

XGBoost 与 SHAP 协同揭示河南省 CH₄ 柱浓度的时空分布及驱动因子

姚伊玲¹, 孙俊英², 张显云^{1*}, 吕佳敏¹, 龚 胜¹

1. 贵州大学矿业学院, 贵州 贵阳 550025
2. 贵州省第二测绘院, 贵州 贵阳 550004

摘要: 为更好地解析河南省甲烷 (CH₄) 柱浓度的时空特征及驱动因子, 首先基于机器学习算法构建了 CH₄ 柱浓度的估算模型, 在此基础上对 CH₄ 柱浓度进行了空间连续性估算以及时间特征和空间分布的揭示, 最后针对机器学习模型可解释性较差的缺陷, 引入 Shapley 加性解释方法 (SHAP) 诊断分析了 CH₄ 柱浓度驱动因子的重要性及作用机制。结果表明: ①在所研究的机器学习模型中 XGBoost (eXtreme Gradient Boosting) 的精度最高, 测试集上的决定系数、均方根误差和平均绝对误差分别为 0.964、 4.78×10^{-9} 、 3.11×10^{-9} 。②2019—2023 年河南省 CH₄ 柱浓度月均值具有一定的增长趋势和季节周期性, 呈春季下降、夏季上升、秋季达峰、冬季回落的季节循环模式, 且季节性强度值为 0.756 8; 空间分布上无论是月均值还是季均值, 均表现为西部低、中部和东南部高的分布特征。③2 m 地面温度、低植被叶面积指数和废弃物源排放量是影响 CH₄ 柱浓度的 3 个主要因素, 其中, 2 m 地面温度对 CH₄ 柱浓度整体起正向作用, 而低植被叶面积指数起负向作用; 高温与密集低矮植被协同促进了 CH₄ 柱浓度的升高, 而高废弃物源排放量与强经向风的交互同样有助于 CH₄ 柱浓度的升高。研究显示, 2019—2023 年河南省 CH₄ 柱浓度虽整体变化趋势不显著, 但受 2 m 地面温度、低植被叶面积指数及废弃物源排放量等因素影响, 表现出明显的季节性波动特征与空间分异。

关键词: 甲烷柱浓度; 时域特征; 空间分布; SHAP; 驱动因子; XGBoost

中图分类号: X705

文章编号: 1001-6929(2026)03-0590-12

文献标志码: A

DOI: 10.13198/j.issn.1001-6929.2026.01.04

Revealing the Spatiotemporal Distribution and Driving Factors of Methane Column Concentrations in Henan Province Using XGBoost and SHAP

YAO Yiling¹, SUN Junying², ZHANG Xianyun^{1*}, LÜ Jiamin¹, GONG Sheng¹

1. School of Mining, Guizhou University, Guiyang 550025, China
2. Guizhou Second Institute of Surveying and Mapping, Guiyang 550004, China

Abstract: To comprehensively investigate the spatiotemporal characteristics and driving factors of methane (CH₄) column concentrations in Henan Province, an estimation model for CH₄ column concentrations was first constructed based on machine learning algorithms. Based on this model, spatial continuity estimation was conducted to characterize the temporal characteristics and spatial distribution of CH₄ column concentrations. To overcome the limited interpretability of conventional machine-learning models, the Shapley Additive Explanations (SHAP) method was applied to diagnose and analyze the relative importance and interaction mechanisms of driving factors influencing CH₄ column concentration. The results showed that: (1) Among the evaluated machine-learning models, eXtreme Gradient Boosting (XGBoost) achieved the highest predictive accuracy, with a coefficient of determination of 0.964, a root mean square error of 4.78×10^{-9} , and a mean absolute error of 3.11×10^{-9} on the test set. (2) The monthly mean CH₄ column concentrations in Henan Province from 2019 to 2023 exhibited an overall increasing trend with pronounced seasonal periodicity, characterized by a 'spring decrease, summer increase, autumn peak, winter decline' cycle and a seasonal strength index of 0.7568. Spatially, both the monthly and seasonal mean concentrations consistently displayed a pattern of lower values in the west and higher values in the central and southeastern regions. (3) The 2 meter temperature, the leaf area index of low vegetation, and waste sources were identified as the 3 dominant factors influencing

收稿日期: 2025-08-07 修订日期: 2026-01-13

作者简介: 姚伊玲(2000-), 女, 贵州锦屏县人, 2771261857@qq.com

* 责任作者: 张显云(1974-), 男, 贵州遵义人, 副教授, 硕士, 主要从事遥感技术理论及应用研究, xyzhang5@gzu.edu.cn

基金项目: 贵州省科技计划项目 (No.黔科合基础-ZK[2024]一般 093); 贵州省科技计划项目 (No.黔科合支撑 [2022]一般 204)

Supported by Guizhou Provincial Science and Technology Program, China (No.Qian Ke He Basic Research-ZK[2024] General 093); Guizhou Provincial Science and Technology Program, China (No.Qian Ke He Basic Support[2022] General 204)

CH₄ column concentrations. Specifically, 2 meter temperature generally exerted a positive effect, while the leaf area index of low vegetation showed a negative effect. The combined influence of high temperatures and dense low vegetation increased CH₄ column concentrations. while the interaction between high waste sources and strong meridional winds also contributed to elevated CH₄ column concentrations. The results indicate that although the overall trend in CH₄ column concentrations in Henan Province was not significant from 2019 to 2023, pronounced seasonal variability and spatial heterogeneity were evident. These patterns were jointly driven by 2 meter temperature, the leaf area index of low vegetation, and waste sources.

Keywords: methane column concentration; time domain features; spatial distribution; SHAP; driving factors; XGBoost

甲烷是继二氧化碳 (CO₂) 之后对全球变暖贡献最大的温室气体,其单位质量的温室效应在 20 年和 100 年时间尺度上分别为 CO₂ 的 84 和 28 倍^[1-2]。尽管在大气中的浓度相对较低,但由于辐射强迫效应强烈^[3]、生命周期适中,甲烷已成为气候变化研究与温室气体减排政策中的关键对象^[4-5]。作为衡量大气中甲烷含量的重要指标,CH₄ 柱浓度可综合反映地表排放与垂直输送过程,是评估区域甲烷排放水平与环境响应的重要参数^[6]。

鉴于 CH₄ 柱浓度在气候系统中的重要作用,相关研究逐渐受到关注^[7-9]。CH₄ 柱浓度的时空分布特征及其驱动因子的正确识别,有助于明确气象因子与人类活动对 CH₄ 柱浓度的影响机制,为高排放区识别、成因解析和减排政策的制定提供科学依据^[10-12]。然而,现有研究主要存在以下 3 方面的问题: ①CH₄ 柱浓度的时空变化受气象因子和人类活动的共同作用^[13-15],然而现有研究多聚焦于单一类型因子的影响^[16-20],缺乏联合气象与人类活动因子的综合建模^[21],难以全面揭示 CH₄ 柱浓度的驱动因子。②其次,卫星遥感技术为 CH₄ 柱浓度的相关研究提供了广覆盖、高频次的数据支撑,但受云层、地表反照率等因素影响,数据常存在缺失问题。这将直接导致排放热点识别不全、变化趋势判断偏差、驱动因子模型训练受限,进而影响后续反演建模和政策分析的准确性^[22-24]。尽管如此,现有关于 CH₄ 时空分布的研究多直接依赖原始数据,未充分考虑数据缺失带来的建模偏差^[25-27]。③当前 CH₄ 柱浓度驱动力研究主要采用传统统计回归、反演模拟与机器学习三类方法。传统统计方法如多元线性回归^[28]和逐步回归^[29]虽可解释性强,但难以处理变量间的非线性关系和多重共线性问题^[30]。反演模拟虽符合大气输送的物理机制,但依赖先验数据且计算复杂^[31]。相比之下,近年来广泛应用的随机森林 (Random Forest, RF)、支持向量机 (Support Vector Machine, SVM) 和梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 在处理非线性关系和多变量交互方面具有显著优势,但其“黑箱”特性限制了模型对变量贡献的准确评估^[32]。为

