

目次

中国城市碳排放强度的时空演变、动态跃迁及收敛趋势 杨清可, 王磊, 朱高立, 李颖, 范业婷, 王雅竹 (1869)

交通运输业碳排放效率时空演变及趋势预测 郑琰, 蒋雪梅, 肖玉杰 (1879)

基于扩展STIRPAT模型LMDI分解的碳排放脱钩因素 张江艳 (1888)

基于LEAP模型的工业园区碳达峰路径:以南京某国家级开发区为例 李慧鹏, 李荔, 殷茵, 何文大, 宿杰, 赵秋月 (1898)

高校碳排放核算与分析:以北京A高校为例 曹睿, 封莉, 张立秋 (1907)

北京市制造业减排降碳协同效应分析和驱动因素 俞珊, 韩玉花, 牟洁, 张双, 张增杰 (1917)

碳排放权交易价格与全要素生产率:来自中国的证据 吴雪萍 (1926)

中国城市PM_{2.5}和PM₁₀时空分布特征和影响因素分析 李江苏, 段良荣, 张天娇 (1938)

2017~2021年苏皖鲁豫交界区域PM_{2.5}和O₃时空变化特征及影响因素 陈伟, 徐学哲, 刘文清 (1950)

疫情管控期西安PM_{2.5}和O₃污染特征及成因分析 原晓红, 张强, 李琦, 谢文豪, 刘跃廷, 樊亭亭, 姜旭朋 (1963)

苏南五市秋冬季PM_{2.5}化学组成特征和空间差异 冯蔚, 丁峰, 尚玥, 谢鸣捷 (1975)

湖北咸宁细颗粒物PM_{2.5}来源 罗怡, 朱宽广, 陈璞琬, 田军, 谢旻, 战杨志豪, 赵润琪 (1983)

邢台氨减排对京津冀PM_{2.5}改善的溢出效益 边泽君, 闻超玉, 郎建垒, 范晓茜, 夏祥晨, 周颖 (1994)

聊城市冬季PM_{2.5}载带金属元素污染特征、风险评价及来源分析 张敬巧, 朱瑶, 曹婷, 燕丽, 王淑兰, 刘铮 (2003)

高原城市拉萨典型VOCs排放源成分谱特征 郭淑政, 叶春翔, 林伟立, 陈熠, 曾立民, 尹晴晴, 刘雪莉 (2011)

北京市工业园区VOCs污染特征及健康风险评估案例:高新技术产业的环境影响 王洁, 姚震, 王敏燕, 陈速敏, 龙腾, 王海滨, 李红, 郭秀锐, 郝江虹, 聂磊 (2019)

高温极端天气影响下的成都平原一次典型臭氧污染过程分析 雷丽娟, 张懿, 罗伊娜, 张潇, 冯森 (2028)

水环境中抗病毒毒物的存在、行为与风险 葛林科, 李璇艳, 曹胜凯, 郑金帅, 张蓬, 朱超, 马宏瑞 (2039)

黄河小浪底水库地表水中重金属的时空变化与概率健康风险 王亮, 邓雪娇, 王潇磊, 李明, 刘奕尧, 姜亚敏, 涂响, 张坤锋 (2054)

藏东多曲河流域总富集水化学特征及控制因素 李敬杰, 连晟, 王明国, 张智印, 张涛 (2067)

西北内陆区降水稳定同位素时空分布特征及其水汽来源 张炎炎, 辛存林, 郭小燕, 张博, 陈宁, 史延飞 (2080)

基于水化学和氢氧同位素的泰安城区岩溶地下水补给来源及演化过程 孟令华 (2096)

丹江流域山区地表水-地下水水化学特征及其影响因素 张子燕, 伏永朋, 王宁涛, 谭建民, 刘亚磊 (2107)

金塔盆地鸳鸯池灌区地下水水化学特征及控制因素 王晓燕, 韩双宝, 张梦南, 尹德超, 吴呈, 安永会 (2118)

德阳市平原区浅层地下水水化学特征与健康风险评价 刘楠, 陈盟, 高东东, 吴勇, 王楠楠 (2129)

典型城市河网沉积物微塑料时空分布特征 许万璐, 范一凡, 钱新 (2142)

抚仙湖流域尺度氮排放清单构建及关键源解析 王延杰, 梁启斌, 王艳霞, 侯磊, 陈奇伯, 王伟, 李晓琳, 高俊淑 (2150)

过氧化钙/海泡石海藻酸钠缓释凝胶复合材料的制备及其对内源磷的控制性能 曲思彤, 单苏洁, 王崇铭, 吴玲子, 李大鹏, 黄勇 (2160)

矿物超细颗粒的形成机制、结构特征及其环境行为和效应 刘振海, 张展华, 袁语欣, 朱盼盼, 陈威, 张彤 (2171)

生物炭固定化菌复合材料在环境修复中的应用研究进展 孙淑玉, 黄梦鑫, 孔强, 张焕新, 刘继伟 (2185)

微塑料对沸石吸附水体氨氮的影响及其机制 练建军, 谢诗婷, 吴培, 孟冠军, 陈波 (2195)

紫外老化作用对纳米生物炭吸附环丙沙星的影响机制 马锋锋, 薛之一, 赵保卫 (2203)

土地利用影响下永定河流域浮游植物群落与环境因子响应 郭善嵩, 胡恩, 丁一桐, 张嘉渭, 孙长顺, 卢悦, 潘保柱 (2211)

鄱阳湖湿地细菌群落多样性和可培养细菌功能基因丰度 喻江, 王淳, 龙永, 刘贵花, 李春杰, 范国权, 于镇华 (2223)

典农河沉积物细菌群落结构特征及其与重金属的关系 刘双羽, 蒙俊杰, 邱小琼, 周瑞娟, 李霖 (2233)

污水处理厂尾水排放对受纳河流细菌和真菌微生物群落的影响 郭有顺, 余仲, 郝文彬, 孟凡刚 (2246)

宏基因组学分析深度处理阶段污水中细菌的赋存特征及其功能 胡健双, 王燕, 周政, 汪雅琴, 王秉政, 李激 (2259)

大别山区生境质量时空特征及自然-人为因素驱动机制 郑亚平, 张俊华, 田惠文, 朱航成, 刘舒, 丁亚鹏 (2268)

基于连续小波变换、SHAP和XGBoost的土壤有机质含量高光谱反演 叶森, 朱琳, 刘旭东, 黄勇, 陈蓓蓓, 李欢 (2280)

秸秆还田、覆膜和施氮对旱地麦田土壤质量的影响 叶子壮, 王松燕, 陆潇, 史多鹏, 吕慎强, 李嘉, 杨泽宇, 王林科 (2292)

长期施用有机肥对土壤微塑料赋存及迁移特征的影响 王长远, 马啸驰, 郭德杰, 刘新红, 马艳, 罗佳 (2304)

土壤盐分变化对N₂O排放影响:基于Meta分析 黄艺华, 余冬立, 史祯琦, 胡磊, 潘永春 (2313)

基于土地利用/覆被动态变化的粤港澳大湾区碳储量评价与预测 郑慧玲, 郑辉峰 (2321)

基于InVEST模型的伏牛山地区生态系统碳储量时空变化模拟 张哲, 时振钦, 朱文博, 孙梓欣, 赵体侠, 邓文萍, 刘志强 (2332)

防护林建设过程中土壤微生物养分限制与有机碳组分之间的关系 徐凤璟, 黄懿梅, 黄倩, 申继凯 (2342)

长期秸秆还田褐土有机碳矿化特征及其驱动力 赵宇航, 殷浩凯, 胡雪纯, 解文艳, 刘志平, 周怀平, 杨振兴 (2353)

冬绿肥覆盖对土壤团聚体及有机碳和AMF多样性的影响 鲁泽让, 陈佳钰, 李智贤, 李永梅, 罗志章, 杨锐, 田明洋, 赵吉霞, 范茂攀 (2363)

辽河流域氮素时空分布及其对土地利用和降雨的响应 周波, 李晓光, 童思陈, 吕旭波, 郭朝臣, 雷坤 (2373)

高强度农业种植区不同景观池塘氧化亚氮排放特征 张欣悦, 肖启涛, 谢晖, 刘臻婧, 邱银国, 罗菊花, 徐向华, 段洪涛 (2385)

生物炭与不同类型氮肥施肥对菜地土壤反硝化细菌群落的影响 柳晓婉, 刘杏认, 高尚洁, 李贵春 (2394)

大岳山不同林龄人工油松林土壤微生物特征 马义淑, 曹亚鑫, 牛敏, 张明昱, 程曼, 文永莉 (2406)

基于稀疏样点的土壤重金属含量模拟方法 张佳琦, 潘瑜春, 高世臣, 赵亚楠, 景胜强, 周艳兵, 郇允兵 (2417)

基于特定源-风险评估模型的兰州黄河风情线绿地土壤重金属污染优先控制源分析 李军, 李旭, 李开明, 焦亮, 台喜生, 臧飞, 陈伟, 脱新颖 (2428)

西南不同类型紫色土pH变化、重金属累积与潜在生态风险评估 张海琳, 张雨, 王顶, 谢军, 张跃强, 张宇亭, 王洁, 石孝均 (2440)

重庆某铁矿周边耕地土壤重金属污染评价及来源解析 廖泽源, 李杰芹, 沈智杰, 李彩霞, 罗程钟, 梅楠, 张成, 王定勇 (2450)

金属矿区周边农田土壤与农作物重金属健康风险评估 魏洪斌, 罗明, 向奎, 查理思 (2461)

赤泥基纳米零价铁对多金属污染土壤修复效果 刘龙宇, 杨世利, 赵黄诗雨, 常凯威, 余江 (2473)

施锌对碱性土壤-小麦幼苗体系累积镉的影响 张瑶, 王天齐, 牛硕, 杨阳, 陈卫平 (2479)

基于稀疏样点的土壤重金属含量模拟方法

张佳琦^{1,2,3}, 潘瑜春^{1,3}, 高世臣², 赵亚楠¹, 景胜强^{4,5}, 周艳兵^{1*}, 郜允兵^{1,3*}

(1. 北京市农林科学院信息技术研究中心, 北京 100097; 2. 中国地质大学(北京)数理学院, 北京 100083; 3. 北京农林科学院智能装备研究中心, 北京 100097; 4. 廊坊市灾害遥感监测重点实验室, 廊坊 065201; 5. 防灾科技学院生态环境学院, 廊坊 065201)

摘要: 土壤重金属受人为和自然因素综合作用, 其空间异质性强, 存在区域均值和方差的非平稳性, 稀疏样本下未知点估计精确度低, 土壤环境质量现状精准估计和风险评估困难. 基于此, 提出了随机森林-序贯高斯模拟混合模型(RF-SGS), 选取多种自然因素和人为因素作为辅助变量, 充分考虑土壤属性指标的空间自相关性以及环境变量属性相似性, 解决传统插值中极端值和空间连续性模式敏感存在的局限性, 为非平稳区域精准估计总体提出可行性方法. 以北京市顺义区采样数据为例, 采用MMSD抽样方法对样点抽稀, 对原始采样数据进行不同采样密度的对比实验, 用随机森林-序贯高斯模拟混合模型(RF-SGS)、序贯高斯模拟模型(SGS)、趋势面-序贯高斯模拟混合模型(TR-SGS)和随机森林模型(RF)对土壤重金属Cd的空间分布进行模拟, 从统计特征和空间结构等方面比较模拟结果, 分析误差产生的原因, 进一步验证方法有效性. 结果表明, 在7种采样密度下, 预测精度由低到高排序为: SGS < TR-SGS < RF < RF-SGS, RF-SGS估算精度最高且Cd含量空间分布也最接近原始数据分布. RF-SGS模型可以作为稀疏样点下土壤重金属空间模拟的一种有效方法.

