prediction accuracy of all time windows increased significantly, with R^2 ranging from 0.413 to 0.477. The reviving and sowing period achieved the highest accuracy, proving that February to May is the optimal time window. When extreme climate variables were further added on the basis of the optimal time

①基金项目: 山东省高等学校青创科技支持计划项目(2024KJH092)和 2021 年山东省高等学校“青创人才引育计划”项目资助。

* 通信作者(ninglx@sdau.edu.cn)

作者简介: 夏迎新(2000—), 男, 山东济南人, 硕士研究生, 主要从事数字土壤制图研究。E-mail: xiayingxin1219@163.com

window, the R^2 of SOM prediction increased to 0.501. The number of warm nights and monthly minimum daily minimum temperature were identified as key variables affecting the spatial distribution of SOM. Focusing on the unique environmental conditions of the North China Plain, this study elucidates the effects of time windows and extreme climate factors, and offers new insights and methods for research in similar regions.

Key words: Time window; Soil organic matter; Digital soil mapping; Extreme weather information

土壤有机质(Soil organic matter, SOM)是土壤的重要组成部分,主要由动植物残体、微生物及其分解产物组成,是衡量土壤质量的关键指标之一^[1]。SOM在植被及农作物的生长发育过程中发挥着关键性作用,同时对农业生态系统内的碳循环有着直接影响,有助于维持生态系统的稳定性。因此,精准预测并绘制SOM空间分布,不但能够揭示区域碳储量的变化趋势,还能够为农业规划及政策制定提供科学依据^[2]。

传统的土壤制图方法主要依赖实地采样和专家描绘,因成本高昂、工作周期长以及单点样本数据的局限性,难以全面精准反映整个区域SOM空间分布状况^[3]。数字土壤制图(Digital soil mapping, DSM)作为一种新兴技术,已逐步成为获取SOM空间分布的主流方法之一^[4]。DSM通过构建土壤属性与众多环境协变量之间的关系模型,实现土壤属性空间分布的有效预测^[5]。近年来,众多研究借助随机森林、回归决策树、深度学习等方法^[6-8],进一步提升了模型对复杂的土壤与环境之间关系的解析能力。相比传统方法,DSM整合了遥感影像、地理信息以及土壤属性等多源数据,构建了多维环境因子协同作用下的预测框架,在预测精度和空间分辨率方面均展现出显著优势。此外,遥感多光谱数据逐渐成为高精度DSM的关键数据来源^[9]。这些数据不仅能够覆盖大面积区域,还能捕捉土壤在不同时空条件下的光谱特征,为SOM的高效预测提供了强有力的数据支撑。

依据SCORPAN理论,SOM空间分布的预测精度不仅取决于模型选择,合理选取并引入更多有效协变量也是提高SOM空间预测精度的关键^[10]。然而,当前大多数研究在处理诸如遥感指数、气候变量等作为输入因子时,普遍采用均一化处理方法,忽视了光谱响应特征以及地物信息的动态变化特性^[11-13]。由于地表覆被变化和植物发育过程的季节动态性,同一地理单元在不同时间的遥感光谱响应存在显著差异,因此选择恰当的时间窗口对提高预测精度尤为关键。最佳时间窗口是指通过多源遥感数据与环境协变量的动态耦合分析,所识别出的土壤属性时空变异信号最显著、受环境干扰影响最小的关键观测时段,可进一步提高土壤属性预测的准确性和可靠性^[14]。相关研

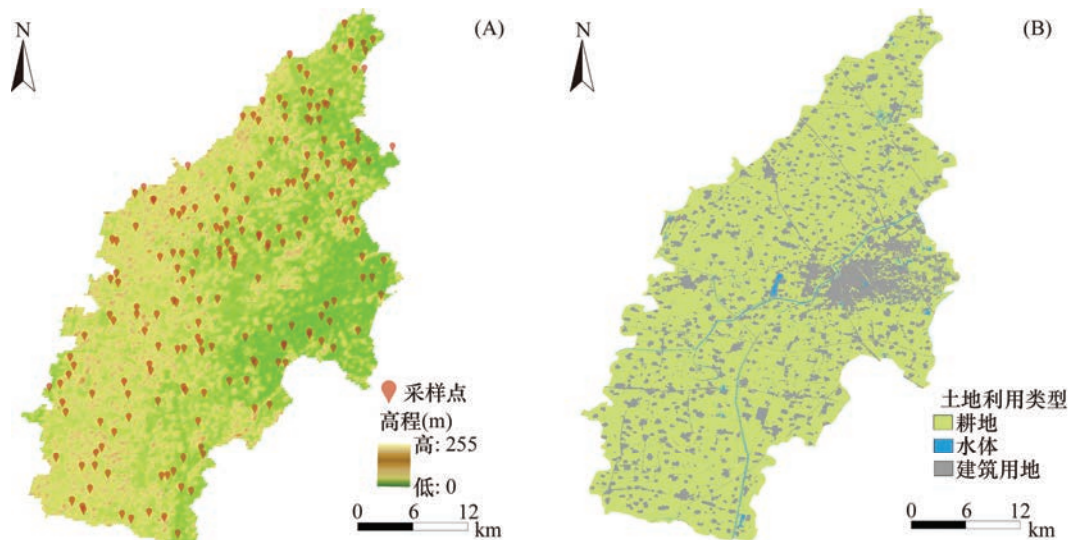
究证实,合理选择最佳时间窗口对于提高预测精度、提升农业生产效率、加强土壤保护措施以及提高环境科学研究的质量具有重要意义。例如,Luo等^[15]成功绘制了黑龙江省的高分辨率作物分布图,并确定5—7月为最佳时间窗口;Ma等^[16]通过划分时间窗口对新疆地区盐渍空间分布进行预测,并确定5—9月为预测最佳时间窗口。然而,现阶段对最佳时间窗口的研究多集中于我国东北和西北地区,华北地区的相关研究较少。华北地区作为我国重要的农业产区,具有独特的地理和气候特点。与东北地区相比,其气候条件更加复杂,降水分布不均,四季分明,温差显著,且华北地区人口密集,农业生产活动频繁而活跃。因此,深入探讨华北平原SOM空间分布及其影响因素并确定最佳时间窗口,对于优化区域农业生产布局及保障粮食安全均具有重要意义。

在全球气候变暖与人类活动协同驱动下,极端气候(Extreme weather, EW)事件的发生频率也不断增加^[17],进而影响土壤的生态平衡以及碳储量。世界气象组织(World meteorological organization, WMO)基于极端气候定义了27个指标,但目前对于极端气候与土壤SOM响应程度的研究相对匮乏。因此,本研究将极端气候作为预测的重要因子纳入建模过程,探讨其对SOM空间分布的响应程度以及作用机制。总体上,本研究的目标如下:①确定禹城市SOM最佳观测时间窗口;②构建禹城市SOM空间分布图;③探索极端气候对SOM预测精度的影响;④对比单一遥感变量与多源协变量的预测精度,探讨多源数据融合在SOM预测中的优势。