提高机器学习模型的可解释性并量化各驱动因子的影响程度,Shapley 加法解释方法 (Shapley Additive Explanations, SHAP)^[33] 近年来逐渐被引入环境遥感领域^[34],该方法能够对复杂模型的输出进行特征贡献分解,全面揭示变量之间的关系,为 CH₄ 研究提供了新的分析工具,但其在 CH₄ 柱浓度驱动力分析中的应用却鲜有报道。

基于此,本研究以河南省为研究区,联合 CH₄ 柱浓度、气象因子和排放清单数据构建了适用于研究区 CH₄ 柱浓度估算的高精度机器学习模型,在实现 CH₄ 柱浓度无间断估算的基础上,对研究区 CH₄ 柱浓度的时空特征进行了分析,并采用 SHAP 诊断分析了 CH₄ 柱浓度驱动因子的重要性及作用机制,以期深化对 CH₄ 时空演变机制的理解、精准识别排放源,以及落实河南省“十四五”生态环境保护规划中的差异化减排政策提供理论依据与科学支撑。

1 研究区域与研究方法

1.1 研究区域

河南省位于我国中部,横跨中国地势第二级与第三级阶梯的过渡带,地理位置为 31°23'N~36°22'N、110°21'E~116°39'E,总面积达到 167 000 km²。该省地形呈现西高东低的特点,西部和南部为山地,东部则是广阔的平原,最高峰老鸦岔海拔 2 413.8 m。近年来,我国 CH₄ 排放呈现东南高、西北低的空间格局,华中地区被识别为 CH₄ 的高排放区域^[35]。河南省作为华中地区经济总量最大、人口与产业高度集聚、能源与农业活动密集的典型省份,研究其 CH₄ 柱浓度时空特征及驱动因子不仅有助于深入揭示该区域 CH₄ 排放的主要来源,还可为制定应对 CH₄ 减排政策提供科学依据。

1.2 数据来源及处理

1.2.1 CH₄ 柱浓度

Sentinel-5P 卫星是欧洲空间局于 2017 年 10 月发射的地球观测卫星,携带有 TROPOMI (Tropospheric Monitoring Instrument) 传感器,能够对地球大气中多种气体进行高精度测量。本研究所用 CH₄ 柱浓度数据是通过 GEE (<http://earthengine.google.com>) 平台下

载的自地表至大气顶的整层大气柱平均干空气体积分数 TROPOMI L3 级 offline CH₄ 柱浓度产品,空间分辨率为 1 113.2 m,时间分辨率为 1 d。由于具备全球覆盖和高精度的优势, Sentinel-5P CH₄ 已成为 CH₄ 反演的重要数据来源^[36-37]。尽管 TROPOMI 实现了全球范围内的 CH₄ 柱浓度观测,其反演数据仍存在较高的缺失率和严格的质量过滤要求,导致部分区域或时段的观测覆盖不足^[38],从而将不利于 CH₄ 柱浓度驱动力及其时空变化规律的系统评估。

1.2.2 排放清单

排放清单数据源于欧盟委员会联合研究中心和荷兰环境评估机构发布的 EDGAR 温室气体数据库中的相关数据^[39]。该数据涵盖了 CH₄ 源排放的不同源和地区,提供有关全球温室气体排放尤其是 CH₄ 的详细信息。本研究所用排放清单数据,涵盖了农业源排放量 (agricultural source emissions)、燃料开采源排放量 (fuel exploitation source emissions)、工业燃烧源排放量 (industrial combustion source emissions)、运输源排放量 (transportation source emissions) 和废弃物排放源排放量 (waste emission source emissions)。

1.2.3 气象数据

第 5 代欧洲再分析数据集 (ERA5) 是由欧洲中期天气预报中心 (European centre for medium-range weather forecasts, ECMWF) 创建和维护的一种全球气象再分析数据集。顾及气象因子在 CH₄ 柱浓度演变过程中的关键作用^[40-41],本研究选用 2019—2023 年 ERA5 气象数据,包括边界层高度 (boundary layer height)、低植被叶面积指数 (low vegetation index)、高植被叶面积指数 (high vegetation index)、气压 (surface

pressure)、2 m 地面温度 (2 metre temperature)、2 m 露点温度 (2 metre dewpoint temperature)、10 m 纬向风分量 (10 metre U wind component)、10 m 经向风分量 (10 metre V wind component) 和总降水量 (total precipitation)。

各类型数据的时间分辨率、空间分辨率及来源等信息如表 1 所示。

1.2.4 数据匹配

鉴于 CH₄ 柱浓度、排放清单及气象数据的时间分辨率及空间分辨率存在差异,本研究采用时空匹配技术对 3 种数据进行整合。时间维度上,排放清单数据以月为分辨率,而 ERA5 气象数据和 Sentinel-5P CH₄ 数据的时间分辨率分别为 1 h 和 1 d,因此对气象数据和 CH₄ 柱浓度按月取平均值,以保持其与排放清单数据具有相同的时间分辨率。空间维度上,采用双线性插值将气象数据重采样至 0.01°,排放清单数据采用面积平均分配法降尺度至相同分辨率,以保持与 CH₄ 柱浓度具有相同的空间分辨率。

1.3 研究方法

1.3.1 XGBoost 模型

XGBoost 是一种基于梯度提升算法的机器学习方法,其通过集成多个决策树来提高模型的性能。这些决策树通过优化参数和利用梯度信息来做出分类或回归决策。2016 年,Chen 等^[42]基于 GBDT 提出了 XGBoost 算法。该算法通过加法训练方式串行构建多棵分类与回归树,其中每棵树均拟合前一轮模型的残差,并在传统梯度提升框架基础上融入了正则化、并行计算、缺失值处理等优化策略。XGBoost 模型的目标函数可参考文献 [43]。

表 1 数据信息

Table 1 Data message

数据类型	名称	简称	时间分辨率	空间分辨率	数据来源
大气甲烷柱浓度	离线甲烷总浓度	CH ₄ 柱浓度	1 d	0.01°	GEE
	边界层高度	BLH	1 h	0.25°	ECMWF
气象数据	2 m地面温度	T2M	1 h	0.25°	ECMWF
	2 m露点温度	D2M	1 h	0.25°	ECMWF
	10 m纬向风分量	U10	1 h	0.25°	ECMWF
	10 m经向风分量	V10	1 h	0.25°	ECMWF
	低植被叶面积指数	LVI	1 h	0.25°	ECMWF
	高植被叶面积指数	HVI	1 h	0.25°	ECMWF
	气压	SP	1 h	0.25°	ECMWF
	总降水量	TP	1 h	0.25°	ECMWF
	排放清单	农业源排放量	AGR	1 m	0.1°
燃料开采源排放量		FE	1 m	0.1°	EDGAR
工业燃烧源排放量		IC	1 m	0.1°	EDGAR
运输源排放量		TRN	1 m	0.1°	EDGAR
废弃物源排放量		WST	1 m	0.1°	EDGAR

