

目次

加速能源转型与产业结构调整的环境健康协同效益评估:以京津冀鲁地区为例	杨玺, 孙奕生, 常世彦, 李胜悦, 郑昊天, 王书肖, 张希良 (3627)
基于不同电力需求的中国减排降碳协同增效路径	向梦宇, 王深, 吕连宏, 张楠, 白梓函 (3637)
城市碳达峰碳中和行动评估方法与应用	张保留, 白梓函, 张楠, 吕连宏, 阳平坚 (3649)
基于可解释性机器学习的城市 O ₃ 驱动因素挖掘	董佳奇, 胡冬梅, 闫雨龙, 彭林, 张辉辉, 牛月圆, 段小琳 (3660)
基于响应曲面法的臭氧生成敏感性分析	朱禹寰, 陈冰, 张雅茹, 刘晓, 李光耀, 舍静, 陈强 (3669)
基于 WRF-CMAQ/ISAM 模型的榆林市夏季 O ₃ 来源解析	王伊凡, 仝纪龙, 陈羽翔, 林鑫, 刘永乐, 敖丛杰, 刘浩天 (3676)
基于轻量级梯度提升机的南京大气臭氧浓度预测	朱珈莹, 安俊琳, 冯悦政, 贺婕, 张玉欣, 王俊秀 (3685)
2022 年广东省冬季一次臭氧污染过程的气象成因及潜在源区分析	李婷苑, 陈靖扬, 龚宇, 沈劲 (3695)
夏季对流层臭氧辐射强迫对华北地区天气和空气质量的影响	杜楠, 陈磊, 廖宏, 朱佳, 李柯 (3705)
气象、本地光化学生成和外围传输对长沙市 2018~2020 年臭氧污染趋势变化影响的识别	杨俊, 杨雷峰, 丁华, 谢丹平, 刘妍妍, 余涛, 吕明, 袁自冰 (3715)
2000~2021 年成渝城市群 PM _{2.5} 时空变化及驱动机制多维探测	徐勇, 郭振东, 郑志威, 戴强玉, 赵纯, 黄雯婷 (3724)
基于 GTWR-XGBoost 模型的四川省 PM _{2.5} 小时浓度估算	吴迪, 杜宁, 王莉, 吴宇宏, 张少磊, 周彬, 张显云 (3738)
2013~2021 年邢台 PM _{2.5} 重污染过程输送和增长特性	江琪, 盛黎, 靳雨晨, 王继康, 尤媛, 王飞 (3749)
基于高分辨率在线观测数据分析上海市城区秋冬季大气有机气溶胶化学特征及污染来源	朱书慧 (3760)
机动车尾气碳质气溶胶排放因子及其稳定碳同位素特征	于鸣媛, 王谦, 付明亮, 戈畅, 谢峰, 曹芳, 章炎麟 (3771)
沈阳工业区夏季 VOCs 组成特征及其对二次污染形成的贡献	关璐, 苏枫枫, 库盈盈, 胡建林, 于兴娜 (3779)
疫情管控期间深圳市城区 VOCs 的变化特征及减排效果评估	云龙, 林楚雄, 李成柳, 邱志诚, 古添发, 李光程, 张明棣, 郭健锋 (3788)
西安市冬季道路扬尘中有机质组成特征及其氧化潜势	王擎雯, 陈庆彩, 王超, 王瑞鹤, 刘国瑞, 李豪, 李艳广 (3797)
淮北孙疃矿区地表尘中多环芳烃类化合物的污染特征及致癌风险评价	徐振鹏, 钱雅慧, 洪秀萍, 罗钟庚, 高秀龙, 梁汉东 (3809)
雄安建设初期白洋淀水质时空差异及其对降雨和土地利用的响应	王子铭, 杨丽虎, 宋献方 (3820)
流域尺度污染溯源模拟-优化防控方法:以铜陵市顺安河流域为例	刘国王辰, 陈磊, 李佳奇, 张钰晨, 赵奕欣, 刘妍琪, 沈珍瑶 (3835)
南昌市浅层地下水水质评价及监测指标优化	郑紫吟, 储小东, 徐金英, 马志飞 (3846)
闪电河流域“三水”氢氧同位素特征及水体转换分析	杨丽娜, 贾德彬, 高瑞忠, 苏文旭, 卢方园, 郝玉胜 (3855)
金沙江丰富类和稀有类浮游真核微生物的分布特征与影响因素	燕炳成, 崔戈, 孙胜浩, 王沛芳, 王超, 吴程, 陈娟 (3864)
嘉陵江不同干扰断面河道沉积物细菌群落多样性	竹兰萍, 张拓, 李佳宁, 王佳颖, 郭伟亮, 徐飞, 张富斌 (3872)
水库水深变化对不同浮游微生物群落及网络互作关键种的影响	王洵, 廖琴, 王沛芳, 袁秋生, 胡斌, 邢小蕾, 徐浩森 (3881)
典型工业区地下水细菌群落多样性特征与环境因子响应初探	吴建强, 张书源, 王敏, 陈敏, 叶文娟, 熊丽君, 黄沈发 (3892)
太湖流域肥料施用策略调整对典型作物系统氮磷流失的影响	俞映倬, 王逸之, 杨根, 杨林章, 段婧婧, 韩雪梅, 薛利红 (3902)
基于机器学习的长江流域农田氮径流流失负荷估算	张有福, 潘哲祺, 陈丁江 (3913)
千岛湖地区上梧溪流域地表径流非点源氮污染分类识别	俞珂, 严璇, 唐张轩, 张方方, 何圣嘉, 姜培坤 (3923)
沱江和涪江水系干支流氮磷营养盐的空间分布特征	李子阳, 周明华, 徐鹏, 陈露, 刘祥龙, 林洪羽, 江南, 任兵, 张博文 (3933)
外秦淮河底泥污染及疏浚效果	张沐, 任增谊, 张曼, 赵琼, 尹洪斌 (3945)
新污染物多环芳烃衍生物的来源、分布与光化学行为	葛林科, 王子宇, 曹胜凯, 车晓佳, 朱超, 张蓬, 马宏瑞 (3957)
镉砷在针铁矿界面共吸附的相互作用机制	苏子贤, 刘赛红, 管玉峰, 陶亮 (3970)
环糊精键合 Fe-TAML 催化剂的制备及其活化 H ₂ O ₂ 氧化水中有机微污染物	刘清泉, 蔡本哲, 蔡喜运 (3978)
Co(II) 活化过一硫酸盐降解氨基三甲膦酸的性能及反应机制	朱敬林, 汪舒 (3990)
微塑料诱导下污泥造粒潜能变化及微生物富集特征	谢晴帆, 俞楠, 张妮, 谢周云, 单珂欣, 吴亦馨, 唐力, 夏静芬, 杨国靖 (3997)
中国农田土壤 Cd 累积分布特征及概率风险评价	王静, 魏恒, 潘波 (4006)
渝东南典型地质高背景区土壤重金属来源解析及污染评价	蒋玉莲, 余京, 王锐, 王佳彬, 李瑜, 余飞, 张云逸 (4017)
基于 PMF 模型的农田土壤重金属源暴露风险综合评价:以浙江省某电子垃圾拆解区为例	方嘉, 何影, 黄乃涛, 支裕优, 傅伟军 (4027)
乌鲁木齐市郊农田土壤及农作物中多环芳烃的污染特征及风险评价	范悦, 曹双瑜, 艾力江·努尔拉, 于芸云江·吗米提敏, 阿不都艾尼·阿不里, 谢宣宣, 古丽斯坦·阿不都拉, 刘河疆 (4039)
荒漠绿洲土壤抗生素抗性基因分布特征及驱动机制	黄福义, 周曙仡, 苏建强, 朱永官 (4052)
近 15 年我国土壤抗生素污染特征与生物修复研究进展	赵晓东, 乔青青, 秦宵睿, 李晓晶, 李永涛 (4059)
改性生物炭修复砷镉复合污染土壤研究进展	吕鹏, 李莲芳, 黄晓雅 (4077)
沼液还田对土壤-作物系统重金属累积的影响:Meta 分析	赵奇志, 杨志敏, 孔凡靖, 熊海灵, 朱康文, 陈玉成 (4091)
两种铁改性生物炭对微碱性砷镉污染土壤的修复效果	梁欣冉, 何丹, 郑墨华, 付庆灵, 胡红青, 朱俊 (4100)
秸秆离田对土壤 Cd 生物有效性及水稻 Cd 积累的影响	王子钰, 周航, 周坤华, 谭文韬, 蒋毅, 唐棋, 伍港繁, 辜娇峰, 曾鹏, 廖柏寒 (4109)
秸秆与脱硫石膏配施改良黄河三角洲盐碱地的理化性质	赵惠丽, 于金艺, 刘涛, 王丽, 赵英 (4119)
基于环境 DNA 的复合污染土壤生物评价和胁迫诊断	黄湘云, 钟文军, 刘训杰, 毕婉娟, 钱林皓, 张效伟 (4130)
地质高背景区土壤-玉米重金属综合质量评价	张传华, 王钟书, 刘力, 刘燕 (4142)
浙江省蔬菜生产系统重金属污染生态健康风险	张述敏, 刘翠玲, 杨桂玲, 邓美华 (4151)
生物炭配施磷肥对土壤养分、酶活性及紫花苜蓿养分吸收的影响	刘鑫裕, 王冬梅, 张泽洲, 张鹏, 樊桐桐 (4162)
低密度聚乙烯微塑料对空心菜生长和生理特征的影响	周颖, 蒋文婷, 刘训悦, 朱高荻, 唐荣贵, 章海波, 蔡延江 (4170)
沼液施用对麦茬口期土壤微生物群落结构特征及功能的影响	乔宇颖, 奚辉, 李娜, 陈喜靖, 沈阿林, 喻曼 (4179)
大同铅锌尾矿不同污染程度土壤细菌群落分析及生态功能特征	刘泽勋, 庄家尧, 刘超, 郑康, 陈玲 (4191)
马尾松采伐迹地火烧黑炭对土壤有机碳组分和碳转化酶活性的影响	姚智, 焦鹏宇, 吴晓生, 严强, 刘先, 胡亚林, 王玉哲 (4201)
《环境科学》征订启事(3648) 《环境科学》征稿简则(3659) 信息(3714, 4150, 4178)	

