

城市交通碳排放的非线性影响机制研究

——基于 XGBoost-SHAP 可解释机器学习方法

纪凯丽

天津商业大学理学院统计学

摘要: 基于 2010—2023 年中国 30 个省（自治区、直辖市）面板数据，采用 IPCC 投入产出法系统核算城市交通碳排放量，并运用 XGBoost-SHAP 可解释机器学习模型分析六大驱动因素的非线性影响。研究表明：（1）城市交通碳排放呈现显著的区域差异与阶段性演化特征；（2）各驱动因素对交通碳排放的影响普遍存在非线性特征与阈值效应；（3）VAT 与 FTK 是推动碳排放增长的核心因素，PC 是抑制碳排放的关键变量，PGDP 呈“倒 U 型”关系，ET 呈正向边际递减特征，RPK 表现为多阶段波动效应。研究可为精准制定城市交通领域低碳政策提供科学支撑。

关键词: 城市交通碳排放；非线性影响；XGBoost-SHAP；低碳转型

DOI: 10.65976/3078-8145.2026.01.028

引言

全球气候变化已成为全人类共同面临的重大挑战，联合国政府间气候变化专门委员会（IPCC）相关报告明确指出，温室气体排放是全球气温持续上升的核心驱动因素，由此引发的极端气候灾害严重威胁生态系统平衡与人类社会可持续发展^[1]。作为全球最大的发展中国家，我国已明确提出“2030 年前碳达峰、2060 年前碳中和”的战略目标（以下简称“双碳”目标），将绿色低碳发展纳入国家发展全局，构建起覆盖多领域的绿色低碳发展体系。

城市交通是温室气体排放的重点领域之一。根据 IPCC 第六次评估报告，全球交通运输部门碳排放以年均约 2% 的速度持续增长^[2]。在我国，交通运输部门碳排放总量占全国碳排放总量的 20%—30%，其中道路机动车尾气排放占城市交通碳排放的 80% 以上，是交通领域碳减排的核心管控对象^[3]。随着城镇化进程持续推进，2023 年我国城镇化率已突破 66%，居民出行需求呈多元化、高频化特征，货物运输需求大幅增长，机动车保有量已超 4.3 亿辆，上述因素共同加剧了城市交通碳排放压力。

当前，学界围绕城市交通碳排放开展了大量研究，主要集中于以下几个方面。一是碳排放测算研究。现有研究多采用 IPCC 排放因子法，基于“自上而下”投入产出法或“自下而上”过程分析法开展测算，并借助 LMDI 分解模型剖析各能源类型的贡献差异^[4]。二是影响因素分析。既有研究引入经济增长、城镇化率、能源强度、交通运输结构等变量，采用 OLS、面板回归、

地理加权回归等计量方法探讨各因素的线性影响^[5-6]。然而，上述研究侧重分析线性关系，忽视了城市交通系统的复杂非线性特征，难以捕捉各驱动因素在不同发展阶段的作用强度变化与阈值效应^[7]。三是机器学习方法的引入。近年来，XGBoost、随机森林等机器学习方法逐步被引入碳排放研究领域，并结合 SHAP 可解释性框架有效揭示了变量间的非线性驱动机制^[8-9]。

基于此，本文借鉴可解释机器学习的研究思路，引入 XGBoost-SHAP 框架，以全国 30 个省（自治区、直辖市）（不含西藏、港澳台地区）2010—2023 年面板数据为基础，系统探究各驱动因素对城市交通碳排放的非线性影响机制，识别各因素的作用阈值与阶段性特征，为城市交通领域精准减碳政策的制定提供科学支撑。本文的主要创新在于：（1）构建了覆盖经济、交通活动、能源结构多维度的驱动因素体系；（2）采用 XGBoost-SHAP 框架系统刻画各因素的非线性影响路径与阈值效应；（3）为区域差异化低碳政策的精准制定提供方法论支撑。

