

李安琪, 杨琳, 蔡言颜, 等. 基于递归特征消除-随机森林模型的江浙沪农田土壤肥力属性制图 [J]. 地理科学, 2024, 44(1): 168-178. [Li Anqi, Yang Lin, Cai Yanyan et al. Digital mapping of soil fertility attributes in croplands in Jiangsu, Zhejiang and Shanghai based on recursive feature elimination-random forest model. Scientia Geographica Sinica, 2024, 44(1): 168-178.] doi: 10.13249/j.cnki.sgs.20220131

# 基于递归特征消除-随机森林模型的江浙沪 农田土壤肥力属性制图

李安琪, 杨琳, 蔡言颜, 张磊, 黄海莉, 吴琪, 王雯琪

(南京大学地理与海洋科学学院, 江苏 南京 210023)

**摘要:** 以江苏省、浙江省、上海市农田为研究区, 选用气候、地形、植被、土壤属性等自然环境协变量, 及农业机械总动力、每公顷农用作肥施用量、农业总产值、农村用电量等农业活动变量, 利用递归特征消除方法(RFE)对环境协变量进行筛选, 基于筛选后的最优变量组合建立随机森林(RF)模型, 进行表层土壤 pH、有机碳、全氮、全磷、全钾、铵态氮、硝态氮、有效磷、速效钾、交换性钙、交换性镁 11 种主要土壤肥力属性的空间分布预测, 并采用 100 次重复的十折交叉验证法进行验证。结果表明: ① 11 个模型筛选出的环境协变量类型主要集中在气候、地形与植被变量, 表征人类农业活动的变量在有机碳、全磷、全钾、铵态氮和有效磷预测中体现重要作用。② 11 个模型的决定系数( $R^2$ )在 0.27~0.53, pH、速效钾、交换性镁和交换性钙的预测模型决定系数( $R^2$ )均在 0.45 以上。本研究表明人类活动变量对于土壤肥力预测具有重要意义, 而递归特征消除-随机森林模型(RFE-RF)可以用于农田主要土壤肥力属性制图, 为农业生产提供准确的土壤肥力属性空间分布信息。

**关键词:** 递归特征消除; 随机森林; 土壤肥力属性; 农田土壤; 数字土壤制图; 江浙沪

**中图分类号:** K903 **文献标识码:** A **文章编号:** 1000-0690(2024)01-0168-11

土壤肥力反映土壤生产力, 对保证作物产量、发展可持续农业起着关键作用<sup>[1-3]</sup>。随着人口不断增加和耕地资源日益紧张, 土壤肥力属性的空间分布研究愈加引起关注。pH 及有机碳、氮、磷、钾等属性含量等是主要土壤肥力指标, 影响作物生长、养分循环与微生物过程<sup>[4-5]</sup>。另外, 交换性钙和交换性镁也被用作农田土壤肥力评价的重要指标<sup>[6]</sup>。这些土壤属性的空间分布信息获取对于土壤健康管理、生态环境建模等具有重要意义。

数字土壤制图(Digital soil mapping, DSM)是一种高效地预测和表达土壤属性空间分布的技术, 已成为过去几十年中获取土壤图的最常见方法。该方法通过创建数学模型, 构建环境协变量与目标土壤变量的关系, 并使用环境变量来预测土壤属性和类型的空间分布<sup>[7]</sup>。McBratney 等<sup>[8-9]</sup>提出了 SCOR-

PAN 模型(S: 土壤, C: 气候, O: 生物, R: 地形, P: 母质, A: 年代和 N: 时间)。国内外利用气候、地形、植被等遥感数据辅助预测土壤属性的相关研究主要集中在土壤有机碳<sup>[10-12]</sup>和全氮<sup>[13]</sup>的预测, 但利用数字土壤制图方法对农田土壤硝态氮、铵态氮、磷和钾及其有效态以及其他中量元素含量空间分布预测的研究并不多。

过去人们常常利用自然环境因子来进行预测制图, 而随着人类活动对土壤属性的影响日益增强, 人类活动因子不可忽视。常见的农业管理措施, 如施肥、灌溉、秸秆还田等, 会影响土壤碳等土壤肥力含量及其空间分布特征。农业影响变量也逐渐被用于土壤肥力属性制图<sup>[14]</sup>。例如, Deng 等<sup>[15]</sup>将县级氮肥施用率加入华东地区农田表土有机碳储量的随机森林预测模型, 结果表明施肥解释了土壤有机碳估

收稿日期: 2022-02-13; 修订日期: 2022-05-01

**基金项目:** 国家自然科学基金项目(41971054)、中央高校基本科研业务费(020914380079)资助。[Foundation: National Natural Science Foundation of China (41971054), Fundamental Research Funds for the Central Universities (020914380079).]

**作者简介:** 李安琪(1997—), 男, 黑龙江哈尔滨人, 硕士研究生, 主要从事数字土壤制图方面的研究。E-mail: [lianqidimu@163.com](mailto:lianqidimu@163.com)

**通信作者:** 杨琳。E-mail: [yanglin@nju.edu.cn](mailto:yanglin@nju.edu.cn)

计值的部分变异来源。Deng 等<sup>[16]</sup>在其关于华东地区农田表层土壤全氮和碳氮比的研究中,将县级氮肥施用率和人口密度加入随机森林模型,结果表明氮肥施用率是 2 个预测模型重要性排名前两名的变量。这些研究表明了人类活动变量在土壤制图中的不可或缺,但整体上这些研究还较少。

由于机器学习具有强大的数据处理能力、能处理复杂环境与土壤间的非线性关系等特点,因此常被用于数字土壤制图。随机森林(Random Forest, RF)是其中精度较高的一种,现已广泛用于预测土壤属性并制图<sup>[17-20]</sup>,与其他土壤制图方法相比,它具有许多优点,包括灵活与输入环境协变量类型(类型变量和连续型变量),对于异常值和噪声的容忍度较高<sup>[21]</sup>,可以评估变量的可变重要性等。递归特征消除(Recursive feature elimination, RFE)是一种用作筛选最佳特征组合的方法<sup>[22]</sup>。RFE 方法的特点是识别对目标变量有显著贡献的环境协变量子集,避免变量子集的冗余<sup>[23]</sup>。以上两者相结合进行土壤制图也被应用在了一部分研究中:如 He 等选用自然变量、物候参数、遥感变量利用 RFE-RF 模型预测了安徽省宣城市的农田土壤有机碳<sup>[23]</sup>。王世航等利用 RFE-RF 模型预测了安徽省土壤 pH<sup>[24]</sup>。Mahboobeh Tayebi 等利用 RFE-RF 模型预测了巴西 Piracicaba 地区的土壤有机碳储量<sup>[25]</sup>。以上研究均得到了较好的预测结果。但目前将 RFE-RF 结合模型用于多种土壤肥力属性的预测和制图的研究还较少,探讨其在土壤肥力属性预测及制图中的可行性具有重要意义。