1 材料与方法

1.1 研究区概况

禹城市位于山东省西北部,德州市南部,地处黄河下游冲积平原地带,总面积约990 km²。市内地势以平原为主,整体呈现自西南向东北逐渐倾斜的趋势,平均海拔约20 m(图1)。地貌类型丰富多样,主要包括河滩高地、缓平坡地和河间浅平洼地等多种类型。气候类型为暖温带大陆性季风气候,四季特征分明,日照条件充足,年平均气温为13.5℃,年平均



(A: 高程和土壤采样点分布; B: 土地利用类型图)

图 1 研究区域概览

Fig. 1 Overview of study area

降水量为 911.5 mm，无霜期相对较长。耕地面积达 62 035 hm²，主要种植粮食、棉花、大豆等作物，是国家重要的商品粮和优质棉花生产基地之一。

1.2 数据来源与处理

1.2.1 土样采集与测定 2018 年，中华人民共和国自然资源部制定了《耕地质量等别调查评价与监测工作方案》，用于获取耕地的土壤性质和管理信息^[18]。本研究采用拉丁超立方抽样法(Latin hypercube sampling, LHS)布设 200 个样点^[19]，于 2020 年参照上述工作方案系统采集研究区 0~20 cm 土层土壤作为样本，并同时由手持 GPS 记录样点经纬度信息。所有土壤样品带回实验室，采用重铬酸钾氧化-外加热

法测定 SOM 含量。

1.2.2 遥感图像采集与处理 本研究选择使用研究区 2017 年全年可用的 Landsat-8 遥感影像数据，其重访周期为 16 d，空间分辨率为 30 m，数据来源于美国地质调查局(United states geological survey, USGS)。首先，通过 QA_PIXEL 波段构建云掩膜函数，剔除影像云层，获得 2017 年各时间窗口的无云图像。然后，选取 Landsat-8 六个原始波段的地表反射率(Band 2~Band 7)作为预测建模输入因子。Band 1 因其对陆表植被参数敏感性较低，故未纳入本次研究。最后，基于原始波段选择并计算多个植被指数，最终获得 16 个遥感变量，具体公式和指标如表 1 所示。

表 1 研究区遥感变量定义
Table 1 Definitions of remote sensing variables in study area

指数	简写	计算公式
6 个原始光谱	B2~B7	B2、B3、B4、B5、B6、B7
归一化植被指数	NDVI	$(B5-B4)/(B5+B4)$
增强型植被指数	EVI	$g \times (B5-B4)/(B5+C1 \times B4-C2 \times B2+L)$
扩展增强型植被指数	EEVI	$2.5 \times (B5+B6-B4)/(B5+2.5 \times (B6+6 \times B5+B47.5 \times B6-B4)B2+1)$
增强型归一化植被指数	ENDVI	$(B5+B7-B4)/(B5+B7+B4)$
土壤调整植被指数	SAVI	$(B5-B4)/(B5+B4+L) \times (1+L)$
差值植被指数	DVI	B5-B4
比值植被指数	RVI	B5/B4
红外百分比植被指数	IPVI	$B5/(B5+B4)$
绿色比率植被指数	GRVI	$B5/(B3-1)$
绿色归一化植被指数	GNDVI	$(B5-B3)/(B5+B3)$

注：计算公式中的 B2~B7 为 Landsat-8 遥感影像第 2 至第 7 波段的地表反射率；g 为增益系数；C1、C2 为气溶胶校正系数；L 为土壤背景调整系数。

1.2.3 地形因子数据 地形主要通过调控地表物质迁移与能量分配,影响土壤空间异质性。本研究采用的数字高程模型(Digital elevation model, DEM)数据,源于 USGS 提供的 SRTM 数据,空间分辨率为 30 m。

1.2.4 气候因子数据 气候条件对土壤的水热环境产生直接影响,进而影响 SOM 的空间分布。本研究的气候因子数据来源于 WorldClim 全球气候数据库(<https://www.worldclim.org/>),并从中选取 1979—2020 年逐月气温与降水数据,空间分辨率为 1 km。为匹配多源数据的空间尺度,将其重采样为 30 m。

1.2.5 土壤湿度数据 本研究土壤数据来源于 ZENODO 平台的中国土壤水分数据集^[20],并从中获得多个统计指标:平均土壤湿度、土壤湿度标准

差、最高土壤湿度、最低土壤湿度等,时间跨度为 2002—2018 年。本研究所采用的地形、土壤湿度、气候以及地表温度等环境协变量如表 2 所示。

1.2.6 极端气候变量 在全球气候变化的背景下,极端天气和气候事件频发,对土壤的碳平衡以及碳储量的影响程度与日俱增。本研究采用中国科学院资源环境科学数据中心发布的《中国极端气温指数栅格数据集(1961—2020, v1.0)》^[21],从中选取 2010—2017 年极端气候数据。具体极端气候变量定义如表 3 所示。

1.3 时间窗口划分

研究区位于温带季风气候区,四季分明,农作物呈现显著的季节性生长规律。根据研究区耕地利用以及作物生长特点^[22-23],本研究将全年划分并定义为 4 个时间窗口(表 4):2—5 月为返青播种期,

表 2 研究区环境协变量定义
Table 2 Definition of environmental variables in the study area

因子类别	指数	简写
气候	年均气温(Mean annual air temperature)	MAAT
	年均降水(Mean annual precipitation)	MAP
土壤湿度	春季平均土壤湿度(Mean soil moisture in spring)	MSMS3
	夏季平均土壤湿度(Mean soil moisture in summer)	MSMS6
	秋季平均土壤湿度(Mean soil moisture in autumn)	MSMA
	冬季平均土壤湿度(Mean soil moisture in winter)	MSMW
	平均土壤湿度(Mean soil moisture)	MSM
	最高土壤湿度(Highest soil moisture)	HSM
	最低土壤湿度(Lowest soil moisture)	LSM
	土壤湿度标准差(Standard deviation of soil moisture)	SDSM
地表温度	春季昼夜平均地表温度(Mean land surface temperature in spring)	MLSTS3
	夏季昼夜平均地表温度(Mean land surface temperature in summer)	MLSTS6
	秋季昼夜平均地表温度(Mean land surface temperature in autumn)	MLSTA
	冬季昼夜平均地表温度(Mean land surface temperature in winter)	MLSTW
地形	海拔高度(Digital elevation model)	DEM
	岩层厚度(Rock layer thickness)	RT