1 研究方法与数据来源

1.1 城市交通碳排放核算

2006 年，IPCC 发布国家温室气体清单指南，为测算交通部门碳排放量提供了两种方法：方法一为“自上而下”投入产出法，通过各类能源消耗量与相应转化参数宏观估算碳排放量；方法二为“自下而上”过程分析法，基于车辆保有量、单位燃料消耗及总里程数等微观参数进行估算^[10]。IPCC 指南明确指出，“自上而下”方法对 CO₂ 估算更为可靠，且所需数据可直

接从《中国能源统计年鉴》中获取。由于“自下而上”方法所需微观数据未被纳入我国现行统计体系,本文选择“自上而下”投入产出法进行测算,具体公式如下:

$$C = \sum_m C_m = \sum_m E_m \times K_m \times F_m \times \frac{44}{12} \quad (1)$$

式中, E_m 为第 m 类交通运输能源实物消费量,当前中国各类统计年鉴中均将交通运输、仓储和邮电业合并统计,由于仓储和邮电业占比较小,且二者相当一部分消耗在交通运输领域,因而该合并数据可以在很大程度上代表交通运输业的实物消费量,故本文交通运输能源实物消费量用该数据表示; K_m 为第 m 类能源折算标准煤参考系数; F_m 为第 m 类能源碳排放系数。

本文共选取国内常用的 20 种能源,排放系数参考 IPCC 指南与国家发改委建议值,主要能源类型的折算系数与排放因子见表 1。

表 1 主要能源标准煤折算系数与 CO₂ 排放因子

能源类型	K_m 标准煤折算系数 (kgce/kg)	F_m CO ₂ 排放因子 (t-C/TJ)
原煤	0.714	0.756
洗精煤	0.900	0.765
焦炭	0.971	0.855
汽油	1.471	0.401
柴油	1.457	0.592
燃料油	1.429	0.618
液化石油气	1.714	0.504
天然气	12.143	0.448
液化天然气	1.757	0.571
电力	1.229	0.290

资料来源: IPCC 国家温室气体清单指南(2006); 国家发展改革委相关规定。

1.2 XGBoost-SHAP 模型

1.2.1 XGBoost 模型

传统回归方法通常侧重于刻画自变量与因变量之间的线性关系,但在实际研究中,交通碳排放的驱动机制往往具有明显的非线性特征,单一线性模型难以全面反映这种复杂关系。为更准确地识别变量间的作用机制,本文引入 XGBoost (eXtreme Gradient Boosting) 模型进行分析。该模型是在梯度提升决策树 (GBDT) 基础上发展而来的集成学习算法,通过逐步构建多棵决策树,以每一轮迭代中前一轮模型的预测误差作为学习目标,不断修正和优化模型,从而提高整体预测精度。

相较于传统 GBDT 模型, XGBoost 进行了多方面

改进:通过引入样本权重机制灵活处理不同样本的重要性;在目标函数中加入二阶导数信息,提升收敛速度与结果稳定性;通过正则化项约束模型复杂度,有效降低过拟合风险;同时支持并行运算,具备对缺失值的较强处理能力。其核心目标函数形式如公式(2-2)所示:

$$ob_j^{(k)} = \sum_{j=1}^T [G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2] + \gamma T \quad (2)$$

式中, $G_j = \sum_{i \in I_j} g_i$, 为所有属于第 j 个叶子结点的样本 g_i 的总和; g_i 为损失函数的一阶导数; $H = \sum_{i \in I_j} h_i + \lambda$, 为所有属于第 j 个叶子结点的样本 h_i 的总和; h_i 为损失函数的二阶导数; $I_j = \{i | q(x_i) = j\}$, $q(x_i) = j$ 表示为第 i 个样本位于第 j 个叶子结点; T 为叶子结点的数量; w_j 为叶子结点的权重值; λ 与 γ 为超参数。

