

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)
环糊精催化 Fe_3O_4 催化剂的制备及其活化 H_2O_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, 张力, 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)	

基于轻量级梯度提升机的南京大气臭氧浓度预测

朱珈莹¹, 安俊琳^{1*}, 冯悦政¹, 贺婕¹, 张玉欣², 王俊秀³

(1. 南京信息工程大学, 中国气象局气溶胶-云-降水重点开放实验室, 南京 210044; 2. 青海省人工影响天气办公室, 西宁 810000; 3. 呼和浩特市气象局, 呼和浩特 010020)

摘要: 采用南京地区 2015 年 1 月至 2016 年 12 月期间的空气质量数据和常规气象资料数据, 分析了南京地区 O_3 浓度变化特征, 建立基于轻量级梯度提升机 (LightGBM) 的 O_3 浓度预测模型, 并将该模型与支持向量机、循环神经网络和随机森林等 3 种在空气质量预测方向上常用的机器学习方法进行了对比, 验证模型的有效性和可行性。结果表明, 南京地区 O_3 浓度变化具有显著的季节性差异, 浓度变化受前期浓度、气象因子和其他空气污染物浓度的共同影响。LightGBM 模型较为准确地预测了南京地区地面 O_3 浓度 ($R^2 = 0.92$), 且该模型的预测精度和计算效率等性能优于其他模型。尤其是在容易出现臭氧污染的高温天气, 该模型预测准确性明显高于其他模型, 模型稳定性较好。LightGBM 具有预测准确度高、稳定性好、有良好的泛化能力和运算时间短等特点, 在 O_3 浓度预测方面具有显著的优势。

关键词: 轻量级梯度提升机 (LightGBM); 地面臭氧; 臭氧浓度预测; 随机森林 (RF); 循环神经网络 (RNN)

中图分类号: X515 文献标识码: A 文章编号: 0250-3301(2023)07-3685-10 DOI: 10.13227/j.hjxx.202208095

Atmospheric Ozone Concentration Prediction in Nanjing Based on LightGBM

ZHU Jia-ying¹, AN Jun-lin^{1*}, FENG Yue-zheng¹, HE Jie¹, ZHANG Yu-xin², WANG Jun-xin³

(1. Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Weather Modification Office of Qinghai Province, Xining 810000, China; 3. Hohhot Meteorological Bureau, Hohhot 010020, China)

Abstract: Based on the air quality data and conventional meteorological data of the Nanjing Region from January 2015 to December 2016, to analyze the characteristics of O_3 concentration changes in the Nanjing Region, a light gradient boosting machine (LightGBM) model was established to predict O_3 concentration. The model was compared with three machine learning methods that are commonly used in air quality prediction, including support vector machine, recurrent neural network, and random forest methods, to verify its effectiveness and feasibility. Finally, the performance of the prediction model was analyzed under different meteorological conditions. The results showed that the variation in O_3 concentration in Nanjing had significant seasonal differences and was affected by a combination of its pre-concentration, meteorological factors, and other air pollutant concentrations. The LightGBM model predicted the ground-level O_3 concentration in the Nanjing area more precisely to a large extent ($R^2 = 0.92$), and the model outperformed other models in prediction accuracy and computational efficiency. In particular, the model showed a significantly higher prediction accuracy and stability than that of other models under a high-temperature condition that was more likely prone to ozone pollution. The LightGBM model was characterized by its high prediction accuracy, good stability, satisfactory generalization ability, and short operation time, which broaden its application prospect in O_3 concentration prediction.

Key words: light gradient boosting machine (LightGBM); ground-level ozone; ozone concentration prediction; random forest (RF); recurrent neural networks (RNN)

臭氧 (ozone, O_3) 是六大空气污染物之一, 主要由光化学反应产生。对流层 O_3 具有强氧化性^[1], 会对人类健康和植被生产力产生危害^[2,3]。此外, O_3 也是一种重要的温室气体, 能影响环境变化^[4]。

随着各项污染防治政策的实施, 我国大部分城市细颗粒物浓度逐渐下降, 然而臭氧浓度持续上升^[5]。尤其是在京津冀、长三角和珠三角等中心城市, 随着化石燃料消耗量的快速增长, 氮氧化物 (NO_x) 和挥发性有机物 (VOCs) 等前体物的排放量急剧增加, 进而导致近地面 O_3 浓度显著上升^[2,6]。2014 ~ 2018 年全国空气质量监测数据表明, 我国中心城市群首要污染物由以细颗粒物为主转变为以臭氧为主, O_3 为首要污染物的天数呈上升趋势, 由 12.9% 增长至 28.8%^[7]。 O_3 污染已经成为污染防治的核心问题, 对 O_3 浓度的预测研究能为 O_3 污染防治提供技术支持。

O_3 预报方法主要分为数值模式预报和统计预

报两大类^[8]。数值模式能够反映污染物的变化过程, 但模式预报对于输入数据要求苛刻, 且计算量大, 计算成本高^[9]。相比于传统模式预报, 基于数学统计方法的统计预测模型具有预报精度高、运算速度快和资料要求较低等优点^[10]。近年来, 也有研究将两种方法结合使用^[11~13]。李梓铭等^[11]采用相似集合预报技术对 RMAPS-CHEM 预报的 O_3 浓度进行了释用订正, 研究表明该方法明显改善了模式预报效果, 统计预报可以有效提高模式预报的准确性。

随着人工智能技术的快速发展, 机器学习方法已广泛应用于空气质量预测^[14]。常用的机器学习方法主要有支持向量机^[15~17]、决策树^[18,19]和人工神经网络^[20,21]等。支持向量机 (support vector machine,

收稿日期: 2022-08-11; 修订日期: 2022-10-06

基金项目: 国家自然科学基金项目 (42075177); 国家重点研发计划项目 (2017YFC0210003); 江苏省高校“青蓝工程”项目

作者简介: 朱珈莹 (2000 ~), 女, 硕士研究生, 主要研究方向为机器学习与大气环境, E-mail: 20211203039@nuist.edu.cn

* 通信作者, E-mail: junlinan@nuist.edu.cn

SVM)能够捕捉预测因子之间的非线性关系^[15,16]. Chelani^[17]将 SVM 应用于 O_3 浓度预测,并获得了较好的预测结果. 随机森林(random forest, RF)是一种基于分类树的算法,能够处理高维数据^[22,23]. 马润美等^[24]基于 RF 方法建立了日最大 O_3 -8h 浓度预测模型,结果表明,模型预测性能较好($R^2 = 0.85$). 循环神经网络(recurrent neural networks, RNN)是常用的神经网络之一,它在处理具有序列特性的数据时表现优异^[20]. Feng 等^[21]通过 RNN 对地面 O_3 浓度进行了预测,结果表明该模型预测精度显著高于其他预测模型.