基于可解释性机器学习的城市 O₃ 驱动因素挖掘

董佳奇¹, 胡冬梅^{1*}, 闫雨龙², 彭林², 张鹏辉¹, 牛月圆¹, 段小琳¹

(1. 华北电力大学环境科学与工程学院, 资源环境系统优化教育部重点实验室, 北京 102206; 2. 北京交通大学环境学院, 北京 100044)

摘要: 受前体物排放和气象条件等因素共同驱动, 大气臭氧(O₃)已成为影响城市夏季环境空气质量的主要污染物。目前物理化学机制驱动的演绎模型在进行 O₃ 污染解析时需要的模型参数众多, 运算时效性较差; 数据驱动的归纳模型运算效率高, 但存在可解释性差等问题。通过建立可解释性数据驱动的 Correlation-ML-SHAP 模型, Correlation 模块挖掘 O₃ 浓度关联影响因素, 机器学习 ML 模块耦合可解释性 SHAP 模块计算各驱动因素对 O₃ 浓度的影响贡献, 实现对驱动因素的定量解析, 并以晋城市 2021 年夏季 O₃ 污染过程为例开展应用研究。结果表明, Correlation-ML-SHAP 模型能够挖掘并利用强驱动因素模拟 O₃ 浓度和量化影响贡献, 其中 ML 模块采用 XGBoost 模型模拟准确度最佳。2021 年夏季晋城市 O₃ 污染强驱动因素为: 气温、日照强度、湿度和前体物排放水平, 贡献权重为: 32.1%、21.3%、16.5% 和 15.6%, 其中气温、日照强度和前体物排放水平贡献在污染天分别提升 3.4%、1.2% 和 1.2%, 前体物排放贡献权重在污染天排名提升至第 3。各驱动因素对 O₃ 浓度呈非线性交互影响, 气温超 24℃ 或湿度低于 70%, 分别有 94.9% 和 94.1% 的概率对 O₃ 污染呈正贡献, 这种气象条件下 $\rho(\text{NO}_2)$ 超 $9 \mu\text{g}\cdot\text{m}^{-3}$ 或 $\rho(\text{CO})$ 超 $0.7 \text{mg}\cdot\text{m}^{-3}$, 分别有 94.9% 和 99.3% 的概率对 O₃ 污染呈正贡献。东南风风速低于 $5.8 \text{m}\cdot\text{s}^{-1}$ 或南风风速低于 $5.3 \text{m}\cdot\text{s}^{-1}$, 均对 O₃ 污染呈正贡献。模型定量解析了各驱动因素对城市 O₃ 浓度的影响贡献, 可为夏季城市大气 O₃ 污染防控提供基础依据。

关键词: O₃ 污染; 驱动因素; 影响贡献; 可解释性; 机器学习

中图分类号: X515 文献标识码: A 文章编号: 0250-3301(2023)07-3660-09 DOI: 10.13227/j.hjxx.202208214

Revealing Driving Factors of Urban O₃ Based on Explainable Machine Learning

DONG Jia-qi¹, HU Dong-mei^{1*}, YAN Yu-long², PENG Lin², ZHANG Peng-hui¹, NIU Yue-yuan¹, DUAN Xiao-lin¹

(1. Key Laboratory of Resources and Environmental System Optimization, College of Environmental Science and Engineering, North China Electric Power University, Beijing 102206, China; 2. School of the Environment, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Driven by precursor emissions, meteorological conditions, and other factors, atmospheric ozone (O₃) has become the main pollutant affecting urban air quality in summer. The current deductive models driven by physical and chemical mechanisms require a large number of parameters for the analysis of O₃ pollution, and the calculation timeliness is poor. The data-driven inductive models are efficient but have problems such as poor explanation. In this study, an explainable model of data-driven Correlation-ML-SHAP was established to reveal the strongly correlated influencing factors of O₃ concentration. Additionally, the machine learning ML module coupled with the explainable SHAP module was used to calculate the contributions of driving factors to O₃ concentration, so as to realize the quantitative analysis of driving factors. The O₃ pollution process in the summer of 2021 in Jincheng City was used as an example to carry out the application research. The results showed that the Correlation-ML-SHAP model could reveal and use strong driving factors to simulate O₃ concentration and quantify influence contribution, and the ML module used the XGBoost model to achieve the best simulation accuracy. Air temperature, solar radiation, relative humidity, and precursor emission level were the strong driving factors of O₃ pollution in Jincheng City in summer 2021, and the contribution weights were 32.1%, 21.3%, 16.5%, and 15.6%. The contribution weights of air temperature, solar radiation, and precursor emission level increased by 3.4%, 1.2%, and 1.2% on polluted days, respectively, and the contribution weights of precursor emission level rose to third place on polluted days. Each driving factor had a nonlinear interaction effect on O₃ concentration. When the air temperature exceeded 24℃, or the relative humidity was lower than 70%, there was a 94.9% and 94.1% probability of positive contribution to O₃ pollution, respectively. Under such meteorological conditions, $\rho(\text{NO}_2)$ exceeded $9 \mu\text{g}\cdot\text{m}^{-3}$, or $\rho(\text{CO})$ exceeded $0.7 \text{mg}\cdot\text{m}^{-3}$, and there was a 94.9% and 99.3% probability of positive contribution to O₃ pollution, respectively. The southeast wind speed was lower than $5.8 \text{m}\cdot\text{s}^{-1}$, or the south wind speed was lower than $5.3 \text{m}\cdot\text{s}^{-1}$, both of which contributed positively to O₃ pollution. The model quantitatively analyzed the influence contribution of various driving factors on urban O₃ concentration, which could provide a basis for the prevention and control of urban atmospheric O₃ pollution in summer.

Key words: O₃ pollution; driving factors; contribution; explanation; machine learning

有研究表明, 受前体物排放^[1~3]和气象条件^[4~7]等因素共同驱动, 大气 O₃ 已成为影响城市夏季环境空气质量的主要污染物^[8,9]。目前对于 O₃ 污染驱动因素的研究, 主要应用物理化学机制驱动的演绎模型^[10,11]和数据驱动的归纳模型^[12,13]。其中 CMAQ 和 WRF-Chem 等应用相对较多。张迎春等^[14]采用 CMAQ 模型对成都市夏季 O₃ 进行模拟评估, 并采用情景分析法得出夏季 O₃ 对气象因子的

灵敏度依次为: 气温、绝对湿度和大气压强。Hu 等^[15]利用 WRF-Chem 模拟四川盆地 O₃ 时空变化特征和传输路径。但这些模型需要输入网格化排放清

收稿日期: 2022-08-23; 修订日期: 2022-10-11

基金项目: 大气重污染成因与治理攻关项目(DQGG202109); 国家重点研发计划项目(2019YFC0214203, 2019YFC0214202); 国家自然科学基金项目(21976053)

作者简介: 董佳奇(1997~), 男, 硕士研究生, 主要研究方向为环境数据分析挖掘, E-mail: dongjiaqi972022@163.com

* 通信作者, E-mail: huhu3057@163.com

单和气象场等众多模型参数,运算时效性较差^[16]. 机器学习 ML 模型善于处理非线性问题,运算效率高^[17]. 蔡旺华^[18]和马润美等^[19]分别采用 XGBoost 和 Random Forest 模型预测 O₃ 浓度,精确度较高. 但这些模型的可解释性差,都较难定量解析各因素对 O₃ 的影响^[20]. 近年来,基于可解释性数据驱动模型,学者们将机器学习 ML 模块耦合可解释性 SHAP 模块,为模型的合理性验证和模拟数据的定量解析提供接口^[21,22]. ML-SHAP 模型应用广泛,如 Xu 等^[23]采用 XGBoost-SHAP 解释驾驶行为和机动车排放水平的关系,廖彬等^[24]采用 XGBoost-SHAP 模型解释决定运动员身价的主要因素. XGBoost-SHAP 模型在定量解析影响因素贡献方面已有一定应用,但该类模型缺乏对各驱动因素和目标变量之间的关联性筛选和优化能力,导致部分影响因素可解释性较差.

本研究建立可解释性数据驱动模型 Correlation-ML-SHAP 模型,Correlation 模块挖掘 O₃ 浓度的关联影响因素,优化数据组合,提高模型可解释性. 机器学习 ML 模块耦合可解释性 SHAP 模块计算各

驱动因素对 O₃ 浓度的贡献,并以晋城市 2021 年夏季 O₃ 污染为例开展应用研究,旨在为夏季城市大气 O₃ 污染防控提供基础依据.

1 材料与方法

1.1 改进的可解释性机器学习研究框架

Correlation-ML-SHAP 研究方法耦合 Correlation 模块、ML 模块和可解释性 SHAP 模块,通过挖掘 O₃ 浓度关联影响因素,实现对模型结果可解释性的优化. 以上 3 个模块分别用于时间关联性挖掘、O₃ 浓度模拟和各驱动因素影响评估. 应用 Correlation 模块遍历不同时段驱动因素和模拟时段 O₃ 浓度的 Pearson 相关系数,搜索 O₃ 浓度关联影响因素优化数据组合. 应用机器学习 ML 模块训练候选模型,结合模型评价指标和可解释性 SHAP 特征图分别评估候选模型的精准度和可解释性,确定最佳模型. 将优化数据组合和最佳模型输入可解释性 SHAP 模块计算任意时刻各驱动因素对 O₃ 浓度的影响贡献. 应用于定量解析 O₃ 浓度驱动因素和交互影响. 技术路线如图 1 所示.