2.2.2 SHAP 可解释性分析

SHAP (Shapley Additive exPlanations) 是一种基于博弈论 Shapley 值原理的可解释性方法,通过量化各特征在模型预测中的边际贡献,揭示变量间的作用关系。对于给定特征集合 N 中的特征 i , 其 Shapley 值计算公式如下:

$$\Phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|} (v(S \cup \{i\}) - v(S)) \quad (3)$$

式中, N 表示所有特征的集合; S 是不包含特征 i 的任何一个特征子集; $|S|$ 是集合 S 的特征数; $v(S)$ 是特征集 S 对模型预测输出的贡献; $v(S \cup \{i\})$ 是包含特征 i 的特征集 $S \cup \{i\}$ 对模型预测输出的贡献。借助 SHAP 偏依赖图可深入刻画单一特征对预测结果的非线性影响路径及阈值效应。

1.3 数据来源与变量说明

本文研究时段为 2010—2023 年,研究对象为全国 30 个省(自治区、直辖市)(不含西藏、港澳台地区)。数据来源于《中国能源统计年鉴》《中国统计年鉴》《中国交通运输统计年鉴》及各省市统计年鉴,数据缺失值采用线性插值法补充。在变量选取方面,以城市交通碳排放量(CE)为因变量,选取六个维度的驱动因素作为自变量,具体变量说明见表 2。

2 研究结果

本文基于 Python 环境构建 XGBoost 模型,利用网格搜索与 5 折交叉验证确定最优超参数组合。模型在测试集上的 R^2 达 0.94, RMSE 为 0.083,拟合效果显著优于 OLS ($R^2 = 0.71$) 和随机森林 ($R^2 = 0.89$) 等基准模型,表明 XGBoost 模型能够有效捕捉各驱动因素对城市交通碳排放的复杂非线性关系。在模型训练完成后,引入 SHAP 方法进行可解释性分析,结果如图 1,

表 2 变量选取与说明

变量	变量名称	符号	单位	理论预期
因变量	城市交通碳排放量	CE	万吨 CO ₂	—
产业规模	交通运输、仓储和邮政业增加值	VAT	亿元	+
经济发展	人均 GDP	PGDP	元/人	+/-
客运活动	客运周转量	RPK	亿人公里	+
货运活动	货运周转量	FTK	亿吨公里	+
能源效率	单位周转量能耗	ET	吨标准煤/亿吨公里	+
能源结构	清洁能源比重	PC	%	-

图 2 所示。

从 SHAP 蜂群图 (图 1) 可以看出, VAT 对碳排放的影响最为显著, 其 SHAP 值分布范围最广; FTK 同样表现出较强的正向影响; ET 呈现明显正向作用但分布较为分散; PC 主要表现为负向贡献; RPK 整体影响相对较弱但存在一定正向关系; PGDP 的 SHAP 值基本集中在零附近, 其边际影响相对有限。

进一步基于 SHAP 单特征依赖图 (2) 对各驱动因素的非线性影响路径进行深度解析, 主要发现如下:

首先, 交通运输、仓储和邮政业增加值 (VAT) 对交通碳排放的拉动作用呈现出“先猛增, 后疲惫”态势。在低水平阶段, 大搞基建和频繁的运输活动直接推高了碳排放, SHAP 值由负迅速转正。但当产业规模扩张到一定程度后, 曲线上升趋势被明显压平。这说明, 产业走向成熟伴随着技术升级与效率优化, 这很大程度抵消了扩张带来的碳排放压力。

人均 GDP (PGDP) 的表现完美契合环境库兹涅茨曲线 (EKC) 假说, 呈现标准的“倒 U 型”。在经济发展初期, 经济活力的提升必然伴随交通需求的暴涨, 碳排放量随之增长。可一旦经济水平跨过那道隐形的“门槛”, 事情就会发生反转: 公众环保意识的觉醒、公交体系的完善以及新能源技术的普及, 受这