江浙沪地区是中国重要的农产品生产基地和著名的商品粮基地。由于江浙沪地区地理环境多样,土壤肥力具有高度的空间异质性和依赖性。目前关于该地区详细土壤肥力的空间分布信息了解较少。如何对该地区农田土壤肥力空间分布信息进行准确预测十分必要。

本研究以江苏省、浙江省、上海市农田土壤为研究对象,选择 pH、有机碳、全氮、全磷、全钾、硝态氮、铵态氮、速效钾、有效磷、交换性钙和交换性镁作为农田表层土壤肥力属性,选择气候、地形、植被、土壤特性、农业活动 5 大类 22 种环境协变量,利用 RFE-RF 模型对研究区域主要肥力属性进行制图。该研究旨在基于样点数据,将自然环境与人类活动变量进行筛选,组成最佳环境协变量组合,实现较精准的土壤肥力属性的空间预测,为大区域

土壤肥力属性预测提供方法参考,为农田土壤管理与决策提供基础信息。

## 1 数据来源与研究方法

### 1.1 研究区概况

江苏省、浙江省和上海市地处长江三角洲,位置介于 116°18'E~123°10'E, 27°02'N~35°20'N。以中国科学院资源与环境数据中心(<https://www.resdc.cn/>)2018 年中国土地利用遥感监测数据中江苏省、浙江省和上海市的一级分类代码 1(耕地)作为研究区。该区具有明显的季风气候特征,整体属于亚热带季风气候,年平均气温在 15~18℃。降水充沛,年均降水量总体在 800~1400 mm,受季风影响,降水集中,光照充足,年日照时数在 1100~2200 h。地貌以平原、丘陵为主。研究区土壤类型主要为初育土、水稻土、半水成土、铁铝土等<sup>[26]</sup>。该区是中国重要的农产品生产基地和著名的商品粮基地,同时也是中国重要的经济中心,人口密度高,人类活动剧烈。该区主要是水稻粳、粳交错区,大面积栽培类型主要有常规粳稻、杂交粳稻和杂交粳稻<sup>[27]</sup>。水稻-小麦是区域内主要的轮作种植制度<sup>[28]</sup>。由于较高的生产集约化程度,长期以来大量的施肥对区域内土壤理化性质造成了影响<sup>[29]</sup>。

### 1.2 采样与调查

采用拉丁超立方布设方法<sup>[30]</sup>设计样点,于 2020 年 9—10 月在研究区进行采样,并用 GPS 记录样点经纬度,获取 52 个样点(图 1)。土壤样品取样采用土钻法,在每个样点建立 10 m×10 m 的方格,进行 0~20 cm 土层五点取样,将样品充分混匀后装袋带回,待风干过筛、清除杂质后进行土壤性质测定。

### 1.3 土壤属性测定

土壤样品经风干磨碎后,测定有机碳、全氮、全磷、全钾、硝态氮、铵态氮、速效钾、有效磷、交换性镁、交换性钙、pH 等属性。

### 1.4 环境协变量数据

农田表层土壤肥力属性受初始土壤条件、自然环境因素和农业活动相互作用的控制。据此,获取 5 类 22 种环境协变量作为初始环境协变量,进行特征筛选后选择最优变量作为环境预测变量。所有环境协变量及土壤属性空间预测结果分辨率均为 1 km。

#### 1.4.1 气候数据

从中国科学院资源环境科学与数据中心(<https://www.resdc.cn/>)获得 2 个气候变量,多年平

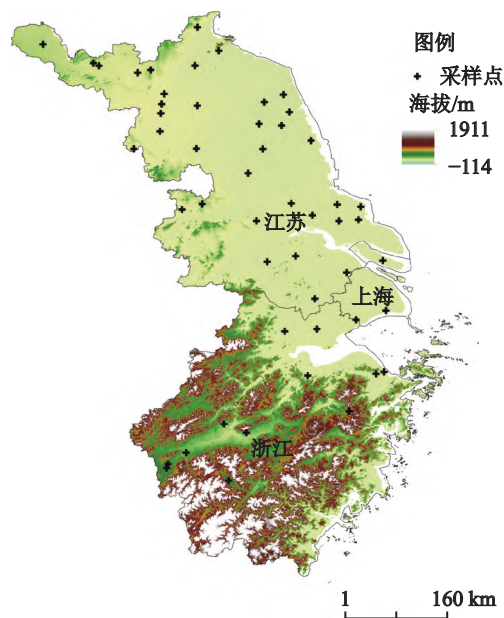


图1 研究区及采样点分布

Fig.1 Study area and distribution of sampling points

均降水量和多年平均温度。由 1980—2015 年的逐年平均值计算生成。从国家科技基础条件平台——国家地球系统科学数据中心([www.resdc.cn/](http://www.resdc.cn/))——黄土高原分中心获得 2020 年当年降水量及平均气温,空间分辨率均为 1 km。

#### 1.4.2 地形数据

从美国宇航局<sup>①</sup>航天飞机雷达地形任务(SRTM)网站获得 90 m 空间分辨率数字高程模型(DEM)。利用 SAGAGIS7.9.0 提取坡度(Slope)、高程(Elevation)、平面曲率(Plan curvature)、剖面曲率(Profile curvature)和地形湿度指数(Topographic Wetness Index, TWI)。

#### 1.4.3 植被数据

归一化植被指数(NDVI)来源于 MODIS 陆地产品 16 d 合成植被指数(MOD13Q1),空间分辨率为 250 m,时间为 2020 年全年,共 23 期,编码为 NDVI2020001-NDVI2020353,为防止该类数据过多对模型构建造成影响,所以在输入模型前根据其土壤属性的相关性对 NDVI 数据进行筛选。相关性结果表明,每期 NDVI 与有机碳和全氮的相关性均不显著,所以对于有机碳和全氮,筛选出相关系数大于 0.15 的 NDVI 期数。对于其他土壤肥力属性,筛选出显著性( $P < 0.05$ )的 NDVI 期数,从筛选

出的 NDVI 期数中删除 NDVI 不同期数之间相关系数  $> 0.65$  且与土壤性质相关系数较小的 NDVI 期数。

净初级生产力(NPP)来源于 MODIS 陆地产品净初级生产力年产量(MOD17A3H),空间分辨率为 500 m,时间为 2019 年。

#### 1.4.4 土壤类型和机械组成

中国土壤质地空间分布数据来源于中国科学院资源环境科学与数据中心,其根据 1:100 万土壤类型图和第二次土壤普查获取到的土壤剖面数据编制而成,是根据砂粒、粉粒、黏粒含量进行土壤质地划分。数据采用国际制,分为砂粒、粉粒、粘粒 3 类,每一类数据均通过百分比来反应不同质地颗粒的含量。