1.3.2 SHAP 解释法

SHAP 是一种用于解释机器学习模型输出的工具,其理论基础为合作博弈论的 Shapley 值,旨在为每个特征分配重要性评分,以此分析各特征对 CH₄ 柱浓度的作用方式和作用效果。Lundberg 等^[44]开发了基于 Python 的模型解释包,使得基于 Python 可便捷地获得不同机器学习算法的 SHAP 值。此外,该解释包还具备良好的可视化功能,能够方便绘制各类分析图表^[45]。

1.3.3 时域特征分析

1.3.3.1 趋势分析

采用基于一元线性回归的斜率 (slope) 分析^[46]进行趋势诊断,以量化 CH₄ 柱浓度月均值时间序列的长期变化趋势及其显著性。斜率、决定系数 (R^2) 和 p 值是 slope 分析的 3 个主要指标,其中,斜率表征时间序列的变化趋势,斜率大于 0 说明时间序列呈上升趋势,反之则为下降趋势; p 值是变化趋势显著性的度量, p 值小于显著性水平,则表明时间序列变化趋势显著。

1.3.3.2 季节性分解

采用季节性分解 (seasonal_decompose) 对 CH₄ 柱浓度时间序列中潜在周期性特征进行诊断分析。seasonal_decompose 方法源于经典的移动平均分解模型,并在关于时间序列分析领域的经典著作^[47]中得到了系统阐述。基于经典的加法分解模型,CH₄ 柱浓度时间序列可分解为趋势项、季节性和残差三部分^[48]。

1.3.4 技术路线

尽管 XGBoost 模型、RF 模型和梯度提升决策树 (Categorical Boosting, CatBoost) 等树模型对特征共线性具有一定鲁棒性,但特征间的高度相关性仍可能导致 SHAP 解释结果失真,因为 SHAP 值会在相关特征间分摊重要性^[49]。因此,本研究采用方差膨胀因子 (Variance Inflation Factor, VIF)^[50]对驱动因子进行了共线性检验,结果表明,2 m 地面温度与 2 m 露点温度之间存在极强的线性相关性,故建模时剔除了 2 m 露点温度。在此基础上,为揭示研究区 CH₄ 柱浓度的时空分布特征及驱动因子,研究内容主要包括三部分,技术路线如图 1 所示。

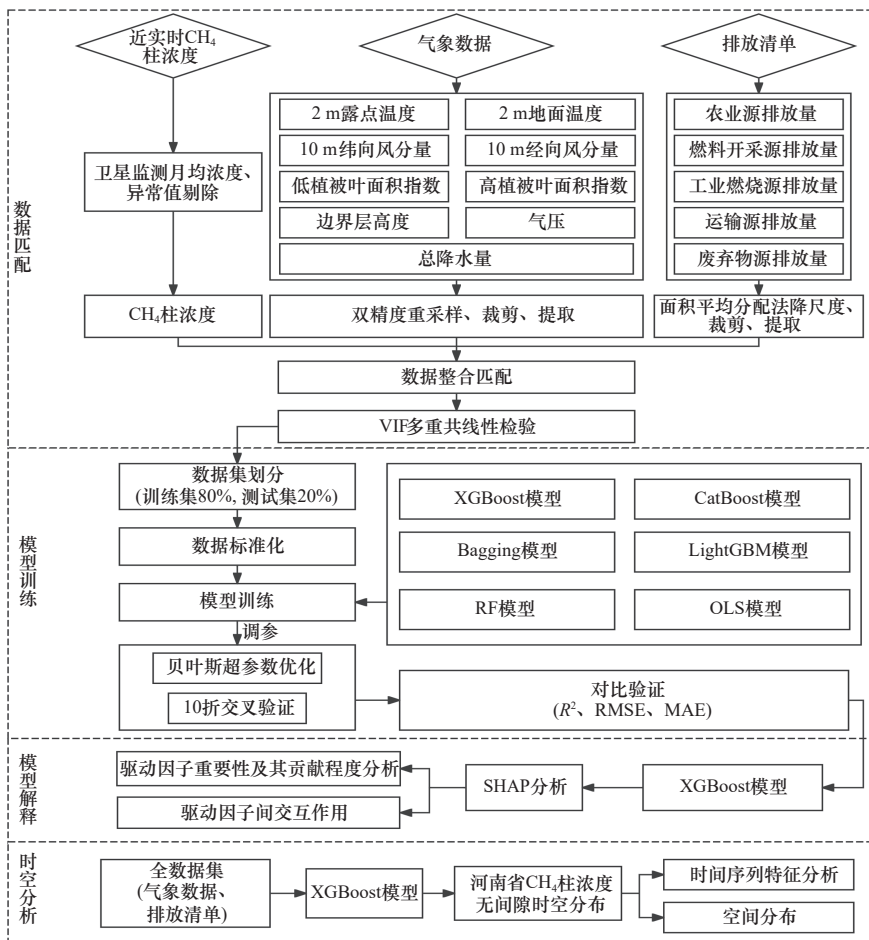


图 1 技术路线

Fig.1 The experimental technical route

CH₄ 柱浓度的无间隙估算: 首先以 CH₄ 柱浓度数据为响应变量, 以通过共线性检验后的气象因子和排放清单为解释变量, 分别采用 XGBoost 模型、CatBoost 模型、普通最小二乘法 (Ordinary Least Squares, OLS)、RF 模型、轻量梯度提升机 (Light Gradient Boosting Machine, LightGBM) 和装袋法 (Bootstrap Aggregating, Bagging) 进行回归建模。为避免原始数据顺序带来的潜在偏差并确保结果的可复现性, 在建模前对数据进行了随机打乱, 同时设置了固定的随机种子参数。为避免模型发生过拟合或欠

拟合的问题, 将数据按 8:2 划分为训练集与测试集, 并采用 10 折交叉验证结合贝叶斯优化^[51] 进行超参数调优, 以确保模型的泛化性能, 其中 XGBoost 模型超参数及其范围如表 2 所示, 贝叶斯优化迭代次数 $n_iter=30$ 。最后基于 R^2 、均方根误差 (Root Mean Square Error, RMSE) 和平均绝对误差 (Mean Absolute Error, MAE) 对模型精度进行评估, 筛选出适用于研究区的高精度 CH₄ 柱浓度估算模型。在此基础上, 将无间隙驱动因子数据作为输入, 利用最优模型进行预测, 得到具有空间连续性的 CH₄ 柱浓度无间隙估算产品。

表 2 XGBoost 模型超参数及其范围

Table 2 The hyperparameters of the XGBoost model along with their respective ranges

超参数名称	超参数含义	数值范围	数值类型
n_estimators	树的数量	100~1 000	整数
max_depth	单棵树最大深度	3~10	整数
learning_rate	学习率	0.01~0.3	实数
subsample	行采样比例(样本子采样)	0.5~1.0	实数
colsample_bytree	列采样比例(特征子采样)	0.5~1.0	实数
reg_lambda	L2正则化强度	1.0~20.0	实数
reg_alpha	L1正则化强度	0.1~10.0	实数
gamma	最小分裂损失减少量	0~5	实数
min_child_weight	叶子节点样本权重	1~10	整数

