张煌帆,谢郑一,温晓娟,等.基于可解释机器学习的交通 $CO_2$ 和 $O_3$ 的驱动机制 [J]. 中国环境科学, 2025,45(11):6058-6067.
Zhang H F, Xie Z Y, Wen X J, et al. Driving mechanisms of traffic-related $CO_2$ and $O_3$ based on interpretable machine learning [J]. China Enviromental Science, 2025,45(11):6058-6067.

# 基于可解释机器学习的交通 $CO_2$ 和 $O_3$ 的驱动机制

张煌帆,谢郑一,温晓娟,王 硕,翁大维,徐艺诺,胡喜生,张兰怡* (福建农林大学交通与土木工程学院.福建 福州 350108)

**摘要**：以华东地区为研究区域,围绕交通产生的 $CO_2$ 以及其对 $O_3$ 的贡献(下文简称“$O_3$”)构建 15 个驱动因素指标体系.首先,采用 Lasso 回归模型筛选关键变量;其次,引入 RF(随机森林)与 XGBoost(极端梯度提升)模型开展特征重要性分析,并结合 SHAP 方法对各驱动因素的贡献度进行可视化解释,构建了系统的驱动因素识别与排序分析框架.研究表明,华东地区交通 $CO_2$ 排放总量持续增长,至 2022 年,上海、江苏与山东三省合计占比达 62.18%; $O_3$ 浓度高值区域主要集中于东北部.其中安徽省 $O_3$ 浓度增幅高达 92.25%.在模型性能方面,RF 模型在变量解释力和稳定性上优于 XGBoost.其中,其 $CO_2$ 排放与 $O_3$ 浓度的 $R^2$ 均大于 0.8,MSE 均为 0.001,MAE 均小于 0.1.SHAP 分析结果显示,交通 $CO_2$ 排放的主要驱动因素包括人口密度、货运周转量、第三产业产值和城镇化水平;而 $O_3$ 浓度则受人口数量、交通运输业产值、载货汽车数量和城市道路长度等因素影响.

**关键词**：交通 $CO_2$ 排放；$O_3$ 浓度；机器学习；重要性评价；驱动因素
**中图分类号**：X734;U491.9 **文献标识码**：A **文章编号**：1000-6923(2025)11-6058-10

**Driving mechanisms of traffic-related $CO_2$ and $O_3$ based on interpretable machine learning.** ZHANG Huang-fan, XIE Zheng-yi, WEN Xiao-juan, WANG Shuo, WENG Da-wei, XU Yi-nuo, HU Xi-sheng, ZHANG Lan-yi* (College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou 350108, China). *China Enviromental Science*, 2025,45(11)：6058~6067

**Abstract**：This study focuses on East China, establishing a 15-indicator system to investigate the driving mechanisms of traffic-related $CO_2$ emissions and their contributions to ozone ($O_3$) concentrations. First, the Lasso regression model is employed to screen key variables. Then, feature importance is assessed using Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models, with the SHAP method used to visualize the marginal contributions of each factor. A systematic framework for identifying and ranking driving factors is thus constructed. Results indicate that traffic-related $CO_2$ emissions in East China have shown a continuous upward trend, with Shanghai, Jiangsu, and Shandong together accounting for 62.18% of total emissions by 2022. High $O_3$ concentration zones are primarily located in the northeastern subregion, with Anhui Province experiencing a remarkable increase of 92.25%. In terms of model performance, the RF model outperforms XGBoost in both explanatory power and stability, achieving $R^2$ values above 0.8 for both $CO_2$ and $O_3$, with MSE values of 0.001 and MAE values below 0.1. SHAP analysis reveals that the main drivers of $CO_2$ emissions include population density, freight turnover, tertiary industry output, and urbanization level; while $O_3$ concentrations are primarily influenced by total population, transport sector output, the number of freight vehicles, and urban road length.

**Key words**：transportation emissions；ozone concentration；machine learning；importance assessment；driving factors

气候变化与环境污染已成为全球可持续发展的主要障碍,交通运输业作为能源消耗和污染物排放的重要来源,其环境影响日益突出[1-4].华东地区位于中国东部沿海,是我国经济最发达、交通运输最活跃的区域之一,依托完善的交通网络、优越的地理区位和多项优惠政策.实现了交通业的快速扩张.然而,随之而来的能源消耗与大气污染问题亦日益严重,亟需从源头进行机制解析与治理路径研究.因此,识别华东地区交通 $CO_2$ 排放以及与交通相关的 $O_3$ 浓度(以下简称“交通 $O_3$”)的关键驱动因素,对于推动区域交通行业绿色低碳发展具有重要现实意义.

明确不同省份交通 $CO_2$ 排放与 $O_3$ 浓度的主要驱动因素,是构建区域差异化减排机制和制定精准治理策略的基础依据.当前,在交通 $CO_2$ 排放研究方面,已有研究主要运用标准差椭圆法、增量动力分析

**收稿日期**：2025-04-27
**基金项目**：福建省自然科学基金项目(2023J01475);福建省社科规划项目(FJ2022B065);福建农林大学“杰出青年科研人才”计划项目(社会科学类)(xjq2020S4);福建农林大学科技创新专项(KFb22101XA);国家级大学生创新创业训练计划项目(202410389030,202410389034)
* 责任作者, 副教授, lyzhang@fafu.edu.cn

和全局莫兰指数等方法分析多尺度下 $CO_2$ 排放影响,并揭示空间分异模式与动态耦合机制[5-6].基于前期研究经验,课题组已应用对数平均迪氏指数法、广义对数迪氏分解法、时空地理加权回归和地理探测器等方法广泛探讨了交通领域 $CO_2$ 排放的驱动因素,并为不同区域的减排路径提供理论参考[7-12];在 $O_3$ 浓度研究方面,学者们常采用如 WRF−Chem 和 CMAQ 等大气输送模型定量分析不同地区 $O_3$ 对关键参数的敏感性[13-14].相较于传统方法,近年来 RF(随机森林)、XGBoost(极端梯度提升)、神经网络等机器学习算法凭借其在相关性识别、变量筛选和模型优化方面的优势,在交通碳排放与大气污染物研究中展现出良好适用性[15-17].然而,现有研究多聚焦于单一污染因子或单一算法,缺乏碳排放与污染物协同视角下、多方法交叉验证下的关键因素识别研究.值得关注的是,$O_3$ 污染不仅危害人体健康,还会通过降低 $CO_2$ 同化速率,削弱植物的碳吸收能力,从而降低生态系统的净初级生产力[18-19],从而加剧温室气体与污染物的耦合效应. 因此,识别交通 $CO_2$ 排放与 $O_3$ 浓度的核心驱动因素,对于实现“减污降碳”协同增效目标具有重要科学价值.