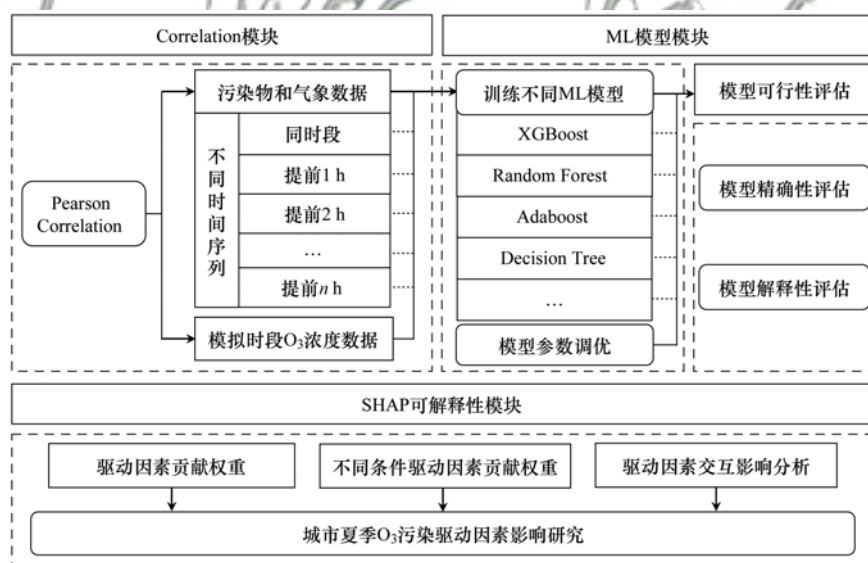


图 1 技术路线

Fig. 1 Technology roadmap

1.2 机器学习 ML 模块

1.2.1 模型训练

对于数据驱动的机器学习 ML 模型,设 O₃ 污染数据集为 $\{X\} = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{is}, y_i\}$. 对于 i 时刻的 O₃ 浓度可由 s 个驱动因素输入候选机器学习 ML 模型计算得到^[25]. 如公式(1):

$$O_{3(i)} = f(x, \alpha) \quad (1)$$

式中, $O_{3(i)}$ 为 i 时刻 O₃ 浓度的模拟值 ($\mu\text{g} \cdot \text{m}^{-3}$); α 为候选 ML 模型的优化参数配置; f 为候选的机器学习 ML 模型.

各驱动因素对夏季城市 O₃ 浓度的影响呈现非线性变化特征^[26,27]. 处理此类数据特征的主流 ML 模型包括 XGBoost、Random Forest、Adaboost、Decision Tree 和 KNN^[28~32]. 采用网格搜索验证法优化以上候选模型的参数配置,如表 1 所示. 以 XGBoost 模型为例, $n_estimator$ 为弱学习器数量,通过生成新的学习器拟合上一步结果的残差以逐步提高模拟精度; max_depth 、 reg_alpha 和 reg_lambda 分别为树的最大深度以及 L1 和 L2 正则化项,这些参数主要用于避免模型发生过拟合.

表 1 模型主要参数配置

Table 1 Configuration of the main parameters of the models

模型名称	参数配置
XGBoost	n_estimators = 210, max_depth = 7, min_child_weight = 6, subsample = 1.0, colsample_bytree = 1.0, gamma = 0.6, reg_alpha = 0.05, reg_lambda = 3, learning_rate = 0.15
Random Forest	n_estimators = 210, max_depth = 22, max_features = 5, min_samples_leaf = 2
Adaboost	n_estimators = 177, learning_rate = 1
Decision Tree	criterion = 'mse', min_samples_split = 2
KNN	n_neighbors = 10, p = 1, weights = distance

1.2.2 模型评估

为定量评价模型模拟结果的准确度,分别采用平均值绝对误差(MAE)、均方根误差(RMSE)和决定系数(R^2)对候选 ML 模型进行定量评价。如公式(2)~(4):

$$MAE = \frac{1}{N} \sum_{i=1}^N |O_{3(i \text{ true})} - O_{3(i)}| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_{3(i \text{ true})} - O_{3(i)})^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (O_{3(i \text{ true})} - O_{3(i)})^2}{\sum_{i=1}^N (O_{3(i \text{ true})} - \bar{O}_{3(i)})^2} \quad (4)$$

式中, $O_{3(i \text{ true})}$ 为*i*时刻 O_3 浓度的真实值($\mu\text{g}\cdot\text{m}^{-3}$); $O_{3(i)}$ 为*i*时刻 O_3 浓度的模拟值($\mu\text{g}\cdot\text{m}^{-3}$)。MAE 和 RMSE 用于评价模拟值和真实值的差异,两种指标越低越准确; R^2 用于评价回归性能,取值越大效果越好。表 2 为候选的不同机器学习 ML 模型模拟结果的评估指标。

表 2 模型评价指标结果

Table 2 Model evaluation index results

模型名称	MAE	RMSE	R^2
XGBoost	15.4	21.7	0.82
Random Forest	16.0	21.9	0.81
Adaboost	22.2	28.2	0.69
Decision Tree	21.8	30.4	0.64
KNN	22.2	30.1	0.65

KNN、Decision Tree 和 Adaboost 模型模拟结果的准确度较低,主要原因在于 KNN 对多维数据的回归和可解释性较差;Decision Tree 和 Adaboost 模型易受异常数据干扰,稳定性较弱。XGBoost 和 Random Forest 模型模拟结果的准确度相对较高,但由于 Random Forest 模型 max_depth 参数值较大,导致模型复杂度高,容易发生过拟合。采用 XGBoost 模型时 MAE、RMSE 和 R^2 这 3 项指标分别为 15.4、21.7 和 0.82,相比其它 ML 模型模拟结果的准确度最高,原因在于该模型通过引入损失函数的二阶泰勒展开和正则项,优化模型复杂度,避免过拟

合^[33,34]。

1.3 可解释性 SHAP 模块

基于合作博弈中 SHAP 理论的归因分析方法,计算机器学习 ML 模型中各驱动因素对 $O_{3(i)}$ 的边际贡献值,结合 SHAP 可加性解释模型量化不同时刻各驱动因素的贡献大小^[35],计算方法如公式(5)~(7):

$$O_{3(i)} = O_{3(\text{base})} + \sum_{j=1}^S \text{shap}(x_{i,j}) \quad (5)$$

$$\text{shap}(x_{i,j}) = \sum_{M \ni j} \beta_M (O_{3(M \cup \{x_{i,j}\})} - O_{3(M)}) \quad (6)$$

$$\beta_M = \frac{|M|! (S - |M| - 1)!}{S!} \quad (7)$$

式中, $O_{3(\text{base})}$ 为 $O_{3(i)}$ 的期望值($\mu\text{g}\cdot\text{m}^{-3}$); $\text{shap}(x_{i,j})$ 为*i*时刻因素*j*对 $O_{3(i)}$ 的贡献值($\mu\text{g}\cdot\text{m}^{-3}$),当 $\text{shap}(x_{i,j}) > 0$ 时,对 $O_{3(i)}$ 呈正贡献; $\text{shap}(x_{i,j}) < 0$ 时,对 $O_{3(i)}$ 呈负贡献; $O_{3(M)}$ 为 ML 模型在子集 *M* 条件下的模拟值($\mu\text{g}\cdot\text{m}^{-3}$);*M* 为不包括因素*j*的特征子集; β_M 为特征子集 *M* 的权重。

环境中不同驱动因素对 $O_{3(i)}$ 的贡献权重不同,定量解析因素*j*对 $O_{3(i)}$ 的贡献权重,可以通过计算因素*j*的 SHAP 特征重要性与全部因素重要性的比值得到^[36]。计算方法如公式(8)和公式(9):

$$I(x_j) = \frac{\alpha(x_j)}{\alpha(x_1) + \alpha(x_2) + \cdots + \alpha(x_j) + \cdots + \alpha(x_S)} \quad (8)$$

$$\alpha(x_j) = n^{-1} \cdot \sum_{i=1}^n |\text{shap}(x_{i,j})| \quad (9)$$

式中, $I(x_j)$ 为因素*j*对 $O_{3(i)}$ 的贡献权重(%); $\alpha(x_j)$ 为因素*j*的 SHAP 特征重要性,以因素*j*的 $|\text{shap}(x_{i,j})|$ 均值计($\mu\text{g}\cdot\text{m}^{-3}$), $\alpha(x_j)$ 取值越大,因素*j*对 $O_{3(i)}$ 的贡献权重越大。

环境中驱动因素通常在其他因素交互影响下共同驱动 $O_{3(i)}$ 变化,分析 $O_{3(i)}$ 驱动因素的交互影响^[37]。计算方法如公式(10)~(12):

$$\text{shap}(x_i, x_j) = \sum_{M^*} \beta^* \delta_{i,j}(M^*) \quad (10)$$

$$\beta^* = \frac{|M^*|! (S - |M^*| - 2)!}{2(S-1)!} \quad (11)$$

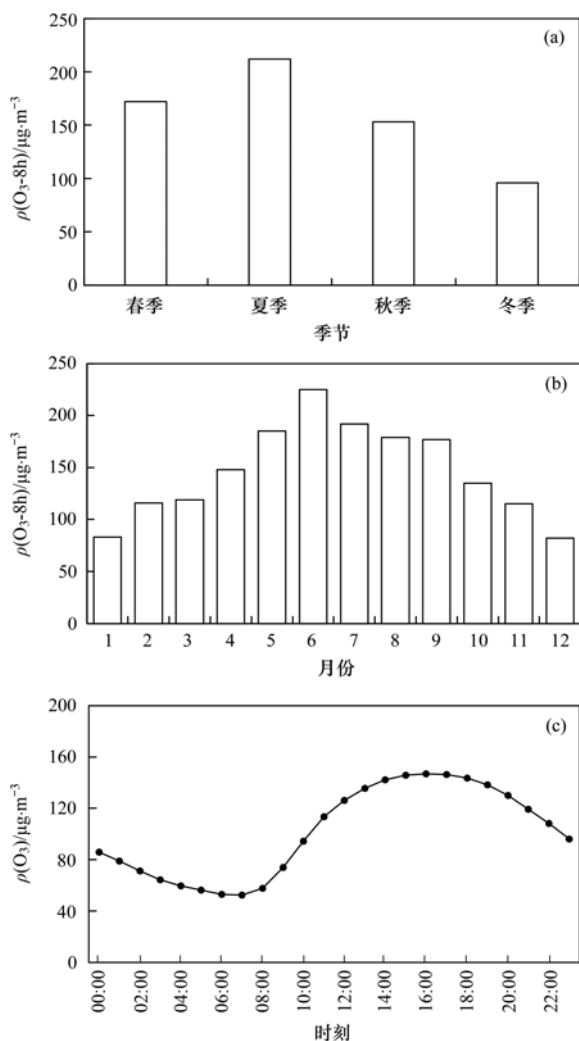
$$\delta_{i,j}(M^*) = (O_{3(M^* \cup \{x_i, x_j\})} - O_{3(M^* \cup \{x_j\})}) - (O_{3(M^* \cup \{x_i\})} - O_{3(M^*)}) \quad (12)$$