表 3 极端气候变量定义
Table 3 Definitions of extreme climate variables

指数	简写	定义	单位
平均温差	DTR	日最高气温与日最低气温的月平均差值	℃
冷夜日数	TN10p	日最低气温小于 2010—2017 年第 10 百分位数的日数	d
暖夜日数	TN90p	日最低气温大于 2010—2017 年第 90 百分位数的日数	d
年月最小日最低气温	TNn	日最低气温的年/月最小值	℃
年月最大日最低气温	TNx	日最低气温的年/月最大值	℃
冷昼日数	TX10p	日最高气温小于 2010—2017 年第 10 百分位数的日数	d
暖昼日数	TX90p	日最高气温大于 2010—2017 年第 90 百分位数的日数	d
年月最小日最高气温	TXn	日最高气温的年/月最小值	℃
年月最大日最高气温	TXx	日最高气温的年/月最大值	℃

表 4 时间窗口的划分
Table 4 Division of time windows

时间窗口	月份	作物生长特征
返青播种期	2—5 月	作物出苗、生根
旺盛生长期	6—8 月	作物快速生长
成熟收获期	9—11 月	作物成熟并收获
休耕养护期	12 月—次年 1 月	作物进入休眠期

冬小麦等越冬作物进入返青期，玉米、棉花等春播作物完成播种并进入幼苗期^[24]；6—8 月为旺盛生长期，各类作物均处于快速生长阶段并逐步进入成熟期；9—11 月为成熟收获期，秋收作物成熟收获后进行耕地整理；12 月—次年 1 月为休耕养护期，低温导致土壤冻结，作物进入越冬休眠状态^[25]。基于前期通过云掩膜处理的无云图像，按照所划分的时间窗口进行中值合成。相关研究表明，中值合成遵循最佳像素的选择策略，能有效抑制大气异常值对光谱信号的干扰^[26]。

1.4 模型选择

随机森林(Random forest, RF)模型通过构建众多决策树，并采用投票(分类任务)或平均(回归任务)的方式，得出最终的预测结果。其在解决非线性问题、特征重要性分析以及处理高维数据等方面表现出色。本研究使用 Python 中的 scikit-learn 软件包进行 RF 建模。对于 RF 模型，两个关键参数为 n_estimators 和 max_depth，本研究通过网格搜索法确定其最佳参数 n_estimators=500，max_depth=7，其他参数为默认值。

1.5 模型精度评估

由于本研究样本点数量有限，采用十折交叉验证(10-fold cross-validation)进行精度评估^[27]。模型精度通过决定系数(R^2)、均方根误差(RMSE)、平均绝对误

差(MAE)以及均方误差(MSE)进行综合评价。 R^2 值越趋近于 1，表示 SOM 预测精度越高。RMSE、MAE 和 MSE 的值越接近于 0，则反映模型预测误差越小，精度越高。为降低在预测过程中样本点的随机性以及预测结果的不确定性，迭代运行 100 次之后，最终以各评价指标的均值作为模型性能的最终结果，以确保结果的稳定性。

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \tag{2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \tag{3}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{4}$$

式中： Y_i 表示第 i 个样本的真实值； $\hat{Y}_i$ 表示第 i 个样本的模型预测值； $\bar{Y}$ 表示所有实际观测值的平均值； n 为样本点总数； $|Y_i - \hat{Y}_i|$ 表示第 i 个样本预测值与真实值的绝对误差。

2 结果与分析

2.1 研究区 SOM 描述性统计

表 5 展示了 200 个土壤采样点的 SOM 描述性统计结果。研究区 SOM 含量范围在 14.94~26.52 g/kg，平均值为 21.42 g/kg，标准偏差为 3.12 g/kg，偏度系数为 -0.13，峰度系数为 -1.05，变异系数为 0.15，属于中等程度变异。表明研究区土壤样本的 SOM 具有较好的稳定性和代表性。

表 5 研究区 SOM 描述性统计
Table 5 Descriptive statistics of SOM in study area

采样点	最大值(g/kg)	最小值(g/kg)	平均值(g/kg)	标准偏差(g/kg)	峰度	偏度	变异系数
200	26.52	14.94	21.42	3.12	-1.05	-0.13	0.15

2.2 基于不同策略的 SOM 预测精度

2.2.1 基于遥感变量的 SOM 预测精度 基于遥感变量的 SOM 预测精度如表 6 所示。在 4 个时间窗口中，返青播种期的 SOM 预测精度最高， R^2 值为 0.330，RMSE 为 2.108 g/kg，MSE 为 4.443 (g/kg)²，MAE 为 1.682 g/kg。相比而言，成熟收获期的预测精度则处于最低水平， R^2 仅为 0.249。对不同时间窗口的 SOM 预测精度进行总体比较，各时间窗口预测精度排序

为：返青播种期>旺盛生长期>休耕养护期>成熟收获期，表明基于 2—5 月遥感变量的 SOM 预测精度最高。然而，即使在预测精度最高的时间段，整体的预测精度仍处于相对较低的水平。

2.2.2 加入环境协变量后 SOM 预测精度 以往研究表明，多维环境协变量的引入能够有效提高各时间窗口的预测精度^[28]。本研究结果表明，引入环境协变量后，基于各时间窗口的 SOM 预测精度显著提升。

以返青播种期为例， R^2 由 0.330 提高至 0.477，提升幅度为 44.55%，RMSE、MSE 与 MAE 分别降低 13.19%、24.62% 和 19.80%(表 6)。基于其余时间窗口的预测精度也实现了不同程度的提高。该结果表明，环境协变量对 SOM 空间分布预测发挥着重要作用，引入的环境变量能够进一步捕捉遥感变量中未能充分体现的影响因素，进一步提升模型预测的性能。

表 6 基于不同时间窗口的 SOM 预测精度
Table 6 Prediction accuracies of SOM based on different time windows

时间窗口	R^2	RMSE	MSE	MAE
返青播种期	0.330	2.108	4.443	1.682
旺盛生长期	0.279	2.206	4.866	1.702
成熟收获期	0.249	2.228	4.962	1.729
休耕养护期	0.257	2.2188	4.921	1.744
返青播种期+ENV	0.477	1.830	3.349	1.356
旺盛生长期+ENV	0.421	1.860	3.459	1.388
成熟收获期+ENV	0.413	1.863	3.472	1.375
休耕养护期+ENV	0.454	1.845	3.405	1.364

注：表中 ENV 表示环境协变量。

2.2.3 加入极端气候变量后 SOM 预测精度 基于预测精度最优水平的返青播种期，进一步融合极端气候变量后的 SOM 预测精度如表 7 所示。引入极端气候因子后，SOM 预测精度得到进一步提升，其中 R_2