关键词: 土壤重金属; 条件模拟; 随机森林(RF); 序贯高斯模拟(SGS); 土壤插值

中图分类号: X53 文献标识码: A 文章编号: 0250-3301(2024)04-2417-11 DOI: 10.13227/j.hjkk.202303262

Simulation of Heavy Metal Content in Soil Based on Sparse Sample Sites

ZHANG Jia-qi^{1,2,3}, PAN Yu-chun^{1,3}, GAO Shi-chen², ZHAO Ya-nan¹, JING Sheng-qiang^{4,5}, ZHOU Yan-bing^{1*}, GAO Yun-bing^{1,3*}

(1. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; 2. School of Mathematics and Physics, China University of Geosciences (Beijing), Beijing 100083, China; 3. Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; 4. Langfang Key Laboratory of Disaster Monitoring by Remote Sensing, Langfang 065201, China; 5. Institute of Disaster Prevention Science and Technology, Langfang 065201, China)

Abstract: Soil heavy metals are affected by the comprehensive action of human and natural factors, and their spatial heterogeneity is strong, there is non-stationarity of regional mean and variance, the accuracy of estimating unknown points under sparse samples is low, and the accurate estimation of soil environmental quality status and risk assessment are difficult. In light of this, this study proposed a random forest-sequential Gaussian simulation hybrid model (RF-SGS), selected a variety of natural factors and human factors as auxiliary variables, fully considered the spatial autocorrelation of soil attribute indicators and the similarity of environmental variable attributes, resolved the limitations of extreme values and spatial continuity mode sensitivity in traditional interpolation, and proposed a feasible method for accurate estimation of nonstationary regions. Taking the sampling data in Shunyi District, Beijing, as an example, the MMSD sampling method was used to thin the sample points, and the original sampling data were compared with different sampling densities. The spatial distribution of soil heavy metal Cd was simulated using the random forest-sequential Gaussian simulation mixed model (RF-SGS), sequential Gaussian simulation (SGS), trend surface-sequential Gaussian simulation hybrid model (TR-SGS), and random forest (RF) model, and the simulation results were compared from the aspects of statistical characteristics and spatial structure. The causes of errors were analyzed to further verify the effectiveness of the method. The results showed that under the seven sampling densities, the prediction accuracy was sorted from low to high as SGS < TR-SGS < RF < RF-SGS, the RF-SGS estimation accuracy was the highest, and the spatial distribution of Cd content was closest to the original data distribution. In conclusion, the RF-SGS model could be used as an effective method for spatial simulation of soil heavy metals under sparse samples.

Key words: heavy metals in soil; condition simulation; random forest(RF); sequential Gaussian simulation(SGS); soil interpolation

土壤是粮食安全、水安全和更广泛的生态系统安全的基础. 随着社会经济高速发展, 长期以来高强度和超负荷的开发利用土地, 造成了耕地土壤环境质量的破坏, 其中土壤重金属污染日益严重, 成为备受关注的全球性问题. 土壤监测调查是准确掌握土壤重金属污染状况及变化趋势的主要技术手段^[1]. 土壤监测调查受控于采样、化验等成本的限制, 相对于研究区总体来讲, 样本量往往相对稀疏, 不能精准刻画土壤污染空间分布特征, 在一定程度上使得土壤治理与风险防控受限^[2]. 母质母岩、

工矿企业排放(大气、水)及农业投入品和交通运输等影响因素与土壤重金属含量分布存在一定的空间同位性和属性相关性^[3]. 基于上述自然和人为因素准确刻画土壤重金属的空间变化特征, 实现稀疏样点下土壤重金属含量高精度估计是当前土壤环境质量

收稿日期: 2023-03-31; 修订日期: 2023-07-12

基金项目: 国家重点研发计划项目(2022YFC3700805, 2023YFD1700105); 国家自然科学基金项目(42077001); 北京市农林科学院科技创新能力建设专项(KJCX20230309)

作者简介: 张佳琦(1998-), 女, 硕士, 主要研究方向为土壤重金属污染空间模拟, E-mail: zhangjiaq98@163.com

* 通信作者, E-mail: zhouyb@nercia.org.cn; gaoyb@nercita.org.cn

关注的热点,在土壤重金属监测监管中具有重要的现实意义。

目前常用的土壤重金属含量估计方法有:基于空间位置相关性的方法,如克里金插值方法和反距离加权法;基于属性相似的方法,如朱阿兴等的地理学第三定理法^[4]和机器学习法;以及基于综合空间相似性和属性相似性的方法^[5],如随机森林克里金法和回归克里金法等。克里金插值方法通过已知样点对未知样点进行线性无偏和最优化估计,样点需满足正态分布及或内蕴平稳假设,同时会对数据结果进行平滑处理,使局部异质点位或变化剧烈区的重要信息丢失,尤其在极值和空间连续性模式敏感的应用中存在一定的局限性^[6]。为克服这一问题,Matheron^[7]提出了转向带法,其主要思路是先根据空间数据的统计特征模拟出一维随机变量,然后将其在空间中旋转得到三维模拟,该方法将数据作为一个整体,引入未知点的最大可能值的概率,最大可能地接近真实的空间分布结果,具有与实测数据相同的频率直方图,且模拟结果与实测结果具有相同的空间自相关关系。随后又衍生出转向带法^[8]、指示模拟法^[9]、矩阵分解法^[10]、蒙特卡罗法^[11]、序贯指示模拟^[12,13]和序贯高斯模拟方法^[14,15]等。赵永存等^[16]以江苏省张家港市土壤镉含量为例对比不同条件模拟方法,结果表明定量预测局部不确定性对于控制土壤空间数据的应用风险具有重要意义。Delbari等^[17]验证了序贯高斯模拟能克服克里格的平滑效果,更能真实地反映土壤有机碳的空间分布。为了提高未知点统计推断结果精度,基于空间相关性方法和条件模拟法往往需增加土壤采样密度,提高样本量^[18]。人工神经网络^[19]、回归树^[20]、随机森林(RF)^[21]和梯度提升决策树^[22]等机器学习模型可以充分利用环境变量间属性特征,通过建立未知点与土壤环境变量的非线性函数关系,实现复杂空间分布下土壤属性预测^[23]。上述机器学习方法因其在处理多维连续型和类别型变量方面的优势,逐渐发展成主流的土壤属性统计推断方法^[24]。其中,随机森林模型是基于多个决策树的集成学习算法,可以防止过拟合且预测效果稳定,是一种有效的土壤预测方法。然而土壤重金属具有多尺度特征^[25,26],其空间分布特征也比较复杂,既有小尺度的细节空间分布特征又具有大尺度的宏观分布特征,将RF模型的预测残差与预测结果相结合的随机森林残差克里格法(RFRK)能提高随机森林模型的预测精度^[27]。

基于上述分析,本文提出了基于随机森林-序贯高斯模拟混合模型,在顾及土壤环境影响结构特征的基础上,利用机器学习模型刻画土壤重金属全

局的非线性和非平稳趋势面,同时将具有空间自相关性的残差部分进行条件模拟,通过结合空间信息和属性信息有效提升模型模拟精度,解决土壤重金属空间分布特征复杂、局部变化异常点预测难的问题,以期为区域土壤重金属含量准确估计提供一种可行的解决方案。

1 材料与方法

1.1 研究区概况

顺义区位于北京市东北方向,距市区30 km,北邻怀柔区和密云区,东界平谷区,南与通州区和河北省三河市接壤,西南和西与昌平区和朝阳区隔温榆河为界。介于北纬 $40^{\circ}00' \sim 40^{\circ}18'$,东经 $116^{\circ}28' \sim 116^{\circ}58'$ 之间,境域东西长45 km,南北宽30 km,总面积1 021 km²,其中耕地面积404.05 km²,占总面的39.56%。地势北高南低,区域内有大小河流20余条,分属北运河、潮白河和蓟运河这3个水系,河道总长232 km。蓟运河水系主要支流金鸡河是顺义东南部的的主要泄洪和排污河道,小中河是顺义西南部的的主要泄洪和排污河道。土壤类型主要为潮土和褐土,且潮土分布于全区大部分地区,褐土主要分布在东北部。种植模式为冬小麦-夏玉米,粮食生产用地占农业用地的56.91%,其中南部是主要的菜地集中区域,东北部则是果园的聚集地,样点的分布见图1。

于2009年8~11月在顺义区农田采用混合采样方法,共采集412个农田表层土壤样品(0~20 cm),同时记录每个样点的经纬度坐标、土地利用方式、土壤类型、质地和施肥等相关信息。每个采样点用木铲采集1.0 kg的土壤样品,样品风干后过2.0 mm的初筛,四分样后选取50 g的样品,使用玛瑙研钵研磨后过0.15 mm的细筛。土壤样品按照EPA Method 3050B的方法使用HNO₃、HCl和H₂O₂混合溶液消解,消解液中的砷、镉、铜、锌和铅使用等离子发射光谱仪(ICP-OES, Thermo iCAP 6300, USA)测定,汞的测定使用1:1的HNO₃:HCl在100℃消解,消解液使用原子荧光光谱仪(AFS, Titan AFS830, China)测定。每个样品设3个平行样,样品处理和和分析的质量控制采用分析土壤标准参考物质GSS-1和GSS-4(中国标准物质中心)。