式中, $\text{shap}(x_i, x_j)$ 为因素 i 和 j 对 $O_{3(i)}$ 的交互影响值 ($\mu\text{g}\cdot\text{m}^{-3}$), $O_{3(M^*)}$ 为 ML 模型在子集 M^* 条件下的模拟值 ($\mu\text{g}\cdot\text{m}^{-3}$); M^* 为不包括因素 i 和 j 的特征子集; β^* 为特征子集 M^* 的权重。

2 结果与讨论

2.1 晋城市 O₃ 时间变化特征

晋城市作为“2+26”城市之一, 位于山西省东南部, 全境被山脊环绕形成盆地地形。如图 2 所示, 2021 年臭氧日最大 8 h 第 90 百分位数 [$\rho(O_3-8h)$] 为 $180 \mu\text{g}\cdot\text{m}^{-3}$, 位居“2+26”城市倒数。 $\rho(O_3-8h)$ 呈“春夏高、秋冬低”分布, 即: 夏季($212 \mu\text{g}\cdot\text{m}^{-3}$)



(a) 2021 年各季节臭氧日最大 8 h 第 90 百分位数; (b) 2021 年各月臭氧日最大 8 h 第 90 百分位数; (c) 2021 年 5 ~ 9 月每小时 O₃ 浓度均值

图 2 晋城市 2021 年 O₃ 浓度时间变化特征

Fig. 2 Temporal variation characteristics of O₃ concentration in Jincheng City in 2021

> 春季($172 \mu\text{g}\cdot\text{m}^{-3}$) > 秋季($153 \mu\text{g}\cdot\text{m}^{-3}$) > 冬季($96 \mu\text{g}\cdot\text{m}^{-3}$)。其中 5 ~ 9 月是污染高发期, $\rho(O_3-8h)$ 分别为 185、225、192、179 和 $177 \mu\text{g}\cdot\text{m}^{-3}$ 。2021 年 5 ~ 9 月每小时 O₃ 浓度均值变化特征呈单峰分布, 07:00 开始浓度持续升高, 12:00 ~ 18:00 处于高值区。主要原因可能是夏季昼间有利的气象条件促进光化学反应向 O₃ 生成方向进行^[38~42]。

以晋城市 2021 年夏季 O₃ 污染过程为例, 应用建立的 Correlation-ML-SHAP 模型定量解析夏季 O₃ 污染驱动因素影响贡献。

2.2 影响 O₃ 的关联驱动因素挖掘

选取晋城市 2021 年夏季 5 ~ 9 月昼间 07:00 ~ 18:00 作为模拟时段。收集 O₃、NO₂、CO、PM₁₀ 和 PM_{2.5} 浓度的小时数据以及气温、湿度、日照强度、风向、风速、大气压和能见度小时数据作为 O₃ 污染数据集。小时浓度数据来源于大气环境科学综合数据采集与共享平台 (<http://106.38.83.12:9002/ahpcp/login/index.html>); 小时气象数据来源于国家气象科学数据中心 (<http://data.cma.cn/>)。数据集中存在的部分异常、缺失值, 采用基于时间序列的拉格朗日插值法填充处理。

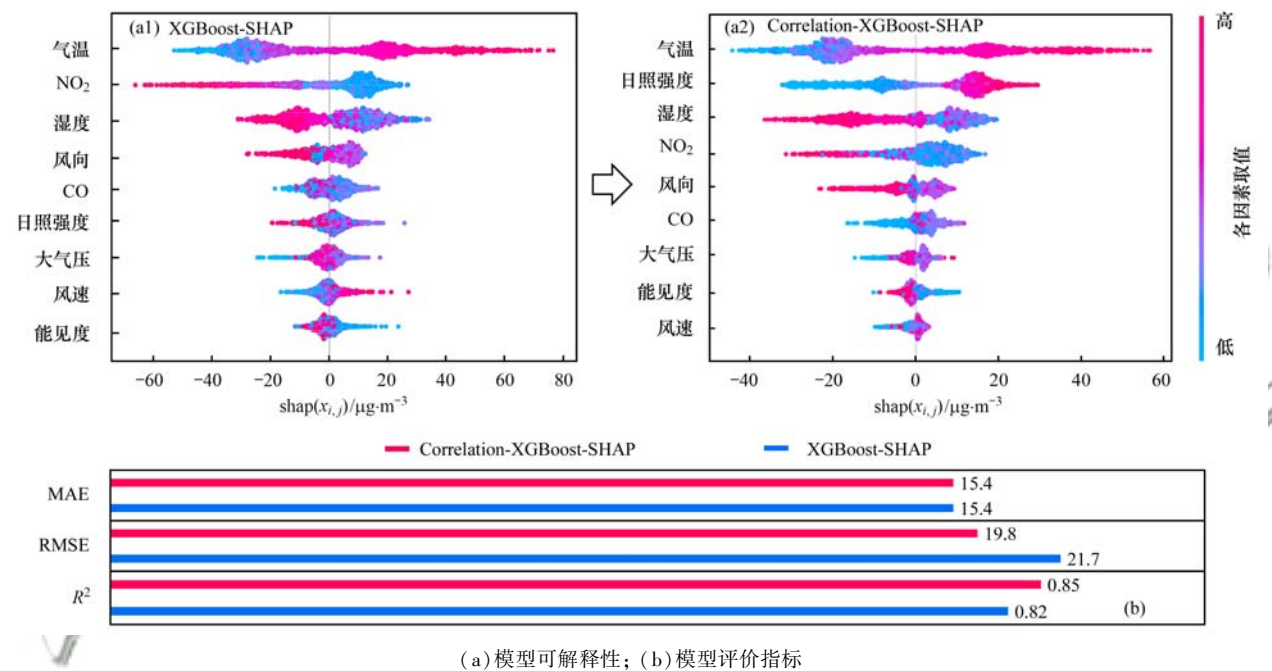
如表 3 所示, 应用 Correlation-XGBoost-SHAP 模型的 Correlation 模块, 遍历不同时段各驱动因素与模拟时段 O₃ 浓度之间的相关性系数, 得到关联因素为: 提前 2 h 的气温、日照强度、湿度以及同时刻的 NO₂、CO、大气压、风速和能见度, Pearson 相关系数分别为 0.777、0.770、-0.611、-0.548、-0.387、-0.346、0.326 和 0.331, 在 0.01 水平上显著相关。数据特征方面, 气温、日照强度、湿度的相关性在提前 2 h 达到最大, 表明 O₃ 浓度变化滞后于上述特征。姜峰等^[43] 在研究日照强度对 O₃ 浓度影响时也发现了这种关系。可知应用数据驱动的方法对 O₃ 浓度进行模拟时, 模型可能更关注提前 2 h 的气温、日照强度和湿度变化。

如图 3 所示, 由于 XGBoost-SHAP 模型缺乏对各驱动因素和 O₃ 浓度之间的时间关联性筛选和优化能力, 导致部分驱动因素的可解释性较差。建立的 Correlation-XGBoost-SHAP 模型关注数据之间的时间关联性, 通过 Correlation 模块挖掘 O₃ 浓度的关联驱动因素, 优化数据组合, 使模型模拟结果的合理性得以验证和改善。从模拟结果的可解释性来看, 该模型优化了影响 O₃ 浓度变化的驱动因素贡献特征, 如该模型的日照强度贡献排名提升至第 2, 对 O₃ 污染呈正贡献。从模拟结果的准确度来看, 该模型的 MAE、RMSE 和 R^2 分别为 15.4、19.8 和 0.85, 在 XGBoost-SHAP 模型的基

表 3 驱动因素与 O₃ 浓度的相关性系数¹⁾

	同时段	提前 1 h	提前 2 h	提前 3 h	提前 4 h	提前 5 h
气温	0.694 **	0.762 **	0.777 **	0.751 **	0.694 **	0.614 **
日照强度	0.539 **	0.711 **	0.770 **	0.757 **	0.707 **	0.642 **
湿度	-0.514 **	-0.583 **	-0.611 **	-0.600 **	-0.558 **	-0.494 **
风速	0.326 **	0.319 **	0.287 **	0.231 **	0.162 **	0.109 **
大气压	-0.346 **	-0.323 **	-0.286 **	-0.245 **	-0.211 **	-0.188 **
能见度	0.331 **	0.344 **	0.339 **	0.324 **	0.291 **	0.235 **
NO ₂	-0.548 **	-0.475 **	-0.372 **	-0.273 **	-0.177 **	-0.058
CO	-0.387 **	-0.278 **	-0.141 **	-0.012	0.096 **	-0.090 **
PM ₁₀	-0.020	-0.012	0.015	0.050	0.094 **	-0.024
PM _{2.5}	-0.017	0.028	0.083 **	0.141 **	0.195 **	0.023

1) ** 表示在 0.01 水平上显著相关



(a) 模型可解释性; (b) 模型评价指标

图 3 优化数据组合的模型性能评估

Fig. 3 Model performance evaluation of optimized data composition

础上准确度有所提高. 由此可知 Correlation-ML-SHAP 模型在解析影响 O₃ 污染的各驱动因素方面具备可行性, 其中 ML 模块采用 XGBoost 模型模拟准确度最佳.