轻量梯度提升机(light gradient boosting machine, LightGBM)是一种集成学习算法,具有训练速度快,内存占用小和精度高等优点,该模型在处理高维数据时有较好的表现^[24~26]. LightGBM 算法已经成为数据挖掘领域的研究热点,经常出现在各大数据挖掘竞赛中^[27]. 在气象领域,已有学者将该模型用于大气能见度预报^[25]、海面风场反演^[26]和 $PM_{2.5}$ 浓度预测^[24,28],但鲜见研究将其应用于 O_3 的预报. 地面 O_3 成因复杂,影响因素多,用传统机器学习方法解决这种高维小样本的回归问题容易过度拟合,且训练时间长^[29,30]. 目前,已有研究证实了 LightGBM 算法在处理回归问题上准确且高效^[31,32]. 如李志生等^[31]和 Wei 等^[32]应用 LightGBM 算法建立 $PM_{2.5}$ 预测模型,结果表明模型预测准确度高且运算时间短. 因此,采用 LightGBM 算法构建地面臭氧浓度预测模型有很好的适用性. 本文应用 LightGBM 算法,以气象因子和空气污染物浓度为输入变量,构建了南京地区地面 O_3 浓度预测模型,并将所提模型与 SVM、RNN 和 RF 这 3 种在空气质量预测方向上常用的机器学习方法进行了对比,以验证该模型的有效性和可行性.

1 材料与方法

1.1 观测站点

本文选择 2015 ~ 2016 年南京市浦口环境监测站(32.08°N, 118.62°E)空气质量监测数据和南京信息工程大学内自动气象站(32.20°N, 118.72°E)的地面气象观测数据作为研究数据,观测站点的位置如图 1 所示. 气象观测站点位于南京江北工业园区下风向地区的南京信息工程大学,其周围环境复杂,站点临近化工厂和交通干道.

1.2 实验数据

本研究主要采用了空气质量监测数据和地面气象观测数据,研究时间为 2015 年 1 月至 2016 年 12 月. 空气质量数据来自中国环境监测总站全国空气



图 1 观测站点位置示意

Fig. 1 Observation site location

质量观测数据网(<http://www.cnemc.cn/>),更新频率为 1 次·h⁻¹. 地面气象观测数据来自 CAWSD600 型自动气象站的连续观测数据,时间分辨率为 1 h. 气象仪器位于南京信息工程大学内的中国气象局综合观测实习基地. 累积降水量数据来自校内地面雨量计数据,时间分辨率为 1 h,雨量分辨率为 0.1 mm.

选取的空气质量因子主要包括 O_3 、 $PM_{2.5}$ 、 PM_{10} 、 NO_2 、CO 和 SO_2 . 气象变量主要包括相对湿度、温度、露点温度、平均气压、最大风速和风向和累积降水量. 本研究所使用的空气质量数据按照《环境空气质量标准》(GB 3095-2012)进行质量控制. 气象数据按照台站级、省级和国家级进行三级质控. 共获得大气环境数据 16 790 组,气象数据 16 770 组,分别占观测期间总样本的 95.70% 和 95.59%,最终得到有效数据共 14005 组. 在剔除异常值和缺失值后,对数据进行归一化处理,使得输入数据的分布符合标准正态分布,有利于提高模型的精度和训练速度. 为评估模型的泛化能力,需要对数据集进行划分,其中 80% 作为训练集用于训练模型,剩下的数据用于验证模型性能. 这一系列数据处理过程,是通过 Python 中的 Sklearn 和 Pandas 库实现.

1.3 模型介绍

SVM 算法是由 Vapnik^[33]提出的,它通过非线性函数,将原本低维空间中的非线性问题转化为高维特征空间的线性回归问题,擅于处理非线性问题. SVM 遵循结构风险最小化原则,有效地避免了过拟合问题,具有良好的泛化能力. 此外,为了减少计算量, SVM 采用核函数来代替高维空间中的内积运

算. 核函数的选择也会影响模型的性能, 本研究选择的是能够处理非线性数据且具有良好的学习能力的 RBF 核函数.

RNN 是一种与时间序列相关的人工神经网络, 它的主要特点为不同层之间的神经元也通过权重函数相互连接. 也就是说, 模型当前输出与前面的输出相关, 网络会记忆前面的信息, 同时将信息传递给下一个网络, 并用于下一时刻输出的计算中^[34]. RNN 的计算公式为:

$$z_t = Uh_{t-1} + Wx_t + b \quad (1)$$

式中, z_t 为当前时刻隐藏单元的值, h_{t-1} 为前一时刻的输出, x_t 为当前时刻的输入, U 和 W 为权值, b 为偏置.

$$h_t = f(z_t) \quad (2)$$

式中, h_t 为当前时刻的输出, $f(x)$ 为激活函数, 主要包括 Relu 函数和 Sigmoid 函数等, 本研究选用 Relu 函数作为激活函数.

RF 是由 Breiman^[35]开发的一种集成学习算法, 它通过 Bootstrap 有放回的采样方法, 随机选取部分训练样本和变量子集生成多棵决策树. 在回归问题中, 模型的最终预测结果为 RF 中各个决策树预测结果的平均值. RF 模型精度高, 具有良好的泛化性, 对于噪声点和离群值更加稳健, 且该算法易于实现.

梯度提升决策树 (gradient boosting decision tree, GBDT) 是基于弱学习器多次迭代得到最终强学习器的算法, 每次迭代得到的新学习器都是针对前一个学习器的残差进行拟合, 最后将所有树的预测相加, 从而完成预测任务^[36].

首先计算残差 r_{ni} :

$$r_{ni} = y_i - f_{n-1}(x_i) \quad (3)$$

式中, y_i 为第 i 个样本的实测值, $f_{n-1}(x_i)$ 为前一轮学习机的预测值.

对残差进行拟合, 得到一个拟合残差模型 $h_n(x)$.

更新回归树:

$$f_n(x) = f_{n-1}(x) + h_n(x) \quad (4)$$

LightGBM 是微软基于 GBDT 框架新开发的机器学习方法, 采用了梯度单边采样 (gradient-based one-side sampling, GOSS) 和互斥特征捆绑 (exclusive feature bundling, EFB) 技术^[37]. LightGBM 算法在精度和计算速度等方面均优于 GBDT 算法^[38]. LightGBM 支持高效的并行训练, 解决了目前机器学习方法所面临的主要挑战, 即对于特征维数多且数据量大的样本计算效率低和泛化性差的问题^[32]. GOSS 在保证信息增益的前提下减少了训练样本, 它排除了大部分对计算信息增益没有帮助的小梯度

样本, 用剩余的样本来估计信息增益. EFB 将相互独立的特征捆绑在一起, 以无损的方式实现了降维, 从而提高了模型的训练速度. 该方法可以处理分类和回归任务, 且具有训练速度快, 内存占用小和精度高等优点, 更适用于高维数据, 因此可应用于地面 O_3 浓度预测.

1.4 模型评估方法

为了比较 4 种预测模型的综合性能, 本文采用以下 3 个统计指标对模型进行评估.