中国土壤类型数据源于中国科学院资源与环境科学与数据中心,该数据根据全国土壤普查办公室 1995 年编制并出版的《1:100 万中华人民共和国土壤图》<sup>[31]</sup> 数字化生成,采用土壤发生分类系统,基本制图单元为亚类。

#### 1.4.5 农田活动变量

农田熟制数据来自于中国科学院资源与环境科学与数据中心的“中国农田熟制遥感监测数据集”<sup>[32]</sup>,该数据通过 1 km SPOT-VGT 旬 NDVI 数据生成作物生长植被指数曲线,并用峰值特征点反演方法提取所得。此外,收集了 2020 年各省统计年鉴<sup>[33-35]</sup> 的县级单元的农业机械动力、每公顷农作物化肥施用量、农村用电量及农业总产值数据,将其赋值到各县得到空间化数据。其中,各县每公顷农作物化肥施用量按当年农业化肥施用量(折纯)除以农作物总播种面积计算。

### 1.5 递归特征消除特征筛选-随机森林土壤肥力属性制图

本文使用 R 语言中的 Caret 软件包和 Random Forest 软件包实现 RFE-RF 模型。利用 Caret 包中的 rfeControl 函数对 RF 模型中 2 个重要参数进行设定。这两个参数分别是 mtry、ntree。mtry 是每个节点上随机选择的协变量的数量,通常设置为环境协变量总数的 1/3; ntree 是 RF 中决策树的数量。RFE-RF 的预测性能对 mtry、ntree 的变化很敏感。为获取最优参数,本研究通过迭代法将 mtry 值定为环境协变量总个数的 1/3, ntree 值以 100 为初

① <https://earthdata.nasa.gov> [2022-01-10]



始值、100 为增量,递增至 20 000,选取袋外数据错误率稳定时的  $ntree$  值。

### 1.6 预测精度评价

因为本研究样点数量有限,为降低模型的随机性,采取十折交叉验证对模型进行评估。十折交叉验证是按随机方法对所有样点进行均分,获取 10 个子集。每一个子集依次作为模型的验证集,剩余的 9 份子集作为模型的训练集,10 次验证精度(评价指标)的均值作一次模型运行的精度估计。为降低样点选取和预测模型的随机性,重复进行 100 次十折交叉验证,取 100 次精度评价指标的平均值作为模型精度结果<sup>[23]</sup>。

模型精度评价选用均方根误差(RMSE)、平均绝对误差(MAE)以及决定系数( $R^2$ )3 个指标,其中,MAE 和 RMSE 越小,表明预测精度越高, $R^2$  越大,表明预测精度越高<sup>[10]</sup>。

## 2 结果与分析

### 2.1 土壤肥力属性统计分析

研究区土壤表层(0~20 cm)肥力属性统计特征,包含每种土壤肥力属性的最大最小值、平均值、标准偏差、偏度和峰度,如表 1 所示。根据全国第二次土壤普查养分分级标准<sup>[36]</sup>,有机碳(10.48 g/kg)和全钾(14.97 g/kg)为 4 级,处于中度贫瘠化水平。全氮(1.39 g/kg)、全磷(0.71 g/kg)和有效磷(18.26 mg/kg)为 3 级,处于中等水平。速效钾(385.07 mg/kg)为 1 级,含量丰富。另外,参照前期相关研究<sup>[37-38]</sup>,交换性钙(>1 g/kg)和交换性镁(>0.12 g/kg)含量都是丰富的。

### 2.2 环境协变量与土壤肥力属性的相关性

通过计算环境协变量与土壤肥力属性的相关性关系得到以下结论,有机碳和全氮与环境因子的相关关系普遍较弱,有机碳与平面曲率呈显著负相关( $P<0.05$ ),全氮与地形湿度指数呈显著正相关( $P<0.05$ )。除有机碳和全氮外,其余土壤肥力属性均与气候变量有显著相关关系( $P<0.05$ )。此外,全磷、全钾、pH、交换性钙、交换性镁都与海拔有显著相关关系( $P<0.01$ ),其中全钾与海拔呈正相关关系。全钾还与剖面曲率呈显著负相关( $P<0.001$ ),与土壤类型、坡度、粘粒含量及净初级生产力呈显著正相关,与农村用电量呈显著负相关( $P<0.05$ )。铵态氮与坡度、地形湿度指数以及净初级生产力呈显著正相关( $P<0.01$ )。pH 与剖面曲率呈显著正相关( $P<0.01$ ),而与坡度及净初级生产力呈显著负相关( $P<0.05$ )。交换性钙则与熟制呈显著正相关( $P<0.05$ )。

### 2.3 RFE-RF 模型筛选所得最佳变量及模型预测精度

RFE-RF 模型筛选出的最佳变量数量及预测精度如表 2 所示,最佳变量组合如表 3 所示。11 个模型的  $R^2$  值在 0.27~0.53。其中 pH、交换性钙、交换性镁的预测模型筛选出的最佳组合变量数量较少,且  $R^2$  可以达到 0.47 以上,模型结果较为准确。而硝态氮的模型  $R^2$  较低(0.27),可能由于其空间分布机制与环境协变量的关系较为复杂,本研究未涉及到与其含量相关性更强的环境协变量。

多年及当年平均气温、多年平均降水、当年降水、海拔、NDVI、净初级生产力、剖面曲率是多个属性排名靠前的最重要变量。其中,铵态氮、速效

表 1 江浙沪农田 0~20 cm 表层土壤样点肥力属性统计

Table 1 Fertility attributes of 0-20 cm surface soil in croplands in Jiangsu, Zhejiang and Shanghai

土壤肥力属性/单位	最小值	最大值	平均值	标准偏差	偏度	峰度	变异系数
pH值	5.26	8.42	7.33	0.97	-0.90	-0.60	0.13
有机碳/(g/kg)	1.81	20.54	10.48	4.12	0.42	-0.37	0.39
全氮/(g/kg)	0.30	3.12	1.39	0.57	1.10	1.67	0.41
全磷/(g/kg)	0.19	1.32	0.71	0.27	0.18	-0.60	0.38
全钾/(g/kg)	9.46	34.51	14.97	4.69	2.25	6.47	0.31
硝态氮/(mg/kg)	0.56	11.64	3.37	2.94	1.27	0.68	0.87
铵态氮/(mg/kg)	1.53	33.73	5.93	6.07	2.89	9.60	1.02
有效磷/(mg/kg)	1.96	66.47	18.26	13.79	1.37	2.12	0.72
速效钾/(mg/kg)	144.56	1 535.75	385.07	243.79	2.73	10.06	0.63
交换性钙/(g/kg)	2.52	32.71	14.42	9.64	0.57	-1.26	0.67
交换性镁/(g/kg)	0.24	3.30	1.06	0.55	1.32	4.17	0.52

表 2 RFE-RF 模型筛选出的最佳变量数量及预测精度

Table 2 The optimal number of factors selected by the RFE-RF model and the prediction accuracy