综上.本文以华东地区为研究对象,围绕交通 $CO_2$ 排放与 $O_3$ 浓度两类关键污染指标,构建 Lasso 回归筛选变量,采用 RF 与 XGBoost 模型量化变量重要性,并引入 SHAP 方法开展可解释性分析.

## 1　材料与方法

### 1.1　驱动因素选取

**表 1　华东地区交通 $CO_2$ 和 $O_3$ 浓度驱动因素**

Table 1　Driving factors of traffic−related $CO_2$ and $O_3$ concentrations in East China

| 序号 | 驱动因素 | 符号 | 序号 | 驱动因素 | 符号 |
|---|---|---|---|---|---|
| 1 | 人口数量 | P | 9 | 城市道路长度 | URL |
| 2 | 地区生产总值 | GDP | 10 | 客运周转量 | PT |
| 3 | 人均国内生产总值 | PGDP | 11 | 货运周转量 | FT |
| 4 | 第三产业生产总值 | TGDP | 12 | 绿化面积 | GA |
| 5 | 民用车数量 | AN | 13 | 城镇化水平 | UL |
| 6 | 载客汽车数量 | PSV | 14 | 交通运输业产值 | TP |
| 7 | 载货汽车数量 | T | 15 | 人口密度 | PD |
| 8 | 社会消费品零售总额 | TSG | | | |

交通运输业的 $CO_2$ 排放和 $O_3$ 浓度受社会经济、交通运输及环境等多重因素的综合驱动,为系统识别二者的驱动机制,本文在梳理国内外有关 $CO_2$ 排放和 $O_3$ 浓度的研究成果的基础上[8,20-24],初步筛选出 15 项具有代表性的潜在驱动因素.变量名称及其符号见表 1.

### 1.2　数据预处理

在模型构建之前,对原始数据进行标准化处理是提升模型学习效率与预测精度的重要前提.标准化的核心目标在于将各特征变量统一至相同的数值尺度,从而消除因量纲差异和取值范围不一致对模型训练过程造成的干扰.采用 Min−Max 归一化方法对原始数据进行预处理,该方法不仅有助于简化数据结构、提高模型的收敛速度,还能有效减弱异常值的不利影响,增强模型的泛化能力[25],表达式如式(1)所示:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

式中:$X$ 是原始数据;$X'$是归一化后的数据;$X_{\min}$ 和 $X_{\max}$ 分别是数据的最小值和最大值.

### 1.3　研究方法

**1.3.1　交通 $CO_2$ 排放量测算**　在本课题组前期研究基础上,采用“自上而下”方法对华东地区各省(市)的交通 $CO_2$ 排放量进行量化核算[10].该方法通过交通能源消费数据与排放因子相结合[9-11,26-28].其中,能源消费数据来源于各省(市)统计年鉴;排放因子数据来源于文献[9−11,26−28];具体计算见式(2):

$$T_p = \sum_{i=1}^{n} E_i \cdot \mathrm{EF}_i \tag{2}$$

式中:$T_P$ 为地区 $P$ 的交通运输 $CO_2$ 排放量;$E_i$ 为第 $i$ 种化石燃料的能源消耗量;$\mathrm{EF}_i$ 为第 $i$ 种化石燃料的 $CO_2$ 排放因子.不同能源类型的 $CO_2$ 排放因子计算如下:

$$\mathrm{EF}_i = \mathrm{NCV}_i \cdot 10^{-9} \cdot \mathrm{CC}_i \cdot \mathrm{OF}_i \cdot 10^{-3} \times 44/12 \tag{3}$$

式中:$\mathrm{NCV}_i$ 为第 $i$ 种化石燃料的平均低位发热量;$\mathrm{CC}_i$ 为第 $i$ 种化石燃料的单位热值含碳量;$\mathrm{OF}_i$ 为第 $i$ 种化石燃料的碳氧化率;44/12 为 $CO_2$ 与碳的相对质量之比.

**1.3.2　驱动因素分析与筛选**　为提升模型预测精度并减少冗余信息干扰,在建模前对初步筛选的 15 个潜在驱动因素进行变量分析与特征筛选.首先,采用 Pearson 相关系数法评估各变量与交通 $CO_2$ 排放及 $O_3$ 浓度之间的相关性,其系数绝对值越接近于 1,

表示变量间相关性越强.

为进一步筛除冗余变量并缓解多重共线性问题,本文引入 Lasso 回归模型进行特征选择.通过引入 $L_1$ 正则化项对模型复杂度进行约束,在实现变量压缩的同时提升模型稳定性.具体流程包括对数据进行归一化处理,将驱动因素作为解释变量、交通 $CO_2$ 排放与 $O_3$ 浓度作为因变量,利用交叉验证确定最优正则化参数 λ,并完成模型拟合.最终保留回归系数大于零的变量作为主要驱动因素.其优化目标函数为:

$$T = \min\left\{\left\|\boldsymbol{X}_W - y\right\|_2^2 + \lambda\left\|w\right\|_1\right\} \tag{4}$$

式中:$\boldsymbol{X}$ 为特征矩阵;$y$ 为预测值;$\lambda$为控制正则化强度的超参数;$w$ 为特征的回归系数.

1.3.3 模型构建 为进一步分析各驱动因素与交通 $CO_2$ 排放及 $O_3$ 浓度之间的关系,分别构建基于 RF 和 XGBoost 两种集成学习算法的预测模型.

(1)RF 模型由多个独立的决策树组成.采用 Bootstrap 采样与 Bagging 聚合策略.通过对多个子模型的预测结果加权平均以提高整体模型的鲁棒性与泛化能力[29-30].

(2)XGBoost 模型在梯度提升框架下迭代构建多个决策树.利用一阶与二阶导数信息对损失函数进行优化.在每轮迭代中修正前一轮残差误差.最终加权集成输出预测结果[31-32].

1.3.4 SHAP 可解释分析 为增强模型输出的可解释性,本文引入 SHAP 方法对模型训练结果进行特征贡献度分析.SHAP 基于博弈论中 Shapley 值的思想.能够量化每个输入变量在不同样本上的边际贡献,实现全局与局部层面的统一解释[33].与传统特征重要性评估方法相比,SHAP 可更有效识别变量间的非线性交互关系.适用于复杂机器学习模型的可视化解释,近年来已被广泛应用于环境科学与智能建模领域.

### 1.4 数据来源

交通能源消费数据来源于华东地区各省(市)历年《中国能源统计年鉴》.涵盖原煤、汽油、煤油、柴油、燃料油、液化石油气与天然气等七类交通能源.碳排放因子所需参数(包括平均低位发热量、单位热值含碳量与碳氧化率参照)《综合能耗计算通则》、《省级温室气体清单编制指南》、《中国陆上交通运输企业温室气体排放核算方法与报告指南》等权威文件,并结合本团队前期研究成果进行统一修正[11].社会经济与交通运输等驱动因素数据来源于各省(市)历年统计年鉴.此外,考虑到交通运输业、仓储和邮电业在统计口径上存在交叉,且后者在总排放中占比较小,故将三者合并处理以保证数据的完整性与可比性[34].$O_3$ 浓度数据主要来源于中国环境监测总站发布的各省(市)实时监测数据,在此基础上,结合相关研究文献[35],提取并估算出交通源相关的年均 $O_3$ 浓度水平,覆盖整个研究时间段的区域尺度特征.