平均绝对差值 (mean absolute error, MAE):

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

均方根误差 (root mean square error, RMSE):

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

决定系数 (squared correlation coefficient, R^2):

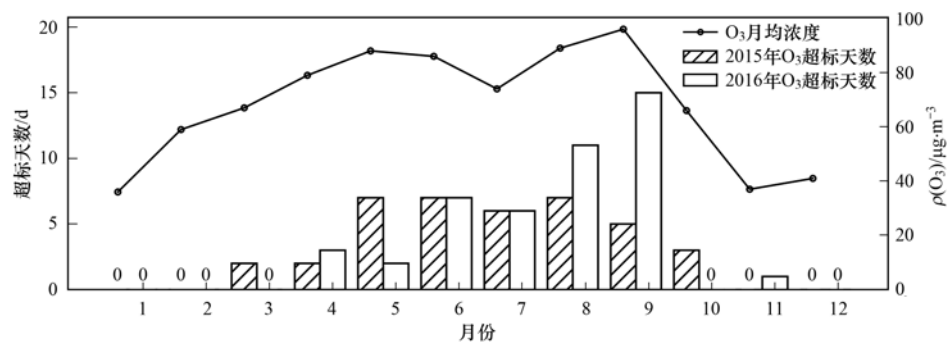
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (7)$$

式中, n 为样本总数量, y_i 、 $\bar{y}_i$ 和 $\hat{y}_i$ 分别为第 i 个样本的实测值、实测值的平均和预测值.

2 结果与讨论

2.1 O_3 浓度变化特征

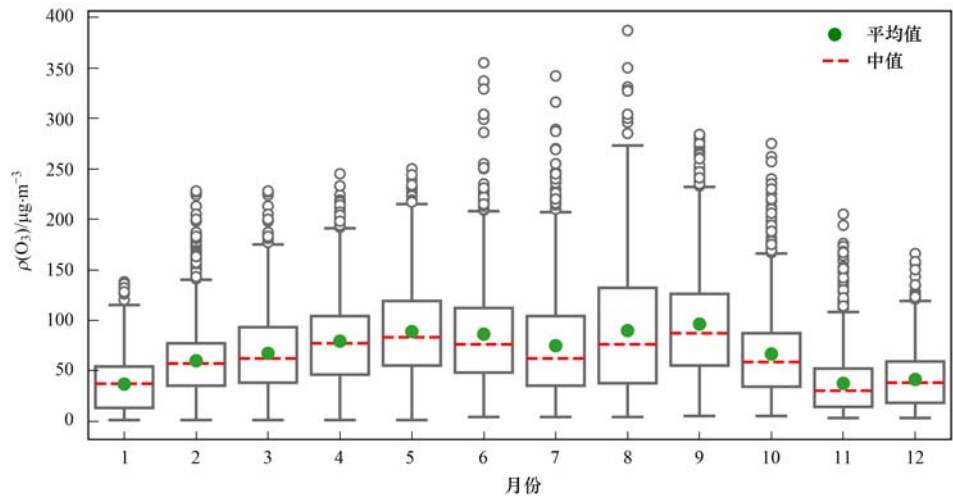
图 2 为南京地区 O_3 浓度的逐月分布特征, 结果表明, 南京地区 O_3 浓度变化具有明显的季节性差异, 呈夏季升高冬季降低的变化趋势, 5~9 月更容易出现 O_3 浓度超标的情况. 2015~2016 年研究区域 O_3 小时浓度平均值超过国家二级标准限值 ($200 \mu\text{g} \cdot \text{m}^{-3}$) 的天数为 84 d, 超标率为 11.5%, 其中 5~9 月超标天数占总体的 86.9%. 图 3 进一步显示了研究区域内 O_3 浓度逐月的分布情况. 从中可知, 各月中位数和平均值变化趋势一致, 总体呈双峰分布, 5~9 月 O_3 浓度处于较高水平. 5 月南京地区气温升高, 表明紫外辐射增强, 光化学反应速率加快, 因此 O_3 浓度在 5 月达第 1 个峰值. 江淮地区的梅雨季节通常出现在 6~7 月, 易出现连续的阴雨天气, 光照强度减弱, 不利于 O_3 生成和累积, 导致 6~7 月 O_3 月均浓度偏低^[39]. 8 月出现了 O_3 浓度的异常高值, 且 O_3 月均浓度在 9 月达到第 2 个峰值, 这可能与 8~9 月南京地区持续的高温天气有关. 南京地区 O_3 污染问题较为严重, 尤其是夏季最为突出. 因此, 为应对该区域 O_3 问题, 应建立准确的 O_3 浓度预测模型, 并分季节和分天气状况细化评估预测模型.



超标天数 0 值表示在对应的时段内没有出现 O₃ 浓度超标的情况

图 2 南京地区 O₃ 浓度逐月分布特征和超标情况

Fig. 2 Monthly distribution characteristics and exceedances of O₃ concentration in Nanjing



箱体的上下限分别表示数据的上四分位数和下四分位数,箱子上方和下方的线分别表示数据的最大值和最小值,白色圆点表示异常值

图 3 南京地区 O₃ 浓度的月变化箱线图

Fig. 3 Boxplots of monthly variation in O₃ concentration in Nanjing

2.2 影响因子分析

本研究在现有数据支持上,建立了 O₃ 浓度变化影响特征,如参与光化学反应的前体物浓度,影响 O₃ 生成和扩散的天气特征,存在相互影响的其他大气污染物特征,以及表征 O₃ 浓度累积的 O₃ 前期浓度特征.研究表明预报时刻前几个小时的历史 O₃ 浓度对预报时刻 O₃ 浓度影响较大^[40].相比于多步预测,将历史 O₃ 浓度作为辅助特征输入到

O₃ 浓度预测模型能够更准确地描述 O₃ 的累积水平.通过模型测试发现选取的前期浓度与预报时刻的时间间隔越小则预测效果越好,但考虑到预测模型的时效性,因此在建模时加入预报时刻前 6 h 的 O₃ 浓度作为输入变量代入模型,从而在保证模型实用性的前提下有效地提高了模型预报性能.为了方便表达,文中模式中的输入特征名称采用缩写的形式表示(表 1).