土壤肥力属性	最佳组合变量数量	RMSE	$R^2$	MAE
pH值	3	0.78	0.47	0.60
有机碳	17	3.76	0.34	3.14
全氮	2	0.58	0.32	0.47
全磷	12	0.22	0.42	0.18
全钾	19	3.77	0.43	2.90
铵态氮	17	5.25	0.41	3.83
硝态氮	6	2.82	0.27	2.32
有效磷	15	12.38	0.33	9.69
速效钾	17	192.10	0.47	137.90
交换性钙	4	7.11	0.53	5.78
交换性镁	4	0.45	0.45	0.35

表 3 RFE-RF 模型筛选出的最佳变量组合

Table 3 The best factor combination selected by the RFE-RF model

土壤肥力属性	最佳组合变量(按照%IncMSE由大到小排序)
pH值	多年平均降水、剖面曲率、多年平均气温
有机碳	砂粒含量、NDVI2020257、当年平均气温、粉粒含量、NDVI2020081、NDVI2020337、粘粒含量、多年平均降水、地形湿度指数、NDVI2020161、土壤类型、多年平均气温、每公顷农用地肥施用量、坡度、农业总产值、海拔、当年降水
全氮	坡度、地形湿度指数
全磷	NDVI2020145、多年平均气温、当年平均气温、多年平均降水、剖面曲率、NDVI2020081、当年降水、农业总产值、农村用电量、NDVI2020289、净初级生产力、海拔
全钾	NDVI2020257、多年平均气温、海拔、多年平均降水、净初级生产力、当年平均气温、粉粒含量、NDVI2020353、农村用电量、粘粒含量、砂粒含量、坡度、土壤类型、剖面曲率、NDVI2020081、NDVI2020305、每公顷农用地肥施用量、当年降水、农业机械总动力
铵态氮	当年降水、净初级生产力、多年平均气温、多年平均降水、剖面曲率、地形湿度指数、NDVI2020305、当年平均气温、NDVI2020289、坡度、海拔、NDVI2020145、每公顷农用地肥施用量、粘粒含量、农业机械总动力、农业总产值、土壤类型
硝态氮	地形湿度指数、NDVI2020289、多年平均气温、多年平均降水、当年平均气温、NDVI2020305
有效磷	多年平均降水、多年平均气温、当年平均气温、当年降水、NDVI2020257、海拔、NDVI2020225、土壤类型、净初级生产力、坡度、农业总产值、NDVI2020081、农业机械总动力、农村用电量、每公顷农用地肥施用量
速效钾	多年平均降水、当年降水、多年平均气温、海拔、NDVI2020081、NDVI2020257、土壤类型、NDVI2020353、粘粒含量、当年平均气温、砂粒含量、粉粒含量、净初级生产力、NDVI2020225、NDVI2020193、剖面曲率、NDVI2020065
交换性钙	多年平均气温、NDVI2020145、当年降水、多年平均降水
交换性镁	当年降水、NDVI2020257、海拔、多年平均降水

钾、有效磷、交换性钙、交换性镁、pH 的变量重要性排序中,气候变量排名首位。全氮预测模型的变量重要性排序中,地形湿度指数和坡度占据前两位。对硝态氮而言,地形湿度指数排在首位,气候变量和植被变量重要性排名在前 5 位。有机碳预测模型的变量重要性排序中,砂粒含量、粉粒含量、NDVI 和当年平均气温排名靠前。全钾的预测模型中,NDVI、净初级生产力、海拔、多年平均气温、多年平均降水排在前列。表征人类农业活动的变量在有机碳、全磷、全钾、铵态氮和有效磷的预测模型中均

有体现,其中,全钾、铵态氮及有效磷模型中有 3 个及以上农业活动变量入选最佳因子组合,主要是每公顷农用地肥施用量和农业机械总动力。

## 2.4 土壤肥力属性空间分布制图结果

选择  $R^2$  值最高的最佳变量组合模型对土壤肥力属性空间分布进行预测,制图结果如图 2 所示。模型预测的有机碳空间分布表现为江苏中部、环太湖地区、浙江杭嘉湖平原、宁绍平原及上海南部预测值较高,浙江西南部、江苏西南部和滨海平原地区较低。模型预测的土壤全氮的空间分布破碎,