2.3 夏季 O₃ 污染驱动因素影响研究

2.3.1 驱动因素的贡献权重

应用 Correlation-XGBoost-SHAP 模型解析晋城市夏季 O₃ 污染因素. 如图 4 所示, 通过可解释性

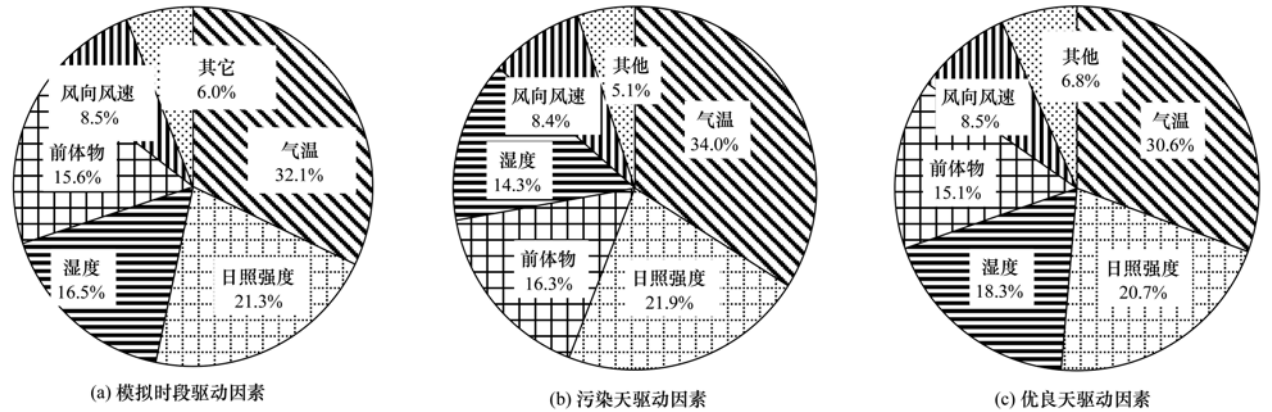


图 4 晋城市夏季 O₃ 浓度的驱动因素贡献权重

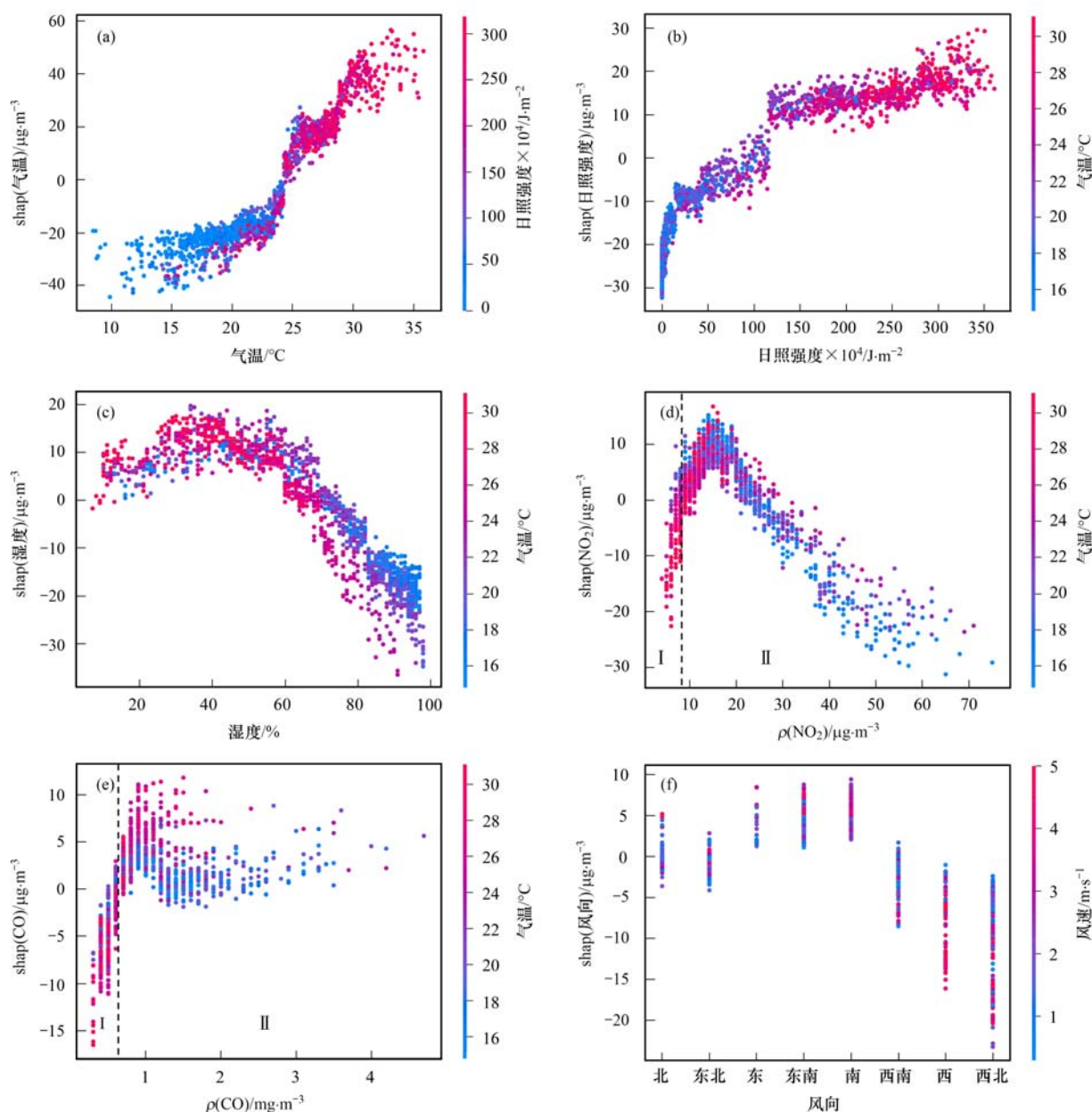
Fig. 4 Contribution weights of driving factors for O₃ concentration in Jincheng City in summer

SHAP 模块公式(8)和公式(9)计算各驱动因素的贡献权重,用 NO_2 和 CO 对 O_3 的贡献权重和代表前体物排放水平的贡献权重. 影响 O_3 浓度变化的各驱动因素贡献权重依次为: 32.1% (气温)、21.3% (日照强度)、16.5% (湿度)、15.6% (前体物的排放水平)、8.5% (风向风速) 和 6.0% (其它). 气温、日照强度、湿度以及前体物对 O_3 污染的贡献权重均高于传输扩散相关的风向风速,表明晋城市夏季 O_3 浓度变化主要受本地光化学反应生成影响. 其中前体物的排放贡献权重在污染天排名提升至第 3, 其它驱动因素的贡献权重排名不变, 但贡献程度不同. 气温、日照强度和前体物排放水平的贡献

权重在污染天分别为 34.0%、21.9% 和 16.3%, 高于优良天 3.4%、1.2% 和 1.2%; 湿度的贡献权重在污染天为 14.3%, 低于优良天 4.0%. 可知夏季城市 O_3 浓度受气温和日照强度影响较大, 污染天前体物排放对 O_3 浓度的影响也不容忽视.

2.3.2 驱动因素交互影响分析

通过 Correlation-XGBoost-SHAP 模型可解释性 SHAP 模块的交互值计算, 进一步探究各驱动因素对 O_3 浓度变化的交互影响. 如图 5 所示, 气象条件方面, 气温、日照强度和 O_3 浓度呈非线性正相关, 随着日照强度的增强, 气温对 O_3 污染的正贡献提升, 当气温超过 $24^{\circ}C$ 时在较强日照的交互影响下有



散点为横坐标上因素 j 在其他因素共同驱动下对 O_3 的贡献值; 色柱为产生交互影响的因素取值; (d) 和 (e) 用虚线将反应条件加以划分: I 阶段为午后的光化学反应, II 阶段为日出后的光化学反应

图 5 晋城市夏季 O_3 浓度的驱动因素交互影响

Fig. 5 Interaction of driving factors on O_3 concentration in Jincheng City in summer

94.9%的概率对 O_3 污染呈正贡献;湿度低于70%时有94.1%的概率对 O_3 污染呈正贡献,但超过70%后在低温的交互影响下对 O_3 污染呈负贡献.传输扩散方面,东南风风速低于 $5.8 \text{ m}\cdot\text{s}^{-1}$ 或南风风速低于 $5.3 \text{ m}\cdot\text{s}^{-1}$ 时,风速增加对 O_3 污染的正贡献提升.由此可知,晋城市夏季 O_3 污染受多方面的气象因素交互影响,强日照条件下高温和低湿是引发 O_3 污染的原因,同时还受到东南部和南部区域的传输影响.与周边城市临汾^[44]、郑州^[45]的 O_3 污染条件基本一致.前体物方面, NO_2 、 CO 和 O_3 浓度的变化相近.高温($\geq 24^\circ\text{C}$)且低湿($\leq 70\%$)的气象条件下, $\rho(\text{NO}_2) \geq 9 \mu\text{g}\cdot\text{m}^{-3}$ 或 $\rho(\text{CO}) \geq 0.7 \text{ mg}\cdot\text{m}^{-3}$, 分别有94.9%和99.3%的概率对 O_3 污染呈正贡献.