## 2 结果与讨论

### 2.1 交通 $CO_2$ 排放测算结果与空气质量分析

华东地区 2003~2022 年交通 $CO_2$ 排放量及 $O_3$ 浓度估算结果如图 1 所示.图 1(a)~(d)展示了 2014~2022 年各年份下交通 $CO_2$ 排放的空间分布演变特征.可以看出,交通 $CO_2$ 排放总量呈持续上升趋势,尤其在山东、江苏和上海等东部沿海省份表现最为显著.其中,2022 年上海交通 $CO_2$ 排放量达到全区域最高值,累计贡献占比达 22.44%,其次为山东(21.99%)与江苏(17.75%),三省合计占比高达 62.18%.相比之下,福建与江西受限于地形地貌、工业基础与交通规模.$CO_2$ 排放水平相对较低.图 1(e)~(h)进一步揭示了 $O_3$ 浓度的空间演化态势.高值区集中于华东北部.尤其是安徽省,其 $O_3$ 浓度在 2014~2022 年间增幅高达 92.95%(图 2(b)),成为增长最快的省份.该现象的成因可能包括:① 人口密度较高.使得民用车保有量增加.2022 年,山东、江苏、安徽与上海四省(市)的民用汽车保有量合计占全区 58%.其中,安徽省在四省(市)的民用车数量总增长速度最为迅速,为 61.29%;② 区域经济与交通网络发达,导致交通活动频繁,从而增加 $O_3$ 前体物排放.安徽省作为华东地区交通枢纽之一,网络布局不断完善,全省高速公路通车总里程较 2012 年增加 60%,交通流量随着显著提升.

图 2(a)显示了各省(市)交通 $CO_2$ 排放的累计占比,其中上海、山东和江苏三地处于领先地位,结构性贡献突出;圆形图反映 $O_3$ 浓度的区域分布特征,整体趋势与 $CO_2$ 排放具有一定空间一致性,但不完全重合.图 2(b)进一步量化了各省(市)2003~2022 年交

通 $CO_2$ 排放与 $O_3$ 增长率,结果显示 $O_3$ 浓度整体增速普遍高于 $CO_2$,尤其在安徽、江苏和山东等北部省份表现最为明显,体现出 $O_3$ 污染问题在华东地区的加剧趋势.福建和浙江则表现出较为温和的增长态势.其中,浙江通过推进产业结构优化与节能减排政策,有效控制了 $O_3$ 浓度的过快上升;福建因其山地丘陵面积占比高达 74.63%,有利于污染物的扩散与稀释,成为区域内 $O_3$ 浓度最低的省份.

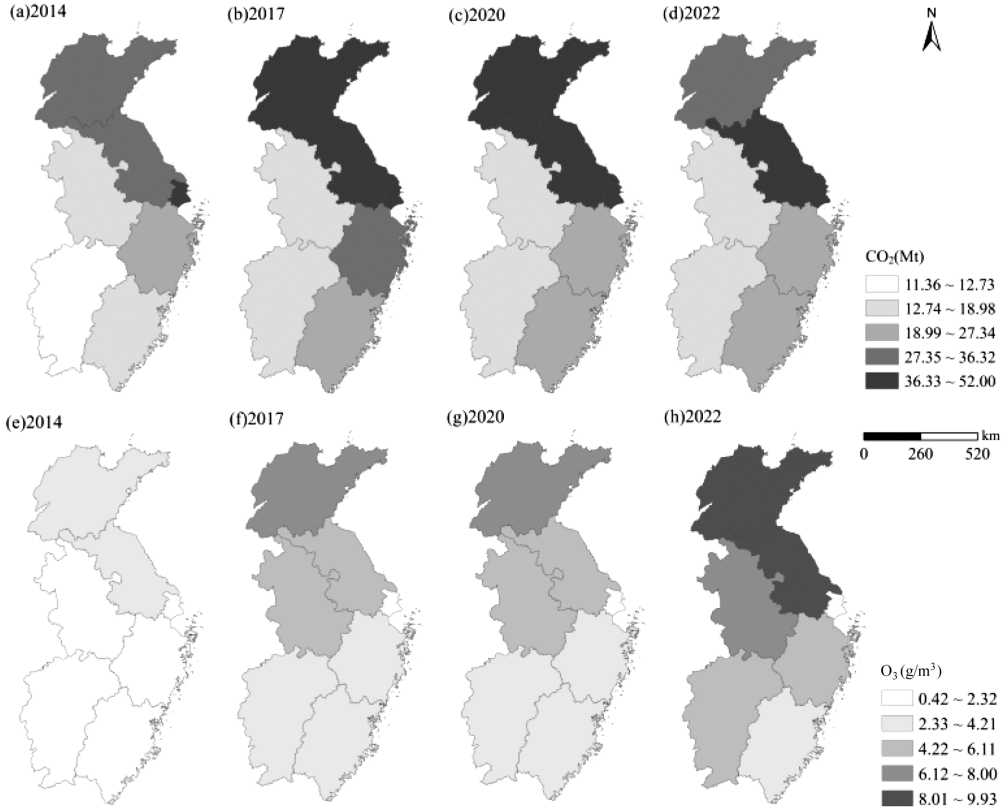



图 1　华东地区交通 $CO_2$ 和 $O_3$ 浓度空间分布(2014~2022 年)

Fig.1　Spatial Distribution of Traffic-Related $CO_2$ Emissions and $O_3$ Concentrations in East China (2014~2022)

审图号:GS(2019)1822

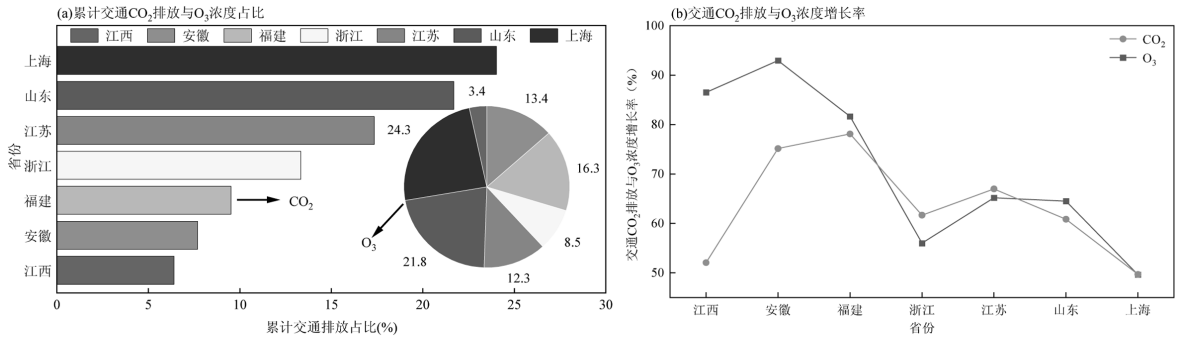



图 2　华东地区交通 $CO_2$ 和 $O_3$ 浓度结构与增长趋势

Fig.2　Structural composition and growth trends of traffic-related $CO_2$ emissions and $O_3$ concentrations in East China