1.2 研究方法

1.2.1 随机森林-序贯高斯模拟方法

随机森林-序贯高斯模拟混合方法(random forest with sequential Gaussian simulation, RF-SGS)是一种结合了随机森林(RF)与序贯高斯模拟(SGS)的混合地统计模型。其中,随机森林(RF)是一种基于多个

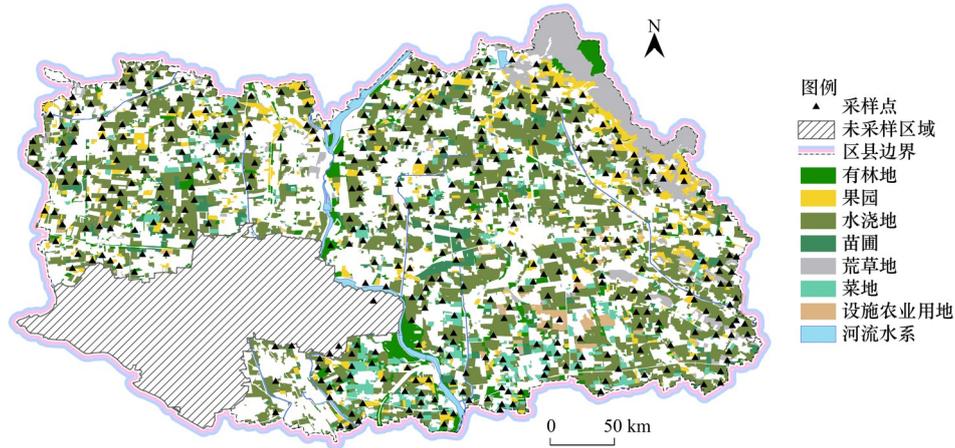


图 1 研究区点位分布示意

Fig. 1 Distribution of sites in the study area

决策树的数据挖掘算法, 通过在训练数据集中应用 bootstrap 方法随机抽取样本构建单独的树模型, 生成大量的树后组成随机森林, 与 CART 决策树模型相比, RF 模型对过拟合不敏感^[28]. 序贯高斯模拟主要是随机生成一条模拟路径, 逐步求取各个待估点处的累积条件分布函数, 并根据该累积条件分布函数求得该点的模拟值^[29]. 本文通过随机森林在随机模型中引入协变量, 即在序贯高斯模拟中添加了确定性模型, 通过引入有效辅助变量来构建变量间的非线性关系. 通常情况残差被认为是误差, 代表模型中无法用确定性成分解释的成分^[30]. 理想情况下, 模型的残差应该是相同且独立分布的. 有研究表明, 事实上残差在某些情况下表现为空间自相关^[31].

Cd 在空间上的分布宜受人类活动影响^[32], 主要的自然和人为因素影响的趋势性部分可以通过随机森林模拟, 对于残差项部分的随机性可以看作随机过程, 多年累积中的随机过程可以用马尔可夫链模型表达^[33]. 马尔可夫链模型是一种随机过程, 假设某一个系统发展过程有 m 个可能状态, 且各状态互不相容, 即 $S(1), S(2), \dots, S(m)$, 初始概率向量 $S(0)$ 表示为:

$$S(0) = [S_1(0), S_2(0), \dots, S_m(0)] \quad (1)$$

经 k 步转移后, 状态 i 的概率向量:

$$S(k) = [S_1(k), S_2(k), \dots, S_m(k)] \quad (2)$$

$S(0)$ 与 $S(k)$ 的关系可表示为:

$$S(k) = S(k-1)P = S(0)P^k$$

$$= S(0) \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \vdots & \vdots & & \vdots \\ P_{m1} & P_{m2} & \dots & P_{mm} \end{bmatrix}^k \quad (3)$$

当 $S(0)$ 和转移概率 P 已知时, 可以用式(1)得到在 k 时刻系统所处的状态, 这就是马尔可夫模型.

用这个模型的关键是确定 P , 再给定系统的初始状态向量 $S(0)$, 就可以得到某种情况下的状态向量 $S(k)$. 序贯高斯模拟条件累积过程可以用马尔可夫状态转移表达, 从而实现对随机性残差的空间分布模拟^[34]. 因此本文在确定性分量上加入残差, 通过地统计条件模拟将空间自相关引入模型, 从而提高模拟精度. 具体步骤如下.

根据 Pearson 相关系数和方差分析探索连续型和分类型因素对土壤重金属含量影响是否显著, 由此选取辅助变量; 利用随机森林拟合模型, 用实测值减去拟合值(即趋势项的影响)得残差值 r ; 对残差值 r 进行序贯高斯模拟实现随机场生成, 得到模拟残差值; 将趋势项部分和模拟残差部分结果进行相加, 得到最后预测值 y . 通过以上步骤, 最终得到随机森林与序贯高斯模拟融合结果.

$$P_{\text{RF-SGS}}(x_0) = P_{\text{RF}}(x_0) + \varepsilon_{\text{SGS}}(x_0) \quad (4)$$

式中, $P_{\text{RF-SGS}}(x_0)$ 为 RF-SGS 模型在 x_0 位置的 Cd 含量预测值, $P_{\text{RF}}(x_0)$ 为趋势项, 通过 RF 模型拟合, 可影响因素 $\varepsilon_{\text{SGS}}(x_0)$ 为误差项, 通过对残差进行序贯高斯模拟拟合.

1.2.2 实验设计与精度评价

本文选取区域土壤重金属 Cd 作为研究对象, 将 412 个高密度样点看作研究“总体”进行抽稀, 基于不同样本密度对本文提出的随机森林-序贯高斯模拟模型(RF-SGS)、序贯高斯模拟模型(SGS)、随机森林模型(RF)和趋势面-序贯高斯模拟模型(TR-SGS)分别建立模型, 并对比 4 种模型预测结果的数值分布和空间分布, 进行估计精度的分析和评价.

为了更好地评估模型精度, 本文以 7:3 的比例划分训练集和验证集, 选取平均绝对误差(MAE)、平均相对误差(MRE)和均方根误差(RMSE)作为评价指标, 对比 SGS 模型、TR-SGS 模型、RF 模型和

RF-SGS模型的估计结果与实际对应的采样点观测数据接近程度,进而评估模型预测结果的准确性,以验证RF-SGS模型是否能有效提高模拟精度.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6)$$

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \quad (7)$$

式中, \hat{y}_i 为预测值, y_i 为观测值; n 为检验样本数.

为进一步验证模拟数据和实际值之间偏差,使用预测结果与样本点的相对误差的绝对值来测度拟合偏差 $\text{RE}^{[35]}$, 设定均值相对误差(relative error of mean, REM)、离散系数相对误差(relative error of dispersion coefficient, RED)、偏度系数相对误差(relative error of skewness, RES)和峰度系数相对误差(relative error of kurtosis, REK)指标,表征拟合结果与实际值之间差异.

$$\text{REM} = \left| \frac{M - M_i}{M} \right| \times 100\% \quad (8)$$

$$\text{RED} = \left| \frac{D - D_i}{D} \right| \times 100\% \quad (9)$$

$$\text{RES} = \left| \frac{S - S_i}{S} \right| \times 100\% \quad (10)$$

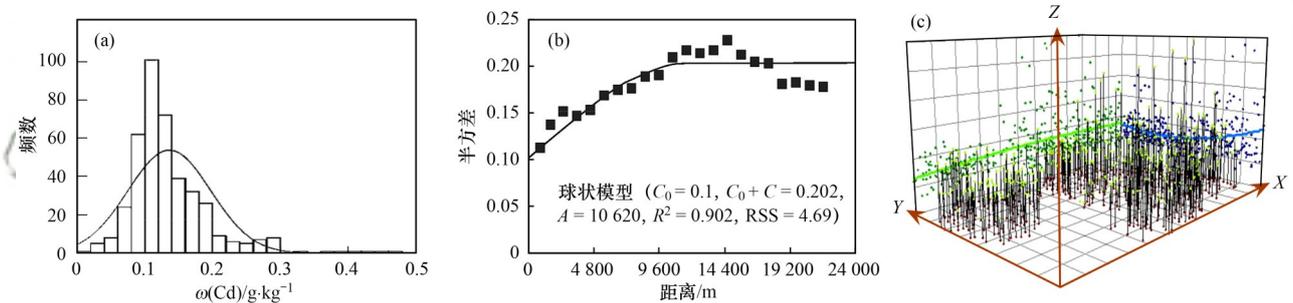
$$\text{REK} = \left| \frac{K - K_i}{K} \right| \times 100\% \quad (11)$$

式中, M 、 D 、 S 和 K 分别为采样点实际均值、离散系数、偏度系数和峰度系数, M_i 、 D_i 、 S_i 和 K_i 分别为4种模拟方法下验证集的均值、离散系数、偏度系数和峰度系数.

2 结果与分析

2.1 土壤重金属空间特征分析

对顺义区412个采样点Cd含量数据进行描述性统计分析,以反映研究区土壤重金属的总体空间分布特征(见表1),其中 $\omega(\text{Cd})$ 范围为0.015~0.469 $\text{g}\cdot\text{kg}^{-1}$,均值为0.136 $\text{g}\cdot\text{kg}^{-1}$,高于背景值0.12 $\text{g}\cdot\text{kg}^{-1}$;标准差为0.06 $\text{g}\cdot\text{kg}^{-1}$,偏度系数为2,峰度系数为6.36,变异系数为0.45,属于中等变异.结合样点的含量直方图[图2(a)],可知顺义区土壤Cd含量受到人类活动影响,呈右偏态分布.



(a) Cd含量直方图, (b) Cd半方差函数及参数, (c) Cd趋势效应; (b)中 C_0 为块金值, C_0+C 为基台值, A 为变程; (c) 中蓝线为重金属Cd含量在南北方向的变化趋势, 绿线为东西方向的变化趋势, 线段高度表示此处Cd含量值

图2 Cd数据特征

Fig. 2 Cd data feature

表1 不同采样方案描述性统计¹⁾

Table 1 Descriptive statistics for different sampling schemes

采样个数	平均值	标准差	最大值	最小值	变异系数	偏度	峰度
50	0.153	0.078	0.430	0.029	0.505	2.622	11.296
100	0.131	0.067	0.447	0.015	0.517	2.196	7.206
150	0.135	0.057	0.409	0.015	0.421	1.509	4.099
200	0.136	0.056	0.469	0.015	0.458	1.909	5.949
250	0.136	0.062	0.469	0.015	0.455	1.951	6.496
300	0.135	0.059	0.469	0.015	0.437	2.838	6.916
350	0.137	0.062	0.469	0.015	0.451	2.069	6.528
412	0.136	0.062	0.469	0.015	0.449	2.000	6.360

1)除变异系数、偏度和峰度无量纲外,其他项目单位均为 $\text{g}\cdot\text{kg}^{-1}$

对研究区土壤Cd进行空间变异函数拟合[图2(b)],其块金系数为0.5,表明在空间上具有中等

程度的变异性,即存在由人为因素和自然因素共同作用产生的空间变异.从土壤Cd含量区域趋势[图

2(c)]也可以看出, 东西方向呈线性或一阶趋势, Cd含量的分布具有一定趋势性; 南北方向呈二阶多项式变化, 也说明Cd含量的分布还存在均值或者方差非平稳特征; 上述特点可能受土壤类型、质地、土地利用和工矿企业排放局部污染等的影响, 适合采用上述方法进行区域土壤含量的建模研究.