由 0.477 提升至 0.501，提升了 5.03%，RMSE 降低了 4.75%，MSE 降低了 9.29%，MAE 降低了 6.71%。该结果表明，极端气候变量与研究区 SOM 空间分布存在一定的相关性，极端气候指数能够进一步反映气候对 SOM 累积或分解的作用程度，从而提升模型预测性能。

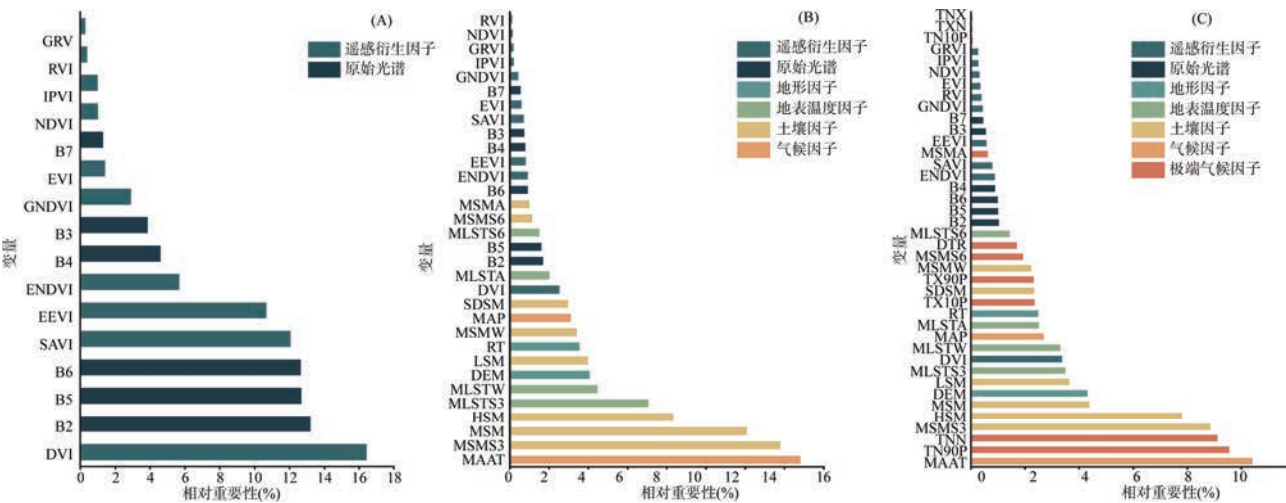
表 7 加入极端气候变量前后的 SOM 预测精度对比
Table 7 Comparison of prediction accuracies of SOM before and after inclusion of extreme weather factors

建模策略	R^2	RMSE	MSE	MAE
返青播种期+ENV	0.477	1.830	3.349	1.356
返青播种期+ENV+EW	0.501	1.743	3.038	1.265

注：表中 EW 表示极端气候变量。

2.3 变量相对重要性分析

以 3 种建模策略各自获得的 SOM 空间分布最优预测结果为例，各变量的相对重要性排序如图 2 所示。对于仅采用遥感变量建模的预测结果，DVI 对 SOM 空间分布的相对重要性高于其他变量，约为 16.42%，原始光谱 B2、B5、B6 及 SAVI 的相对重要性较为接近，分别为 13.20%、12.67%、12.63% 和 12.05%。且遥感衍生因子与原始光谱对 SOM 空间变异的累积解释力较为均衡，相对重要性分别为 50.31% 和 49.69%。



(A: 返青播种期; B: 返青播种期+ENV; C: 返青播种期+ENV+EW)

图 2 不同建模策略变量相对重要性

Fig. 2 Relative importance of variables under different prediction models

引入环境协变量后，MAAT 对 SOM 预测的相对重要性居首位，达 14.80%。土壤因子中，MSMS3、MSM 和 HSM 对 SOM 预测的相对重要性依次为 13.77%、12.08% 和 8.31%。相较于未引入环境协变量的预测结果，遥感衍生因子和原始光谱的相对重要

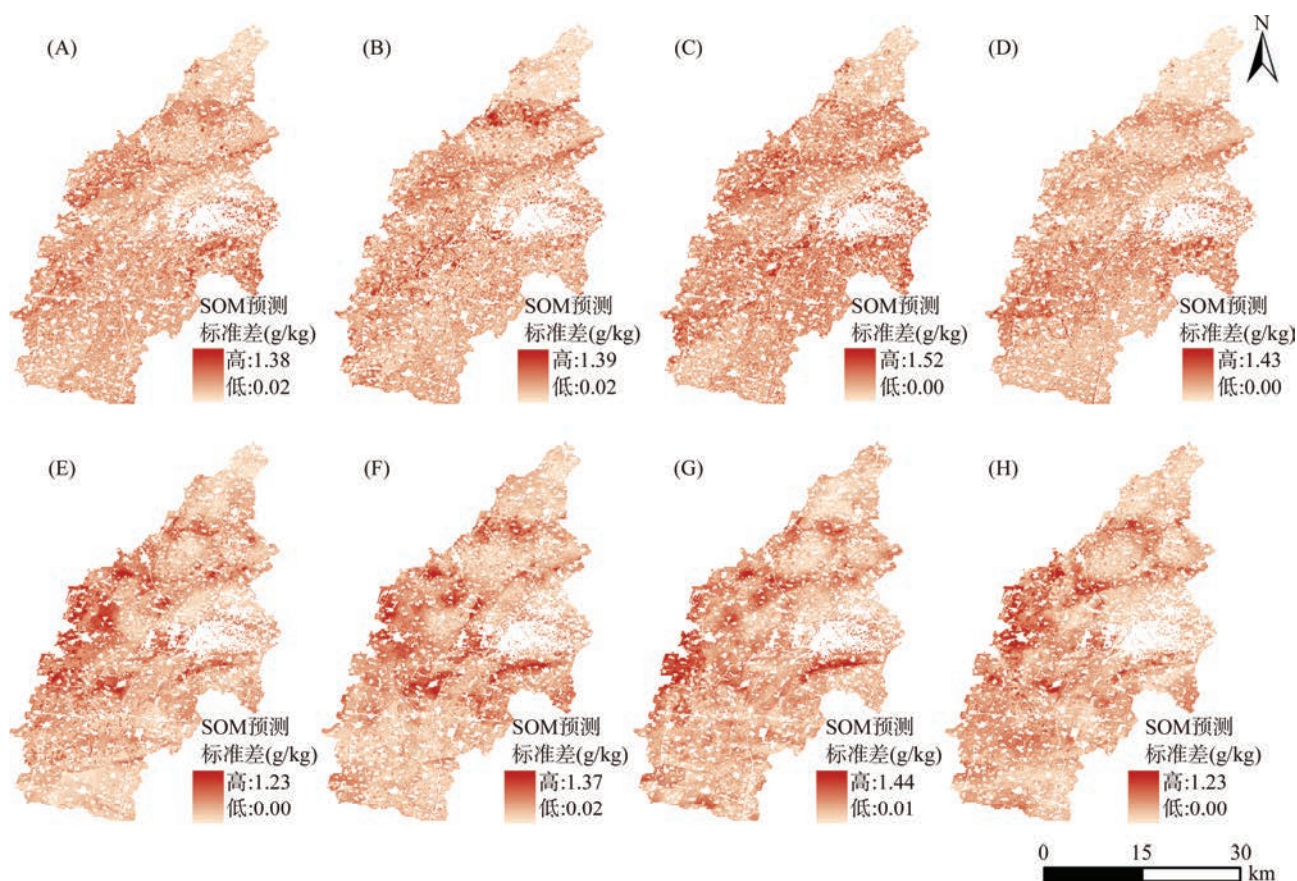
性显著下降。对各类环境协变量因子进一步分析，土壤因子可解释 SOM 空间变化的 46.66%，气候因子和地表温度因子的解释程度分别约为 17.91% 和 15.02%，地形因子、遥感衍生因子和原始光谱的解释程度相对有限，分别为 7.61%、6.56% 和 6.26%。