3 结论

(1) 基于可解释性数据驱动思想,以城市 O_3 污染过程为研究对象,建立 Correlation-ML-SHAP 模型,模型能够挖掘并利用强驱动因素模拟 O_3 浓度和量化影响贡献,其中机器学习 ML 模块采用 XGBoost 模型模拟准确度最佳,模型可为夏季城市大气 O_3 污染防控提供基础依据.

(2) 晋城市2021年夏季5~9月 $\rho(O_3-8\text{h})$ 分别为185、225、192、179和177 $\mu\text{g}\cdot\text{m}^{-3}$.影响夏季 O_3 浓度的强驱动因素依次为气温、日照强度、湿度和前体物排放水平,贡献权重分别为32.1%、21.3%、16.5%和15.6%.

(3) 气温、日照强度、前体物排放水平在污染天对 O_3 浓度的贡献权重分别为34.0%、21.9%和16.3%,相比优良天分别提升3.4%、1.2%和1.2%,前体物排放贡献权重排名提升至第3.

(4) 各驱动因素对 O_3 浓度呈非线性交互影响,气温超 24°C 或湿度低于70%时分别有94.9%和94.1%的概率对 O_3 污染呈正贡献,这种气象条件下 $\rho(\text{NO}_2) \geq 9 \mu\text{g}\cdot\text{m}^{-3}$ 或 $\rho(\text{CO}) \geq 0.7 \text{ mg}\cdot\text{m}^{-3}$, 分别有94.9%和99.3%的概率对 O_3 污染呈正贡献.东南风风速低于 $5.8 \text{ m}\cdot\text{s}^{-1}$ 或南风风速低于 $5.3 \text{ m}\cdot\text{s}^{-1}$, 均对 O_3 污染呈正贡献.

参考文献:

- [1] 刘建, 吴兑, 范绍佳, 等. 前体物与气象因子对珠江三角洲臭氧污染的影响[J]. 中国环境科学, 2017, **37**(3): 813-820.
- [2] Liu J, Wu D, Fan S J, et al. Impacts of precursors and meteorological factors on ozone pollution in Pearl River Delta[J]. China Environmental Science, 2017, **37**(3): 813-820.
- [3] 潘本锋, 程麟钧, 王建国, 等. 京津冀地区臭氧污染特征与来源分析[J]. 中国环境监测, 2016, **32**(5): 17-23.
- [4] Pan B F, Cheng L J, Wang J G, et al. Characteristics and source attribution of ozone pollution in Beijing-Tianjin-Hebei Region[J]. Environmental Monitoring in China, 2016, **32**(5): 17-23.
- [5] 王玫, 郑有飞, 柳艳菊, 等. 京津冀臭氧变化特征及与气象要素的关系[J]. 中国环境科学, 2019, **39**(7): 2689-2698.
- [6] Wang M, Zheng Y F, Liu Y J, et al. Characteristics of ozone and its relationship with meteorological factors in Beijing-Tianjin-Hebei Region[J]. China Environmental Science, 2019, **39**(7): 2689-2698.
- [7] 曹庭伟, 吴锴, 康平, 等. 成渝城市群臭氧污染特征及影响因素分析[J]. 环境科学学报, 2018, **38**(4): 1275-1284.
- [8] Cao T W, Wu K, Kang P, et al. Study on ozone pollution characteristics and meteorological cause of Chengdu-Chongqing urban agglomeration[J]. Acta Scientiae Circumstantiae, 2018, **38**(4): 1275-1284.
- [9] 赵伟, 高博, 刘明, 等. 气象因素对香港地区臭氧污染的影响[J]. 环境科学, 2019, **40**(1): 55-66.
- [10] Zhao W, Gao B, Liu M, et al. Impact of meteorological factors on the ozone pollution in Hong Kong [J]. Environmental Science, 2019, **40**(1): 55-66.
- [11] 余钟奇, 马井会, 毛卓成, 等. 2017年上海臭氧污染气象条件分析及臭氧污染天气分型研究[J]. 气象与环境学报, 2019, **35**(6): 46-54.
- [12] Yu Z Q, Ma J H, Mao Z C, et al. Study on the meteorological conditions and synoptic classifications of O_3 pollution in Shanghai in 2017[J]. Journal of Meteorology and Environment, 2019, **35**(6): 46-54.
- [13] 曾贤刚, 阮芳芳, 姜艺婧. 中国臭氧污染的空间分布和健康效应[J]. 中国环境科学, 2019, **39**(9): 4025-4032.
- [14] Zeng X G, Ruan F F, Jiang Y J. Spatial distribution and health effects of ozone pollution in China [J]. China Environmental Science, 2019, **39**(9): 4025-4032.
- [15] 王纯璐, 王毅一, 史之浩, 等. 基于多源融合数据评估2014—2018年中国地表大气臭氧污染变化及其健康影响[J]. 大气科学学报, 2021, **44**(5): 737-745.
- [16] Wang C L, Wang Y Y, Shi Z H, et al. Assessment of surface ozone pollution change and its health effect in China from 2014 to 2018 based on multi-source fusion data [J]. Transactions of Atmospheric Sciences, 2021, **44**(5): 737-745.
- [17] 马强, 陈天然, 陈翠红, 等. 嘉兴市大气污染特征及区域传输影响研究[J]. 环境污染与防治, 2020, **42**(4): 467-471, 476.
- [18] Ma Q, Chen T R, Chen C H, et al. Study on the atmosphere pollution characteristics and regional transport in Jiaxing [J]. Environmental Pollution and Control, 2020, **42**(4): 467-471, 476.
- [19] 何涛, 乔利平, 徐圃青, 等. 常州市臭氧污染传输路径和潜在源区[J]. 中国环境监测, 2017, **33**(4): 77-83.
- [20] He T, Qiao L P, Xu P Q, et al. Study of transport pathways and potential sources of ozone pollution in Changzhou [J]. Environmental Monitoring in China, 2017, **33**(4): 77-83.
- [21] Vu T V, Shi Z B, Cheng J, et al. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique[J]. Atmospheric Chemistry and Physics, 2019, **19**(17): 11303-11314.
- [22] Xing J, Zheng S X, Ding D, et al. Deep learning for prediction of the air quality response to emission changes [J].

- Environmental Science & Technology, 2020, **54** (14): 8589-8600.
- [14] 张迎春, 付虹, 李迪, 等. 基于 CMAQ 模型分析成都市 O₃ 对气象因子的灵敏度[J]. 中国资源综合利用, 2021, **39** (9): 31-35.
- Zhang Y C, Fu H, Li D, *et al.* Sensitivity analysis of O₃ to meteorological factors in Chengdu City based on CMAQ model [J]. China Resources Comprehensive Utilization, 2021, **39** (9): 31-35.
- [15] Hu J, Zhao T L, Liu J, *et al.* Exploring the ozone pollution over the western Sichuan Basin, Southwest China: the impact of diurnal change in mountain-plains solenoid [J]. Science of the Total Environment, 2022, **839**, doi: 10.1016/j.scitotenv.2022.156264.
- [16] 胡冬梅. 基于时空数据挖掘的城市空气污染分析与预测[D]. 北京: 清华大学, 2018.
- Hu D M. Research on the analysis and prediction of urban air pollution through spatio-temporal data mining [D]. Beijing: Tsinghua University, 2018.
- [17] Reyes J M, Xu Y D, Vizuete W, *et al.* Regionalized PM_{2.5} community multiscale air quality model performance evaluation across a continuous spatiotemporal domain [J]. Atmospheric Environment, 2017, **148**: 258-265.
- [18] 蔡旺华. 运用机器学习方法预测空气中臭氧浓度[J]. 中国环境管理, 2018, **10** (2): 78-84.
- Cai W H. Using machine learning method for predicting the concentration of ozone in the air [J]. Chinese Journal of Environmental Management, 2018, **10** (2): 78-84.
- [19] 马润美, 张亚一, 班婕, 等. 基于随机森林模型的京津冀地区近地面臭氧站点浓度预测[J]. 环境与健康杂志, 2019, **36** (11): 954-957.
- Ma R M, Zhang Y Y, Ban J, *et al.* Prediction of ground ambient ozone site concentration in Beijing-Tianjin-Hebei Region based on random forest model [J]. Journal of Environment and Health, 2019, **36** (11): 954-957.
- [20] 纪守领, 李进锋, 杜天宇, 等. 机器学习模型可解释性方法、应用与安全研究综述[J]. 计算机研究与发展, 2019, **56** (10): 2071-2096.
- Ji S L, Li J F, Du T Y, *et al.* Survey on techniques, applications and security of machine learning interpretability [J]. Journal of Computer Research and Development, 2019, **56** (10): 2071-2096.
- [21] Lundberg S M, Erion G, Chen H, *et al.* From local explanations to global understanding with explainable AI for trees [J]. Nature Machine Intelligence, 2020, **2** (1): 56-67.
- [22] Ribeiro M T, Singh S, Guestrin C, *et al.* "Why should I trust you?": explaining the predictions of any classifier [A]. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. San Francisco: ACM, 2016. 1135-1144.
- [23] Xu J S, Saleh M, Hatzopoulou M. A machine learning approach capturing the effects of driving behaviour and driver characteristics on trip-level emissions [J]. Atmospheric Environment, 2020, **224**, doi: 10.1016/j.atmosenv.2020.117311.
- [24] 廖彬, 王志宁, 李敏, 等. 融合 XGBoost 与 SHAP 模型的足球运动员身价预测及特征分析方法[J]. 计算机科学, 2022, **49** (2): 195-204.
- Liao B, Wang Z N, Li M, *et al.* Integrating XGBoost and SHAP model for football player value prediction and characteristic analysis [J]. Computer Science, 2022, **49** (2): 195-204.
- [25] 卢亚灵, 李勃, 范朝阳, 等. 空气质量预测模拟技术演变与发展研究[J]. 中国环境管理, 2021, **13** (4): 84-92.
- Lu Y L, Li B, Fan Z Y, *et al.* Evolution and development of air quality prediction and simulation technology [J]. Chinese Journal of Environmental Management, 2021, **13** (4): 84-92.
- [26] 易睿, 王亚林, 张殷俊, 等. 长江三角洲地区城市臭氧污染特征与影响因素分析[J]. 环境科学学报, 2015, **35** (8): 2370-2377.
- Yi R, Wang Y L, Zhang Y J, *et al.* Pollution characteristics and influence factors of ozone in Yangtze River Delta [J]. Acta Scientiae Circumstantiae, 2015, **35** (8): 2370-2377.
- [27] Diao L L, Bi X H, Zhang W H, *et al.* The characteristics of heavy ozone pollution episodes and identification of the primary driving factors using a generalized additive model (GAM) in an industrial megacity of northern China [J]. Atmosphere, 2021, **12**, doi: 10.3390/atmos12111517.
- [28] 杨正理, 史文, 陈海霞, 等. 大数据背景下采用互信息与随机森林算法的空气质量预测[J]. 环境工程, 2019, **37** (3): 180-185.
- Yang Z L, Shi W, Chen H X, *et al.* Air quality forecasting with mutual information and random forests based on big [J]. Environmental Engineering, 2019, **37** (3): 180-185.
- [29] 徐旭冉, 涂娟娟. 基于决策树算法的空气质量预测系统[J]. 电子设计工程, 2019, **27** (9): 39-42.
- Xu X R, Tu J J. Air quality forecast system based on decision tree algorithm [J]. Electronic Design Engineering, 2019, **27** (9): 39-42.
- [30] 莫欣岳. 基于人工智能的空气质量决策支持系统建模研究[D]. 兰州: 兰州大学, 2021.
- Mo X Y. Research on modeling of air quality decision support system based on artificial intelligence [D]. Lanzhou: Lanzhou University, 2021.
- [31] Zhang Z C, Xu B, Xu W M, *et al.* Machine learning combined with the PMF model reveal the synergistic effects of sources and meteorological factors on PM_{2.5} pollution [J]. Environmental Research, 2022, **212**, doi: 10.1016/j.envres.2022.113322.
- [32] 孙苏琪, 王式功, 罗彬, 等. 应用机器学习算法的成都市冬季空气污染预报研究[J]. 气象与环境学报, 2020, **36** (2): 98-104.
- Sun S Q, Wang S G, Luo B, *et al.* Air pollution forecast in winter based on machine learning method in Chengdu [J]. Journal of Meteorology and Environment, 2020, **36** (2): 98-104.
- [33] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [A]. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. San Francisco: ACM, 2016. 785-794.
- [34] 康俊锋, 谭建林, 方雷, 等. XGBoost-LSTM 变权组合模型支持下短期 PM_{2.5} 浓度预测——以上海为例[J]. 中国环境科学, 2021, **41** (9): 4016-4025.
- Kang J F, Tan J L, Fang L, *et al.* Short-term PM_{2.5} concentration prediction based on XGBoost and LSTM variable weight combination model: a case study of Shanghai [J]. China Environmental Science, 2021, **41** (9): 4016-4025.
- [35] Lundberg S M, Lee S L. A unified approach to interpreting model predictions [A]. In: Proceedings of the 31st International Conference on Neural Information Processing Systems [C]. Long Beach: Curran Associates Inc., 2017. 4768-4777.
- [36] 许博, 徐晗, 赵焕, 等. 机器学习耦合受体模型揭示驱动因素对 PM_{2.5} 的影响[J]. 环境科学研究, 2022, **35** (11): 2425-2434.
- Xu B, Xu H, Zhao H, *et al.* Machine learning coupled with