2.2 环境变量选取及实验方案生成

为了表征土壤重金属受自然因素和人为活动的影响, 即反映模型中 $P_{RF}(x_0)$ 项, 本文选取表2中的土壤、气候、人类活动以及地理位置等10个代理变量, 对研究区土壤重金属Cd进行模拟. 通过搜集对

比近5年人为影响因素变化情况, 发现土地利用类型、 PM_{10} 和距污染企业距离因素变化较小, 因此本文采集2009年数据作为代理变量进行建模研究. 其中土壤类型和土壤质地、酸碱度用于反映土壤的差异性和趋势性; 年均降水量和归一化植被指数反映成土条件及成土过程等因素差异性和趋势性; 经度和纬度代表调查点地理位置空间自相关性; 距离污染企业距离、可吸入颗粒物 PM_{10} 、土地利用方式代表企业、农业投入品强度等点源和面源等人类活动影响因素的空间异质性; 使用ArcMap 10.5中多值提取到点工具将各辅助变量值提取到对应采样点, 生成实验数据集.

表2 数据类型与来源¹⁾
Table 2 Data type and source

序号	数据名	缩写	类型	数据来源
1	土壤类型	STP	CD	https://www.resdc.cn/
2	土壤质地	ST	CD	https://www.resdc.cn/
3	土地利用方式	LUT	CD	https://www.resdc.cn/
4	经度	LON	LD	ArcGIS 几何计算
5	纬度	LAT	LD	ArcGIS 几何计算
6	距污染企业距离	DTPE	LD	ArcGIS 几何计算
7	年均降水量	MAP	LD	https://www.worldclim.org/
8	酸碱度	pH	LD	http://www.gisrs.cn/
9	归一化植被指数	NDVI	LD	https://adsweb.modaps.eosdis.nasa/
10	可吸入颗粒物 PM_{10}	PM_{10}	LD	https://www.aqistudy.cn/

1) CD 为类别型变量, LD 为连续型变量

使用皮尔逊相关系数计算连续型变量与Cd含量之间的关系, 使用方差分析^[37]判别类别型变量对土壤重金属含量是否有显著影响, 选择与Cd存在较高相关性的要素作为预测模型的输入. 表3为连续型变量与Cd含量之间的相关性. 从中可知,

Cd含量与DTPE、MAP、pH、LON和 PM_{10} 呈正相关关系, 与LAT和NDVI呈现负相关关系. 同时, Cd含量与LAT、DTPE和pH均呈现较强的相关性关系. 基于上述分析选择合适的辅助变量进行Cd空间分布预测.

表3 连续型变量与Cd含量的相关性¹⁾
Table 3 Analysis of correlation for continuous variables and Cd concentration

项目	纬度	降水量	距污染企业距离	酸碱度	经度	归一化植被指数	可吸入颗粒物
r	-0.26**	0.20**	0.16**	0.13*	0.12*	-0.11*	0.11*

1)*表示显著相关($P < 0.05$), **表示极显著相关($P < 0.01$)

表4为类别型变量与Cd含量的方差分析检验结果. 从中可知, LUT、STP和ST这3个分类变量与Cd含量的方差分析检验结果均在0.05的水平下显著($P < 0.05$), 说明土地利用方式、土壤类型和土壤质地在该研究区内对土壤重金属Cd的含量具有一定的影响.

为了验证模拟的稳健性, 以研究区412个高密度土壤采样为基础, 通过平均最短距离最小化准则

表4 类别型变量方差分析检验结果

Table 4 Spatial heterogeneity test results for categorical variables

项目	土地利用类型	土壤类型	土壤质地
P	0.046	0.000	0.027

(minimization of the mean of the shortest distance, MMSD)对样本进行抽稀^[38], 模拟实际采样的真实情况. 该算法的核心思想是使采样点在整个空间分布尽可能均匀, 从而使抽样空间内任意一个空间位置点与它最邻近点之间的距离的期望最小^[39], 该方法在土壤调查和土壤制图中得到广泛应用, 其公式见式(12).

$$\Phi_{\text{MMSD}}(S) = \min \frac{1}{A} \iint \|\vec{X} - V_s(\vec{X})\| \quad (12)$$

式中, S 为抽样样本; A 为抽样总体; \vec{X} 为抽样总体在空间上的任意点; $V_s(\vec{X})$ 为 S 内与 \vec{X} 最近邻的点.

MMSD 采样个数分别为 50、100、150、200、250、300 和 350 的样本点, 模拟不同密度的采样情况. 为降低采样随机性的影响, 本文对不同采样密

度的样本分别进行 200 次重复, 并取 200 次计算结果的平均值作为最终结果, 抽稀过程使用 R 语言实现, 如图 3 所示.

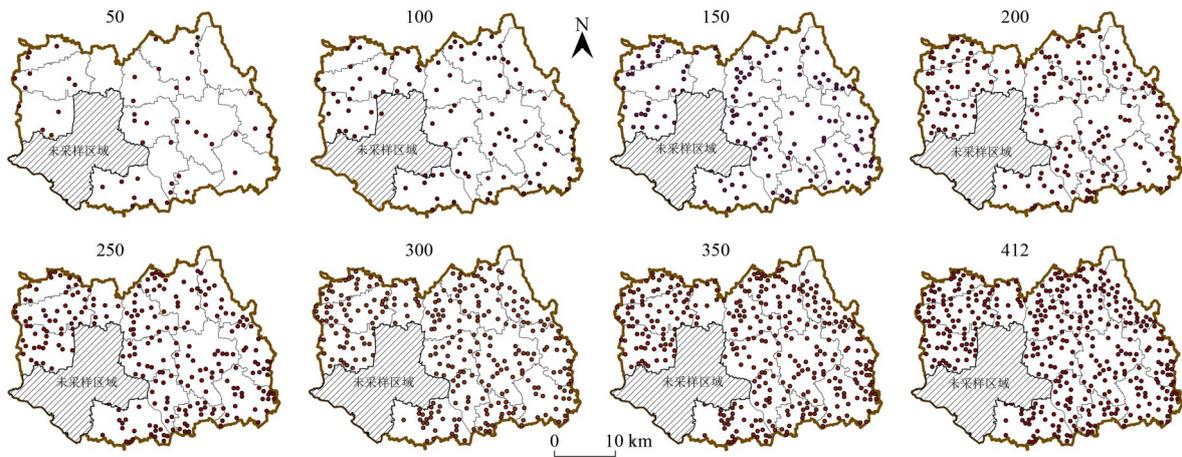


图 3 不同采样密度方案设计

Fig. 3 Different sampling density scheme design

样本抽稀后各统计值见表 1, 不同抽样密度下统计量值有微小浮动, 随着样本个数增加, 各密度下样本统计指标也更接近原始数据, 可见抽稀后样本存在良好的稳定性.

2.3 模拟结果对比分析

对比 4 种模型对重金属 Cd 含量的预测精度, 结果如图 4 所示. 在不同抽稀密度下, SGS 模型、TR-SGS 模型、RF 模型和 RF-SGS 模型的 MAE 均值分别为 0.045、0.045、0.038 和 0.023 $\text{g}\cdot\text{kg}^{-1}$, RMSE 均值分别为 0.067、0.063、0.054 和 0.035 $\text{g}\cdot\text{kg}^{-1}$, MRE 均值分别为 37.38%、38.34%、33.67 和 16.42%. 可见, 在未采用辅助数据时, TR-SGS 模型整体要优于 SGS 模型, 但该方法在不同采样密度下不够稳健, 在样点

密度小时其误差大于 SGS 模型, 模型优化效果不够显著. 在采用辅助数据时, RF 模型基于引入辅助变量构建非线性拟合, 其拟合效果优于 SGS 模型与 TR-SGS 模型. RF-SGS 模型在 MAE、RMSE、MRE 这 3 个模型精度评价指标上的表现均显著优于 SGS 模型、TR-SGS 模型和 RF 模型, 进而验证了 RF-SGS 模型的准确性和稳健性, 这种能力归因于它在 RF 模型中处理重金属 Cd 含量和环境辅助变量之间非线性关系的能力, 以及 SGS 模型残差中处理未解释信息的能力^[40]. 综上, 由于样本抽稀过程 RF-SGS 模型模拟效果均优于传统模型, 证实 RF-SGS 模型可应用于不同采样密度下的土壤重金属空间分布模拟.