## 2.2　驱动因素分析与筛选

**2.2.1　变量相关性分析**　为探究驱动因素与交通 $CO_2$ 排放与 $O_3$ 的相关性系数,采用 Pearson 相关系数方法来衡量这些因素之间的关系.系数的取值范围介于-1~1 之间,绝对值越接近 1,表示变量间的相关性越显著.相关性分析结果如图 3 所示,其中图 3(a)对应交通 $CO_2$ 排放,图 3(b)对应 $O_3$ 浓度.图中点的大小表示相关系数的绝对值,颜色深浅则反映相关方向与强度.

结果显示, 交通 $CO_2$ 排放和 $O_3$ 与 GDP、TGDP、

GA、TP、URL、TSG 等因素相关系数为 0.5~0.8.其中,GA 与 $CO_2$ 相关性最强(0.64),PSV 与 $O_3$ 极强相关性(0.8),表明它们之间存在极强的正相关性,表明随着经济活动与交通流量的增长,交通 $CO_2$ 排放和 $O_3$ 浓度亦呈上升趋势,反映了交通系统对区域排放水平的放大效应.

此外, $CO_2$ 与 PGDP、FT、PT、UL 和 PD 呈正相关性,但 $O_3$ 与上述因素的相关性则相对较弱或呈负相关.其中 FT 与 $CO_2$ 的相关系数为 0.78,而与 $O_3$ 的相关系数为−0.35;UL 与 $CO_2$ 的相关系数为 0.71,与 $O_3$ 的相关系数为−0.35,出现差异可能与 $O_3$ 浓度受到气象条件、地形因素以及光化学反应过程等多重因素共同影响有关.

需要强调的是,Pearson 相关性分析反映的是变量间的相关关系,并不代表因果关系.为进一步提升模型的解释精度与稳健性,后续将引入 Lasso 回归模型进行变量选择与降维处理,以剔除冗余信息并缓解多重共线性问题.

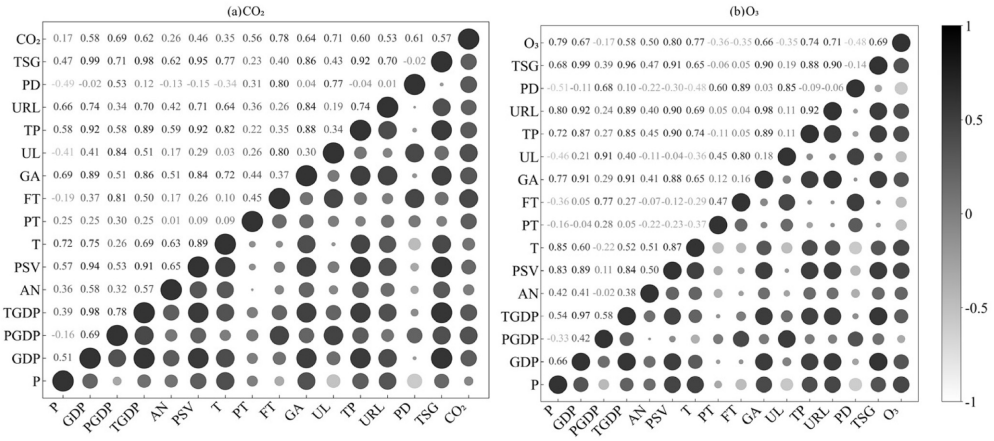



图 3 交通 $CO_2$ 和 $O_3$ 浓度与驱动因素相关系数热力

Fig.3 Heatmaps of correlation coefficients between traffic-related $CO_2$/$O_3$ concentrations and driving factors

2.2.2 Lasso 回归变量筛选 通过 Lasso 回归模型对 15 个驱动因素进行变量压缩与筛选后,交通 $CO_2$ 排放与 $O_3$ 浓度在关键驱动因素上呈现一定的共性与差异性,具体如图 4 所示.

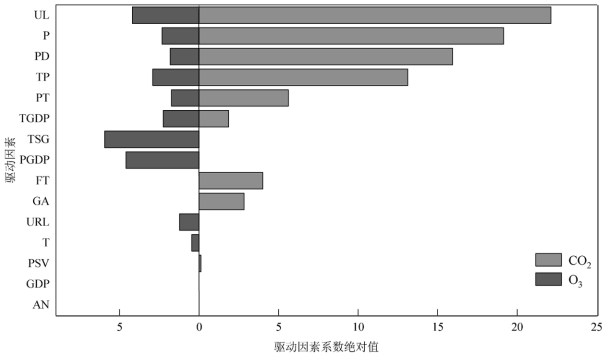



图 4 基于 Lasso 回归模型的交通 $CO_2$ 和 $O_3$ 浓度驱动因素筛选结果

Fig.4 Selection results of driving factors for traffic-related $CO_2$ and $O_3$ concentrations based on the Lasso regression model

在共性方面,交通 $CO_2$ 排放与 $O_3$ 均筛选出 UL、P、PD、TP、PT 和 TGDP 等为共同驱动因素,表明这些变量在解释温室气体与 $O_3$ 污染方面均具有较强的解释力.

在差异性方面,交通 $CO_2$ 排放模型中独有的关键变量包括 FT、GA 和 PSV;而在 $O_3$ 浓度模型中,仅保留了 TSG、PGDP、URL 和 T,说明 $CO_2$ 排放和 $O_3$ 浓度在驱动机制上具有一定差异性,尤其在反映排放结构和空间分布维度方面.此外,GDP 和 AN 在两个模型中均未被选入,说明其对目标变量的预测贡献较小.具体而言,TP 是直接反应交通运输强度的重要指标,是 GDP 的重要组成结构,并且 TP 与 GDP 之间存在显著的共线性特征(相关系数>0.86),在 Lasso 模型中优先保留 TP 而剔除 GDP,有助于提升模型稳定性与解释能力.另一方面,随着新能源汽车的大力推广,传统燃油汽车在交通排放中的占比逐年下降,导致 AN 变量对排放的解释能力逐渐减弱,最终未被纳入模型.