些因素的影响, SHAP 值重回负区间, 碳排放强度随之回落。

客运周转量 (RPK) 和货运周转量 (FTK) 这对指标, 则走出了截然不同的轨迹。客运周转量的影响颇具戏剧, 经历了“先升、中平、后翘尾”的三个阶段。初期客流增加直接推高排放; 到了中段, 由于公共交通占比不断提升, 碳排放增速达到平稳; 但如果客运规模继续膨胀、突破系统承载极限, 就会引发拥堵, 这种“拥挤效应”会导致排放量再次非现行飙升。相比之下, 货运周转量就“强势”得多, 它几乎是一路向上, 绝对是推高碳排放的核心力量。虽然在后期受物流效率提高的影响, 增速略微放缓, 但整体正向拉动的格局并未改变。

再聚焦于能源供给侧。单位周转量能耗 (ET) 与碳排放正相关, 但表现出有趣的“边际递减”特征。在能源较低时, 稍微增加一点能耗, 排放就会快速上升; 但当能耗本身已经处于高位时, 进一步增加的边际影响反而变弱了, 这说明在极端高耗能状态下, 现有的节能管理措施发挥约束保底作用。

PC 对碳排放的影响呈显著负向关系, 且在低水平阶段下降较快。当清洁能源占比较低时, 其比例提升可显著降低 SHAP 值, 说明能源结构优化对减排效果

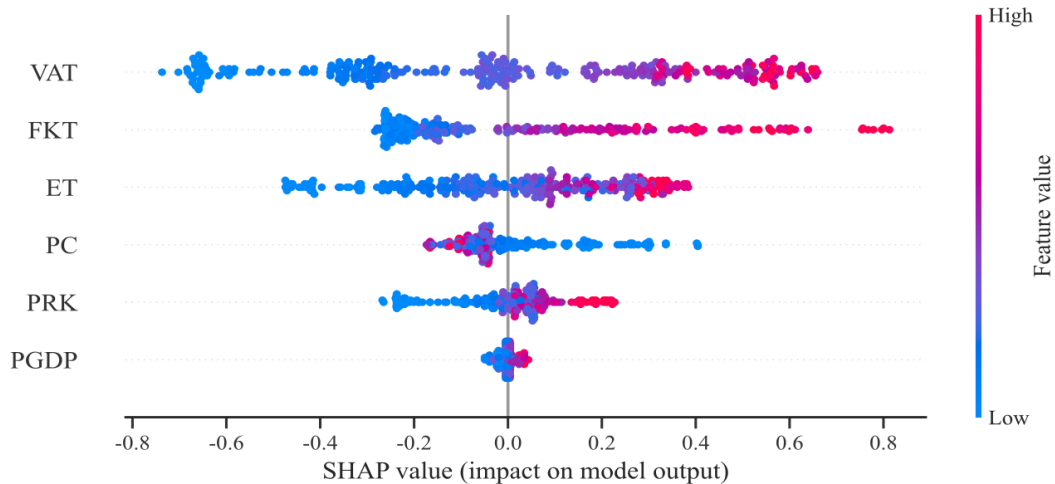


图 1 SHAP 蜂窝图

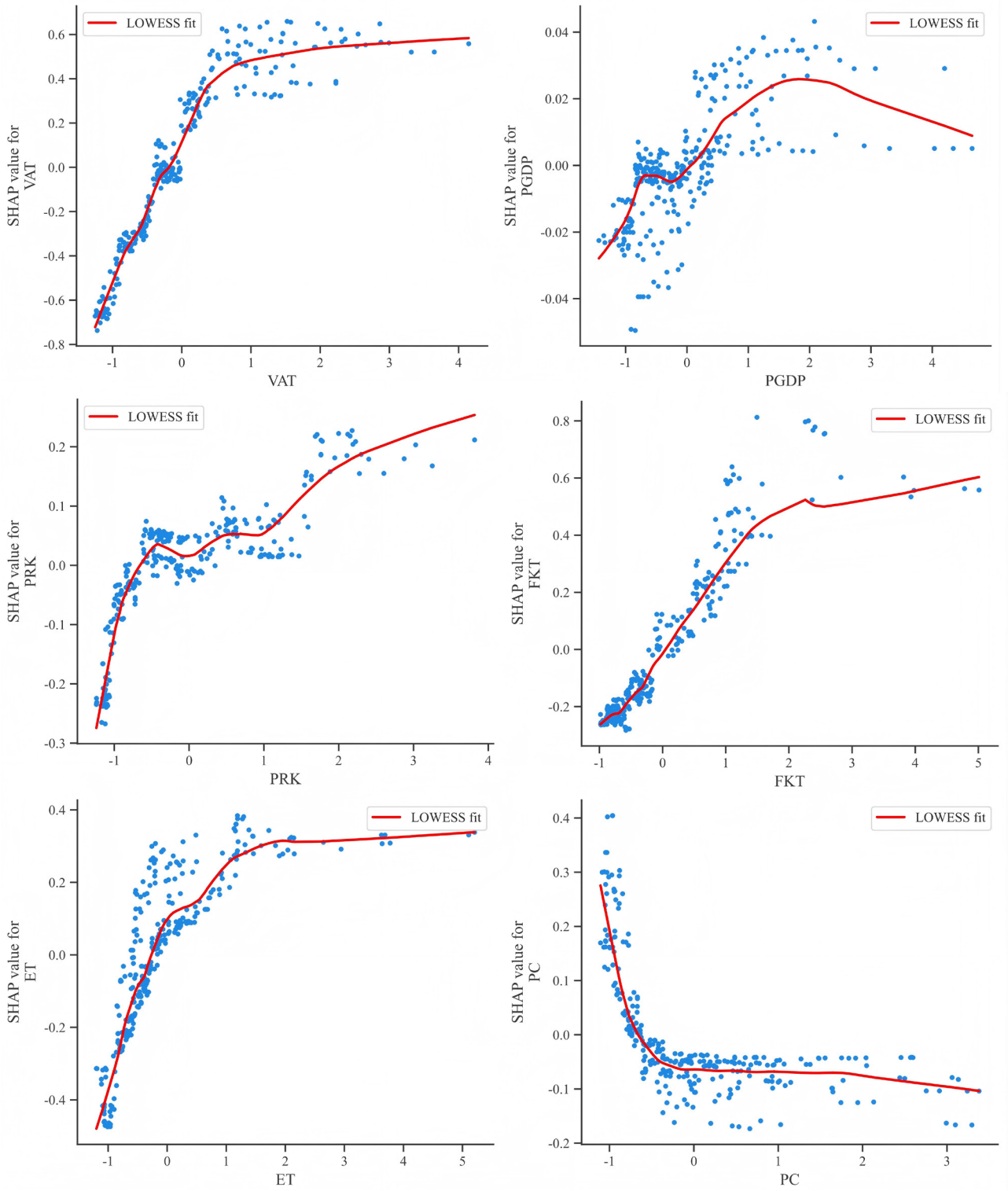


图 2 SHAP 单特征依赖图

十分显著；在较高水平阶段，曲线趋于平缓，边际减排效应逐渐减弱，表明清洁能源替代边际减排潜力逐步释放完毕。

3 结论与政策建议

3.1 研究结论

本文基于 2010—2023 年全国 30 个省（自治区、直辖市）（不含西藏、港澳台地区）面板数据，采用

IPCC “自上而下”投入产出法核算城市交通碳排放量，运用 XGBoost-SHAP 框架系统探究各驱动因素的非线性影响机制，主要得出以下结论：

第一，城市交通碳排放呈现显著的时序演化与空间异质性特征。研究期间碳排放总量总体上升，呈“东高西低”的空间分布格局，东部地区绝对量领先，中西部地区增速较快，区域差异较为突出。