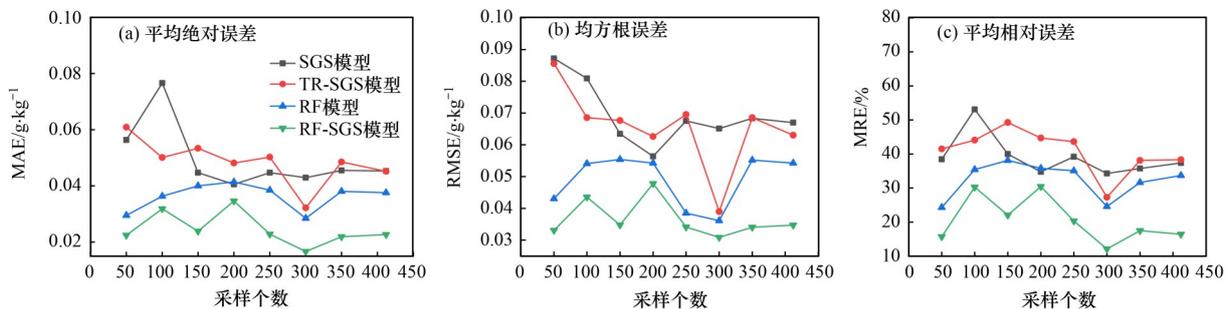


图 4 不同模型 Cd 模拟精度对比

Fig. 4 Comparison of Cd simulation accuracy of different models

2.4 模拟结果空间分布

分别使用 4 种预测模型绘制研究区土壤重金属 Cd 空间分布, 如图 5 所示, 可见研究区 4 种预测模型的土壤重金属 Cd 预测空间分布趋势整体表现一致, Cd 的空间分布存在异质性, 与空间变异性分析吻合. Cd 含量均表现为高值集中在南部和东北部,

中部地区的 Cd 整体含量偏低, 局部含量分布不均、呈现高低错落, 与趋势分析走势大致相符. 预测精度相对较低的 SGS 模型绘制的空间分布图存在边界较明显的块状区域, 且相比其他 3 种方法 SGS 模型在中部存在区域高值. 预测精度最优的 RF-SGS 模型预测结果中无边界明显的块状区域, 高值区中存在

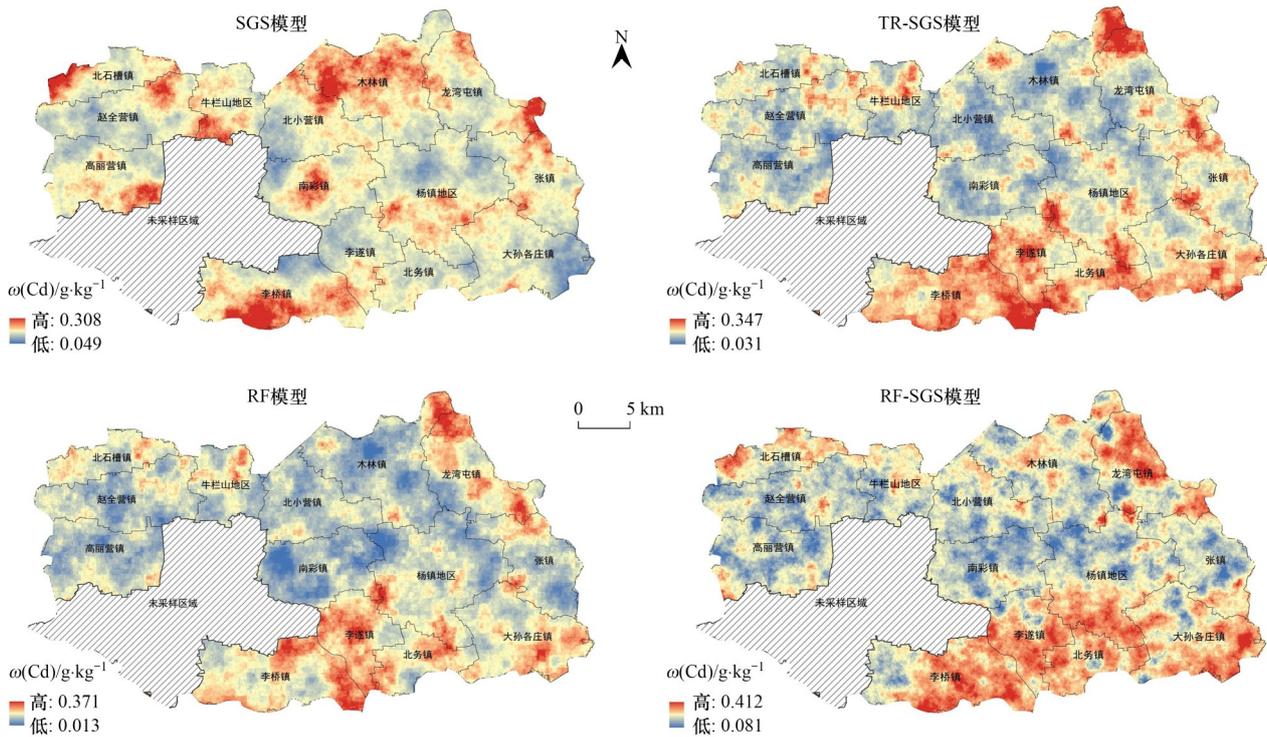


图5 4种模型模拟结果空间分布

Fig. 5 Spatial distribution of simulation results of four models

零星低值, 在低值区域存在零星高值, 并且值域范围最大, 与统计结果和研究区Cd空间分布更接近.

3 讨论

3.1 模拟效果分析

从土壤重金属空间特征分析可知, 顺义区Cd既存在空间相关性, 又存在局部空间异质性, 影响因素对土壤重金属分布有一定影响. 当未引入辅助变量时, 仅基于空间相关性的SGS模型, 其RMSE为 $0.067 \text{ g}\cdot\text{kg}^{-1}$, MAE为 $0.045 \text{ g}\cdot\text{kg}^{-1}$, MRE为37.38%, 能够达到较为可靠的空间拟合精度. SGS模型适用于平稳区域或者较密集采样点的情况, 或者在协变量难以收集或相关性不高时, 可以降低精度要求, 采用仅依赖空间相关性的模拟方法.

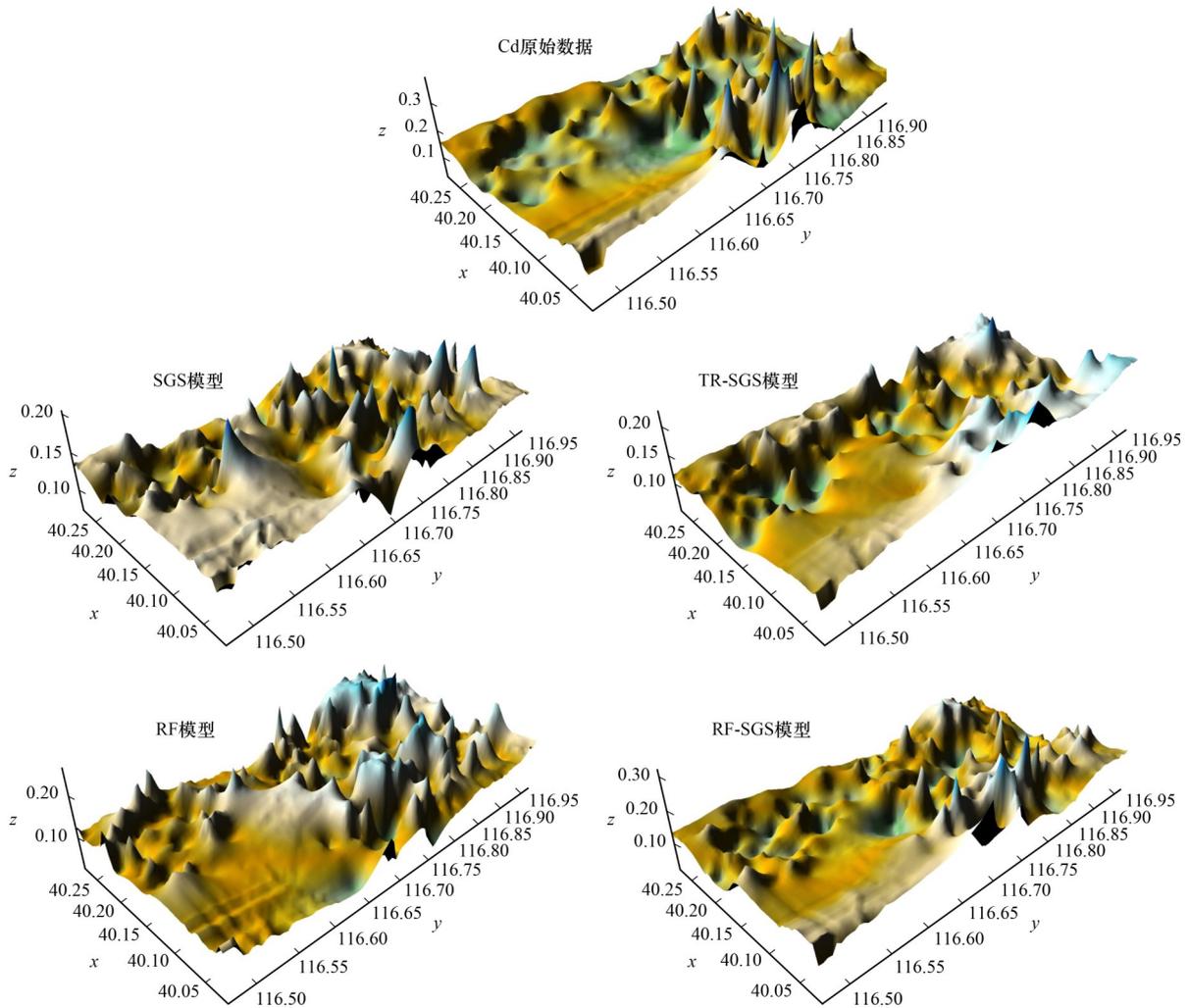
TR-SGS模型仅采用多项式构建确定性趋势面、再引入残差空间影响的方法, 适用于研究区具有明显一阶或多阶趋势分布的情况, 但在趋势的选择上通过直观判别可能会存在偏差, 选择符合度高的多项式拟合趋势面可以提高模拟精度, 本研究区TR-SGS模型的RMSE为 $0.063 \text{ g}\cdot\text{kg}^{-1}$, MAE为 $0.045 \text{ g}\cdot\text{kg}^{-1}$, MRE为38.34%, 对比SGS模拟精度提升不明显, 且在样本抽稀过程该方法稳健性较差, 50点和200点时拟合误差高于SGS模型.

引入辅助变量的RF模型拟合精度优于SGS模型和TR-SGS模型(图4). 其原因在于土壤重金属与地

理位置、距离因子、气候因子、植被因子等环境因素以及用地类型等人类活动密切相关, 具有较强的空间异质性^[41,42]. 通过引入与Cd相关的NDVI、 PM_{10} 和年均降水量等多种环境变量, 利用属性相似性建模, 能够取得高于仅考虑空间相关性方法的精度. 因此, 能够处理多维多类型协变量的机器学习方法在土壤重金属反演中应用较多^[43]. RF-SGS模型同时考虑了空间相关性和属性相似性, 对两者分别进行建模. 在50~412个采样点数量下其MAE和RMSE误差曲线均较其他3种方法效果更优, 且通过模拟结果对比(图6)得到RF-SGS模型的模拟结果更接近真实情况. 因此, RF-SGS模型的模拟精度比SGS模型、TR-SGS模型和RF模型有所改进, 适用于非平稳和稀疏样点下的空间分布预测. 兼顾空间异质性和相关性进行建模的思想在相关领域也得到了学者的认可, 刘艳芳等^[44]发现使用考虑残差空间结构的回归克里格模型预测土壤有机碳密度的表现优于其它模型.