进一步引入极端气候变量后, MAAT 的重要性依旧居首位, 但相对重要性由 14.80% 降至 10.39%。极端气候变量 TN90p 和 TNn 的重要性超过 MSMS3、HSM 和 MSM, 相对重要性分别为 9.53% 和 9.11%。引入极端气候因子后, 各类变量因子相对重要性总体呈现下降趋势。土壤因子的相对重要性由 46.66% 降低至 31.73%, 地表温度因子的相对重要性由 15.02% 下降至 11.74%, 气候因子的相对重要性由 17.91% 下降至 10.74%。极端气候因子的相对重要性为 25.20%, 在 7 类环境变量中仅次于土壤因子, 表明极端气候因

子对研究区域 SOM 空间分布具有显著的驱动作用。

2.4 模型精度与不确定性评价

通过 100 次迭代计算预测结果的平均值和标准差, 评估基于不同时间窗口的 SOM 预测精度差异, 结果如图 3 所示。可见, 基于返青播种期遥感数据的预测结果的不确定性处于最低水平, 而基于旺盛生长期遥感数据的预测结果不确定性达到最大值。引入环境协变量后, 基于返青播种期和休耕养护期的预测结果不确定性显著降低, 但基于成熟收获期的预测结果仍处于较高水平。



(A: 返青播种期; B: 旺盛生长期; C: 成熟收获期; D: 休耕养护期; E: 返青播种期+ENV; F: 旺盛生长期+ENV; G: 成熟收获期+ENV; H: 休耕养护期+ENV)

图 3 基于不同时间窗口的 SOM 预测不确定性

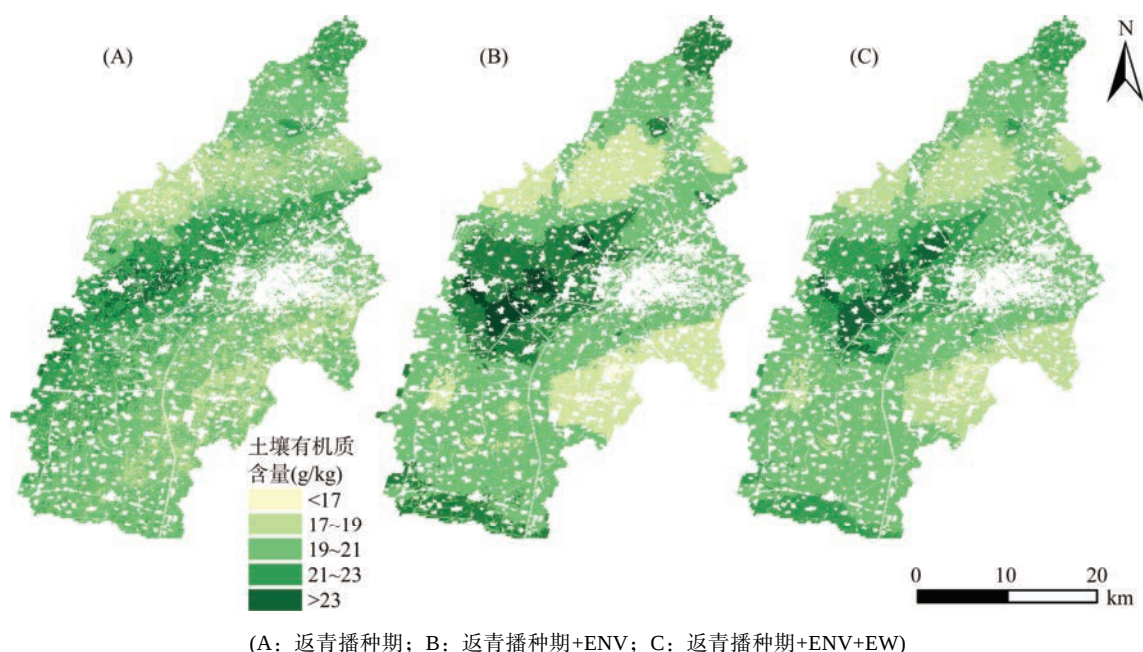
Fig. 3 SOM forecast uncertainties based on different time windows

仅使用遥感数据时, SOM 预测不确定性主要集中在研究区的中部和北部。引入环境协变量后, SOM 预测不确定性主要位于研究区西部。导致这一现象的可能原因包括以下几个方面: 第一, 本研究预测未考虑人类活动因子的影响, 如施肥制度、耕作方式的差异等; 第二, 未考虑农田灌溉条件和土地利用状态的影响; 第三, 不确定性较高的地区可能 SOM 空间变异性较强; 第四, 采样点数量较少且缺乏代表性。这

些因素共同作用导致对于 SOM 的预测出现波动。

2.5 研究区 SOM 空间分布特征

如图 4 所示, 以 3 种建模策略各自获得的 SOM 空间分布最优预测结果为例, 3 种建模策略预测结果整体趋势基本一致, 均呈现中部高、北部和东南部低的特征。由于古黄河泛滥的影响, 禹城市中部土壤结构较复杂^[29], 且该地区农田管理措施优化政策也在一定程度上提高了土壤持水保肥能力, 为 SOM 的积



(A: 返青播种期; B: 返青播种期+ENV; C: 返青播种期+ENV+EW)