## 2.3 模型构建与评价

2.3.1　模型对比分析　为评估不同机器学习算法在交通 $CO_2$ 排放与 $O_3$ 浓度中的性能表现,基于华东地区各省(市)数据,构建 RF 和 XGBoost 两类预测模型.具体步骤如下:①以全部 15 个驱动因素作为自变量;②采用 Lasso 回归模型筛选后的主要变量作为输入特征.模型训练过程中,样本数据按 75%:25%的比例划分为训练集与测试集,并通过网格搜索结合交叉验证对模型的超参数进行优化,提高模型的鲁棒性与泛化能力.模型性能评估指标包括 $R^2$(拟合优度)、MSE(均方误差)和 MAE(均绝对误差).其中,$R^2$ 越接近 1 表明模型的解释能力越强,MSE 和 MAE 越小则代表预测误差越低,模型性能越稳定.RF 和 XGBoost 模型在不同预测任务下的表现如表 2 所示.

从表 2 可知,在交通 $CO_2$ 排放预测方面,RF 模型表现出显著优势,其 $R^2$ 达到 0.96,明显优于 XGBoost 模型的 0.91,同时 MSE 和 MAE 分别为 0.002 和 0.028,均低于 XGBoost 的 0.004 和 0.036,表明其在拟合精度与稳定性方面更具优势.在 $O_3$ 浓度预测方面,两种模型的 $R^2$ 均为 0.84,整体拟合效果相当.然而,RF 模型在误差控制方面依然略胜一筹,其 MSE 为 0.008,MAE 为 0.071,分别低于 XGBoost 模型的 0.009 和 0.073,进一步验证了其泛化能力与稳健性.

**表 2　不同机器学习模型在交通相关 $CO_2$ 和 $O_3$ 浓度方面的性能比较**

Table 2　Performance comparison of different machine learning models in estimating traffic-related $CO_2$ and $O_3$ concentrations

| 模型 | 交通 $CO_2$/$O_3$ | $R^2$ | MSE | MAE |
|---|---|---|---|---|
| RF | $CO_2$ | 0.96 | 0.002 | 0.028 |
| | $O_3$ | 0.84 | 0.008 | 0.071 |
| XGBoost | $CO_2$ | 0.91 | 0.004 | 0.036 |
| | $O_3$ | 0.84 | 0.009 | 0.073 |

综上,RF 模型在研究数据集上的表现优于 XGBoost 模型,无论是拟合度还是误差控制方面,均具有更高可靠性.因此,选定 RF 模型作为最终的预测模型,模型拟合效果如图 5 所示,其中图 5(a)为 $CO_2$ 排放模型结果,图 5(b)为 $O_3$ 浓度模型结果.

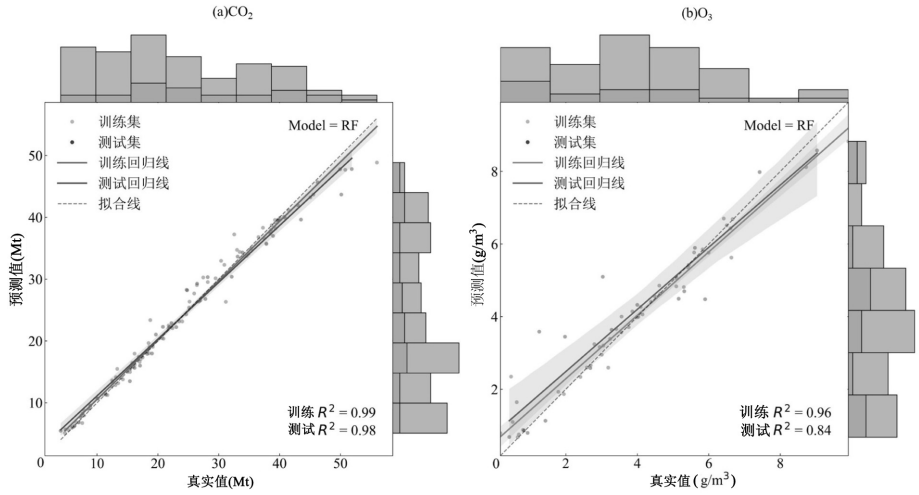



图 5　RF 模型在交通相关 $CO_2$ 与 $O_3$ 浓度中的拟合效果

Fig.5　Fitting performance of the RF model for traffic-related $CO_2$ and $O_3$ concentrations

2.3.2　华东地区 RF 模型精度对比分析　为进一步评估 RF 模型在区域层面的适用性,本文对华东地区的交通 $CO_2$ 排放与 $O_3$ 浓度分别进行建模,结果如表 3 所示.除江西省 $O_3$ 浓度模型外,其余各地区的 $R^2$ 值均超过 0.75,在 0.76~0.98 之间,说明 RF 模型在大多数省(市)中具备较强的解释能力与稳定性.从 $CO_2$ 排放模型结果来看,江苏省表现最为优越,$R^2$ 高达 0.98,表明模型对其交通 $CO_2$ 排放演化趋势具备极高的适应性与拟合能力.安徽、福建、山东等地模型也均表现良好,$R^2$ 均在 0.95 以上,验证了 RF 模型在多区域背景下的普适性.在 $O_3$ 浓度方面,安徽省模型精度最高,$R^2$ 达到 0.98,表明 RF 模型能够较为准确地刻画其 $O_3$ 浓度变化规律.福建和山东的 $R^2$ 分别为 0.92 和 0.94,模型效果亦处于较高水平.相较之下,江西($R^2$=

0.54)、上海($R^2$=0.76)和浙江($R^2$=0.77)等地的模型精度略显不足,尤其是江西,其模型解释力明显偏低,可能与其复杂的地理与气候条件有关.江西地势呈三山环绕、中部丘陵与平原交错分布,交通排放空间分布较为离散,加之气象条件差异大,导致模型对 $O_3$ 浓度的预测难度较大.综上,RF 模型在华东大部分省(市)的交通 $CO_2$ 排放与 $O_3$ 浓度中表现稳定可靠.

**表 3 各省(市)交通相关 $CO_2$ 和 $O_3$ 浓度 RF 模型精度对比**

Table 3 Comparative accuracy of the RF model for traffic-related $CO_2$ and $O_3$ concentrations across provinces (cities)

| 省份(城市) | $CO_2$ ($R^2$) | MSE | MAE | $O_3$($R^2$) | MSE | MAE |
|---|---|---|---|---|---|---|
| 安徽 | 0.95 | 0.01 | 0.09 | 0.98 | 0.08 | 0.25 |
| 福建 | 0.95 | 0.01 | 0.08 | 0.92 | 0.02 | 0.11 |
| 江苏 | 0.98 | 0.006 | 0.06 | 0.85 | 0.09 | 0.23 |
| 江西 | 0.96 | 0.006 | 0.05 | 0.54 | 0.05 | 0.2 |
| 山东 | 0.96 | 0.006 | 0.07 | 0.94 | 0.05 | 0.17 |
| 上海 | 0.82 | 0.044 | 0.14 | 0.76 | 0.03 | 0.3 |
| 浙江 | 0.87 | 0.032 | 0.12 | 0.77 | 0.01 | 0.1 |

## 2.4 变量重要性分析

为提升模型可解释性,本文基于已构建的 RF 模型,引入 SHAP 方法对交通 $CO_2$ 排放与 $O_3$ 浓度的驱动机制进行量化分析.当 SHAP 值为正时,表示该变量对预测结果具有正向促进作用;反之,则表示具有负向抑制作用.图 6(a)与(b)分别展示了交通 $CO_2$ 排放与 $O_3$ 浓度模型中各变量的 SHAP 值分布及变量的重要性排序.图中每个散点代表一个样本,点的颜色反映特征取值大小.