3.2 空间格局分析

以样本量为412时的全部模拟结果为例, 从模拟后空间分布基本趋势与原始数据的接近程度来检验模拟效果^[45,46], 从Cd实测数据、SGS模型、RF模型以及RF-SGS模型模拟结果的三维立体图可以看出(图6), RF-SGS模型模拟结果的空间分布趋势与实测数据是最接近的. 未采用辅助数据的SGS模型



x 和 y 轴表示经纬度($^{\circ}$), z 轴表示重金属Cd含量($\text{g}\cdot\text{kg}^{-1}$)

图6 Cd原始数据和模拟结果的对比

Fig. 6 Comparison of Cd's original data and simulated results

预测结果在区域内整体偏高; TR-SGS模型由于趋势面拟合时没有引入辅助数据且仅采用多项式函数拟合, 模拟结果在高值和低值区刻画都不够准确, 出现低值偏低和高值偏高的情况. 加入辅助数据后的RF模型整体拟合趋势更接近实际情况, 但在东部高值区存在拟合偏高的误差, 其原因可能是模型拟合时仅考虑属性相关, 没有融合空间相关性建模. RF-SGS模型可通过残差空间性对RF模型拟合中偏离实际值的空间点位含量实现修正, 从而达到全局拟合效果最接近于实际的模拟结果.

从采样数量为412时4种模型全部模拟结果与原始数据的统计量相对误差对比可知(表5), TR-SGS模型相对SGS模型REM、RED、RES和REK这4个误差统计量下降不够明显, 引入辅助变量的RF模型相对SGS模型, TR-SGS模型误差有显著下降, RF-SGS模型4个统计量的相对误差均为最小, 结合图6空间分布可得, RF-SGS模型拟合效果最好, 整

体数据分布与原始数据最接近.

表5 统计量相对误差对比

Table 5 Statistical relative error comparison

项目	SGS模型	TR-SGS模型	RF模型	RF-SGS模型
REM	0.064 2	0.069 2	0.035 4	0.009 4
RED	0.587 3	0.429 6	0.344 1	0.197 4
RES	0.613 7	0.720 1	0.396 6	0.138 5
REK	0.765 9	0.972 8	0.256 2	0.148 6

综上, RF-SGS模型能够有效提升未知空间模拟精度, 其模拟结果与原始数据中Cd含量更接近、空间分布格局特征更吻合, 证明了该方法的有效性与可行性, 可用于空间非平稳态下土壤重金属的空间模拟. 相较其他3种方法RF-SGS模型的优势有以下3点: 首先, RF-SGS模型中RF部分具有非线性映射能力, 其考虑了包含自然和人为方面的多源辅助变量的确定性影响, 能以较高精度逼近任意连续函数; 其次, RF-SGS模型的SGS部分将具有空间相关

性的随机残差表示为随机变量, 对 RF 模型预测的不确定性部分进行条件模拟, 能够在尊重原始数据的基础上较好地反映土壤 Cd 含量空间变化; 最后, RF-SGS 模型可以改善单一机器学习与地统计模型面临的不稳定性, 单一机器学习容易受研究对象空间相关程度影响, 机器学习忽略空间性只考虑属性相关, RF-SGS 模型将属性相关和空间相关相结合, 可更精确地描述土壤 Cd 含量的空间变化特征。

本文在对土壤重金属 Cd 进行建模时, 通过模型精度对比分析发现, 重金属的样点密度、样点空间结构以及辅助变量数据质量对于模拟效果有一定影响。第一, 在非平稳区域中 RF-SGS 模型在不同密度下均保持有效性及稳健性。第二, 部分人为活动因素的辅助变量数据未纳入建模过程, 如果将更多的人为活动因素, 比如耕作制度、施肥类型、施用量等作为协变量纳入模型, 可能会在一定程度上提高模型模拟精度。

4 结论

(1) 对于存在空间异质性的研究区, 基于自然和人类活动等影响因素建模, 利用 RF 模型和 RF-SGS 模型对区域土壤重金属含量模拟, 比未采用辅助数据的 SGS 模型和 TR-SGS 模型其未知点估算的精度有所提升, 在稀疏样本下运用有效辅助变量模拟效果更优。

(2) 将机器学习与地统计建模融合的 RF-SGS 模型, 可以兼顾环境变量的属性相似性和人为活动局部影响的空间相关性, 在不同采样密度下, 其 RMSE、MAE 和 MRE 较其它 3 种模型均为最低, 且模拟结果与实际数据具有更相似的空间分布, 该方法可以用于空间非平稳、样本稀疏下未知点的土壤环境质量的估算。

参考文献:

- [1] Sun X F, Zhang L X, Lv J S. Spatial assessment models to evaluate human health risk associated to soil potentially toxic elements[J]. *Environmental Pollution*, 2021, **268**, doi: 10.1016/j.envpol.2020.115699.
- [2] Qu M K, Guang X, Liu H B, *et al.* Additional sampling using *in-situ* portable X-ray fluorescence (PXRF) for rapid and high-precision investigation of soil heavy metals at a regional scale[J]. *Environmental Pollution*, 2022, **292**, doi: 10.1016/j.envpol.2021.118324.
- [3] 傅婷婷. 区域土壤和水稻镉含量相关分析与估测研究[D]. 杭州: 浙江大学, 2021.
- Fu T T. Regional research of correlation analysis and estimation of Cd content in soil and rice [D]. Hangzhou: Zhejiang University, 2021.
- [4] 朱阿兴, 闫国年, 周成虎, 等. 地理相似性: 地理学的第三定律[J]. *地球信息科学学报*, 2020, **22**(4): 673-679.
- Zhu A X, Lv G N, Zhou C H, *et al.* Geographic Similarity: Third Law of Geography?[J]. *Journal of Geo-information Science*, 2020, **22**(4): 673-679.
- [5] 王雨雪, 杨柯, 高秉博, 等. 基于两点机器学习方法的土壤有机质空间分布预测[J]. *农业工程学报*, 2022, **38**(12): 65-73.
- Wang Y X, Yang K, Gao B B, *et al.* Prediction of the spatial distribution of soil organic matter based on two-point machine learning method [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, **38**(12): 65-73.
- [6] 谢云峰, 陈同斌, 雷梅, 等. 空间插值模型对土壤 Cd 污染评价结果的影响[J]. *环境科学学报*, 2010, **30**(4): 847-854.
- Xie Y F, Chen T B, Lei M, *et al.* Impact of spatial interpolation methods on the estimation of regional soil Cd [J]. *Acta Scientiae Circumstantiae*, 2010, **30**(4): 847-854.
- [7] Matheron G. The intrinsic random functions and their applications [J]. *Advances in Applied Probability*, 1973, **5**(3): 439-468.
- [8] 李少龙, 张家发, 张伟, 等. 表层土渗透系数空间变异与随机模拟研究[J]. *岩土力学*, 2009, **30**(10): 3168-3172.
- Li S L, Zhang J F, Zhang W, *et al.* Study of spatial variability and stochastic modeling of surface soil permeability. *Rock and Soil Mechanics*, 2009, **30**(10): 3168-3172.
- [9] 杨颖, 宋英强, 胡月明, 等. 基于序贯指示模拟的农田土壤重金属风险区域识别[J]. *应用生态学报*, 2018, **29**(5): 1695-1704.
- Yang H, Song Y Q, Hu Y M, *et al.* Using sequential indicator simulation method to define risk areas of soil heavy metals in farmland [J]. *Chinese Journal of Applied Ecology*, 2018, **29**(5): 1695-1704.
- [10] Qiao P W, Dong N, Yang S C, *et al.* Quantitative analysis of the main sources of pollutants in the soils around key areas based on the positive matrix factorization method [J]. *Environmental Pollution*, 2021, **273**, doi: 10.1016/j.envpol.2021.116518.
- [11] 佟瑞鹏, 杨校毅. 基于蒙特卡罗模拟的土壤环境健康风险评估: 以 PAHs 为例[J]. *环境科学*, 2017, **38**(6): 2522-2529.
- Tong R P, Yang X Y. Environmental health risk assessment of contaminated soil based on Monte Carlo method: a case of PAHs [J]. *Environmental Science*, 2017, **38**(6): 2522-2529.
- [12] 徐辰星. 宜兴市土壤重金属分布特征与不确定性分析[D]. 南京: 南京大学, 2018.
- Xu C X. Distribution and uncertainties evaluation of soil heavy metals in Yixing, Yangtze River delta, China [D]. Nanjing: Nanjing University, 2018.
- [13] 孙雪菲, 张丽霞, 董玉龙, 等. 典型石化工业城市土壤重金属源解析及空间分布模拟[J]. *环境科学*, 2021, **42**(3): 1093-1104.
- Sun X F, Zhang L X, Dong Y L, *et al.* Source apportionment and spatial distribution simulation of heavy metals in a typical petrochemical industrial city [J]. *Environmental Science*, 2021, **42**(3): 1093-1104.
- [14] 宋志廷, 赵玉杰, 周其文, 等. 基于地质统计及随机模拟技术的天津武清区土壤重金属源解析[J]. *环境科学*, 2016, **37**(7): 2756-2762.
- Song Z T, Zhao Y J, Zhou Q W, *et al.* Applications of geostatistical analyses and stochastic models to identify sources of soil heavy metals in Wuqing district [J]. *Environmental Science*, 2016, **37**(7): 2756-2762.
- [15] 张泽浦, 王学军. 土壤微量元素含量空间分布的条件模拟[J]. *土壤学报*, 1998, **35**(3): 423-429.
- Zhang Z P, Wang X J. Conditional simulation for spatial distribution of trace elements in soils [J]. *Acta Pedologica Sinica*, 1998, **35**(3): 423-429.