表 1 模型中输入变量的缩写名称及其含义

Table 1 Abbreviated names of input variables in the model and their descriptions			
模式中缩写名称	含义	模式中缩写名称	含义
PM _{2.5} /μg·m ⁻³	预报时刻的 PM _{2.5} 浓度	RH/%	预报时刻的相对湿度
PM ₁₀ /μg·m ⁻³	预报时刻的 PM ₁₀ 浓度	T/℃	预报时刻的气温
NO ₂ /μg·m ⁻³	预报时刻的 NO ₂ 浓度	T _d /℃	预报时刻的露点温度
NO ₂ _1h/μg·m ⁻³	预报时刻前 1 h 的 NO ₂ 浓度	p/hPa	预报时刻的平均气压
CO/mg·m ⁻³	预报时刻的 CO 浓度	WD_max/(°)	预报时刻的最大风速的风向
SO ₂ /μg·m ⁻³	预报时刻的 SO ₂ 浓度	WS_max/m·s ⁻¹	预报时刻的最大风速
O ₃ _6h/μg·m ⁻³	预报时刻前 6 h 的 O ₃ 浓度	PRE/mm	预报时刻的降水量

为了解特征与预测量之间的相关关系,本研究通过 RF 模型获得特征相对重要性,结果如图 4

所示.空气质量因子中 NO₂_1h 对 O₃ 浓度预测结果影响最大.研究区域内的 NO₂ 主要来自观测站

点周围的工业区和交通干道,NO₂ 作为主要的前体物之一,在阳光照射下发生一系列光化学反应生成 O₃. 此外,由于 NO 对 O₃ 有“滴定作用”,NO₂ 作为该反应的产物,其浓度水平能够反映 O₃ 的浓度变化. 从图 4 中可以看出,NO₂_1h 的贡献明显大于 NO₂ 的贡献,这主要是与 O₃ 生成的化学反应速率有关^[41]. 因此,将预报时刻前 1 h 的 NO₂ 浓度加入模型能提高模型预测准确度. 在气象因子中

T、RH 和 *p* 对 O₃ 浓度预测贡献较大. 将这 3 个对 O₃ 浓度预测贡献较大的气象因子与 O₃ 进行相关性分析,相关系数分别为 0.42、-0.35 和 -0.34. 这表明在高温、低湿和低压的天气状况下更有利于 O₃ 的生成和累积. 预报时刻的降水量贡献最小,仅占 0.01%,这主要与降水量数据的不连续性有关. 为提高模型的计算效率,去除冗余特征,故将其剔除.

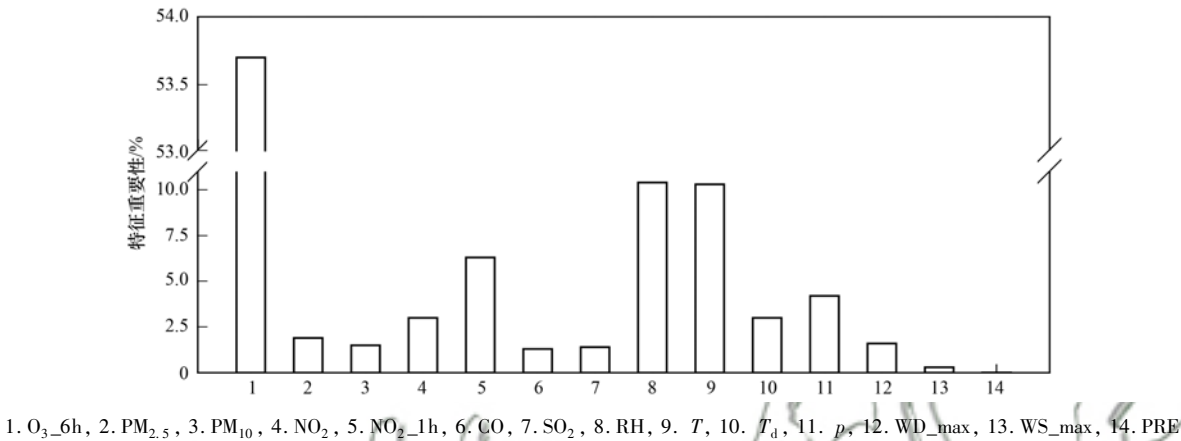


图 4 O₃-8h 预测的 RF 模型中特征重要性
Fig. 4 Feature importance in the RF model for O₃-8h prediction

2.3 模型精度分析及对比

为检验 4 种机器学习模型的预测效果,采用残差(模型预测值与实测值的差值)对模型进行评估,结果如图 5 所示. 其中,SVM 模型的残差波动范围最大,在 -172.4 ~ 66.8 μg·m⁻³波动,该模型预测值与观测值偏差较大,模型预测值低于观测值的情况居多. RNN 模型的残差在 0 值下方波动,波动范围为 -130.3 ~ 82.5 μg·m⁻³,且存在较多突变值,说明该模型普遍低估了 O₃ 浓度水平,模型预测效果较差. LightGBM 与 RF 模型的残差均在 0 值附近波动,这两种预测模型都能够较好地捕捉 O₃ 浓度的变化趋势. LightGBM 残差波动范围最小,为 -91.2 ~ 82.3 μg·m⁻³,说明 LightGBM 模型的性能略优于 RF 模型. 综合以上分析,LightGBM 与 RF 模型预测性能最好,其次是 RNN 模型,SVM 模型预测性能最差.

采用 MAE、RMSE 和 *R*² 这 3 种统计指标和模型运算时间进一步对 4 种机器学习模型的性能进行

了对比,结果如表 2 所示. 各模型的 *R*² 处于 0.79 ~ 0.92 之间,RF 和 LightGBM 模型的 *R*² 均 >0.9,预测结果较为理想,LightGBM 模型的 *R*² 值最大,为 0.92. 对比各模型的 RMSE 值,LightGBM 模型的 RMSE 值最低为 12.72 μg·m⁻³,RNN 的 RMSE 值最高为 25.41 μg·m⁻³. 对于 MAE 指标,LightGBM 模型同样是最低的,为 8.96 μg·m⁻³,相比于 SVM 预测模型降低了 42.8%,其次为 RF 模型,而 RNN 模型的 MAE 最大. 综合对比 4 个机器学习模型,LightGBM 模型的 *R*² 最大且 RMSE 和 MAE 值最小,LightGBM 模型预测性能最好,其次为 RF 模型,而 SVM 模型在各模型中表现最差. 在计算效率方面,LightGBM 运行时间最快(4.18 s),RNN 运行时间最慢(407.81 s). 图 6 为各预测模型的实测值和预测值的散点图,显示了二者之间的拟合程度. 从中可以看出,LightGBM 模型的拟合情况明显优于其他模型.