**2.4.1 变量重要性排序与差异性分析** 在交通 $CO_2$ 排放模型中,最具影响力的变量依次为:PD、FT、TGDP、UL、PT、P、TP、GA 和 PSV,其中,PD 的 SHAP 值最高,表明对模型结果产生最大的影响.而 TP、GA 和 PSV 的 SHAP 值较低,三者累加的 SHAP 平均绝对值值仅为 0.93,说明对预测结果影响较小.在 $O_3$ 浓度模型中,排名前列的变量重要性排序为:P、TP、T、URL、PT、UL、PD、PGDP、TSG 和 TGDP,其中,P、TP、T、URL、PT 和 UL 的预测贡献度较高,而 PD、PGDP、TSG 和 TGDP 的大多数 SHAP 值均在 0 的附近,说明对预测贡献度极低.

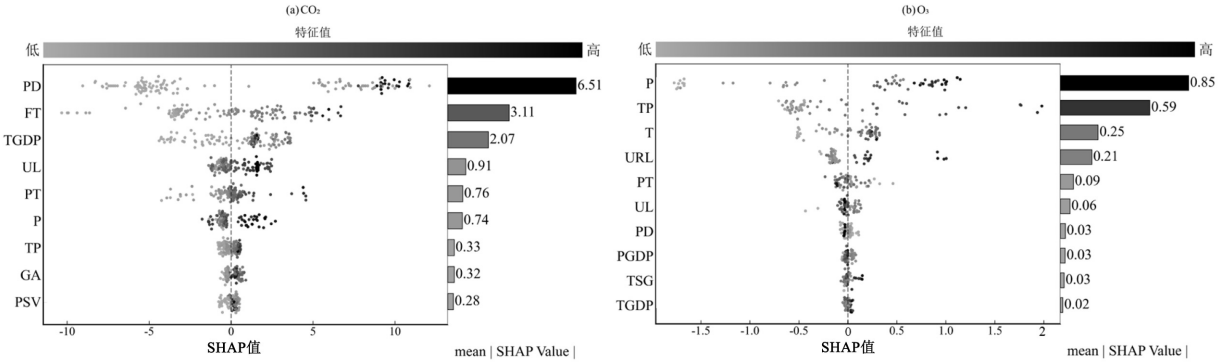



图 6 基于 SHAP 方法的交通 $CO_2$ 与 $O_3$ 浓度驱动因素重要性排序

Fig.6 Feature importance ranking of driving factors for traffic-related $CO_2$ and $O_3$ concentrations based on the SHAP method

**2.4.2 共同变量的双重影响机制** 在两个模型中,UL、P 和 PT 都表现出较高的 SHAP 值,说明它们对“减污降碳”政策具有显著的影响,需要特别关注.其中,UL 与交通 $CO_2$ 排放存在相关关系,即城镇化水平越高,交通 $CO_2$ 排放也越高.具体而言,城镇化水平高的省份往往具有经济发达、基础设施完善和能源需求增加等特点,从而导致交通排放增加;而 UL 与 $O_3$ 浓度呈相反的趋势,即当城镇化水平越低时,$O_3$ 浓度越高,反之亦然.其次,P 在所有驱动因素中对 $CO_2$ 排放和 $O_3$ 浓度的影响分别为第 6 和第 1,其可能原因是机动车保有量大,使得交通流量也相应的过得变大,而大部分车辆为为低排放车辆,因此交通运输排放量也会较低.最后,PT 与交通 $CO_2$ 排放主要为正相关关系并排名第 5,但其与 $O_3$ 浓度主要为负相关关系,在 $O_3$ 的 SHAP 模型中排名第 5,出现这样的差异主要是因为 $O_3$ 主要是二次反应生成物,主要受气象条件影响,局部气象条件对 $O_3$ 浓度变化及前体物浓度有重要影响.

**2.4.3 差异性变量的成因推测** 在交通 $CO_2$ 排放模型中,FT、GA 与 PSV 为具有代表性的差异性变

量.FT 的平均 SHAP 值为 3.11,表明其是模型中最具影响力的变量之一.然而,FT 对 $CO_2$ 排放的预测贡献存在一定的不确定性,这可能与货运结构不均衡以及车辆满载率偏低等因素密切相关.相比之下,GA 和 PSV 的 SHAP 值相对较低,但它们仍在一定程度上反映了区域绿色基础设施建设水平和传统燃油车辆保有量对 $CO_2$ 排放的辅助性影响.

在 $O_3$ 浓度模型中,T、URL、PGDP 和 TSG 展现出相对突出的解释力.T 的平均 SHAP 值为 0.25,其影响机制较为复杂.在部分区域,较低的 T 值反而伴随较高的 $O_3$ 浓度,这可能源于运输路线不合理、车辆运行效率较低所导致的污染物滞留和局部累积效应.URL 与 $O_3$ 浓度呈现正相关关系,其 SHAP 值为 0.21.这表明在 URL 较高的区域,机动车通行频率往往更高, $O_3$ 前体物排放更为密集,从而加剧了 $O_3$ 污染水平.

## 3　讨论

3.1　“减碳”和“控臭氧”的动态平衡性.在政策制定时,应充分考虑城镇化水平、人口数量、人口密度、交通运输业产值、客运周转量和第三产业生产总值等驱动因素,这些因素对交通 $CO_2$ 排放与 $O_3$ 浓度均有显著影响.值得注意的是,从 SHAP 分析可知,城镇化水平、人口数量和客运周转量影响尤为突出,但其作用趋势并不相同.因此,在针对这些因素制定相关政策时,需特别注意其动态平衡性.

3.2　可持续城镇化的有效性.城镇化水平与交通 $CO_2$ 排放与 $O_3$ 浓度的关系较为复杂,城镇化水平高既会增加交通 $CO_2$ 排放,然而,其也具备降低 $O_3$ 浓度的潜力.在城镇化进程中,需要平衡经济发展与环境保护之间的关系.因此,应优化城镇交通布局,合理控制城镇建筑总规模,同时加大对城市交通基础设施和智能交通系统的建设投入,以改善城市环境,推动城镇可持续发展.