图4 基于不同建模策略预测的 SOM 空间分布

Fig. 4 Spatial distributions of SOM prediction under different prediction models

累提供了良好的条件^[30]。而北部和东南部地区可能由于黄河泥沙堆积形成沙质土壤,持水保肥能力相对较弱,导致 SOM 含量较低。

尽管3种建模策略的 SOM 预测结果在整体上具有相似性,但在特定区域仍然存在差异。例如,返青播种期+ENV 预测模型在中部和南部对 SOM 空间分布的预测精度更高,在此基础上,引入极端气候变量后,研究区中部 SOM 空间分布的预测精度进一步提高,且与实测值更为接近。

3 讨论

3.1 不同时间窗口对 SOM 预测精度的影响

在 SOM 空间分布预测研究中,最佳时间窗口的研究大多聚焦于东北地区,主要归因于东北地区独特的气候和耕作条件^[31-32]。东北地区属温带季风区中温带及小部分寒温带,冬季漫长且严寒,降水主要集中在夏季,年降水量相对较少,具有显著的大陆性气候特征,且气温年较差和日较差较大,种植方式一般为一年一熟制^[33]。在确定最佳预测时间窗口时,大多在裸土期即4—5月这一特定时间段。例如,Luo 等^[34]通过不同图像合成方法,预测得出5月是东北松嫩平原地区 SOM 预测最佳时间窗口;Luo 等^[35]通过绘制东北地区旱地和水田 SOM 空间分布图,得出水田的最佳预测时间窗口为4月,旱地为5月,且加入环境协变量后,提高了 SOM 预测的准确性;Luo 等^[36]基于多年合成图像和分区算法对松嫩平原 SOM 进行预

测,结果得出5月为其预测的最佳时间窗口。

华北地区作为典型暖温带大陆性季风气候区,与东北地区相比,冬季相对较温和,气温年较差和日较差相对较小,裸土时期极短。虽然在休耕养护期这一阶段,植被覆盖度相较于其他时间窗口处于较低水平,但与东北地区裸土期相比,华北地区在这一时间段雨雪天气较为频繁,导致土壤表层积雪覆盖时间较长,使得遥感卫星难以精确获取能够真实反映该时段土壤特性的光谱信息。且该时间段云层较厚,雾霾天气较为严重,进一步降低了遥感卫星穿透云层直接扫描到土壤表层的可能性。

而在返青播种期,天气逐渐回暖,冻土开始融化,大量作物相继进入萌芽期或生根阶段。此阶段植被虽在土壤表面有一定程度的生长,但仍处于发育的初始阶段,植株整体较为矮小,对土壤表层的遮挡十分有限。且相较于休耕养护期,此阶段的雨雪天气明显减少,遥感卫星能够有效地对土壤表层不同波段的反射进行勘测。因此,无论是单纯使用遥感变量构建预测模型,还是在此基础上进一步引入环境协变量进行 SOM 空间分布的预测,在这一时间段内的预测精度均能够达到4个时间窗口中相对较高的水平。

另外,在旺盛生长期,受季风气候影响,降水频繁且土壤湿度较高,作物茎叶逐渐覆盖土壤表面,作物生长接近饱和,作物茎叶大面积覆盖,会对光线的吸收、反射和散射产生显著影响;在成熟收获期,夏玉米等农作物成熟收获后,残留秸秆的覆盖及其与湿

润土壤混合的影响^[37], 会极大地干扰遥感卫星对土壤表层反射光谱的准确接收与测量。这些因素使遥感卫星难以获取土壤表层完整、真实且准确的光谱信息, 从而导致在建模预测时, 所依据的数据缺乏足够的精度和可靠性, SOM 的预测精度相较于其他时期较低。

3.2 不同建模策略对 SOM 预测精度的影响

为评估不同建模策略对 SOM 空间预测精度的影响, 本研究采用递进式分析方法对不同时间窗口的预测精度进行评价。

基于遥感变量的建模策略: 以原始波段地表反射率计算的植被指数及光谱特征作为输入变量, 构建 SOM 预测模型, 结果显示, 在返青播种期这一时间窗口, SOM 预测精度达到最高值, R^2 为 0.330。但整体而言, 4 个时间窗口的 SOM 预测精度差异较小, 且总体精度处于较低水平。这可能是由于研究区域空间范围较小, 整体光谱特征变化幅度有限, 导致遥感变量对 SOM 空间变异的解释能力受限。

引入多元环境协变量的建模策略: 在基于遥感变量建模的基础上, 进一步引入多源环境协变量, 构建 SOM 预测模型, 结果显示, 各时间窗口 SOM 预测精度均显著提升, 且 MSE 明显减小, R^2 显著增大。而在返青播种期的预测精度依然保持最优。由此可见, 引入环境协变量后, SOM 预测结果的精度和准确性更出色。

基于最佳时间窗口引入极端气候因子的建模策略: 在预测精度最佳的时间窗口, 即返青播种期进一步引入极端气候因子, 构建 SOM 预测模型, 结果表明, 引入极端气候因子后, SOM 预测精度进一步提高, R^2 由 0.477 提升至 0.501, 模型精度提升约 5.03%。这表明极端气候变量不仅能够与常规变量因子产生协同效应, 且能进一步捕捉常规气候变量无法体现的信息。

3.3 影响 SOM 空间分布的主要环境协变量

通过各变量的贡献程度进行分析可知, 年均气温(MAAT)是影响 SOM 空间分布的关键驱动因子。太阳辐射作为地表主要热源, 其能量输入直接影响表层土壤温度变化。地表温度的升高能直接或间接地作用于土壤内部的水分含量, 进而影响 SOM 的累积与分解。另外, 适宜的温度能够有效调控微生物活性, 并促使微生物的群落结构发生变化, 而微生物的代谢和活动对 SOM 的空间分布也有直接影响^[38]。

除年均气温外, 土壤湿度因子对 SOM 空间分布的影响也表现出较高的贡献率。作为微生物生存及进

行各类代谢活动的关键因素之一, 土壤湿度对 SOM 的分解、转化及积累等一系列进程均发挥着重要作用。例如, 各类微生物体内众多酶促反应的正常进行, 均需要以水作为介质来保障, 而这些酶促反应又直接参与 SOM 的分解代谢过程, 进而对 SOM 的含量及空间分布产生影响。此外, 适宜的土壤湿度条件能够在微生物和植物根系的共同作用下促进的土壤团聚体形成, 能够在一定程度上延缓 SOM 分解, 从而为 SOM 在土壤中的存储提供有利条件^[39]。在土壤湿度适宜、团聚体稳定性较好的区域, SOM 的含量相对更高^[40], 分布更为集中, 例如本研究区的中部地区。