- receptor model to reveal the effect of driving factors on $PM_{2.5}$ [J]. *Research of Environmental Sciences*, 2022, **35** (11): 2425-2434.
- [37] Hou L L, Dai Q L, Song C B, *et al.* Revealing drivers of haze pollution by explainable machine learning [J]. *Environmental Science & Technology Letters*, 2022, **9**(2): 112-119.
- [38] 张莹, 倪长健, 冯鑫媛, 等. 基于 GAMs 模型分析成都市气象因子交互作用对 O_3 浓度变化的影响 [J]. *环境科学*, 2021, **42**(11): 5228-5238.
- Zhang Y, Ni C J, Feng X Y, *et al.* Interactive effects of the influencing factors on the changes of O_3 concentrations based on GAMs model in Chengdu [J]. *Environmental Science*, 2021, **42** (11): 5228-5238.
- [39] 黄小刚, 邵天杰, 赵景波, 等. 基于 GAM 模型的西安市 O_3 浓度影响因素解析 [J]. *环境科学*, 2020, **41** (4): 1535-1543.
- Huang X G, Shao T J, Zhao J B, *et al.* Influencing factors of ozone concentration in Xi'an based on generalized additive models [J]. *Environmental Science*, 2020, **41** (4): 1535-1543.
- [40] 花丛, 江琪, 迟茜元, 等. 我国中东部地区 2015—2020 年夏半年 $PM_{2.5}$ 和臭氧复合污染气象特征分析 [J]. *环境科学研究*, 2022, **35**(3): 650-658.
- Hua C, Jiang Q, Chi X Y, *et al.* Meteorological characteristics of air combined pollution in central and Eastern China in the summer half years of 2015-2020 [J]. *Research of Environmental Sciences*, 2022, **35**(3): 650-658.
- [41] 薛莲, 徐少才, 孙萌, 等. 气象要素及前体物对青岛市臭氧浓度变化的影响 [J]. *中国环境监测*, 2017, **33** (4): 179-185.
- Xue L, Xu S C, Sun M, *et al.* Effect of meteorological factors and precursors on the change of ambient air ozone concentrations in Qingdao [J]. *Environmental Monitoring in China*, 2017, **33** (4): 179-185.
- [42] 张晓, 张艳品, 铃伟妙. 石家庄市主城区臭氧污染特征及气象成因分析 [J]. *环境科学研究*, 2021, **34**(2): 245-253.
- Zhang X, Zhang Y P, Qian W M. Characteristics of ozone pollution and its meteorological factors in Shijiazhuang urban area [J]. *Research of Environmental Sciences*, 2021, **34**(2): 245-253.
- [43] 姜峰, 王福伟. 夏季城市臭氧浓度变化规律分析 [J]. *环境与可持续发展*, 2016, **41**(1): 62-64.
- Jiang F, Wang F W. Analysis of the changes of urban ozone concentration in summer [J]. *Environment and Sustainable Development*, 2016, **41**(1): 62-64.
- [44] 宋晓伟, 郝永佩, 朱晓东, 等. 临汾市臭氧污染变化特征、气象影响及输送源分析 [J]. *中国环境科学*, 2022, **42**(8): 3626-3634.
- Song X W, Hao Y P, Zhu X D, *et al.* Variations in ozone pollution and their meteorological influences and transmission sources in Linfen City of China [J]. *China Environmental Science*, 2022, **42**(8): 3626-3634.
- [45] 王旭东, 尹沙沙, 杨健, 等. 郑州市臭氧污染变化特征、气象影响及输送源分析 [J]. *环境科学*, 2021, **42**(2): 604-615.
- Wang X D, Yin S S, Yang J, *et al.* Characteristics, meteorological influences, and transport source of ozone pollution in Zhengzhou City [J]. *Environmental Science*, 2021, **42**(2): 604-615.