采用 slope 分析对时间序列数据进行趋势识别, 并结合季节分解技术提取不同时间尺度上的周期性变化特征, 揭示研究区 CH₄ 柱浓度的时域变化特征以及季节性波动规律, 通过制图进一步分析 CH₄ 的空间分布格局。

CH₄ 柱浓度驱动因子的重要性及作用机制。XGBoost 模型、CatBoost 模型、LightGBM 模型和 RF 模型等机器学习模型本身属于“黑箱”模型, 可解释性较差, 其预测结果难以揭示各特征变量对响应变量的具体作用机制。本研究引入 SHAP 方法, 在 Python 环境中利用 SHAP 对 XGBoost 模型的输出结果进行可解释性分析, 从全局和局部两个层面定量各影响因子的相对重要性、作用方向、非线性响应特征, 以及影响因子间的交互效应。

2 结果与讨论

2.1 模型在测试集上的精度

各机器学习模型在测试集上的精度统计情况如图 2 可见。由图 2 可见, OLS 模型的精度最低, R^2 、RMSE 和 MAE 分别为 0.285、 21.27×10^{-9} 和 16.87×10^{-9} 。精度较高的 3 个模型分别是 XGBoost 模型、CatBoost 模型和 LightGBM 模型, 其中 XGBoost 模型最高。相较于 LightGBM 模型, XGBoost 模型的精度得到了不

同程度的改善, R^2 提高了 7.23%, RMSE 和 MAE 分别降低了 40.32% 和 46.19%; 相比于 CatBoost 模型, XGBoost 模型的 R^2 提高了 5.13%, RMSE 和 MAE 则分别降低了 34.25% 和 39.73%。因此, 本研究选取 XGBoost 模型对研究区 CH₄ 柱浓度进行无间隙估算。

2.2 CH₄ 柱浓度的时空演变特征分析

2.2.1 CH₄ 柱浓度的时域变化特征

2.2.1.1 时域趋势分析

为揭示 CH₄ 柱浓度的整体趋势及其显著性, 以 CH₄ 柱浓度月均值为因变量, 以时间为自变量, 采用 slope 趋势分析方法对 2019 年 1 月—2023 年 12 月河南省 CH₄ 柱浓度月均值进行了诊断分析, 结果如图 3 所示。由图 3 可见, 回归方程系数 $R^2(0.06)$ 大于 0, 表明 2019 年 1 月—2023 年 12 月河南省 CH₄ 柱浓度有一定的上升, 但上升趋势并不显著 ($p>0.05$)。

2.2.1.2 季节性特征分析

在季节性特征分析中, 将研究区 CH₄ 柱浓度月均值构建为时间序列, 利用 seasonal_decompose 方法以时间为索引进行分解, 提取季节项和残余项。基于 Python 中 statsmodels 库的 seasonal_decompose 函数对 CH₄ 柱浓度月均值序列进行分解^[52], 并设定周期为 12, 以捕捉 CH₄ 柱浓度的季节性特征。研究区

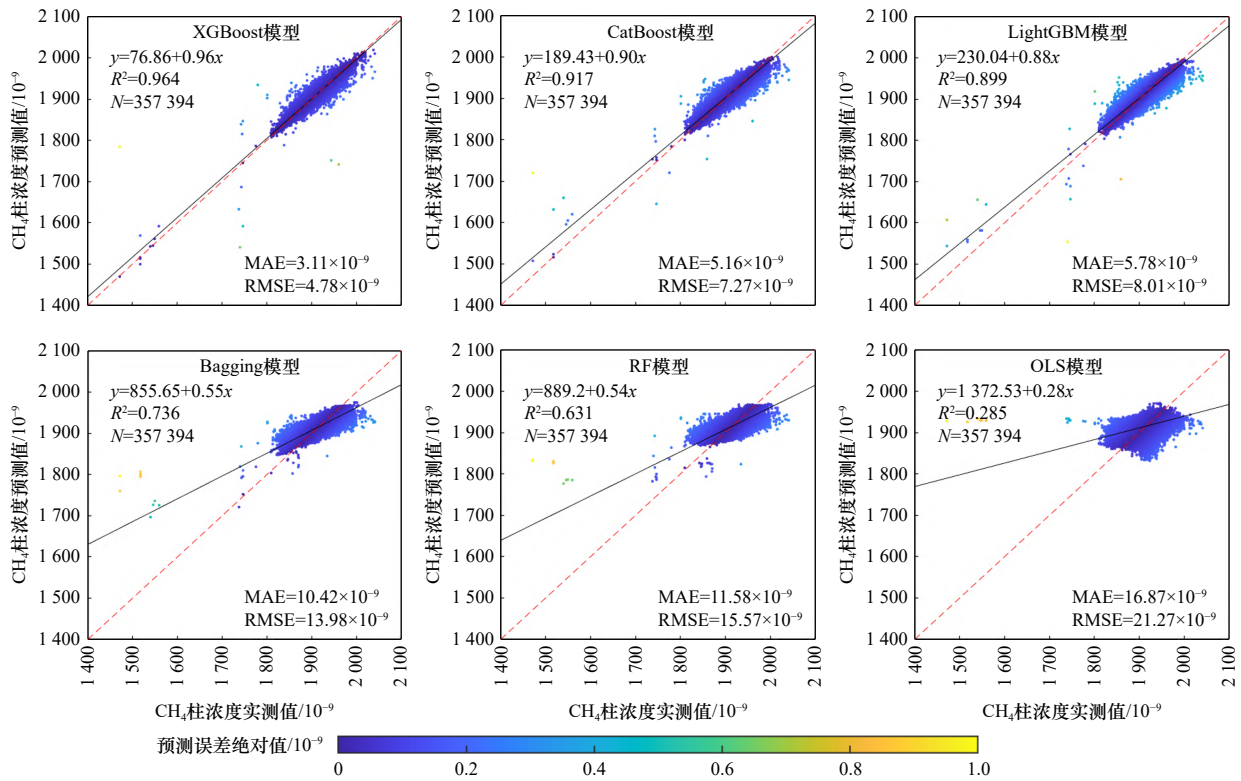


图2 测试集精度对比

Fig.2 Comparison of test set accuracy

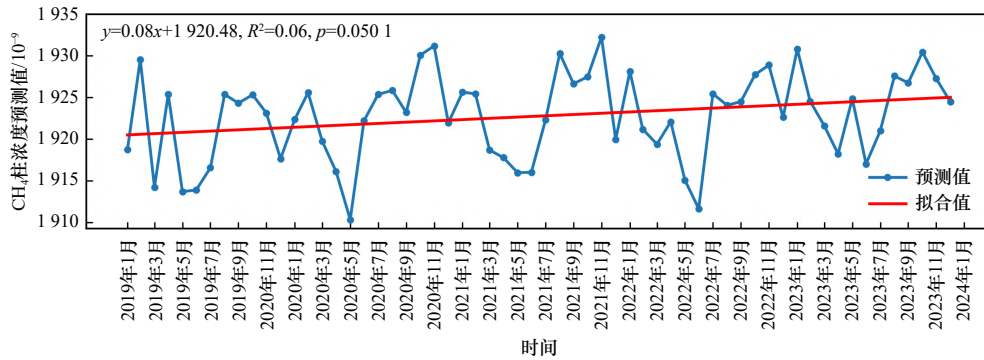


图3 2019—2023年河南省CH₄柱浓度的时域趋势

Fig.3 Temporal trend of CH₄ column concentration in Henan Province from 2019 to 2023