地表温度因子对 SOM 的存储和分布同样起着重要作用。地表热状况直接决定了土壤所处的温度环境, 从而间接地对微生物的代谢活动产生影响^[41]。当地表温度较为适宜时, 微生物新陈代谢速率相对较快, 能够更为高效地分解 SOM, 而对于地表温度较低的区域, 微生物活性受到抑制, 从而致使 SOM 更易于在土壤中累积^[42]。其次, 地表温度能够直接或间接地对土壤湿度产生作用, 例如当温度升高时, 会促使土壤表层的水分子运动加剧, 从而加快其蒸发速度, 而温度较低时则相反。

对于地形因子, 尽管本研究仅采用了海拔高度(DEM)和岩层厚度(RT)两个地形变量, 但其仍占据一定的 SOM 解释能力。研究区域地形呈现西高东低的趋势, 随着海拔的升高, 可能会对小区域的气候产生一定的影响^[43], 致使其温度和降水随之发生相应的变动, 进而直接或者间接地影响土壤微生物的活动。

对于遥感变量, 差值植被指数(DVI)在 SOM 空间预测的解释程度始终居于首位, 这表明其在遥感变量中具有较强的代表性和解释能力。这可能是因为 DVI 能够有效捕捉地表植被状况与 SOM 空间变化的内在联系^[44]。此外, 相较于其他遥感变量, DVI 对于本研究区域其响应特性可能更为稳定, 不易出现饱和或失真等现象。

3.4 极端气候因子对 SOM 空间预测的影响

本研究变量相对重要性分析表明, 在环境协变量的基础上, 进一步引入极端气候变量能够有效提升 SOM 空间分布的预测精度。其中, 暖夜日数(TN90p)和年月最小日最低气温(TNn)在变量重要性中分别位居第二、第三位, 主要原因可能在于: 在植被返青播种期, 较低气温会导致植被萌芽时间延迟, 这种延迟效应会进一步干扰植物的光合作用, 进而影响 SOM 的分布趋势。持续高温胁迫则可能导致植物早期生长受限, 减少根系分泌物的输入, 从而间接影响 SOM

的空间分布^[45]。从微生物活性的视角来看, 极端低温会抑制微生物的代谢和繁殖, 分解和转化 SOM 的细菌活性和种类在低温下受到一定限制^[46]。当气温回升时, 对微生物活性的解冻效应显著, TNn 较高的地区微生物会更早恢复活性, 进而影响 SOM 的分布趋势。较高的 TN90p 可能意味着长时间高温和相对较少的降水, 可能导致研究区局部土壤干燥, 抑制微生物的活性, 从而间接影响 SOM 的积累和分布。

春季作为极端气候事件的高发期, 寒潮现象的发生可能会导致不同地区极端低温事件的强度和频率呈现明显差异, 并直接作用于 SOM 的分解速率, 使得不同区域 SOM 分解出现差异。而暖流引发的高温事件可能进一步加剧 SOM 分布的不均匀性。以往研究表明, 年均气温与 SOM 含量呈负相关, 而某些极端气候事件对 SOM 含量及空间分布具有促进作用^[47]。这一现象说明引入极端气候变量能够进一步捕捉常规气候因子无法体现的关键气候信息, 例如气候强度的变化幅度、持续时间等。这也是加入极端气候变量后年均气温相对重要性降低的原因。

4 结论

本研究以德州禹城市为研究区域, 将一年划分为4个时间窗口, 利用随机森林模型基于不同建模策略对 SOM 空间分布进行预测并评价其精度。研究结果显示, 仅使用遥感变量时, 基于返青播种期遥感变量的 SOM 预测精度显著高于其他时间窗口, 但整体预测精度相对较低。引入环境协变量后, 基于各时间窗口的 SOM 预测精度大幅提升, R^2 由 0.257~0.330 提升至 0.413~0.477, 其中基于返青播种期(2—5月)的预测精度最高。在最佳时间窗口(返青播种期)基础上引入极端气候变量, SOM 预测精度进一步提高, R^2 达 0.501。在众多环境协变量中, MAAT、TN90p 和 TNn 是影响 SOM 空间分布的重要因子。研究区 SOM 空间分布呈中部高、北部和东南部低的特征。本研究结果表明, 基于多源数据融合的 SOM 预测比基于单一遥感变量的预测更具优势, 且返青播种期为本研究区预测的最佳时间窗口。

参考文献:

- [1] 王清奎, 田鹏, 孙兆林, 等. 森林土壤有机质研究的现状与挑战[J]. 生态学杂志, 2020, 39(11): 3829–3843.
- [2] 杨晋帆, 杨阳, 刘婷婷, 等. 基于神经网络优化算法的云南土壤有机质含量数字制图——以景洪市为例[J]. 土壤通报, 2023, 54(2): 273–285.
- [3] 谭溪晗, 冯文兰, 秦鱼生, 等. 川西南山区县域数字土

- 壤制图研究[J]. 西南农业学报, 2024, 37(9): 2086–2095.
- [4] 周洋, 赵小敏, 郭熙. 基于多源辅助变量和随机森林模型的表层土壤全氮分布预测[J]. 土壤学报, 2022, 59(2): 451–460.
- [5] Wadoux A M J C, Samuel-Rosa A, Poggio L, et al. A note on knowledge discovery and machine learning in digital soil mapping[J]. European Journal of Soil Science, 2020, 71(2): 133–136.
- [6] 江叶枫, 郭熙, 叶英聪, 等. 省域尺度土壤有机质空间分布的神经网络法预测[J]. 江苏农业学报, 2017, 33(4): 828–835.
- [7] 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(4): 887–899.
- [8] 祝元丽, 冯向阳, 闫庆武, 等. 基于 GBDT 的望奎县农田土壤有机碳主控因子研究[J]. 中国环境科学, 2024, 44(3): 1407–1417.
- [9] 刘雅璇, 于慧, 罗勇. 基于辅助变量的紫色土耕地土壤有机质空间预测[J]. 土壤, 2024, 56(4): 857–865.
- [10] 江叶枫, 郭熙, 叶英聪, 等. 基于辅助变量和神经网络模型的土壤有机质空间分布模拟[J]. 长江流域资源与环境, 2017, 26(8): 1150–1158.
- [11] 韩杏杏, 陈杰, 王海洋, 等. 基于随机森林模型的耕地表层土壤有机质含量空间预测——以河南省辉县市为例[J]. 土壤, 2019, 51(1): 152–159.
- [12] 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.
- [13] 尹群, 郭纪敏, 张世文, 等. 基于辅助变量的县域土壤有机质预测[J]. 江苏农业科学, 2020, 48(24): 267–273.
- [14] Wang Y A, Luo C, Zhang W Q, et al. Remote sensing prediction model of cultivated land soil organic matter considering the best time window[J]. Sustainability, 2023, 15(1): 469.
- [15] Luo C, Liu H J, Lu L, et al. Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine[J]. Journal of Integrative Agriculture, 2021, 20(7): 1944–1957.
- [16] Ma S L, He B Z, Ge X Y, et al. Spatial prediction of soil salinity based on the Google Earth Engine platform with multitemporal synthetic remote sensing images[J]. Ecological Informatics, 2023, 75: 102111.
- [17] 蔡子怡, 游庆龙, 吴芳营, 等. 全球碳中和背景下中国气候与极端气候变化[J]. 科技导报, 2024, 42(19): 73–84.
- [18] Ning L X, Cheng C X, Lu X, et al. Improving the prediction of soil organic matter in arable land using human activity factors[J]. Water, 2022, 14(10): 1668.
- [19] 李维友, 段良霞, 谢红霞, 等. 基于条件拉丁超立方抽样的县域耕地土壤有机质空间插值合理样本密度的确定[J]. 土壤通报, 2022, 53(3): 505–513.
- [20] Meng X J, Mao K B, Meng F, et al. A fine-resolution soil moisture dataset for China in 2002–2018[J]. Earth System Science Data, 2021, 13(7): 3239–3261.
- [21] 陈秋沅, 张宇, 刘潇喻, 等. 基于 ETCCDI 的中国极端气温指数栅格数据集研发及评估[J]. 全球变化数据学报