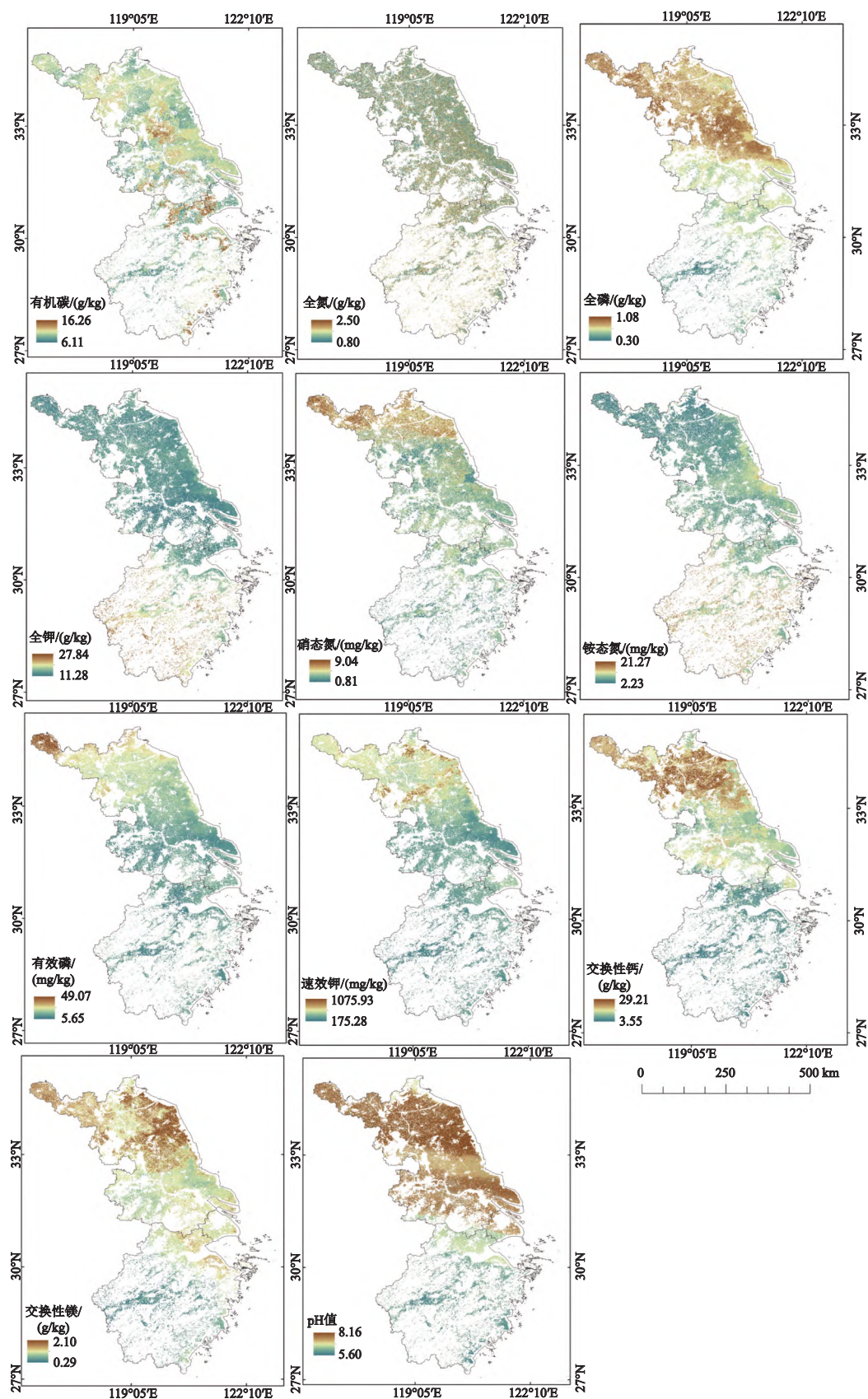


图2 研究区表层土壤肥力属性预测结果

Fig.2 Prediction results of surface soil fertility attributes



且与地形湿度指数的变化特征大致相符。模型预测的土壤全钾的空间分布表现出较强的南北差异:在江苏省、上海及浙江北部的区域含量较低,在浙江西部与东南部的宁绍平原含量较高。模型预测的土壤全磷的空间分布表现为中北部高,南部低,在金衢盆地达到最小值,且与多年平均气温和多年平均降水的变化特征大致相符。土壤铵态氮空间分布,大致表现为中北部高,南部低,但金衢盆地较低。土壤硝态氮的空间分布表现为北部高,中南部低。土壤有效磷的空间分布表现为北部高,中南部低,黄淮平原北部达到最大。土壤速效钾表现为由北向南降低,其中江淮平原北部和东部滨海平原北部的速效钾含量较高。土壤交换性钙的空间分布表现为中北部高,南部低。土壤交换性镁的空间分布表现为中北部高,南部低,且均与多年平均气温和多年平均降水的变化特征大致相符。 $\text{pH}$ 值大致呈现“南酸北碱”的分布特征,江苏中部的江淮平原偏中性。

### 3 讨论

#### 3.1 土壤肥力环境变量的重要性

已有研究表明,土壤属性的重要环境变量因地区和尺度而异。本研究的结果也表明了不同土壤属性起重要环境因子的差异性。交换性钙、交换性镁和  $\text{pH}$  的重要变量主要是气候和地形,全氮与地形变量中的坡度和地形湿度指数关系密切。此外,计算了其他土壤肥力属性的气候、地形、土壤、农业活动和植被 5 类环境因子的相对重要性(表 4),即通过将所有变量的相对重要性之和达到 100% 来换算<sup>[39]</sup>。可以看出,除有机碳、全钾外,气候变量的相对重要性都突破了 40%,此外植被和地形具有重要作用,而土壤自身性质对有机碳、全钾和速效钾具

有重要作用,特别是其对有机碳重要性要高于其他类型变量。表征人类农业活动的变量在有机碳、全磷、全钾、铵态氮和有效磷的预测模型中体现了一定作用。但是,本研究有关农业活动的数据来自于各地统计年鉴区县级统计数据,相对粗糙。若未来研究中有更为准确细致的农业活动数据(如:较多数量的农场记录等),或可更好地进行表层土壤肥力属性的推测制图。

过去在该区的研究以土壤有机碳或有机质为主,也涉及土壤速效钾、有效磷等其他较为重要的肥力属性。例如,Guo 等<sup>[40]</sup>对于中国长江中下游平原驱动稻田土壤有机质变化的环境和人为因素的研究中发现,土壤特性、气候变量、地形变量和农业管理措施是该区域稻田土壤有机质变化的主要影响因素。刘书田在中国农田土壤有机碳时空分布规律及影响因素研究<sup>[41]</sup>中也指出,温度、土壤粘粒、砂粒、粉粒含量是影响农田土壤有机碳时空分布的重要变量,这与本研究的结论较为一致。气候变量和地形变量是本研究区土壤有效磷含量空间分布最重要的影响因素,这与曹佳萍<sup>[42]</sup>、杨之江<sup>[43]</sup>等研究结论一致。

#### 3.2 模型性能

RFE-RF 模型不仅确定了每个土壤属性的最佳影响变量子集,而且通过对环境因子筛选有效地降低了模型处理时间。该模型为基于稀疏样点的江苏、浙江和上海地区的农田土壤肥力属性空间预测制图提供了有效的方法。不同土壤肥力属性的最佳预测模型的环境变量有 2~17 个。 $\text{pH}$ 、速效钾、交换性镁和交换性钙的预测模型决定系数( $R^2$ )均在 0.45 以上。王世航等利用递归特征消除等不同特征挖掘方法结合随机森林等模型预测了安徽省土壤  $\text{pH}$ <sup>[24]</sup>,结果显示 RFE-RF 模型的训练集  $R^2$  为 0.62,验证集

表 4 环境协变量在预测研究区土壤肥力属性的最准确模型中的相对重要性

Table 4 Relative importance of environmental covariates in the most accurate models for predicting soil fertility properties in the study area

土壤肥力属性	气候	地形	土壤	农业活动	植被
有机碳	21.15%	9.10%	33.16%	6.66%	29.92%
全磷	48.78%	10.85%	—	8.94%	14.51%
全钾	27.40%	15.84%	19.13%	7.72%	29.91%
硝态氮	45.35%	22.04%	—	—	32.61%
铵态氮	42.77%	24.99%	1.24%	3.53%	27.47%
速效钾	43.93%	9.98%	17.42%	—	28.67%
有效磷	60.45%	11.14%	3.68%	3.42%	21.31%

注:—为无此项。

$R^2$  为 0.73。He 等选用自然变量、物候参数、遥感变量利用 RFE-RF 模型预测了安徽省宣城市的农田土壤有机碳<sup>[23]</sup>, 预测结果显示不同的环境协变量组合对应模型的  $R^2$  也不同(0.14~0.38)。由此可以看出, 在不同的研究区域选择不同的环境协变量预测不同土壤肥力属性时, 模型的  $R^2$  也会出现较大差异。本研究中部分模型(例如, 全氮和硝态氮)的决定系数( $R^2$ )较低可能是所选环境因子分辨率较低或土壤属性具有其他更为重要的影响因子。研究区样点数量偏少也是一个可能原因。