第二,城市交通碳排放的驱动机制具有显著的非线性特征。各驱动因素的影响均存在明确的阈值效应与阶段性变化,传统线性模型难以精准捕捉这种复杂关系,XGBoost-SHAP 框架能够有效弥补这一不足。

第三,不同驱动因素的非线性演化路径存在明显差异:VAT 呈“先增强后趋缓”特征,PGDP 呈“倒 U 型”关系,RPK 表现为多阶段波动特征,FTK 呈持续增强的正向影响,ET 呈正向边际递减特征,PC 呈负向影响且初期减排效果更为突出。FTK 与 VAT 是推动碳排放增长的核心因素,PC 是抑制碳排放的关键变量。

3.2 政策建议

上述非线性特征不仅是数据上的规律,更是政策制定的靶点。基于这些发现,我们提出以下对策:

第一,产业结构得因地制宜,避免“一刀切”。既然 VAT 的影响“先强后弱”,东部发达地区就该赶紧带着交通产业往智能化、绿色化转;而中西部还在发力期,扩张步伐得稳住,提前规划绿色路径,千万别再踩一遍高排放的老坑。

第二,顺应经济发展阶段,打好“差异牌”。还没跨过 PGDP“倒 U 型”拐点的地方,当务之急是控制私家车野蛮生长,把公共交通的底子打牢;过了拐点、财大气粗的地区,就该砸重金推广新能源,把经济高质量发展的减排红利吃干榨净。

第三,给公路运输“减负”,死磕能源效率。既然货运(FTK)拉动排放这么猛,公转铁、公转水必须得提速,强行优化运输结构。同时,别看能耗(ET)的边际影响在递减,节能技术的研发绝不能停,得持续向技术要减排空间。

第四,抢抓清洁能源的“初期红利”。既然清洁能源(PC)在低水平阶段的减排性价比极高,现在就是大力铺设充电桩、加氢站的最好窗口期。特别是中西部地区,完全可以守着丰富的风光资源,借清洁能源替代实现弯道超车。

最后,打破省域壁垒,搞好“传帮带”。单打独斗解决不了整体减排问题,必须建立跨区域协同机制,让东部的绿色技术和管理经验向中西部流动。全国一盘棋,才能真正带着中西部跨过“先污染后治理”的泥潭,实现交通系统的整体低碳转型。

参考文献:

- [1]IPCC.Climate Change 2023:Synthesis Report[R]. Geneva:IPCC,2023.
- [2]IEA.Global EV Outlook 2023[R].Paris:International Energy Agency,2023.
- [3]Wang H,Ou X,Zhang X.Mode,technology,energy consumption,and resulting CO₂ emissions in China's transport sector up to 2050[J].Energy Policy,2017,109:719-733.
- [4]Zhang X,Li Y,Wang S,et al.An analysis of the decomposition and driving force of carbon emissions in transport sector in China[J].Scientific Reports,2024,14:28743.
- [5]Cheng Z,Li L,Liu J.Industrial structure,technical progress and carbon intensity in China's provinces[J].Renewable and Sustainable Energy Reviews,2018,81:2935-2946.
- [6]Wang Y,Liao M,et al.Unearthing the impact of urban sprawl on transport carbon dioxide emissions:fresh evidence from 274 Chinese cities[J].Journal of Geographical Sciences,2024,34(11):2173-2196.
- [7]Grossman G M,Krueger A B.Economic growth and the environment[J].Quarterly Journal of Economics,1995,110(2): 353-377.
- [8]Chen T,Guestrin C.XGBoost:A scalable tree boosting system[C].KDD 2016:785-794.
- [9]Xing M,Wang Z,et al.An XGBoost-SHAP analysis of the driving factors of carbon emissions in China's first-tier cities[J].Scientific Reports,2025,15:8601.
- [10]IPCC.2006 IPCC Guidelines for National Greenhouse Gas Inventories[R].Hayama:IGES,2006.