CH₄ 柱浓度月均值的季节性分解结果如图 4(a) 所示, CH₄ 柱浓度月均值为 $-5 \times 10^{-9} \sim 5 \times 10^{-9}$, 具有明显的年周期特性, 整体呈现以下特点: 春季至夏季 (5—8 月) CH₄ 柱浓度持续上升; 秋季 (9—11 月) 浓度达到年度峰值, 其中 9—10 月出现明显的浓度拐点; 冬季 (12 月—翌年 2 月) 则回落至较低水平。结合 Seasonal Strength 的数值 (季节性强度为 0.756 8) 进一步分析可知, 季节成分可解释约 75.68% 的时间序列波动, 说明研究区 CH₄ 柱浓度序列受季节性变化的显著影响, 具有稳定且周期性强的年内波动规律。需要说明的是, 季节性分量在不同年份呈现完全一致

的形态, 这是 seasonal_decompose 方法的固有特性 (该方法假设季节模式在时间序列中保持稳定, 通过计算同一月份在所有年份的平均季节性效应来实现分解); 尽管 seasonal_decompose 方法还可以提取时间序列的趋势项, 但趋势强度只是一个方差比例指标, 并不是统计检验, 不能判断趋势的显著性, 故本研究未采用 seasonal_decompose 提取趋势项。

在 seasonal_decompose 时间序列分解中残差 (Residual) 指原始序列去除趋势项 (Trend) 和季节项 (Seasonal) 后剩余的部分。采用 seasonal_decompose 进行季节分解后的残差分布如图 4(b) 所示, 残差围

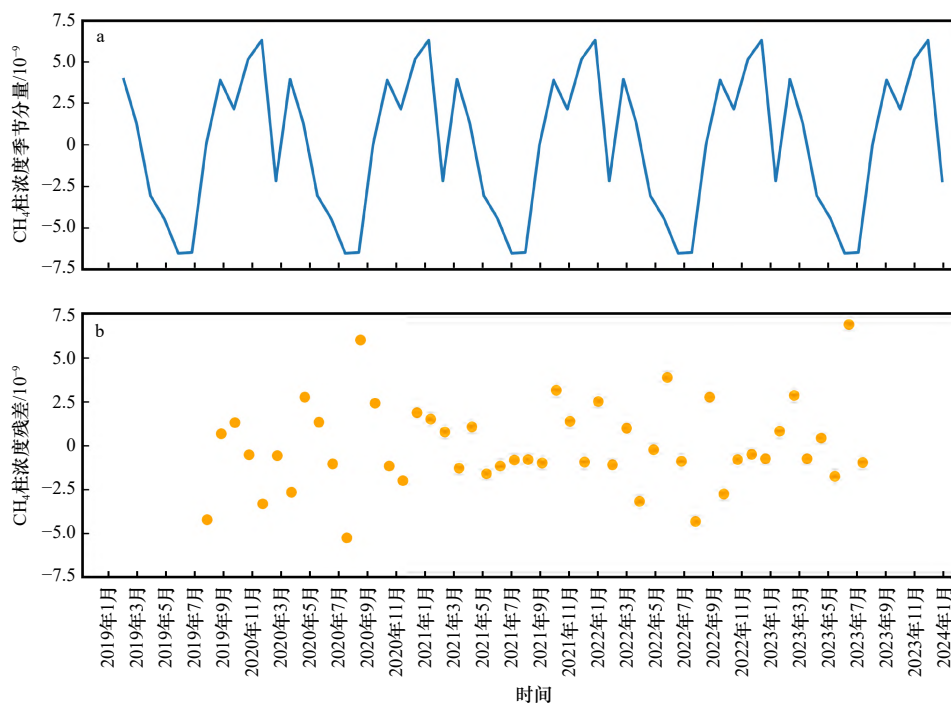


图 4 河南省 CH_4 柱浓度季节特征及模型残差

Fig.4 Seasonal characteristics of CH_4 column concentration and model residuals in Henan Province

绕 0 值上下波动,表现出偶然误差的特性,表明 CH_4 柱浓度月均值的主要信息可被趋势项和季节项较好地捕捉。

2.2.2 CH_4 柱浓度空间分布

2.2.2.1 CH_4 柱浓度月均值空间分布

2019—2023 年河南省 CH_4 柱浓度月均值的空间分布如图 5 所示。由图 5 可见:1 月河南省东南大部分区域 CH_4 柱浓度维持较高水平,其余西部地区整体处于低值状态;3 月起 CH_4 柱浓度呈现减弱趋势,至 6 月全域 CH_4 柱浓度基本维持在中等偏低水平;进入 7 月后, CH_4 柱浓度虽有小幅抬升,但总体仍属污染低位波动阶段;8 月, CH_4 柱浓度空间格局发生显著转变,形成以西部、中部为核心的低值区,而东部、南部及东北部则成为污染较高区域;9 月出现 CH_4 柱浓度攀升态势,尤以东部区域增幅最为显著;10—11 月构成全年污染峰值时段,除西部高海拔地区受地形因素影响 CH_4 柱浓度相对较低外,大部分监测区域达到年度最高污染等级;进入冬季后,12 月 CH_4 柱浓度开始梯度下降,但南部及东部区域仍维持大面积高值分布。总体上,该时空分布与文献 [35] 中我国 CH_4 柱浓度呈东南高、西北低的空间格局相吻合,即河南省东部、南部及北部在夏秋季更易出现高值,而西部高海拔区全年相对低值。进一步研究表明,河南作为华中典型省份,其高值区受工业与城镇源、农业源及气象条件等叠加的综合作用,而西部低

值区与地形屏障和较低排放量有关。

2.2.2.2 CH_4 柱浓度季均值空间分布

基于模型预测的逐月数据结果按季节取平均值,得到 CH_4 柱浓度季节尺度空间分布情况(见图 6)。由图 6 可见:2019—2023 年河南省 CH_4 柱浓度季节性差异性显著,其中,春季 CH_4 柱浓度整体处于中等偏低污染水平;进入夏季后 CH_4 柱浓度除西部外,大部分地区均有不同程度的上升;随着秋季气温下降、农业生产活动增加以及植被生长状态的变化, CH_4 柱浓度的空间分布格局呈现“西低东高”分布态势,西部区域 CH_4 柱浓度整体偏低,而东部及中部大部分地区则处于较高污染水平,构成全年污染最严重的时段;冬季 CH_4 柱浓度呈梯度式衰减特征,大部分区域污染较秋季均有不同程度的减弱,污染核心区域向东南地区转移。

2.3 CH_4 柱浓度驱动因子分析

2.3.1 CH_4 柱浓度驱动因子的影响分析

为深入理解研究区 CH_4 柱浓度驱动因子的重要性及其贡献程度,采用 SHAP 方法对各驱动因子的贡献进行量化分析。分析数据为 2019—2023 年 CH_4 柱浓度的月均值, CH_4 柱浓度模型中各驱动因子 SHAP 值的分布情况如图 7 所示。

由图 7(a) 可见,2 m 地面温度、低植被叶面积指数和废弃物源排放量是 CH_4 柱浓度的 3 个最重要的驱动因子,其中 2 m 地面温度高值对模型的输出主要