3.3　社会民众参与的重要性.政府需积极引导并采取措施,可通过鼓励市民选择绿色低碳的出行方式,增加非机动化、公共化及共享化出行的比例.特别是在污染集中的区域,应优化公共交通网络布局,加密公交线路,提高运营频次,从而促进绿色出行方式的普及.

3.4　低排放客运的重要性.加快构建绿色出行体系,因地制宜构建以城市轨道交通和快速公交为骨干、常规公交为主体的公共交通出行体系,强化“轨道+公交+慢行”网络融合发展.深化国家公交都市建设,提升城市轨道交通服务水平,持续改善公共交通出行体验.开展绿色出行创建行动,改善绿色出行环境,提高城市绿色出行比例.完善城市慢行交通系统,提升城市步行和非机动车的出行品质,构建安全、连续和舒适的城市慢行交通体系.

3.5　增加 $O_3$ 的拟合效果.由于 $O_3$ 是二次生成污染物,在很大程度上受到气象条件,如辐射、温度、湿度和风的影响.然而,本研究在选取驱动因素时,主要集中于人为因素,尚未充分纳入气象条件.鉴于此,在未来的研究中,考虑增加人为因素与气象因素的交互作用,以更全面、准确地揭示 $O_3$ 浓度变化的驱动机制,进一步提高模型的拟合效果.

## 4　结论

4.1　在中国经济高速发展的背景下,华东地区交通 $CO_2$ 排放与 $O_3$ 浓度整体呈持续上升趋势,且在空间分布上存在明显差异,东部沿海地区排放强度尤为突出.其中,山东、江苏和上海三地 $CO_2$ 排放占比达 62.18%,山东和江苏累计 $O_3$ 浓度占整体 46%.

4.2　交通 $CO_2$ 排放与 $O_3$ 浓度的驱动机制存在差异性.Lasso 回归模型分别为两个预测任务筛选出最优变量集合,其中 $CO_2$ 排放模型中成功剔除 6 个冗余因素,保留 9 个核心变量;$O_3$ 模型中共识别出 10 个重要变量.两者的相同变量包括:城镇化水平、人口数量、人口密度、交通运输业产值、客运周转量和第三产业产值,初步揭示上述因素在“减污降碳”方面具有协同性.

4.3　在模型性能方面,RF 模型在交通 $CO_2$ 及其 $O_3$ 建模中表现出较强的解释能力与稳健性.基于 RF 的 $CO_2$ 排放模型的 $R^2$ 值高达 0.96,MSE 和 MAE 分别为 0.002 和 0.028;其次,基于 RF 的 $O_3$ 浓度模型的 $R^2$ 为 0.84,MSE 和 MAE 分别为 0.008 和 0.071.

4.4　基于 SHAP 分析的变量贡献度排序结果显示,交通 $CO_2$ 排放模型中变量重要性依次为:人口密度、货运周转量、第三产业生产总值、城镇化水平、客运周转量和人口数量;而在 $O_3$ 浓度模型中,主要驱动因素包括人口数量、交通运输业产值、载货汽车数量、城市道路长度、客运周转量和城镇化水平.综合

来看,城镇化水平、人口数量和客运周转量三个因素对交通 $CO_2$ 和 $O_3$ 浓度均具备高解释力.

**参考文献:**


[1] He H, Lu D, Zhao H, et al. Characterizing $CO_2$ and $NO_x$ emission of vehicles crossing toll stations in highway [J]. Transportation Research Part D: Transport and Environment, 2024,126:104024.

[2] Zhang S, Xiong Y, Liang X, et al. Spatial and cross-sectoral transfer of air pollutant emissions from the fleet electrification in China by 2030 [J]. Environmental Science & Technology, 2023,57(50):21249-21259.

[3] Weng D, Zhang H, Wen X, et al. Co-benefits of carbon and pollutant emission reduction in urban transport: Sustainable pathways and economic efficiency [J]. Urban Climate, 2025,60:102348.

[4] 张兰怡,徐艺诺,王 硕,等.碳达峰背景下沿海城市群道路交通节能减排潜力 [J]. 交通运输系统工程与信息, 2024,24(5):45-55.
Zhang L, Xu Y, Wang S, et al. Energy saving and emission reduction potential of road traffic in coastal urban agglomerations under background of carbon peak [J]. Journal of Transportation Systems Engineering and Information Technology, 2024,24(5):45-55.

[5] Huang Y, Wang Y, Peng J, et al. Can China achieve its 2030 and 2060 $CO_2$ commitments? Scenario analysis based on the integration of LEAP model with LMDI decomposition [J]. Science of the Total Environment, 2023,888:164151.

[6] Shangguan Z. Investigating the spatiotemporal dynamics and interplay mechanisms between population urbanization and carbon dioxide emissions in China [J]. Journal of Geochemical Exploration, 2024,266: 107571-107571.

[7] 王 硕,徐艺诺,翁大维,等.中国道路交通碳排放驱动因素及碳达峰 [J]. 生态学报, 2025,45(3):1315-1327.
Wang S, Xu Y, Weng D, et al. Driving factors and peak analysis of road traffic carbon emissions in China [J]. Acta Ecologica Sinica, 2025,45(3):1315-1327.

[8] 翁大维,谢郑一,温晓娟,等.福建省城市群道路交通碳排放时空异质性 [J]. 环境科学, 2025,46(8):4826-4838.
Weng D, Xie Z, Wen X, et al. Spatial and temporal heterogeneity of road transportation carbon emissions in Urban Agglomerations of Fujian Province [J]. Environmental Science, 2025,46(8):4826-4838.

[9] 张兰怡,翁大维,王 硕,等.赣闽粤地区交通碳排放时空演化及驱动力 [J]. 环境科学, 2025,46(5):2886-2896.
Zhang L, Weng D, Wang S, et al. Spatio-temporal evolution and driving forces of transportation carbon emissions in Jiangxi, Fujian, and Guangdong regions [J]. Environmental Science, 2025,46(5): 2886-2896.

[10] 张兰怡,徐艺诺,翁大维,等.经济高质量发展背景下中国省域物流业碳排放时空分异 [J]. 环境科学, 2024,45(9):5086-5096.
Zhang L, Xu Y, Weng D, et al. Spatiotemporal differentiation of carbon emissions from logistics industry at provincial scale in China under the background of high-quality economic development [J]. Environmental Science, 2024,45(9):5086-5096.

[11] Zhang L, Weng D, Xu Y, et al. Spatio-temporal evolution characteristics of carbon emissions from road transportation in the mainland of China from 2006 to 2021 [J]. Science of the Total Environment, 2024,917:170430.

[12] 戈秋虞,张兰怡,邱荣祖,等.基于 GDIM 法的福建省交通碳排放驱动因素分析 [J]. 福建农林大学学报(自然科学版), 2022,51(1):131-136.
Ge Q, Zhang L, Qiu R, et al. Analysis on driving factors of transportation carbon emissions in Fujian Province based on GDIM [J]. Journal of Fujian Agriculture and Forestry University (Natural Science Edition), 2022,51(1):131-136.