- [16] 赵永存, 孙维侠, 黄标, 等. 不同随机模拟方法定量土壤镉含量预测的不确定性研究[J]. 农业环境科学学报, 2008, **27**(1): 139-146.
Zhao Y C, Sun W X, Huang B, *et al.* Comparison of sequential indicator simulation and sequential indicator co-simulation for quantifying the local uncertainty of soil Cd content [J]. Journal of Agro-Environment Science, 2008, **27**(1): 139-146.
- [17] Delbari M, Loiskandl W, Afrasiab P. Uncertainty assessment of soil organic carbon content spatial distribution using geostatistical stochastic simulation [J]. Australian Journal of Soil Research, 2010, **48**(1): 27-35.
- [18] Han J Q, Mao K B, Xu T R, *et al.* A soil moisture estimation framework based on the CART algorithm and its application in China [J]. Journal of Hydrology, 2018, **563**: 65-75.
- [19] Dharumarajan S, Lalitha M, Niranjana K V, *et al.* Evaluation of digital soil mapping approach for predicting soil fertility parameters—a case study from Karnataka Plateau, India [J]. Arabian Journal of Geosciences, 2022, **15**(5), doi: 10.1007/s12517-022-09629-8.
- [20] Gardin L, Chiesi M, Fibbi L, *et al.* Mapping soil organic carbon in Tuscany through the statistical combination of ground observations with ancillary and remote sensing data [J]. Geoderma, 2021, **404**, doi: 10.1016/j.geoderma.2021.115386.
- [21] Were K, Bui D T, Dick Ø B, *et al.* A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape [J]. Ecological Indicators, 2015, **52**: 394-403.
- [22] Chen S C, Liang Z Z, Webster R, *et al.* A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution [J]. Science of the Total Environment, 2019, **655**: 273-283.
- [23] Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author) [J]. Statistical Science, 2001, **16**(3): 199-231.
- [24] Zeraatpisheh M, Ayoubi S, Jafari A, *et al.* Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran [J]. Geoderma, 2019, **338**: 445-452.
- [25] Sekulić A, Kilibarda M, Heuvelink G B M, *et al.* Random forest spatial interpolation [J]. Remote Sensing, 2020, **12**(10), doi: 10.3390/rs12101687.
- [26] 刘伟, 郜允兵, 周艳兵, 等. 农田土壤重金属空间变异多尺度分析——以北京顺义土壤Cd为例 [J]. 农业环境科学学报, 2019, **38**(1): 87-94.
Liu W, Gao Y B, Zhou Y B, *et al.* Multi scale analysis of spatial variability of heavy metals in farmland soils: Case study of soil Cd in Shunyi District of Beijing, China [J]. Journal of Agro-Environment Science, 2019, **38**(1): 87-94.
- [27] Wang L, Wu W, Liu H B. Digital mapping of topsoil pH by random forest with residual kriging (RFRK) in a hilly region [J]. Soil Research, 2019, **57**(4): 387-396.
- [28] 刘焕军, 张美薇, 杨昊轩, 等. 多光谱遥感结合随机森林算法反演耕作土壤有机质含量 [J]. 农业工程学报, 2020, **36**(10): 134-140.
Liu H J, Zhang M W, Yang H X, *et al.* Inversion of cultivated soil organic matter content combining multi-spectral remote sensing and random forest algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, **36**(10): 134-140.
- [29] 刘兴业, 李景叶, 陈小宏, 等. 联合多点地质统计学与序贯高斯模拟的随机反演方法 [J]. 地球物理学报, 2018, **61**(7): 2998-3007.
Liu X Y, Li J Y, Chen X H, *et al.* A stochastic inversion method integrating multi-point geostatistics and sequential Gaussian simulation [J]. Chinese Journal of Geophysics, 2018, **61**(7): 2998-3007.
- [30] 许尔琪, 张红旗. 伊犁新垦区土壤有机质的克里金插值和条件模拟多尺度分析 [J]. 土壤, 2013, **45**(1): 91-98.
Xu E Q, Zhang H Q. Multi-scale Analysis of kriging interpolation and conditional simulation for soil organic matters in newly reclaimed area in Yili [J]. Soils, 2013, **45**(1): 91-98.
- [31] Tan Z, Yang Q, Zheng Y. Machine learning models of groundwater arsenic spatial distribution in Bangladesh: Influence of Holocene sediment depositional history [J]. Environmental Science & Technology, 2020, **54**(15): 9454-9463.
- [32] 刘志鹏, 邵明安, 王云强. 区域尺度下黄土高原土壤全钾含量的空间模拟 [J]. 农业工程学报, 2012, **28**(22): 132-140.
Liu Z P, Shao M A, Wang Y Q. Spatial simulation of soil total potassium in regional scale for Loess Plateau Region [J]. Transactions of the Chinese Society of Agricultural Engineering, 2012, **28**(22): 132-140.
- [33] 王绍武, 沈润平, 陈萍萍, 等. 基于马尔可夫链模型的华中地区土壤质地空间分布模拟研究 [J]. 河南农业大学学报, 2019, **53**(2): 282-288.
Wang S W, Shen R P, Chen P P, *et al.* Simulation study on spatial distribution of soil texture based on Markov-chain model in central China [J]. Journal of Henan Agricultural University, 2019, **53**(2): 282-288.
- [34] 李丹凤, 邵明安. 一维马尔可夫链模拟黑河中游流域土壤质地垂向变异 [J]. 农业工程学报, 2013, **29**(5): 71-80.
Li D F, Shao M A. One-dimensional Markov chain simulation of vertical change of soil texture in middle reaches of Heihe river northwest China [J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, **29**(5): 71-80.
- [35] 任旭红, 潘瑜春, 高秉博, 等. 类别辅助变量参与下的土壤无偏采样布局优化方法 [J]. 农业工程学报, 2014, **30**(21): 120-128.
Ren X H, Pan Y C, Gao B B, *et al.* Optimization method of unbiased soil sampling and layout using categorical auxiliary variables information [J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, **30**(21): 120-128.
- [36] 陈同斌, 郑袁明, 陈煌, 等. 北京市土壤重金属含量背景值的系统研究 [J]. 环境科学, 2004, **25**(1): 117-122.
Chen T B, Zheng Y M, Chen H, *et al.* Background concentrations of soil heavy metals in Beijing [J]. Environmental Science, 2004, **25**(1): 117-122.
- [37] Li B H, Tan C H. Distribution and risk evaluation of heavy metal in the soil of typical grassed swale [J]. MATEC Web of Conferences, 2016, **59**, doi: 10.1051/mateconf/20165904001.
- [38] 刘明杰, 徐卓揆, 郜允兵, 等. 基于机器学习的稀疏样本下的土壤有机质估算方法 [J]. 地球信息科学学报, 2020, **22**(9): 1799-1813.
Liu M J, Xu Z K, Gao Y B, *et al.* Estimating soil organic matter based on machine learning under sparse sample [J]. Journal of Geo-information Science, 2020, **22**(9): 1799-1813.
- [39] 姜成晟, 王劲峰, 曹志冬. 地理空间抽样理论研究综述 [J]. 地理学报, 2009, **64**(3): 368-380.
Jiang C S, Wang J F, Cao Z D. A review of geo-spatial sampling theory [J]. Acta Geographica Sinica, 2009, **64**(3): 368-380.
- [40] 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 Y, Zhao X M, Guo X. Prediction of total nitrogen distribution in surface soil based on multi-source auxiliary variables and random forest approach[J]. *Acta Pedologica Sinica*, 2022, **59**(2): 451-460.
- [41] Zhang W C, Wan H S, Zhou M H, *et al.* Soil total and organic carbon mapping and uncertainty analysis using machine learning techniques[J]. *Ecological Indicators*, 2022, **143**, doi: 10.1016/j.ecolind.2022.109420.
- [42] 高中原, 肖荣波, 王鹏, 等. 融合自然-人为因子改进回归克里格对土壤镉空间分布预测[J]. *环境科学*, 2021, **42**(1): 343-352.
Gao Z Y, Xiao R B, Wang P, *et al.* Improved regression Kriging prediction of the spatial distribution of the soil cadmium by integrating natural and human factors[J]. *Environmental Science*, 2021, **42**(1): 343-352.
- [43] 谢恩泽, 赵永存, 陆访仪, 等. 不同方法预测苏南农田土壤有机质空间分布对比研究[J]. *土壤学报*, 2018, **55**(5): 1051-1061.
Xie E Z, Zhao Y C, Lu F Y, *et al.* Comparison Analysis of methods for prediction of spatial distribution of soil organic matter contents in farmlands south Jiangsu, China [J]. *Acta Pedologica Sinica*, 2018, **55**(5): 1051-1061.
- [44] 刘艳芳, 宋玉玲, 郭龙, 等. 结合高光谱信息的土壤有机碳密度统计模型[J]. *农业工程学报*, 2017, **33**(2): 183-191.
Liu Y F, Song Y L, Guo L, *et al.* Geostatistical models of soil organic carbon density prediction based on soil hyperspectral reflectance[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, **33**(2): 183-191.
- [45] Facchinelli A, Sacchi E, Mallen L. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils [J]. *Environmental Pollution*, 2001, **114**(3): 313-324.
- [46] Chi Y, Sun J K, Li T, *et al.* Spatial simulations of soil content, storage, and quality indices in an archipelago off the Yangtze River Estuary, China [J]. *Ecological Indicators*, 2023, **146**, doi: 10.1016/j.ecolind.2022.109774.