CONTENTS

Assessing the Environmental and Health Co-benefits of Accelerated Energy Transition and Industrial Restructuring: A Case Study of the BTHS Region	YANG Xi, SUN Yi-sheng, CHANG Shi-yan, <i>et al.</i> (3627)
Synergistic Paths of Reduced Pollution and Carbon Emissions Based on Different Power Demands in China	XIANG Meng-yu, WANG Shen, LÜ Lian-hong, <i>et al.</i> (3637)
Evaluation Method and Application for Urban Carbon Peaking & Neutrality Performance	ZHANG Bao-liu, BAI Zi-han, ZHANG Nan, <i>et al.</i> (3649)
Revealing Driving Factors of Urban O ₃ Based on Explainable Machine Learning	DONG Jia-qi, HU Dong-mei, YAN Yu-long, <i>et al.</i> (3660)
Sensitivity Analysis of Ozone Formation Using Response Surface Methodology	ZHU Yu-huan, CHEN Bing, ZHANG Ya-ru, <i>et al.</i> (3669)
Analysis of O ₃ Sources in Yulin City in Summer Based on WRF-CMAQ/ISAM Model	WANG Yi-fan, TONG Ji-long, CHEN Yu-xiang, <i>et al.</i> (3676)
Atmospheric Ozone Concentration Prediction in Nanjing Based on LightGBM	ZHU Jia-ying, AN Jun-lin, FENG Yue-zheng, <i>et al.</i> (3685)
Meteorological Formation Mechanisms and Potential Sources of an Ozone Pollution Process in Winter of 2022 in Guangdong Province	LI Ting-yuan, CHEN Jing-yang, GONG Yu, <i>et al.</i> (3695)
Impact of Summer Tropospheric Ozone Radiative Forcing on Meteorology and Air Quality in North China	DU Nan, CHEN Lei, LIAO Hong, <i>et al.</i> (3705)
Identification of Impacts from Meteorology and Local and Transported Photochemical Generation on Ozone Trends in Changsha from 2018 to 2020	YANG Jun, YANG Lei-feng, DING Hua, <i>et al.</i> (3715)
Spatio-temporal Variation and Multi-dimensional Detection of Driving Mechanism of PM _{2.5} Concentration in the Chengdu-Chongqing Urban Agglomeration from 2000 to 2021	XU Yong, GUO Zhen-dong, ZHENG Zhi-wei, <i>et al.</i> (3724)
Estimation of PM _{2.5} Hourly Concentration in Sichuan Province Based on GTWR-XGBoost Model	WU Di, DU Ning, WANG Li, <i>et al.</i> (3738)
Transmission and Growth Characteristics of Severe PM _{2.5} Pollution Events from 2013 to 2021 in Xingtai, Hebei	JIANG Qi, SHENG Li, JIN Yu-chen, <i>et al.</i> (3749)
Chemical Characteristics and Source Apportionment of Organic Aerosols in Urban Shanghai During Cold Season Based on High Time-resolution Measurements of Organic Molecular Markers	ZHU Shu-hui (3760)
Emission Factors of Carbonaceous Aerosol and Stable Carbon Isotope for In-use Vehicles	YU Ming-yuan, WANG Qian, FU Ming-liang, <i>et al.</i> (3771)
Composition Characteristics of Volatile Organic Compounds and Associated Contributions to Secondary Pollution in Shenyang Industrial Area in Summer	GUAN Lu, SU Cong-cong, KU Ying-ying, <i>et al.</i> (3779)
Characteristics of VOCs and Assessment of Emission Reduction Effect During the Epidemic Lockdown Period in Shenzhen Urban Area	YUN Long, LIN Chu-xiong, LI Cheng-liu, <i>et al.</i> (3788)
Characteristics of Organic Matter Composition and Oxidation Potential in Road Dust in Winter in Xi'an	WANG Qing-wen, CHEN Qing-cai, WANG Chao, <i>et al.</i> (3797)
Contamination Characteristics and Risk Assessment of Polycyclic Aromatic Compounds in Surface Dust of Suntuan Mining Area in Huaibei	XU Zhen-peng, QIAN Ya-hui, HONG Xiu-ping, <i>et al.</i> (3809)
Spatial-temporal Variation in Water Quality and Its Response to Precipitation and Land Use in Baiyangdian Lake in the Early Stage of the Construction of Xiong'an New Area	WANG Zi-ming, YANG Li-hu, SONG Xian-fang (3820)
Simulation of Pollution Apportionment and Optimization of Control Methods in Watershed Scale: A Case Study of the Shun'an Watershed in Tongling City	LIU Guo-wangchen, CHEN Lei, LI Jia-qi, <i>et al.</i> (3835)
Evaluation of Shallow Groundwater Quality and Optimization of Monitoring Indicators in Nanchang	ZHENG Zi-yin, CHU Xiao-dong, XU Jin-ying, <i>et al.</i> (3846)
Hydrogen and Oxygen Isotopic Characteristics and Influencing Factors of "Three Waters" in Shandian River Basin	YANG Li-na, JIA De-bin, GAO Rui-zhong, <i>et al.</i> (3855)
Distribution Characteristics and Influencing Factors of Abundant and Rare Planktonic Microeukaryotes in Jinsha River	YAN Bing-cheng, CUI Ge, SUN Sheng-hao, <i>et al.</i> (3864)
Bacterial Community Diversity in Channel Sediments of Different Disturbance Sections of the Jialing River	ZHU Lan-ping, ZHANG Tuo, LI Jia-ning, <i>et al.</i> (3872)
Effects of Reservoir Water Depth on Different Plankton Communities and Keystone Species of Network Interaction	WANG Xun, LIAO Qin, WANG Pei-fang, <i>et al.</i> (3881)
Correlation Between the Diversity Characteristics of Groundwater Bacterial Community and Environmental Factors in Typical Industrial Areas	WU Jian-qiang, ZHANG Shu-yuan, WANG Min, <i>et al.</i> (3892)
Effects of Fertilizer Application Strategy Adjustments on Nitrogen and Phosphorus Loss from Typical Crop Systems in Taihu Lake Region	YU Ying-liang, WANG Yi-zhi, YANG Bei, <i>et al.</i> (3902)
Estimation of Cropland Nitrogen Runoff Loss Loads in the Yangtze River Basin Based on the Machine Learning Approaches	ZHANG Yu-fu, PAN Zhe-qi, CHEN Ding-jiang (3913)
Classification and Identification of Non-point Source Nitrogen Pollution in Surface Flow of the Shangwu River Watershed in the Qiandao Lake Region	YU Ke, YAN Yan, TANG Zhang-xuan, <i>et al.</i> (3923)
Spatial Distribution of Nitrogen and Phosphorus Nutrients in the Main Stream and Typical Tributaries of Tuojiang River and Fujiang River	LI Zi-yang, ZHOU Ming-hua, XU Peng, <i>et al.</i> (3933)
Sediment Pollution and Dredging Effect of Waqinhui River	ZHANG Mu, REN Zeng-yi, ZHANG Man, <i>et al.</i> (3945)
Critical Review on Environmental Occurrence and Photochemical Behavior of Substituted Polycyclic Aromatic Hydrocarbons	GE Lin-ke, WANG Zi-yu, CAO Sheng-kai, <i>et al.</i> (3957)
Cadmium and Arsenic Interactions During Co-adsorption onto Goethite	SU Zi-xian, LIU Sai-hong, GUAN Yu-feng, <i>et al.</i> (3970)
Preparation of Catalyst Cyclodextrin-Fe-TAML to Activate H ₂ O ₂ and Oxidize Organic Micropollutants in Water	LIU Qing-quan, CAI Ben-zhe, CAI Xi-yun (3978)
Performance and Reaction Mechanism of Co(II) Mediated Activation of Peroxymonosulfate for Degrading Nitrilotris (Methylene Phosphonic Acid)	ZHU Jing-lin, WANG Shu (3990)
Change in Granulation Potential and Microbial Enrichment Characteristics of Sludge Induced by Microplastics	XIE Qing-fan, YU Nan, ZHANG Ni, <i>et al.</i> (3997)
Accumulation Characteristics and Probabilistic Risk Assessment of Cd in Agricultural Soils Across China	WANG Jing, WEI Heng, PAN Bo (4006)
Source Analysis and Pollution Assessment of Soil Heavy Metals in Typical Geological High Background Area in Southeastern Chongqing	JIANG Yu-lian, YU Jing, WANG Rui, <i>et al.</i> (4017)
Integrated Analysis on Source-exposure Risk of Heavy Metals in Farmland Soil Based on PMF Model: A Case Study in the E-waste Dismantling Area in Zhejiang Province	FANG Jia, HE Ying, HUANG Nai-tao, <i>et al.</i> (4027)
Pollution Characteristics and Risk Assessment of Polycyclic Aromatic Hydrocarbons in Farmland Soil and Crops in the Suburbs of Urumqi	FAN Yue, CAO Shuang-yu, Nuerla Ailijiang, <i>et al.</i> (4039)
Distribution and Driving Mechanisms of Antibiotic Resistance Genes in Desert-Oasis Continuum	HUANG Fu-yi, ZHOU Shu-yidan, SU Jian-qiang, <i>et al.</i> (4052)
Characteristics of Antibiotic Contamination of Soil in China in Past Fifteen Years and the Bioremediation Technology: A Review	ZHAO Xiao-dong, QIAO Qing-qing, QIN Xiao-rui, <i>et al.</i> (4059)
Modified Biochar for Remediation of Soil Contaminated with Arsenic and Cadmium: A Review	LÜ Peng, LI Lian-fang, HUANG Xiao-ya (4077)
Effect of Biogas Slurry Return to Field on Heavy Metal Accumulation in Soil-crop System: A Meta-analysis	ZHAO Qi-zhi, YANG Zhi-min, KONG Fan-jing, <i>et al.</i> (4091)
Remediation Effect of Two Iron-modified Biochars on Slightly Alkaline Arsenic and Cadmium Contaminated Soil	LIANG Xin-ran, HE Dan, ZHENG Zhao-hua, <i>et al.</i> (4100)
Effects of Straw Removal Measure on Soil Cd Bioavailability and Rice Cd Accumulation	WANG Zi-yu, ZHOU Hang, ZHOU Kun-hua, <i>et al.</i> (4109)
Application of Desulphurized Gypsum with Straw to Improve Physicochemical Properties of Saline-alkali Land in Yellow River Delta	ZHAO Hui-li, YU Jin-yi, LIU Tao, <i>et al.</i> (4119)
Biological Evaluation and Key Stress Factor Diagnosis of Compound Contaminated Soil Based on Environmental DNA	HUANG Xiang-yun, ZHONG Wen-jun, LIU Xun-jie, <i>et al.</i> (4130)
Comprehensive Quality Assessment of Soil-Maize Heavy Metals in High Geological Background Area	ZHANG Chuan-hua, WANG Zhong-shu, LIU Li, <i>et al.</i> (4142)
Ecological Risk and Health Risk of Heavy Metal Pollution in Vegetable Production System of Zhejiang Province	ZHANG Shu-min, LIU Cui-ling, YANG Gui-ling, <i>et al.</i> (4151)
Effect of Biochar with Phosphorus Fertilizer on Soil Nutrients, Enzyme Activity, and Nutrient Uptake of <i>Alfalfa</i>	LIU Xin-yu, WANG Dong-mei, ZHANG Ze-zhou, <i>et al.</i> (4162)
Effects of Low-density Polyethylene Microplastics on the Growth and Physiology Characteristics of <i>Ipomoea aquatica</i> Forsk	ZHOU Ying, JIANG Wen-ting, LIU Xun-jie, <i>et al.</i> (4170)
Effects of Biogas Slurry Application on Soil Microbial Communities Structure and Function During Wheat-rice Stubble Period	QIAO Yu-ying, XI Hui, LI Na, <i>et al.</i> (4179)
Analysis of Soil Bacterial Community Structure and Ecological Function Characteristics in Different Pollution Levels of Lead-zinc Tailings in Datong	LIU Ze-xun, ZHUANG Jia-yao, LIU Chao, <i>et al.</i> (4191)
Effect of Fire-deposited Charcoal on Soil Organic Carbon Pools and Associated Enzyme Activities in a Recently Harvested <i>Pinus massoniana</i> Plantation Subjected to Broadcast Burning	YAO Zhi, JIAO Peng-yu, WU Xiao-sheng, <i>et al.</i> (4201)