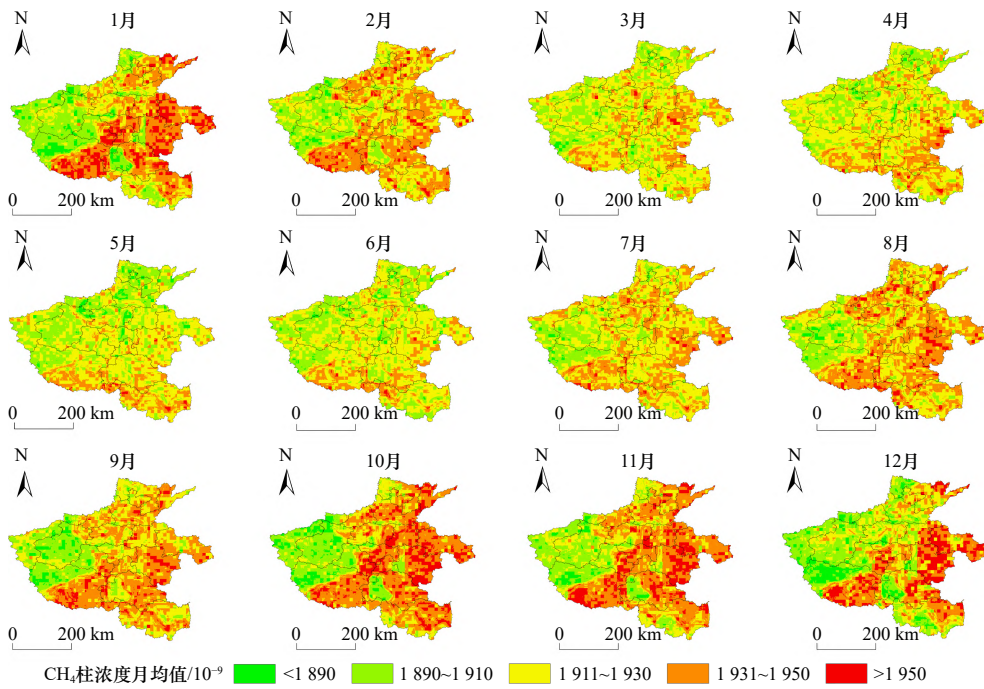


图5 河南省 CH₄ 柱浓度月均值空间分布特征

Fig.5 Spatial distribution characteristics of monthly mean CH₄ column concentration in Henan Province

起提升作用,而低值对模型的输出抑制作用显著。这一结论与实际情况相契合,温度高的地区更容易形成有利于 CH₄ 释放的厌氧环境,从而促进 CH₄ 的释放^[53]。低植被叶面积指数表示区域内低矮植被的分布密度,主要包括农田、草地等非林地类型,是农业活动强度的重要代理驱动因子^[54-55]。低植被叶面积指数的低值所对应的 SHAP 值主要集中在正区间,表明较低的低植被叶面积指数对 CH₄ 柱浓度具有正向推动作用,这与非灌溉排放源排放量或人为干扰更相关;低植被叶面积指数高值集中在 SHAP 值为负或接近 0 的区间,表明密集的低矮植被(如健康草地、农田)并不一定促进 CH₄ 柱浓度的上升,这可能得益于植被覆盖较好的区域一般具有更好的生态调节作用。

各驱动因子的平均 SHAP 值分布如图 7(b) 所示,用以展示各驱动因子在 SHAP 分析中对模型预测结

果的总体贡献度及其影响方向。由图 7(b) 可见, 2 m 地面温度的平均 SHAP 值最高,表明其在模型中占主导地位,且对 CH₄ 柱浓度具有显著的正向驱动作用。排放清单贡献程度由高到低分别为废弃物源排放量、工业燃烧源排放量、农业源排放量、燃料开采源排放量和运输源排放量,从影响方向来看,废弃物源排放量、工业燃烧源排放量和农业源排放量均对 CH₄ 柱浓度具有正向影响,而燃料开采源排放量和运输源排放量对 CH₄ 柱浓度增加的作用有限。这一现象可能与排放的时空特征及大气扩散过程有关。燃料开采源排放量与运输源排放量通常具有明显的局地性,其排放集中于特定区域或短时效事件中,可能导致近地层 CH₄ 浓度的瞬时升高^[56]。然而,在更大空间范围或长时间平均尺度上,由于边界层混合、风场输送等气象过程的稀释作用,局地 CH₄ 被迅速扩散并趋于背景浓度水平,削弱了这两类排放源排放量对整层

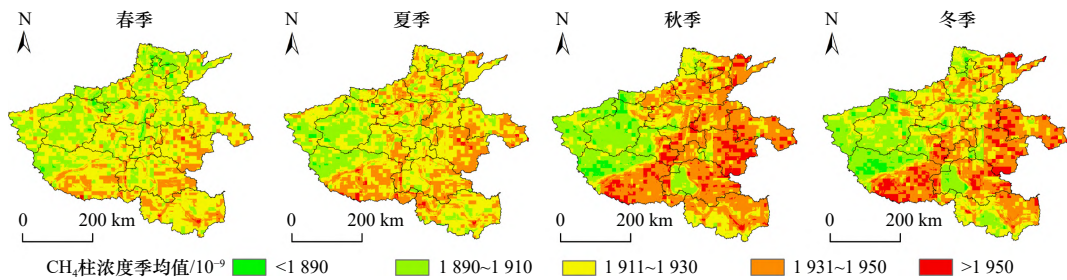
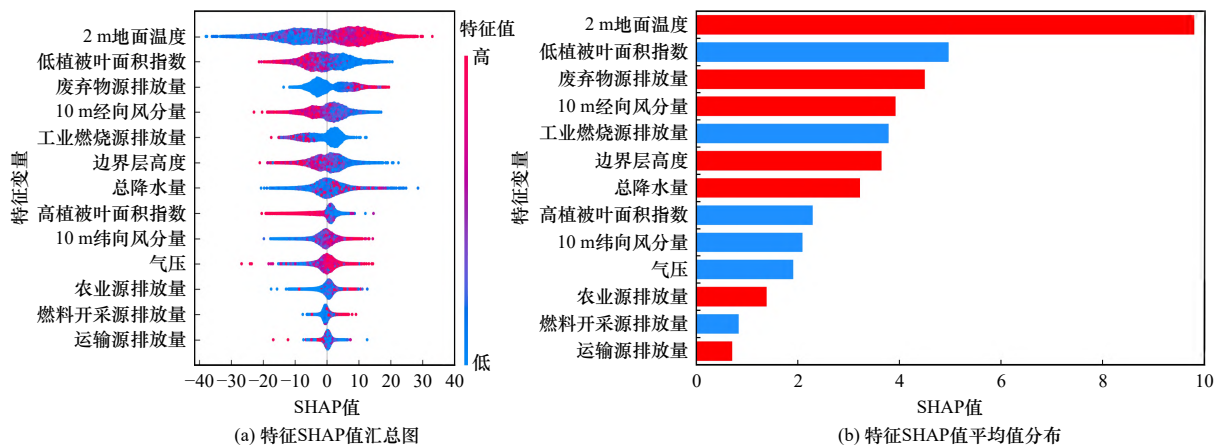


图6 河南省 CH₄ 柱浓度季节尺度空间分布特征

Fig.6 Spatial distribution characteristics of seasonal CH₄ column concentration in Henan Province