[13] He Z, Liu P, Zhao X, et al. Responses of surface $O_3$ and $PM_{2.5}$ trends to changes of anthropogenic emissions in summer over Beijing during 2014~2019: A study based on multiple linear regression and WRF-Chem [J]. Science of the Total Environment, 2022,807(P2):150792-150792.

[14] Shi Z, Huang L, Li J, et al. Sensitivity analysis of the surface ozone and fine particulate matter to meteorological parameters in China [J]. Atmospheric Chemistry and Physics, 2020,20(21):13455-13466.

[15] 陈 坚,陈嘉果,陈 琦,等.山地城市道路交通碳排放影响因素非关系模型 [J]. 交通运输系统工程与信息, 2024,24(5):237-245.
Chen J, Chen J, Chen Q, et al. Nonlinear relationship model of factors influencing carbon emissions in mountainous [J]. Journal of Transportation Systems Engineering and Information, 2024,24(5): 237-245.

[16] 陈 镇,刘 润,罗 征,等.基于机器学习的珠三角秋季臭氧浓度预测 [J]. 环境科学, 2024,45(1):1-7.
Chen Z, Liu R, Luo Z, et al. Prediction of autumn ozone concentration in the Pearl River Delta based on machine learning [J]. Environmental Science, 2024,45(1):1-7.

[17] Vigneshkumar B, Vinothkumar B, Jia C. Importance of ozone precursors information in modelling urban surface ozone variability using machine learning algorithm [J]. Scientific Reports, 2022,12(1): 5646-5646.

[18] Ainsworth A, Yendrek R, Sitch S, et al. The effects of tropospheric Ozone on net primary productivity and implications for climate change [J]. Annual Review of Plant Biology, 2012,63(1):637-661.

[19] Lombardozzi D, Levis S, Bonan G, et al. The influence of chronic Ozone exposure on global carbon and water cycles [J]. Journal of Climate, 2014,28(1):292-305.

[20] 曾晓莹,邱荣祖,林丹婷,等.中国交通碳排放及影响因素时空异质性 [J]. 中国环境科学, 2020,40(10):4304-4313.
Zeng X, Qiu R, Lin D, et al. Spatio-temporal heterogeneity of transportation carbon emissions and its influencing factors in China [J]. China Environmental Science, 2020,40(10):4304-4313.

[21] 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, Zhang Y, et al. Pollution characteristics and influence factors of Ozone in Yangtze River Delta [J]. Journal of Environmental Sciences, 2015,35(8):2370-2377.

[22] 陈玉红.中国臭氧污染时空分布特征及影响因素研究 [D]. 大连:大连理工大学, 2021.
Chen Y. Research on the spatiotemporal distribution characteristics and influencing factors of ozone pollution in China [D]. Dalian: Dalian University of Technology, 2021.

[23] Wu R, Xie S. Spatial distribution of ozone formation in China derived from emissions of speciated volatile organic compounds [J]. Environmental Science & Technology, 2017,51(5):2574−2583.

[24] Ding S, He J, Liu D. Investigating the biophysical and socioeconomic determinants of China tropospheric $O_3$ pollution based on a multilevel analysis approach [J]. Environmental Geochemistry and Health, 2021, 43(8):2835−2849.

[25] 杨寒雨,赵晓永,王　磊.数据归一化方法综述 [J]. 计算机工程与应用, 2023,59(3):13−22.
Yang H, Zhao X, Wang L, et al. Review of data normalization methods [J]. Computer Engineering and Applications, 2023,59(3): 13−22.

[26] Shan Y, Guan D, Zheng H, et al. China $CO_2$ emission accounts 1997~2015 [J]. Scientific Data 5, 2018:17021.

[27] Liu M, Wen Y, Wu X, et al. Province−level decarbonization potentials for China's road transportation sector [J]. Environmental Science & Technology, 2024,58(41):18213−18221.

[28] Zhang L, Weng D, Wen X, et al. Unveiling intra−provincial cities carbon gap: Exploring transportation emissions in ecological Vanguard [J]. Transportation Research Part D: Transport and Environment, 2025,144,104799.

[29] 杨晓彤,康　平,王安怡,等.基于随机森林模型的四川盆地臭氧污染预测 [J]. 环境科学, 2024,45(5):2507−2515.
Yang X, Kang, P, Wang A, et al. Prediction of Ozone pollution in Sichuan Basin based on random forest model [J]. Environmental Science, 2024,45(5):2507−2515.

[30] 卢鋆镆,曾穗平,曾　坚,等.基于随机森林的高分辨率 $PM_{2.5}$ 浓度时空变化模拟—以中原城市群核心区为例 [J]. 中国环境科学, 2023, 43(7):3299−3311.
Lu Y, Zeng H, Zeng J, et al. High resolution simulation of temporal and spatial variation of $PM_{2.5}$ concentration based on random forest — A case study of central plains urban agglomeration core area [J]. China Environmental Science, 2023,43(7):3299−3311.

[31] 赵　楠,卢毅敏.基于 XGBoost 算法的近地面臭氧浓度遥感估算 [J]. Journal of Environmental Sciences, 2022,42(5):95−108.
Zhao N, Lu Y. Remote−sensing estimation of near−surface Ozone concentration based on XGBoost [J]. 环境科学学报, 2022,42(5): 95−108.

[32] 陈启香,吕　斌,李显林.站域建成环境对出租车−地铁组合出行的非线性影响 [J]. 交通运输工程学报, 2024,24(5):285−300.
Chen Q, Lu Q, Li X. Nonlinear effect of station−area built environment on taxi−metro combined travel [J]. Journal of Traffic and Transportation Engineering, 2024,24(5):285−300.

[33] 阎　莹,王玉莹,周　旋,等.基于混合模型的自动驾驶车辆事故严重程度影响因素分析 [J]. 交通运输工程学报, 2025,25(1):184−196.
Yan Y, Wang Y, Zhou X, et al. Analysis of autonomous vehicle accident severity factors based on hybrid model [J]. Journal of Traffic and Transportation Engineering, 2025,25(1):184−196.

[34] 宋永朝,舒　秦,金程容,等.中国区域交通碳排放预测与碳达峰路径规划 [J]. 环境科学, 2025,46(4):1995−2008.
Song Y, Shu Q, Jin C, et al. Regional transport carbon emission forecasting and peak carbon pathway planning in China [J]. Environmental Science, 2025,46(4):1995−2008.

[35] Silva R, Adelman Z, Fry M, et al. The impact of individual anthropogenic emissions sectors on the global burden of human mortality due to ambient air pollution [J]. Environmental health perspectives, 2016,124(11):1776−1784.

**作者简介：**张煌帆(1998−),男,福建永泰人,福建农林大学硕士研究生,主要研究方向为交通环境影响评价.zhang_huangf@163.com.