表 2 SVM、RNN、RF 和 LightGBM 的预测结果

Table 2 Evaluation indicators of SVM, RNN, RF, and LightGBM prediction results				
项目	MAE/μg·m ⁻³	RMSE/μg·m ⁻³	<i>R</i> ²	运算时间/s
SVM	15.67	22.77	0.79	12.73
RNN	18.98	25.41	0.83	407.81
RF	9.65	13.81	0.90	5.71
LightGBM	8.96	12.72	0.92	4.18

南京地区的 O₃ 浓度变化具有明显的季节性差异,为评估各模型的预测稳定性,应对各模型在不同

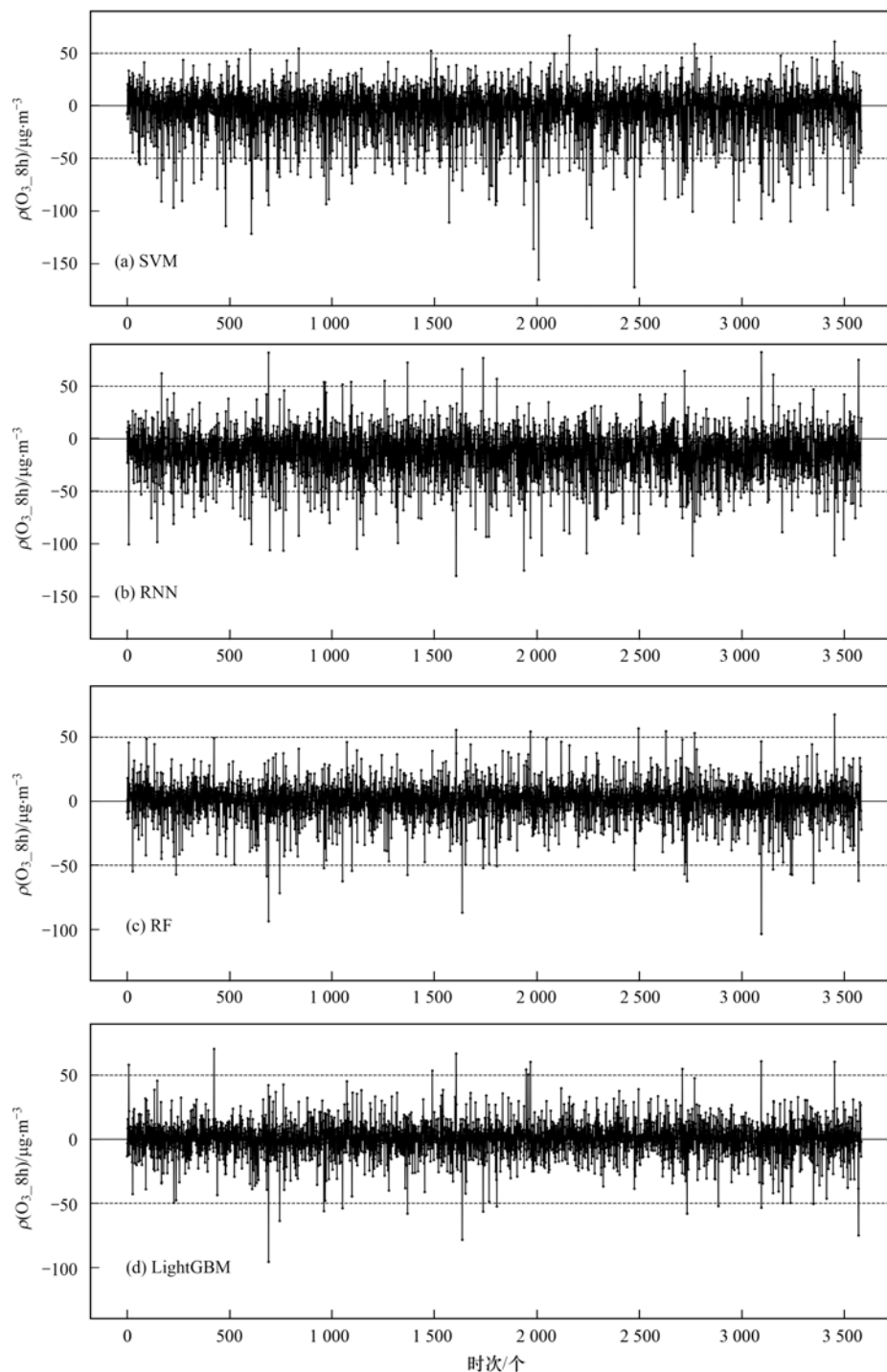
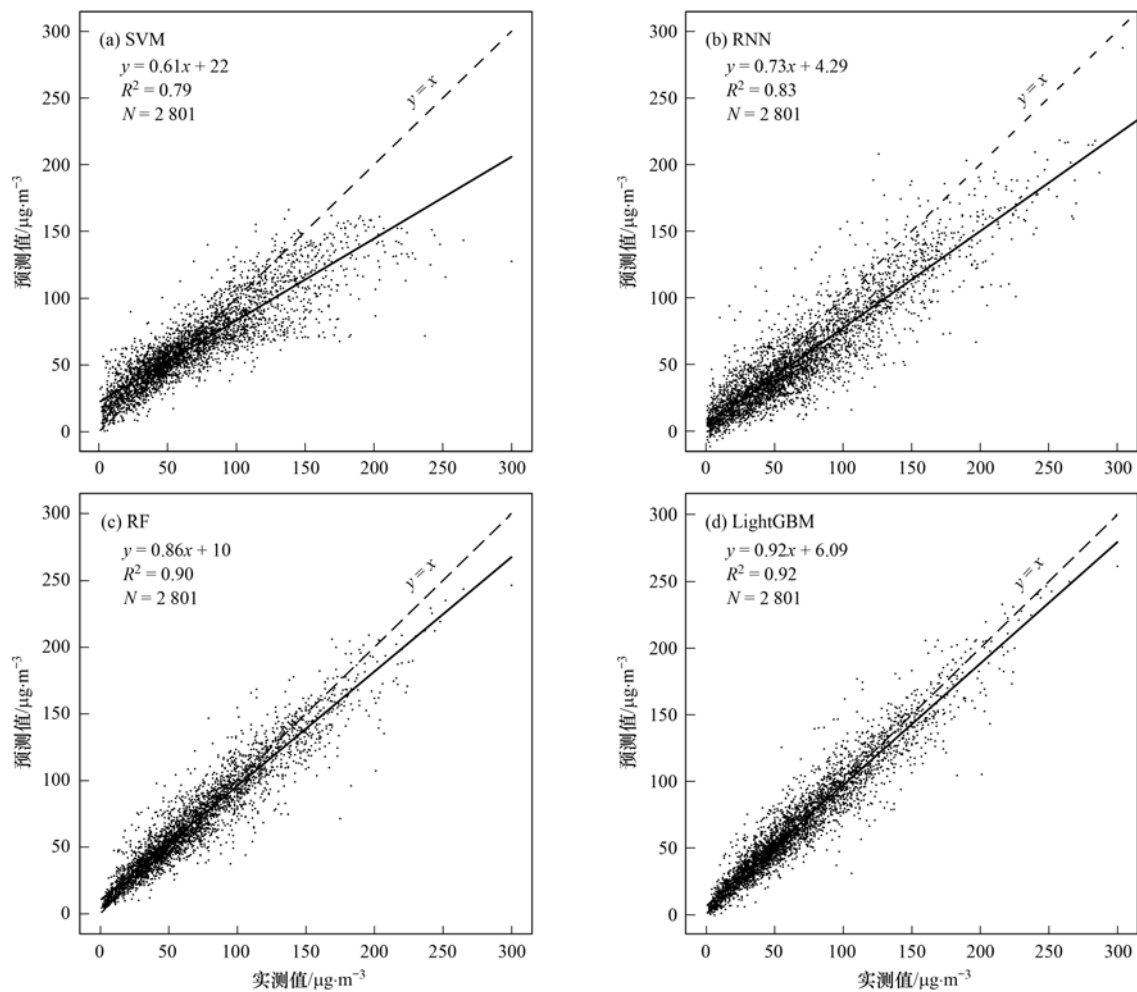