注: 图 (b) 中红色柱形表示对 CH_4 柱浓度的正向 (升高预测) 影响, 蓝色柱形表示负向 (拉低预测) 影响。

图7 河南省 CH_4 柱浓度驱动因子 SHAP 值分析图

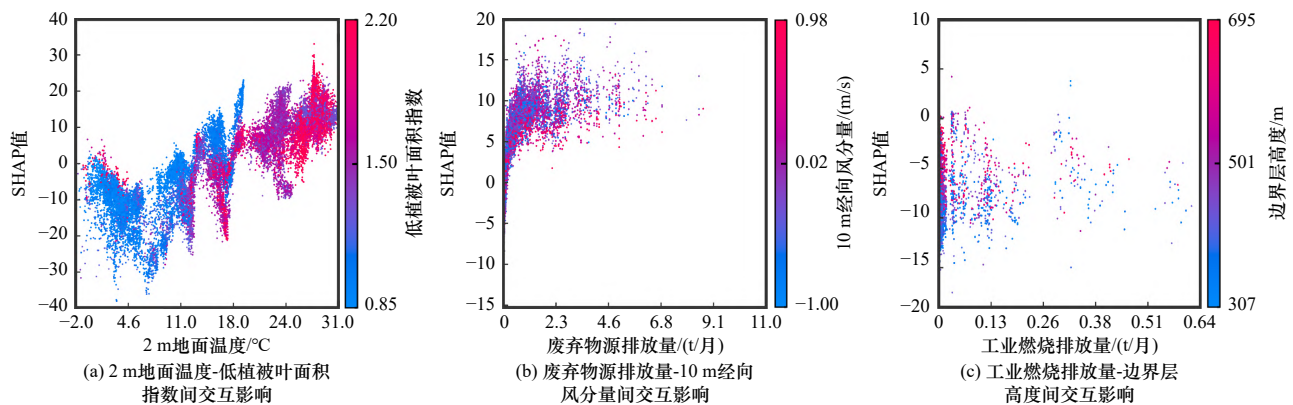
Fig.7 SHAP value analysis plot of driving factors for CH_4 column concentration in Henan Province

CH_4 柱浓度的总体提升效应。因此, 燃料开采源排放量与运输源排放量在模型中的平均 SHAP 值为负。

2.3.2 驱动因子间的交互作用

以 2 m 地面温度、低植被叶面积指数、废弃物源

排放量、10 m 经向风分量、工业燃烧源排放量和边界层高度前 6 个重要驱动因子为例, 进一步揭示各驱动因子间非线性作用对 CH_4 柱浓度形成的复合作用机制, 结果如图 8 所示。



注: 6 个驱动因子数值均为标准化值映射还原后的数值。

图8 6 个重要因子交互作用

Fig.8 Interaction plot of six key factors

2 m 地面温度与低植被叶面积指数之间的交互作用关系如图 8(a) 所示。当 2 m 地面温度 (24~31 °C) 与低植被叶面积指数 (>0.5) 同时较高时, SHAP 值显著为正, 表明高温与密集低矮植被共同增强了 CH_4 柱浓度的释放; 相反, 在 2 m 地面温度值 (-2~11 °C) 与低植被叶面积指数 (<0.5) 均较低时, SHAP 值整体偏低, 说明低温稀疏低矮植被环境不利于 CH_4 释放。

废弃物源排放量与 10 m 经向风分量之间的交互作用对 CH_4 柱浓度的影响如图 8(b) 所示。当废弃物源排放量处于极低值 (约 0.00 t/月) 时, SHAP 值呈现正负分布特征, 说明在废弃物 CH_4 排放较少的情况下, CH_4 柱浓度易受 10 m 经向风的影响; 随着废弃物源排放量升高, 10 m 经向风分量的变化范围明显增

大, 且 SHAP 值整体呈上升趋势, 显示出二者交互对 CH_4 柱浓度的正向推动效应逐渐增强。

由图 8(c) 可见, 工业燃烧源排放量与边界层高度之间存在明显的交互效应。高边界层高度对应的 SHAP 值普遍较大, 表明在强对流和混合条件下, 工业燃烧源排放量对 CH_4 柱浓度的正向贡献更为显著。这一结果反映出柱浓度的垂直积分特性, 即较高的边界层高度增强了地表 CH_4 的上输能力, 使其在对流层中上层积聚并被卫星敏感波段探测到, 从而提升了整层柱浓度的反演值; 相反, 在边界层高度较低条件下, CH_4 主要滞留于近地层, 虽局地浓度较高, 但由于垂直传播受限, 卫星观测的整层柱浓度贡献有限, 对应的 SHAP 值较低。

3 结论

a) XGBoost 模型较 OLS 模型、RF 模型、Bagging 模型、LightGBM 模型和 CatBoost 模型更能刻画河南省 CH₄ 柱浓度与气象因子和人类活动间的复杂关系,在测试集上的 R^2 、RMSE 和 MAE 分别为 0.964、 4.78×10^{-9} 、 3.11×10^{-9} 。

b) 河南省 CH₄ 柱浓度月均值整体虽呈现出一定的上升,但上升趋势并不显著,并叠加了较为稳定的年周期性特征(季节性强度为 0.756 8)。这种季节循环表现为“春季下降、夏季上升、秋季达峰、冬季回落”模式:CH₄ 柱浓度自 3 月起持续升高,于 9—10 月达到年度峰值,随后在冬季降至低位(谷值多出现在 2—3 月)。空间维度上,污染格局呈现明显的季节分异,春季整体柱浓度较低,仅东南局部地区偏高;夏季除西部地区外,大部分区域柱浓度逐步上升;秋季达到全年峰值,空间上呈现典型的“西低东高”格局;冬季柱浓度普遍较秋季低,高值区收缩并转移至东南地区。因此,建议于东南部工业密集区强化能源结构优化和工业过程管控,推动绿色低碳发展。

c) 2 m 地面温度、低植被叶面积指数和废弃物源排放量是影响 CH₄ 柱浓度的重要驱动因子。高温显著促进 CH₄ 释放,低植被叶面积指数低值区对浓度具有正向作用,而低植被叶面积指数高值区因生态调节效应正向驱动效应减弱。在主要排放源中,废弃物源排放量、运输源排放量和农业源排放量是正向驱动的主力;可能受大气扩散稀释作用的影响,工业燃烧源排放量和燃料开采源排放量的平均 SHAP 值为负,对 CH₄ 柱浓度的区域贡献减弱。此外,驱动因子间的交互作用机制显著,其中,2 m 地面温度与低植被覆盖度协同可促进 CH₄ 柱浓度的生成,在高温密集低植被覆盖度条件下交互效应更加突出;废弃物源排放量与经向风分量协同提升浓度;而工业燃烧源排放量与边界层高度则存在拮抗效应,低边界层高度在工业燃烧高排放时会影响 CH₄ 的垂直扩散,导致卫星反演的柱浓度偏低。该分析揭示了自然因素与人为排放源排放量对区域 CH₄ 浓度的差异化贡献及复杂交互关系。

参考文献 (References):

[1] SAUNIOS M,MARTINEZ A,POULTER B,et al.Global methane budget 2000-2020[J].Earth System Science Data,2025,17(5): 1873-1958.

[2] VOULGARAKIS A,NAIK V,LAMARQUE J F,et al.Analysis of present day and future OH and methane lifetime in the ACCMIP simulations[J].Atmospheric Chemistry & Physics,2013,13(5):

2563-2587.