环 境 科 学

CONTENTS

Spatio-temporal Evolution, Dynamic Transition, and Convergence Trend of Urban Carbon Emission Intensity in China	YANG Qing-ke, WANG Lei, ZHU Gao-li, <i>et al.</i> (1869)
Spatio-temporal Evolution and Trend Prediction of Transport Carbon Emission Efficiency	ZHENG Yan, JIANG Xue-mei, XIAO Yu-jie (1879)
Research on Carbon Emission Decoupling Factors Based on STIRPAT Model and LMDI Decomposition	ZHANG Jiang-yan (1888)
Carbon Peak Pathways of Industrial Parks Based on the LEAP Model: A Case Study of a National Development Zone in Nanjing	LI Hui-peng, LI Li, YIN Yin, <i>et al.</i> (1898)
Accounting and Analysis of Carbon Emissions in Universities: A Case Study of Beijing A University	CAO Rui, FENG Li, ZHANG Li-qiu (1907)
Synergy Effects and Driving Factors of Pollution and Carbon Emission Reduction in Manufacturing Industry in Beijing	YU Shan, HAN Yu-hua, MU Jie, <i>et al.</i> (1917)
Carbon Emissions Trading Prices and Total Factor Productivity: Evidence from China	WU Xue-ping (1926)
Analysis of Spatio-temporal Distribution Characteristics and Influencing Factors of PM _{2.5} and PM ₁₀ in Chinese Cities	LI Jiang-su, DUAN Liang-rong, ZHANG Tian-jiao (1938)
Spatial-temporal Characteristics and Influencing Factors of PM _{2.5} and Ozone in the Border Area of Jiangsu, Anhui, Shandong, and Henan from 2017 to 2021	CHEN Wei, XU Xue-zhe, LIU Wen-qing (1950)
Analysis of the Characteristics and Causes of PM _{2.5} and O ₃ Pollution in Xi'an During the Epidemic Lockdown Period	YUAN Xiao-hong, ZHANG Qiang, LI Qi, <i>et al.</i> (1963)
PM _{2.5} Chemical Composition and Spatial Variability in the Five Cities of Southern Jiangsu During Fall and Winter	FENG Wei, DING Feng, SHANG Yue, <i>et al.</i> (1975)
Source Apportionment of Fine Particles in Xianning, Hubei	LUO Yi, ZHU Kuan-guang, CHEN Pu-long, <i>et al.</i> (1983)
Spillover Benefits of Ammonia Emission Reduction in Xingtai Area to the Improvement of Atmospheric PM _{2.5} Concentration in the Beijing-Tianjin-Hebei Region	BIAN Ze-jun, WEN Chao-yu, LANG Jian-lei, <i>et al.</i> (1994)
Characteristics, Risk Assessment, and Sources of the Polluted Metallic Elements in PM _{2.5} During Winter in Liaocheng City	ZHANG Jing-qiao, ZHU Yao, CAO Ting, <i>et al.</i> (2003)
Composition Characteristics of Typical VOCs Sources in the Highland City of Lhasa	GUO Shu-zheng, YE Chun-xiang, LIN Wei-li, <i>et al.</i> (2011)
VOCs Pollution Characteristics and Health Risk Assessment in Typical Industrial Parks in Beijing: Environmental Impact of High and New Technology Industries	WANG Jie, YAO Zhen, WANG Min-yan, <i>et al.</i> (2019)
Analysis of a Typical Ozone Pollution Process in the Chengdu Plain Under the Influence of High Temperature Extremes	LEI Li-juan, ZHANG Yi, LUO Yi-na, <i>et al.</i> (2028)
Presence, Behavior, and Risk of Antiviral Drugs in the Aqueous Environment	GE Lin-ke, LI Xuan-yan, CAO Sheng-kai, <i>et al.</i> (2039)
Spatio-temporal Variation and Probability Health Risk of Heavy Metals in Surface Water of Xiaolangdi Reservoir in the Yellow River	WANG Liang, DENG Xue-jiao, WANG Xiao-lei, <i>et al.</i> (2054)
Hydrochemical Characteristics and Control Factors of Strontium Enrichment of the Duoqu River Basin in Eastern Xizang	LI Jing-jie, LIAN Sheng, WANG Ming-guo, <i>et al.</i> (2067)
Characteristics of Stable Isotopes in Precipitation and Its Moisture Sources in the Inland Regions of Northwest China	ZHANG Yan-yan, XIN Cun-lin, GUO Xiao-yan, <i>et al.</i> (2080)
Recharge Source and Evolution Process of Karst Groundwater in Tai'an Urban Area Based on Hydrochemistry and Hydrogen and Oxygen Isotopes	MENG Ling-hua (2096)
Hydrochemical Characteristics and Influencing Factors of Surface Water and Groundwater in the Mountainous Area of Danjiang River Basin	ZHANG Zi-yan, FU Yong-peng, WANG Ning-tao, <i>et al.</i> (2107)
Hydrochemical Characteristics and Control Factors of Groundwater in Yuanyangchi Irrigation Area, Jinta Basin	WANG Xiao-yan, HAN Shuang-bao, ZHANG Meng-nan, <i>et al.</i> (2118)
Hydrochemical Characteristics and Health Risk Assessment of Shallow Groundwater in the Plain Zone of Deyang City	LIU Nan, CHEN Meng, GAO Dong-dong, <i>et al.</i> (2129)
Spatial and Temporal Distribution of Microplastics in the Sediments of Typical Urban River Network	XU Wan-lu, FAN Yi-fan, QIAN Xin (2142)
Construction of Nitrogen Emission Inventory at Sub-basin Scale and Analysis of Key Sources in Fuxian Lake Watershed	WANG Yan-jie, LIANG Qi-bin, WANG Yan-xia, <i>et al.</i> (2150)
Preparation of Lanthanum Crosslinked Calcium Peroxide/Sepiolite/Sodium Alginate Composite Hydrogels and Their Elimination Performance for Endogenous Phosphorus	QU Si-tong, SHAN Su-jie, WANG Chong-ming, <i>et al.</i> (2160)
Formation Mechanism, Structural Characteristics of Ultrafine Mineral Particles, and Their Environmental Effects	LIU Zhen-hai, ZHANG Zhan-hua, YUAN Yu-xin, <i>et al.</i> (2171)
Research Progress in Application of Biochar-immobilized Bacteria Composites in Environmental Remediation	SUN Shu-yu, HUANG Meng-xin, KONG Qiang, <i>et al.</i> (2185)
Effect of Microplastics on Ammonia Nitrogen Adsorption by Zeolite in Water and Its Mechanism	LIAN Jian-jun, XIE Shi-ting, WU Pei, <i>et al.</i> (2195)
Mechanism of Ultraviolet Aging Effect on the Adsorption of Ciprofloxacin by Nano-biochar	MA Feng-feng, XUE Zhi-yi, ZHAO Bao-wei (2203)
Response of Phytoplankton Communities and Environmental Factors Under the Influence of Land Use in the Wuding River Basin	GUO Shan-song, HU En, DING Yi-tong, <i>et al.</i> (2211)
Bacterial Community Diversity and Functional Gene Abundance of Culturable Bacteria in the Wetland of Poyang Lake	YU Jiang, WANG Chun, LONG Yong, <i>et al.</i> (2223)
Bacterial Community Structure and Its Relationship with Heavy Metals in Sediments of Diannong River	LIU Shuang-yu, MENG Jun-jie, QIU Xiao-cong, <i>et al.</i> (2233)
Impacts of Treated Wastewater on Bacterial and Fungal Microbial Communities in Receiving Rivers	GUO You-shun, YU Zhong, HAO Wen-bin, <i>et al.</i> (2246)
Metagenomics Reveals the Characteristics and Functions of Bacterial Community in the Advanced Wastewater Treatment Process	HU Jian-shuang, WANG Yan, ZHOU Zheng, <i>et al.</i> (2259)
Spatio-temporal Characteristics of Habitat Quality and Natural-human Driven Mechanism in Dabie Mountain Area	ZHENG Ya-ping, ZHANG Jun-hua, TIAN Hui-wen, <i>et al.</i> (2268)
Hyperspectral Inversion of Soil Organic Matter Content Based on Continuous Wavelet Transform, SHAP, and XGBoost	YE Miao, ZHU Lin, LIU Xu-dong, <i>et al.</i> (2280)
Effects of Straw Retention, Film Mulching, and Nitrogen Input on Soil Quality in Dryland Wheat Field	YE Zi-zhuang, WANG Song-yan, LU Xiao, <i>et al.</i> (2292)
Effects of Long-term Application of Organic Fertilizer on the Occurrence and Migration Characteristics of Soil Microplastics	WANG Chang-yuan, MA Xiao-chi, GUO De-jie, <i>et al.</i> (2304)
Effect of Different Soil Salinities on N ₂ O Emission: A Meta-analysis	HUANG Yi-hua, SHE Dong-li, SHI Zhen-qi, <i>et al.</i> (2313)
Assessment and Prediction of Carbon Storage Based on Land Use/Land Cover Dynamics in the Guangdong-Hong Kong-Macao Greater Bay Area	ZHENG Hui-ling, ZHENG Hui-feng (2321)
Simulation of Temporal and Spatial Changes in Ecosystem Carbon Storage in Funiu Mountains Based on InVEST Model	ZHANG Zhe, SHI Zhen-qin, ZHU Wen-bo, <i>et al.</i> (2332)
Relationship Between Microbial Nutrient Limitation and Soil Organic Carbon Fraction During Shelterbelts Construction	XU Feng-jing, HUANG Yi-mei, HUANG Qian, <i>et al.</i> (2342)
Characteristics and Driving Forces of Organic Carbon Mineralization in Brown Soil with Long-term Straw Returning	ZHAO Yu-hang, YIN Hao-kai, HU Xue-chun, <i>et al.</i> (2353)
Effects of Winter Green Manure Mulching on Soil Aggregates, Organic Carbon, and AMF Diversity	LU Ze-rang, CHEN Jia-yan, LI Zhi-xian, <i>et al.</i> (2363)
Spatial and Temporal Distribution of Nitrogen in the Liaohu River Basin and Its Responses to Land Use and Rainfall	ZHOU Bo, LI Xiao-guang, TONG Si-chang, <i>et al.</i> (2373)
Nitrous Oxide Emissions from Ponds in Typical Agricultural Catchment with Intensive Agricultural Activity	ZHANG Xin-yue, XIAO Qi-tao, XIE Hui, <i>et al.</i> (2385)
Effects of Biochar Combined with Different Types of Nitrogen Fertilizers on Denitrification Bacteria Community in Vegetable Soil	LIU Xiao-wan, LIU Xing-ren, GAO Shang-jie, <i>et al.</i> (2394)
Investigation of Soil Microbial Characteristics During Stand Development in <i>Pinus tabulaeformis</i> Forest in Taiyue Mountain	MA Yi-shu, CAO Ya-xin, NIU Min, <i>et al.</i> (2406)
Simulation of Heavy Metal Content in Soil Based on Sparse Sample Sites	ZHANG Jia-qi, PAN Yu-chun, GAO Shi-chen, <i>et al.</i> (2417)
Identification Priority Source of Heavy Metal Pollution in Greenspace Soils Based on Source-specific Ecological and Human Health Risk Analysis in the Yellow River Custom Tourist Line of Lanzhou	LI Jun, LI Xu, LI Kai-ming, <i>et al.</i> (2428)
Heavy Metal Accumulation and Assessment of Potential Ecological Risk Caused by Soil pH Changes in Different Types of Purple Soils in Southwest China	ZHANG Hai-lin, ZHANG Yu, WANG Ding, <i>et al.</i> (2440)
Assessment and Source Analysis of Heavy Metal Pollution in Arable Land Around an Iron Ore Mining Area in Chongqing	LIAO Ze-yuan, LI Jie-qin, SHEN Zhi-jie, <i>et al.</i> (2450)
Health Risk Assessment of Heavy Metals in Farmland Soils and Crops Around Metal Mines	WEI Hong-bin, LUO Ming, XIANG Lei, <i>et al.</i> (2461)
Effect of Red Mud-based Nano Zero-valent Iron on Remediation of Polymetallic Contaminated Soil	LIU Long-yu, YANG Shi-li, ZHAO Huang-shi-yu, <i>et al.</i> (2474)
Effects of Zinc Application on Cadmium Accumulation in Alkaline Soil-Wheat Seedling Systems	ZHANG Yao, WANG Tian-qi, NIU Shuo, <i>et al.</i> (2480)