图5 SVM、RNN、RF 和 LightGBM 的残差

Fig. 5 Prediction difference of SVM, RNN, RF, and LightGBM

季节的预测结果进行讨论. 但由于南京的冬、夏两个季节的持续时间明显比春、秋两季长^[42], 因此根据气温将样本划分为高温 ($\geq 28^{\circ}\text{C}$) 和低温 ($< 28^{\circ}\text{C}$) 两种情况进行讨论更合理. 利用泰勒图来评估在高温和低温两种不同的天气状况下各模型的预测性能, 图7中散点表示各模型的误差统计量. 泰勒图由预测值与实测值的相关系数 (R)、标准差 (σ) 和均方根误差 (RMSE) 组成, 图7(a)为低温情况, 其中4个散点均离基准点较近, 除 SVM 模型外

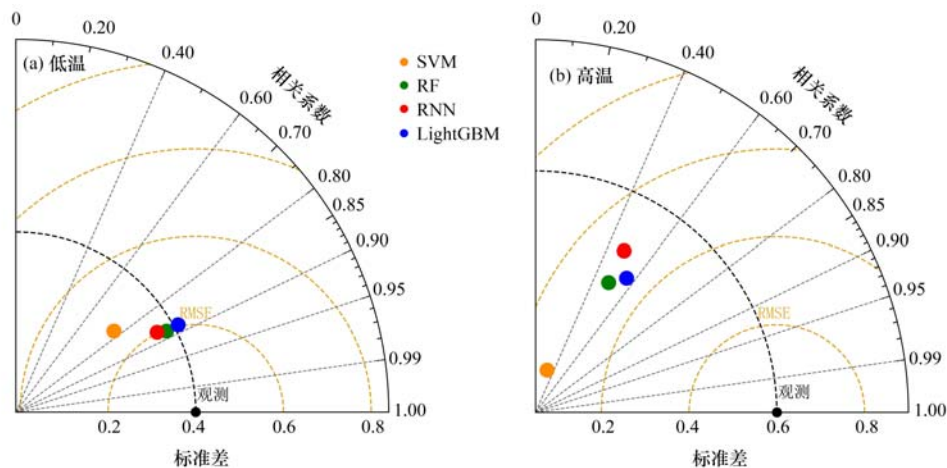
其余各点之间的间距很小. 说明在低温情况下, 4个预测模型的预测结果均较为理想, 其中 SVM 模型表现最差, 另外3种模型预测误差相近. 图7(b)为高温情况, 其中4个散点比较分散, 说明各模型的性能存在明显差异, 其中蓝色散点 (LightGBM 模型) 最接近基准点, 说明在高温情况下 LightGBM 模型性能最好. 表3为在高温和低温天气状况下模型预测结果的 R^2 、RSME 和 MAE, 可直观地看到4个机器学习模型的预测精度在高温情况下均有不同程度地下



N 为样本数量, 黑色虚线表示预测值与实测值相等, 拟合直线越靠近黑色虚线表示模型的拟合优度越高

图 6 各个模型的预测值和实测值的散点图

Fig. 6 Scatter plots of predicted and observed values for each model



辐射线表示模型预测值与实测值相关系数的大小, 横坐标表示标准差, 中间的黄色虚线表示均方根误差的大小, 黑色圆点为基准点 ($R=1, \sigma=0, RMSE=0$), 散点越靠近基准点表示模型的预测准确度越高

图 7 各模型在高温和低温两种情况下的预测结果泰勒图

Fig. 7 Taylor diagram of prediction results for each model at high and low temperature conditions

降. LightGBM、RF、RNN 和 SVM 模型预测值与观测值的 R^2 降低比例分别为 36.36%、44.83%、44.32% 和 66.23%, RMSE 分别上升了 22.23、24.77、24.71 和 27.38 $\mu\text{g}\cdot\text{m}^{-3}$, MAE 分别上升了

16.88、18.08、18.60 和 23.79 $\mu\text{g}\cdot\text{m}^{-3}$, 这主要与 2016 年的异常高温天气有关. 长三角地区的研究表明, 2016 年是显著的高温热浪年, 观测到的最高温度达 41°C [43]. 由于缺少这种特殊天气状况下的训

表 3 在高温和低温两种情况下模型的预测结果

Table 3 Evaluation indicators of the model prediction results under high temperature and low temperature

项目	R^2		RMSE/ $\mu\text{g}\cdot\text{m}^{-3}$		MAE/ $\mu\text{g}\cdot\text{m}^{-3}$	
	低温	高温	低温	高温	低温	高温
SVM	0.77	0.26	19.66	47.04	13.80	37.59
RNN	0.88	0.49	14.86	39.57	10.76	29.36
RF	0.87	0.48	14.23	39.00	10.28	28.23
LightGBM	0.88	0.56	14.14	36.37	10.24	27.12

练数据,导致模型预测结果较差.

综合以上分析,LightGBM 模型性能最优,其次是 RF 模型,RNN 和 SVM 模型预测精度较差.LightGBM 和 RF 模型都属于集成学习算法,RNN 和 SVM 模型属于单一的机器学习算法.已有研究表明^[37,44,45],相比于单一机器学习模型,集成学习模型往往具有更高的准确性、稳定性和泛化性.因此 LightGBM 和 RF 模型的预测结果更为准确.相比于 RF 模型,LightGBM 模型泛化性更好,且计算效率更高,运算时间更短,这主要与算法设计有关.RF 模型是基于 Bagging 算法发展而来,而 LightGBM 模型是基于 Boosting 算法.这两种算法的模型效果不同.Bagging 采用有放回的采样方法,能增加弱学习器之间的独立性,最终预测结果取的是所有弱学习器预测结果的平均,从而得到一个方差最小的模型.Boosting 算法中的各弱学习器之间具有强相关性,最终训练得到偏差最小的模型.因此,对于噪声小的数据,LightGBM 模型具有更高的预测准确度.LightGBM 模型通过直方图算法和带深度限制的叶子生长算法等一系列优化算法,有效地提高了计算效率,减少计算时间.LightGBM 模型较为准确地预测了南京地区地面 O_3 浓度,尤其是在容易出现臭氧污染的高温天气,该模型预测准确性明显高于其他模型.该模型具有预测准确度高、稳定性好、有良好的泛化能力和运算时间短等特点.除高温热浪等异常天气状况外模型预测效果较好,这说明 LightGBM 模型稳定性较好,在 O_3 浓度预测方面具有良好的应用前景.但本文构建的基于 LightGBM 的 O_3 浓度预测模型还存在一些不足:输入模型的预测因子不够全面;LightGBM 算法在高维稀疏特征中难以进行有效的特征空间划分;评估模型预测稳定性时仅考虑了温度的影响.未来研究将加入 NO 和 VOCs 等因素,采用特征提取方法来提高数据质量,并评估湿度、局地输送和特殊天气过程等条件对模型稳定性的影响.