### 3.3 土壤肥力属性空间分布特征分析

模型预测的土壤有机碳、全磷、全钾及 pH 的空间分布与 Liu 等<sup>[44]</sup> 所得表层(0~20 cm)土壤有机碳、全磷、全钾及 pH 的空间分布趋势大致相同。模型预测的有机碳空间分布表现出较强的空间异质性, 主要是因为气候、地形、土壤质地等多因素影响本区域农田土壤有机碳分布。土壤全磷空间分布表现为中北部高, 南部低, 这主要与农田中的磷(尤其是表层磷)极大地受到气候和地形因素的影响有关。研究区南部高温和降水使得土壤中的磷容易风化并释放出来<sup>[45]</sup>。土壤中的钾通常来自于风化。另外, 土壤粘粒含有大量富含钾的粘土矿物, 也影响全钾含量<sup>[46]</sup>。研究区南部的全钾值较高, 与南部年均气温较高、粘粒含量略多有关。模型预测的土壤硝态氮和铵态氮的空间分布, 也与多年平均气温和多年平均降水的变化特征大致相符, 表现为北部高, 中南部低。主要是由于研究区南部相比于北部年均温更高, 高温会促进土壤微生物活动, 进而促进土壤的氮矿化<sup>[47]</sup>。同时, 南部降水更多、海拔较高, 淋溶作用更加显著, 尤其是硝态氮易溶于水的性质, 导致其极易随水下渗至土壤环境之外<sup>[48-49]</sup>。模型推测土壤有效磷高值区主要分布在年均降水量低且地势平坦的江苏北部。研究区南部年均温较高, 加快了矿物风化速率和脱硅富铁铝作用, 但年均降水量较多地势较高, 造成风化速率小于淋失速率, 使得施肥中磷素易淋失, 故有效磷含量较低。模型预测的土壤速效钾的空间分布与张丽君等<sup>[50]</sup> 的研究发现较为一致, 耕地土壤速效钾含量随海拔下降而增加, 研究区中北部大范围的平原区域的土壤速效钾值较高。另外, 研究区南部高温高湿的环境也使其土壤中的速效钾含量较少。模型预测的土壤交换性钙和交换性镁的空间分布, 与多年平均气温和多年平均降水的分布大致相同, 浙江省内也可以观察出其随

海拔的升高而降低。这与危锋<sup>[51]</sup> 和曹榕彬<sup>[38]</sup> 等的研究结果相似。在较高海拔的耕地土壤的交换性钙、交换性镁, 易随水向低海拔或者地形较为平缓区域迁移。而江苏省中北部地区的年均温度较低, 降水量较少且地形较为平坦, 故交换性钙镁的含量较高。模型预测的 pH 大致呈现“南酸北碱”的分布特征, 由于湿润气候条件下土壤普遍发生自然酸化<sup>[52]</sup>。

## 4 结论与展望

基于实测表层土壤样点, 利用 RFE-RF 模型对江苏、浙江和上海农田表层(0~20 cm)的土壤肥力属性含量进行了预测制图研究。发现 11 个模型筛选出的环境协变量主要集中在气候、地形与植被, 表征人类农业活动的变量也筛选入有机碳、全磷、全钾、铵态氮和有效磷的预测模型中。利用 RFE-RF 预测土壤肥力的决定系数( $R^2$ )均在 0.27~0.53。其中, pH、速效钾、交换性镁和交换性钙的预测模型决定系数( $R^2$ )较高, 均在 0.45 以上。表明建立的预测模型可为基于稀疏采样点的江浙沪地区农田表层土壤肥力属性含量进行有效预测, 还可望用于快速而较准确绘制全国其他地区省市尺度表层土壤肥力属性空间分布图。

后续工作中可以使用遥感时序数据结合多年份的统计数据等方法, 生成更丰富的农业活动影响变量数据, 为农田土壤肥力和碳模拟提供更为丰富的人为活动变量; 并增加采样点, 以期获取更多土壤特征信息, 使得土壤肥力属性空间分布预测更加准确可靠。

## 参考文献(References):

- [1] Merrill S D, Liebig M A, Tanaka D L et al. Comparison of soil quality and productivity at two sites differing in profile structure and topsoil properties[J]. *Agriculture, Ecosystems & Environment*, 2013, 179: 53-61.
- [2] 姚利鹏, 黄标, 孙维侠. 鄱阳湖流域典型区限制水稻产量的土壤因素分析[J]. *土壤*, 2015, 47(4): 674-681. [Yao Lipeng, Huang Biao, Sun Weixia. Limiting soil factors on rice yield in typical area of Poyang Lake Region, China. *Soil*, 2015, 47(4): 674-681.]
- [3] Kaniu M I, Angeyo K H. Challenges in rapid soil quality assessment and opportunities presented by multivariate chemometric energy dispersive X-ray fluorescence and scattering spectroscopy[J]. *Geoderma*, 2015, 241-242: 32-40.
- [4] 陈鑫, 刘勤, 张刚. 太湖地区不同轮作模式对土壤肥力和水稻产量的影响[J]. *江苏农业学报*, 2021, 37(4): 874-883. [Chen Xin,