- (中英文), 2024, 8(1): 67–75, 180–188.
- [22] Zhang Z M, Lu C H. Identification of maize yield trend patterns in the North China Plain[J]. *International Journal of Plant Production*, 2021, 15(1): 125–137.
- [23] 陈述悦, 李俊, 陆佩玲, 等. 华北平原麦田土壤呼吸特征[J]. *应用生态学报*, 2004, 15(9): 1552–1560.
- [24] 李美萱, 霍治国, 孔瑞, 等. 黄淮海冬小麦春季低温灾害等级指标构建[J]. *应用气象学报*, 2024, 35(1): 45–56.
- [25] 赵俊芳, 谢鸿飞, 杨嘉琪, 等. 1981—2020 年黄淮海冬小麦干旱时空分异规律及演变趋势[J]. *生态学杂志*, 2024, 43(10): 3104–3112.
- [26] Roy D P, Wulder M A, Loveland T R, et al. Landsat-8: Science and product vision for terrestrial global change research[J]. *Remote Sensing of Environment*, 2014, 145: 154–172.
- [27] Wong T T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation[J]. *Pattern Recognition*, 2015, 48(9): 2839–2846.
- [28] Luo C, Zhang W Q, Zhang X L, et al. Mapping of soil organic matter in a typical black soil area using Landsat-8 synthetic images at different time periods[J]. *Catena*, 2023, 231: 107336.
- [29] 张贝尔, 黄标, 张晓光, 等. 近 30 年华北平原粮食主产区土壤肥力质量时空演变分析——以山东禹城市为例[J]. *土壤*, 2012, 44(3): 381–388.
- [30] 李裕瑞, 李峰, 范朋灿, 等. 平原农区空心村典型土壤的压实程度及利用导向——以山东省禹城市为例[J]. *中国农业资源与区划*, 2018, 39(2): 176–182.
- [31] Luo C, Zhang W Q, Zhang X L, et al. Mapping soil organic matter content using Sentinel-2 synthetic images at different time intervals in Northeast China[J]. *International Journal of Digital Earth*, 2023, 16(1): 1094–1107.
- [32] Luo C, Zhang X L, Wang Y H, et al. Regional soil organic matter mapping models based on the optimal time window, feature selection algorithm and Google Earth Engine[J]. *Soil and Tillage Research*, 2022, 219: 105325.
- [33] 李克南, 杨晓光, 刘志娟, 等. 全球气候变化对中国种植制度可能影响分析Ⅲ. 中国北方地区气候资源变化特征及其对种植制度界限的可能影响[J]. *中国农业科学*, 2010, 43(10): 2088–2097.
- [34] Luo C, Zhang X L, Meng X T, et al. Regional mapping of soil organic matter content using multitemporal synthetic Landsat 8 images in Google Earth Engine[J]. *Catena*, 2022, 209: 105842.
- [35] Luo C, Zhang W Q, Meng X T, et al. Mapping the soil organic matter content in Northeast China considering the difference between dry lands and paddy fields[J]. *Soil and Tillage Research*, 2024, 244: 106270.
- [36] Luo C, Wang Y A, Zhang X L, et al. Spatial prediction of soil organic matter content using multiyear synthetic images and partitioning algorithms[J]. *Catena*, 2022, 211: 106023.
- [37] 胡实, 彭娜, 谢小立, 等. 农田秸秆覆盖保墒研究[J]. *中国农业气象*, 2007, 28(1): 49–53.
- [38] 夏迎新, 肖二龙, 李道诚, 等. 基于物候信息的华北平原农田土壤有机碳密度空间分布预测[J]. *节水灌溉*, 2025(5): 1–8.
- [39] 徐英德, 汪景宽, 王思引, 等. 玉米残体分解对不同肥力棕壤团聚体组成及有机碳分布的影响[J]. *中国生态农业学报*, 2018, 26(7): 1029–1037.
- [40] 温丽娜, 周波, 王立, 等. 淤地坝沟道植被覆盖/土地利用和土壤水分的分析[J]. *草原与草坪*, 2021, 41(3): 41–47.
- [41] Qi R M, Li J, Lin Z A, 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.
- [42] Yang L, Cai Y Y, Zhang L, et al. A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 102: 102428.
- [43] 朱阿兴, 杨琳, 樊乃卿, 等. 数字土壤制图研究综述与展望[J]. *地理科学进展*, 2018, 37(1): 66–78.
- [44] Aldana-Jague E, Heckrath G, MacDonald A, et al. UAS-based soil carbon mapping using VIS-NIR (480–1000 nm) multi-spectral imaging: Potential and limitations[J]. *Geoderma*, 2016, 275: 55–66.
- [45] Zhelnina K, Louie K B, Hao Z, et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly[J]. *Nature Microbiology*, 2018, 3(4): 470–480.
- [46] Tsui C C, Chen Z S, Hsieh C F. Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan[J]. *Geoderma*, 2004, 123(1/2): 131–142.
- [47] 周琪清, 赵小敏, 郭熙, 等. 基于物候与极端气候信息的耕地土壤有机碳空间分布预测研究[J]. *土壤学报*, 2024, 61(3): 648–661.

(责任编辑: 于 飞)