3 结论

(1)南京地区 O_3 浓度变化具有显著的季节性差异,浓度变化受前期浓度、气象因子和其他空气

污染物浓度的共同影响.

(2)LightGBM 模型较为准确地预测了南京地区地面 O_3 浓度,且计算效率高,该模型的性能优于 SVM、RNN 和 RF 等 3 种常用的机器学习模型,4 种机器学习模型的预测精度从高到低依次为:LightGBM > RF > RNN > SVM. R^2 分别为 0.92、0.90、0.83 和 0.79,RMSE 分别为 12.72、13.81、25.41 和 22.77 $\mu\text{g}\cdot\text{m}^{-3}$.

(3)在容易出现臭氧污染的高温天气,各模型的预测精度均有不同程度地下降,LightGBM、RF、RNN 和 SVM 模型的预测值与观测值的 RMSE 分别上升了 22.23、24.77、24.71 和 27.38 $\mu\text{g}\cdot\text{m}^{-3}$,其中 LightGBM 模型的预测精度下降幅度最小,模型稳定性较好.LightGBM 模型具有预测准确度高、稳定性好、有良好的泛化能力和运算时间短等特点,在 O_3 浓度预测方面有良好的应用前景.

致谢:本文使用了中国环境监测总站的空气质量监测数据和南京信息工程大学中国气象局综合观测实习基地的地面气象观测数据,在此表示感谢.

参考文献:

[1] Zheng X Y, Orellano P, Lin H L, *et al.* Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: a systematic review and meta-analysis [J]. *Environment International*, 2021, **150**, doi: 10.1016/j.envint.2021.106435.

[2] Feng Z Z, Agathokleous E, Yue X, *et al.* Emerging challenges of ozone impacts on Asian plants: actions are needed to protect ecosystem health [J]. *Ecosystem Health and Sustainability*, 2021, **7**(1), doi: 10.1080/20964129.2021.1911602.

[3] Lu X, Zhang L, Wang X L, *et al.* Rapid increases in warm-season surface ozone and resulting health impact in China since 2013[J]. *Environmental Science & Technology Letters*, 2020, **7**(4): 240-247.

[4] Zhang J F, Wei Y J, Fang Z F. Ozone pollution: a major health hazard worldwide[J]. *Frontiers in Immunology*, 2019, **10**, doi: 10.3389/fimmu.2019.02518.

[5] Jia P C, Cao N W, Yang S B. Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence to sequence model [J]. *Atmospheric Environment*, 2021, **244**, doi: 10.1016/j.atmosenv.2020.117917.

[6] Tang X, Gao X, Li G L, *et al.* Study on spatiotemporal distribution of airborne ozone pollution in subtropical region considering socioeconomic driving impacts: a case study in Guangzhou, China[J]. *Sustainable Cities and Society*, 2020,

- 54, doi: 10.1016/j.scs.2019.101989.
- [7] 王未来, 李明艳, 张美航, 等. 我国 31 个重点城市 2014—2018 年空气质量变化特征分析[J]. 中国公共卫生管理, 2020, 36(4): 441-445.
- Wang W L, Li M Y, Zhang M H, *et al.* Analysis on variation characteristics of air quality in 31 key cities from 2014 to 2018 in China[J]. Chinese Journal of Public Health Management, 2020, 36(4): 441-445.
- [8] 朱晏民, 徐爱兰, 孙强. 基于深度学习的空气质量预报方法新进展[J]. 中国环境监测, 2020, 36(3): 10-18.
- Zhu Y M, Xu A L, Sun Q. New progress for air quality forecasting methods based on deep learning[J]. Environmental Monitoring in China, 2020, 36(3): 10-18.
- [9] 石佳超. 基于 CMAQ 与前馈神经网络的区域大气污染物浓度快速响应研究[D]. 杭州: 浙江大学, 2019.
- Shi J C. Study on rapid response of regional air pollution concentration based on CMAQ and feed forward neural network [D]. Hangzhou: Zhejiang University, 2019.
- [10] 董红召, 王乐恒, 唐伟, 等. 融合时空特征的 PCA-PSO-SVM 臭氧(O_3)预测方法研究[J]. 中国环境科学, 2021, 41(2): 596-605.
- Dong H Z, Wang L H, Tang W, *et al.* Research on PCA-PSO-SVM ozone prediction considering spatial-temporal features[J]. China Environmental Science, 2021, 41(2): 596-605.
- [11] 李梓铭, 赵秀娟, 孙兆彬, 等. 基于相似集合预报技术的臭氧预报释用研究[J]. 中国环境科学, 2020, 40(2): 475-484.
- Li Z M, Zhao X J, Sun Z B, *et al.* Research on the interpretation and correction of numerical ozone forecast based on Analog Ensemble[J]. China Environmental Science, 2020, 40(2): 475-484.
- [12] 熊亚军, 徐敬, 孙兆彬, 等. 基于数据挖掘算法和数值模拟技术的大气污染减排效果评估[J]. 环境科学学报, 2019, 39(1): 116-125.
- Xiong Y J, Xu J, Sun Z B, *et al.* Air pollution reduction effect evaluation based on data mining algorithm and numerical simulation technology[J]. Acta Scientiae Circumstantiae, 2019, 39(1): 116-125.
- [13] Sayeed A, Choi Y, Eslami E, *et al.* A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance[J]. Scientific Reports, 2021, 11(1), doi: 10.1038/s41598-021-90446-6.
- [14] Cheng Y, He L Y, Huang X F, *et al.* Development of a high-performance machine learning model to predict ground ozone pollution in typical cities of China[J]. Journal of Environmental Management, 2021, 299, doi: 10.1016/j.jenvman.2021.113670.
- [15] Fan J L, Wu L F, Ma X, *et al.* Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions[J]. Renewable Energy, 2020, 145: 2034-2045.
- [16] 苏筱倩. 基于核极限学习机和小波变换的臭氧预测及比较研究[D]. 南京: 南京信息工程大学, 2020.
- Su X Q. Research on ozone concentration prediction and method comparison based on kernel extreme learning machine and wavelet transform [D]. Nanjing: Nanjing University of Information Science & Technology, 2020.
- [17] Chelani A B. Prediction of daily maximum ground ozone concentration using support vector machine[J]. Environmental Monitoring and Assessment, 2010, 162(1-4): 169-176.
- [18] Jumin E, Zaini N, Ahmed A N, *et al.* Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction [J]. Engineering Applications of Computational Fluid Mechanics, 2020, 14(1): 713-725.
- [19] Eslami E, Salman A K, Choi Y, *et al.* A data ensemble approach for real-time air quality forecasting using extremely randomized trees and deep neural networks [J]. Neural Computing and Applications, 2020, 32(11): 7563-7579.
- [20] Kim K, Kim D K, Noh J, *et al.* Stable forecasting of environmental time series via long short term memory recurrent neural network[J]. IEEE Access, 2018, 6: 75216-75228.
- [21] Feng R, Zheng H J, Zhang A R, *et al.* Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: a case study in Hangzhou, China[J]. Environmental Pollution, 2019, 252: 366-378.
- [22] Scornet E, Biau G, Vert J P. Consistency of random forests[J]. The Annals of Statistics, 2015, 43(4): 1716-1741.
- [23] 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.
- [24] Zhang Y, Wang Y H, Gao M H, *et al.* A predictive data feature exploration-based air quality prediction approach [J]. IEEE Access, 2019, 7: 30732-30743.
- [25] 余东昌, 赵文芳, 聂凯, 等. 基于 LightGBM 算法的能见度预测模型[J]. 计算机应用, 2021, 41(4): 1035-1041.
- Yu D C, Zhao W F, Nie K, *et al.* Visibility forecast model based on LightGBM algorithm [J]. Journal of Computer Applications, 2021, 41(4): 1035-1041.
- [26] 骆黎明, 白伟华, 孙越强, 等. 基于树模型机器学习方法的 GNSS-R 海面风速反演[J]. 空间科学学报, 2020, 40(4): 595-601.
- Luo L M, Bai W H, Sun Y Q, *et al.* GNSS-R sea surface wind speed inversion based on tree model machine learning method [J]. Chinese Journal of Space Science, 2020, 40(4): 595-601.
- [27] 刘新伟, 黄武斌, 蒋盈沙, 等. 基于 LightGBM 算法的强对流天气分类识别研究[J]. 高原气象, 2021, 40(4): 909-918.
- Liu X W, Huang W B, Jiang Y S, *et al.* Study of the classified identification of the strong convective weathers based on the LightGBM algorithm[J]. Plateau Meteorology, 2021, 40(4): 909-918.
- [28] 黄烈星. 基于 Stacking 融合模型的 $PM_{2.5}$ 浓度时空变化分析及预测[D]. 赣州: 江西理工大学, 2020.
- Huang L X. $PM_{2.5}$ concentration based on Stacking model spatio-temporal variation analysis and prediction [D]. Ganzhou: Jiangxi University of Science and Technology, 2020.
- [29] Wu Q, Liu W Y, Yang Y H. Time series online prediction algorithm based on least squares support vector machine [J]. Journal of Central South University of Technology, 2007, 14(3): 442-446.
- [30] 孙荣基, 赵松, 张小琴, 等. 一种基于 BP 神经网络改进算法的 $PM_{2.5}$ 预测方法[J]. 四川环境, 2015, 34(4): 85-90.
- Sun R J, Zhao S, Zhang X Q, *et al.* An improved $PM_{2.5}$ prediction model based on BP neural network [J]. Sichuan Environment, 2015, 34(4): 85-90.
- [31] 李志生, 梁锡冠, 金宇凯, 等. 基于树模型的北京市 $PM_{2.5}$ 预测效果对比分析[J]. 环境工程, 2021, 39(6): 106-113.
- Li Z S, Liang X G, Jin Y K, *et al.* A comparative study on