- Liu Qin, Zhang Gang. Effects of different crop rotation modes on soil fertility and rice yield in Taihu Region. *Jiangsu Journal of Agricultural Sciences*, 2021, 37(4): 874-883.]
- [5] Zhang Huan, Wang Dengfeng, Su Baowei et al. Distribution and determinants of organic carbon and available nutrients in tropical paddy soils revealed by high-resolution sampling[J]. *Agriculture, Ecosystems & Environment*, 2021, 320: 107580.
- [6] Wang Huifeng, Yao Lipeng, Huang Biao et al. An integrated approach to exploring soil fertility from the perspective of rice (*Oryza sativa* L.) yields[J]. *Soil and Tillage Research*, 2019, 194: 104332.
- [7] Budiman Minasny, McBratney A B. Digital soil mapping: A brief history and some lessons[J]. *Geoderma*, 2016, 264: 301-311.
- [8] Marine Lacoste, Budiman Minasny, Alex McBratney et al. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape[J]. *Geoderma*, 2014, 213: 296-311.
- [9] Bratney A B M, Mendonça Santos M L, Minasny B. On digital soil mapping[J]. *Geoderma*, 2003, 117(1-2): 3-52.
- [10] Yang Lin, Cai Yanyan, Zhang Lei 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.
- [11] Alexandre M. J.-C. Wadoux. Using deep learning for multivariate mapping of soil with quantified uncertainty[J]. *Geoderma*, 2019, 351: 59-70.
- [12] Paul S S, Dowell L, Coops N C et al. Tracking changes in soil organic carbon across the heterogeneous agricultural landscape of the Lower Fraser Valley of British Columbia[J]. *Science of the Total Environment*, 2020, 732: 138994.
- [13] 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. [Zhou Yang, Zhao Xiaomin, Guo Xi. Prediction of total nitrogen distribution in surface soil based on multi-source auxiliary variables and random forest approach. *Acta Pedologica Sinica*, 2022, 59(2): 451-460.]
- [14] 俞海, 黄季焜, Scott Rozelle, 等. 土壤肥力变化的社会经济影响因素分析[J]. *资源科学*, 2003, 25(2): 63-72. [Yu Hai, Huang Jikun, Scott Rozelle et al. Impact of socio-economic factors on soil fertility over time. *Resources Science*, 2003, 25(2): 63-72.]
- [15] Deng Xunfei, Chen Xiaojia, Ma Wanzhu et al. Baseline map of organic carbon stock in farmland topsoil in East China[J]. *Agriculture, Ecosystems & Environment*, 2018, 254: 213-223.
- [16] Deng Xunfei, Ma Wanzhu, Ren Zhouqiao et al. Spatial and temporal trends of soil total nitrogen and C/N ratio for croplands of East China[J]. *Geoderma*, 2020, 361: 114035.
- [17] Brandon Heung, Bulmer C E, Margaret G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach[J]. *Geoderma*, 2014, 214-215: 141-154.
- [18] Hengl T, Mendes de Jesus J, Heuvelink G B M et al. Soil grids 250 m: Global gridded soil information based on machine learning[J]. *PLoS One*, 2017, 12(2): e0169748.
- [19] Zhang Lei, Yang Lin, Ma Tianwu et al. A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data[J]. *Geoderma*, 2021, 384: 114809.
- [20] Zhang Lei, Yang Lin, Cai Yanyan et al. A multiple soil properties oriented representative sampling strategy for digital soil mapping[J]. *Geoderma*, 2022, 406: 115531.
- [21] Leo Breiman. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [22] Chen Qi, Meng Zhaopeng, Liu Xinyi et al. Decision variants for the automatic determination of optimal feature subset in RF-RFE[J]. *Genes*, 2018, 9(6): 301.
- [23] He Xianglin, Yang Lin, Li Anqi et al. Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images[J]. *Catena*, 2021, 205: 105442.
- [24] 王世航, 卢宏亮, 赵明松, 等. 基于不同特征挖掘方法结合广义提升回归模型估测安徽省土壤pH[J]. *应用生态学报*, 2020, 31(10): 3509-3517. [Wang Shihang, Lu Hongliang, Zhao Mingsong et al. Assessing soil pH in Anhui Province based on different features mining methods combined with generalized boosted regression models. *Chinese Journal of Applied Ecology*, 2020, 31(10): 3509-3517.]
- [25] Tayebi Mahboobeh, Fim Rosas Jorge Tadeu, Mendes Wandererson de Sousa et al. Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series[J]. *Remote Sensing*, 2021, 13(11): 2223.
- [26] 杜国华, 张甘霖, 龚子同. 长江三角洲水稻土主要土种在中国土壤系统分类中的归属[J]. *土壤*, 2007(5): 684-691. [Du Guohua, Zhang Ganlin, Gong Zitong. Placement of paddy soils of the Yangtze Delta in the Chinese soil taxonomy. *Soils*, 2007(5): 684-691.]
- [27] 刘明, 李岩, 郭贵华, 等. 长江下游不同类型水稻分蘖期耐淹能力比较[J]. *应用生态学报*, 2015, 26(5): 1373-1381. [Liu Ming, Li Yan, Guo Guihua et al. Comparison on submergence tolerance of different type rice at tillering stage in lower reaches of Yangtze River. *Chinese Journal of Applied Ecology*, 2015, 26(5): 1373-1381.]
- [28] 张志毅, 范先鹏, 夏贤格, 等. 长三角地区稻麦轮作土壤养分对秸秆还田响应Meta分析[J]. *土壤通报*, 2019, 50(2): 401-406. [Zhang Zhiyi, Fan Xianpeng, Xia Xiangge et al. Meta-analysis of soil basic nutrients response to straw returning in rice-wheat rotation system. *Chinese Journal of Soil Science*, 2019, 50(2): 401-406.]
- [29] 吴新民, 潘根兴, 李恋卿. 长江三角洲土壤质量演变趋势分析[J]. *地理与地理信息科学*, 2006, 21(3): 88-91. [Wu Xinmin, Pan Genxing, Li Lianqing. Study on soil quality change in the Yangtze River Delta. *Geography and Geo-Information Science*, 2006, 21(3): 88-91.]

- [30] Budiman Minasny, McBratney A B. Mechanistic soil-landscape modelling as an approach to developing pedogenetic classifications[J]. *Geoderma*, 2006, 133(1-2): 138-149.
- [31] 刘菊兰. 中华人民共和国1:100万土壤图[M]. 北京: 中国出版年鉴社, 1995. [Liu Julan. 1:1 000 000 soil map of the People's Republic of China. Beijing: China Yearbook Press, 1995.]
- [32] 徐新良. 中国农田熟制遥感监测数据集[M]. 北京: 中国科学院资源环境科学数据中心, 2017. [Xu Xinliang. Remote sensing monitoring data set of farmland maturity in China. Beijing: Resource and Environmental Science Data Center of Chinese Academy of Sciences, 2017.]
- [33] 江苏省统计局, 国家统计局江苏调查总队. 江苏统计年鉴, 2020[M]. 北京: 中国统计出版社, 2020. [Jiangsu Provincial Bureau of Statistics, Survey Office of the National Bureau of Statistics in Jiangsu. Jiangsu Statistical Yearbook, 2020. Beijing: China Statistics Press, 2020.]
- [34] 浙江省统计局, 国家统计局浙江调查总队. 浙江统计年鉴, 2020[M]. 北京: 中国统计出版社, 2020. [Zhejiang Provincial Bureau of Statistics, Survey Office of the National Bureau of Statistics in Zhejiang. Zhejiang statistical yearbook—2020. Beijing: China Statistics Press, 2020.]
- [35] 上海市统计局, 国家统计局上海调查总队. 上海统计年鉴, 2020[M]. 北京: 中国统计出版社, 2020. [Shanghai Municipal Statistics Bureau, Survey Office of the National Bureau of Statistics in Shanghai. Shanghai statistical yearbook, 2020. Beijing: China Statistics Press, 2020.]
- [36] 全国土壤普查办公室. 中国土壤[M]. 北京: 中国农业出版社, 1998. [National Soil Survey Office. Soil of China. Beijing: China Agricultural Press, 1998.]
- [37] 张璐, 蔡泽江, 王慧颖, 等. 中国稻田土壤有效态中量和微量元素含量分布特征[J]. *农业工程学报*, 2020, 36(16): 62-70. [Zhang Lu, Cai Zejiang, Wang Huiying et al. Distribution characteristics of effective medium and micronutrient element contents in paddy soils of China. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(16): 62-70.]
- [38] 曹榕彬. 耕地土壤中微量元素含量空间分布及施肥对策[J]. *土壤通报*, 2018, 49(3): 646-652. [Cao Rongbin. Spatial Distribution and fertilization of medium and trace elements in cultivated land. *Chinese Journal of Soil Science*, 2018, 49(3): 646-652.]
- [39] Yang Lin, He Xianglin, Shen Feixue 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.
- [40] Guo Naijia, Shi Xuezheng, Zhao Yongcun et al. Environmental and anthropogenic factors driving changes in paddy soil organic matter: A case study in the middle and lower Yangtze River Plain of China[J]. *Pedosphere*, 2017, 27(5): 926-937.
- [41] 刘书田. 中国农田土壤有机碳时空分布规律及影响因素研究[D]. 长春: 吉林农业大学, 2016. [Liu Shutian. The rule of temporal and spatial distribution of soil organic carbon of crop land and its influencing factors in China. Changchun: Jilin Agricultural University, 2016.]
- [42] 曹佳萍, 张黎明, 邱龙霞, 等. 基于稀疏样点的南方丘陵地区耕地土壤有效磷制图[J/OL]. *中国生态农业学报(中英文)*. <http://kns.cnki.net/kcms/detail/13.1432.S.20211130.1518.002.html>, 2022-01-08(1-13). [Cao Jiaping, Zhang Liming, Qiu Longxia et al. Mapping soil available phosphorus of cultivated land in hilly region of southern China based on sparse samples. *Chinese Journal of Eco-Agriculture*. <http://kns.cnki.net/kcms/detail/13.1432.S.20211130.1518.002.html>, 2022-01-08(1-13).]
- [43] 杨之江, 陈效民, 景峰, 等. 基GIS和地统计学的稻田土壤养分与重金属空间变异[J]. *应用生态学报*, 2018, 29(6): 1893-1901. [Yang Zhijiang, Chen Xiaomin, Jing Feng et al. Spatial variability of nutrients and heavy metals in paddy field soils based on GIS and geostatistics. *Chinese Journal of Applied Ecology*, 2018, 29(6): 1893-1901.]
- [44] Liu Feng, Wu Huayong, Zhao Yuguo et al. Mapping high resolution National Soil Information Grids of China[J]. *Science Bulletin*, 2022, 67(3): 328-340.
- [45] Lin Jinshi, Shi Xuezheng, Lu Xixi et al. Storage and spatial variation of phosphorus in paddy soils of China[J]. *Pedosphere*, 2009, 19(6): 790-798.
- [46] 王箴, 鲁剑巍, 张文君, 等. 田间土壤钾素有效性影响因素及其评估[J]. *土壤*, 2012, 44(6): 898-904. [Wang Zheng, Lu Jianwei, Zhang Wenjun et al. Influential factors on soil available potassium evaluation in agriculture. *Soils*, 2012, 44(6): 898-904.]
- [47] Rustad L, Campbell J, Marion G et al. A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming[J]. *Oecologia*, 2001, 126(4): 543-562.
- [48] 靖彦, 陈效民, 李秋霞, 等. 生物炭对红壤中硝态氮和铵态氮的影响[J]. *水土保持学报*, 2013, 27(6): 265-269. [Jing Yan, Chen Xiaomin, Li Qiuxia et al. Effects of biochar on ammonium nitrogen and nitrate nitrogen in red soil. *Journal of Soil and Water Conservation*, 2013, 27(6): 265-269.]
- [49] Delgado J A. Quantifying the loss mechanisms of nitrogen[J]. *Journal of Soil and Water Conservation*, 2002, 57(6): 389-398.
- [50] 张丽君, 麻万诸, 项佳敏, 等. 浙江省耕地土壤速效钾状况及影响因素分析[J]. *浙江农业科学*, 2020, 61(4): 607-611. [Zhang Lijun, Ma Wanzhu, Xiang Jiamin et al. Status and influencing factors of available potassium in cultivated soil in Zhejiang Province. *Journal of Zhejiang Agricultural Sciences*, 2020, 61(4): 607-611.]
- [51] 危锋, 郝明德. 长期施氮磷化肥对不同种植体系土壤交换性钙分布与累积的影响[J]. *干旱地区农业研究*, 2012, 30(4): 65-69. [Wei Feng, Hao Mingde. Effect of long-term NP chemical fertilization on soil exchangeable Ca distribution and accumulation with different cropping systems in the Loess Plateau. *Agricultural Research in the Arid Areas*, 2012, 30(4): 65-69.]
- [52] 朱小琴, 孙维侠, 黄标, 等. 长江三角洲城乡交错区农业土壤pH特征及影响因素探讨——以江苏省无锡市为例[J]. *土壤学报*,

2009, 46(4): 594-602. [Zhu Xiaoqin, Sun Weixia, Huang Biao et al. Characteristics and influencing factors of agricultural soil

pH in the Yangtze River Delta: A case study of Wuxi City, Jiangsu Province. *Acta Pedologica Sinica*, 2009, 46(4): 594-602.]

## Digital mapping of soil fertility attributes in croplands in Jiangsu, Zhejiang and Shanghai based on recursive feature elimination-random forest model

Li Anqi, Yang Lin, Cai Yanyan, Zhang Lei, Huang Haili, Wu Qi, Wang Wenqi

(School of Geography and Ocean Science, Nanjing University, Nanjing 210023, Jiangsu, China)

**Abstract:** Obtaining quantitatively characterized spatial information of soil fertility is of great significance for improving soil quality, precision agricultural management and sustainable agricultural development. In this study, we selected croplands in Jiangsu Province, Zhejiang Province and Shanghai as the research areas, climate, topography, vegetation and soil properties were selected as natural environment variables. The total power of agricultural machinery, the amount of agricultural chemical electricity consumption is used as agricultural activity variables, combining the above two types of environmental covariates are used as environmental covariates. The original environmental covariates were screened by the recursive feature elimination (RFE) method, and the optimal variables combination after screening was used as the independent variable to establish the spatial distribution prediction model based on random forest (RF) of surface soil pH, organic carbon, total nitrogen, total phosphorus, total potassium and nitrogen, nitrate nitrogen, ammonium nitrogen, available phosphorus, available potassium, exchangeable calcium and exchangeable magnesium in the study area. The importance of environmental covariates was ranked and digital soil mapping was performed, validated with 100 replicates of ten-fold cross-validation. The results show that: 1) The types of environmental covariates screened out by the 11 models mainly focus on climate, topography and vegetation variables, while the variables representing human agricultural activities play an important role in the prediction of organic carbon, total phosphorus, total potassium, ammonium nitrogen and available phosphorus. 2) The coefficients of determination ( $R^2$ ) of the 11 models were between 0.27 and 0.53, and the coefficients of determination ( $R^2$ ) of the prediction models for pH, available potassium, exchangeable magnesium and exchangeable calcium were all above 0.45. The recursive feature elimination-random forest model (RFE-RF) proposed in this article can be used to map the main soil fertility attributes in croplands and provide the necessary basis for the spatial distribution of soil fertility attributes for agricultural production.

**Key words:** recursive feature elimination; Random Forest; soil fertility attributes; croplands; digital soil mapping; Jiangsu, Zhejiang and Shanghai