- edictive effect of PM_{2.5} in Beijing based on tree models [J]. Environmental Engineering, 2021, **39**(6): 106-113.
- [32] Wei J, Li Z Q, Pinker R T, *et al.* Himawari-8-derived diurnal variations in ground-level PM_{2.5} pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM) [J]. Atmospheric Chemistry and Physics, 2021, **21**(10): 7863-7880.
- [33] Vapnik V N. The nature of statistical learning theory (2nd ed.) [M]. New York: Springer, 2000.
- [34] Lu H, Xie M, Liu X R, *et al.* Adjusting prediction of ozone concentration based on CMAQ model and machine learning methods in Sichuan-Chongqing region, China [J]. Atmospheric Pollution Research, 2021, **12**(6), doi: 10.1016/j.apr.2021.101066.
- [35] Breiman L. Random forests [J]. Machine Learning, 2001, **45**(1): 5-32.
- [36] Zhang T N, He W H, Zheng H, *et al.* Satellite-based ground PM_{2.5} estimation using a gradient boosting decision tree [J]. Chemosphere, 2021, **268**, doi: 10.1016/j.chemosphere.2020.128801.
- [37] Ke G L, Meng Q, Finley T, *et al.* LightGBM: a highly efficient gradient boosting decision tree [A]. In: Proceedings of the 31st International Conference on Neural Information Processing Systems [C]. Long Beach: Curran Associates Inc., 2017. 3149-3157.
- [38] Zou Z Y, Wang L, Chen J, *et al.* Research on peanut variety classification based on hyperspectral image [J]. Food Science and Technology, 2022, **42**, doi: 10.1590/1st.18522.
- [39] 方欣, 王敏, 胡顺星. 合肥梅雨季节大气污染物分布特征的激光雷达探测 [J]. 中国激光, 2019, **46**(1), doi: 10.3788/CJL201946.0110003.
- Fang X, Wang M, Hu S X. Distribution characteristics of atmospheric pollutions in Meiyu season observed by Lidar over Hefei [J]. Chinese Journal of Lasers, 2019, **46**(1), doi: 10.3788/CJL201946.0110003.
- [40] Yafouz A, AlDahoul N, Birima A H, *et al.* Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction [J]. Alexandria Engineering Journal, 2022, **61**(6): 4607-4622.
- [41] 苏筱倩, 安俊琳, 张玉欣, 等. 支持向量机回归在臭氧预报中的应用 [J]. 环境科学, 2019, **40**(4): 1697-1704.
- Su X Q, An J L, Zhang Y X, *et al.* Application of support vector machine regression in ozone forecasting [J]. Environmental Science, 2019, **40**(4): 1697-1704.
- [42] 潘航. 近 60a 来南京季节变化特征分析 [J]. 气象科学, 2011, **31**(6): 742-746.
- Pan H. The analysis of seasonal transition characteristics in Nanjing in recent 60 years [J]. Journal of the Meteorological Sciences, 2011, **31**(6): 742-746.
- [43] 蒲茜. 长三角高温热浪期间臭氧的形成机理及辐射效应研究 [D]. 南京: 南京大学, 2017.
- Pu X. Analysis of elevated ozone formation mechanism and radiative effect during heat waves in Yangtze River Delta [D]. Nanjing: Nanjing University, 2017.
- [44] Mohan S, Saranya P. A novel bagging ensemble approach for predicting summertime ground-level ozone concentration [J]. Journal of the Air & Waste Management Association, 2019, **69**(2): 220-233.
- [45] Li J M, Xu G Y, Cheng X Z. Combining spatial pyramid pooling and long short-term memory network to predict PM_{2.5} concentration [J]. Atmospheric Pollution Research, 2022, **13**(3), doi: 10.1016/j.apr.2021.101309.

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)