- [3] ETMINAN M,MYHRE G,HIGHWOOD E J,et al.Radiative forcing of carbon dioxide,methane,and nitrous oxide:a significant revision of the methane radiative forcing[J].Geophysical Research Letters,2016,43(24):12614-12623.
- [4] PÉREZ-DOMÍNGUEZ I,del PRADO A,MITTENZWEI K,et al.Short- and long-term warming effects of methane may affect the cost-effectiveness of mitigation policies and benefits of low-meat diets[J].Nature Food,2021,2(12):970-980.
- [5] HOPE C.The climate change benefits of reducing methane emissions[J].Climatic Change,2005,68(1):21-39.
- [6] USSIRI D,LAL R.Historical and contemporary global methane cycling[M].Cham:Springer International Publishing,2017:227-285.
- [7] ZHANG K,KOGELSCHATZ U,ELIASSON B.Conversion of greenhouse gases to synthesis gas and higher hydrocarbons[J].Energy & Fuels,2001,15(2):395-402.
- [8] FILONCHYK M,PETERSON M P,ZHANG L F,et al.Greenhouse gases emissions and global climate change:examining the influence of CO₂,CH₄,and N₂O[J].Science of the Total Environment,2024,935:173359.
- [9] BEERLING D,BERNER R A,MACKENZIE F T,et al.Methane and the CH₄ related greenhouse effect over the past 400 million years[J].American Journal of Science,2009,309(2):97-113.
- [10] YUSUF R O,NOOR Z Z,ABBA A H,et al.Methane emission by sectors:a comprehensive review of emission sources and mitigation methods[J].Renewable and Sustainable Energy Reviews,2012,16(7):5059-5070.
- [11] KHOLOD N,EVANS M,PILCHER R C,et al.Global methane emissions from coal mining to continue growing even with declining coal production[J].Journal of Cleaner Production,2020, 256:120489.
- [12] WEI Y Y,YANG X J,QIU X T,et al.Spatio-temporal variations of atmospheric methane and its response to climate on the Tibetan Plateau from 2010 to 2022[J].Atmospheric Environment,2023, 314:120088.
- [13] REAY D S,SMITH P,CHRISTENSEN T R,et al.Methane and global environmental change[J].Annual Review of Environment and Resources,2018,43:165-192.
- [14] SKYTT T,NIELSEN S N,JONSSON B G.Global warming potential and absolute global temperature change potential from carbon dioxide and methane fluxes as indicators of regional sustainability:a case study of Jämtland,Sweden[J].Ecological Indicators,2020,110:105831.
- [15] BAMBERGER I,STIEGER J,BUCHMANN N,et al.Spatial variability of methane:attributing atmospheric concentrations to emissions[J].Environmental Pollution,2014,190:65-74.
- [16] XIA L J,ZHANG G,ZHAN M J,et al.Seasonal variations of atmospheric CH₄ at Jingdezhen Station in Central China: understanding the regional transport and its correlation with CO₂ and CO[J].Atmospheric Research,2020,241:104982.

- [17] LESTARI D O, ERNI, ISKANDAR I. Temporal characteristic of the methane (CH_4) concentration observed in Bukit kototabang, West Sumatra during 2004-2013 [J]. American Institute of Physics Conference Series, 2017, 1801: 060001.
- [18] 徐燕飞, 陈永春, 李静, 等. 煤电基地 CO_2 和 CH_4 遥感监测及时空特征分析 [J]. 煤田地质与勘探, 2024, 52(6): 79-90.
XU Y F, CHEN Y C, LI J, et al. Remote sensing monitoring and spatiotemporal characteristics of CO_2 and CH_4 concentrations in coal-electricity production bases [J]. Coal Geology & Exploration, 2024, 52(6): 79-90.
- [19] 李隆, 杜宁, 邓小东, 等. 基于梯度提升算法的山东省 2019 年 SO_2 污染物时空变化分析 [J]. 环境科学研究, 2025, 38(6): 1241-1251.
LI L, DU N, DENG X D, et al. Estimation of near-surface SO_2 concentration in Shandong Province, 2019 using the gradient boosting algorithm [J]. Research of Environmental Sciences, 2025, 38(6): 1241-1251.
- [20] KARAKURT I, AYDIN G, AYDINER K. Sources and mitigation of methane emissions by sectors: a critical review [J]. Renewable Energy, 2012, 39(1): 40-48.
- [21] 宋晓伟, 孔媛, 郝永佩, 等. 山东省 $\text{PM}_{2.5}$ 与 O_3 浓度演变特征及影响因素分析 [J]. 环境科学研究, 2025, 38(9): 1906-1919.
SONG X W, KONG Y, HAO Y P, et al. Evolutionary characteristics and influencing factors of $\text{PM}_{2.5}$ and O_3 concentrations in Shandong Province [J]. Research of Environmental Sciences, 2025, 38(9): 1906-1919.
- [22] ZHENG B, COHEN J B, LU L, et al. How can we trust TROPOMI based methane emissions estimation: calculating emissions over unidentified source regions [J]. EGUSphere, 2025, 2025: 1-24.
- [23] BORSORFF T, MARTINEZ-VELARTE M C, SNEEP M, et al. Random forest classifier for cloud clearing of the operational TROPOMI XCH_4 product [J]. Remote Sensing, 2024, 16(7): 1208.
- [24] BARRÉ J, ABEN I, AGUSTÍ-PANAREDA A, et al. Systematic detection of local CH_4 anomalies by combining satellite measurements with high-resolution forecasts [J]. Atmospheric Chemistry and Physics, 2021, 21(6): 5117-5136.
- [25] WU X D, ZHANG X Y, CHUAI X W, et al. Long-term trends of atmospheric CH_4 concentration across China from 2002 to 2016 [J]. Remote Sensing, 2019, 11(5): 538.
- [26] 肖钟湧, 林晓凤, 高翔, 等. 2010—2019 年中国 CH_4 时空变化及影响因素分析 [J]. 环境科学与技术, 2023, 46(10): 147-155.
XIAO Z Y, LIN X F, GAO X, et al. Study on the temporal and spatial variation of CH_4 and its driving factors over China from 2010 to 2019 [J]. Environmental Science & Technology (China), 2023, 46(10): 147-155.
- [27] RIBEIRO I O, de SOUZA R A F, ANDREOLI R V, et al. Spatiotemporal variability of methane over the Amazon from satellite observations [J]. Advances in Atmospheric Sciences, 2016, 33(7): 852-864.
- [28] LI L M, LEI L P, SONG H, et al. Spatiotemporal geostatistical analysis and global mapping of CH_4 columns from GOSAT observations [J]. Remote Sensing, 2022, 14(3): 654.
- [29] PARK S, MOON K, JEOM H J, et al. Machine learning-based prediction of ambient CO_2 and CH_4 concentrations with high temporal resolution in Seoul metropolitan area [J]. Environmental Pollution, 2025, 376: 126362.
- [30] CORTINA J M. Interaction, nonlinearity, and multicollinearity: implications for multiple regression [J]. Journal of Management, 1993, 19(4): 915-922.
- [31] LIU M, LEI L P, LIU D, et al. Geostatistical analysis of CH_4 columns over monsoon Asia using five years of GOSAT observations [J]. Remote Sensing, 2016, 8(5): 361.
- [32] LIAO B, ZHOU T, LIU Y P, et al. Tackling the wildfire prediction challenge: an explainable artificial intelligence (XAI) model combining extreme gradient boosting (XGBoost) with SHapley additive exPlanations (SHAP) for enhanced interpretability and accuracy [J]. Forests, 2025, 16(4): 689.
- [33] ANTWARG L, MILLER R M, SHAPIRA B, et al. Explaining anomalies detected by autoencoders using Shapley Additive Explanations [J]. Expert Systems with Applications, 2021, 186: 115736.
- [34] VEGA GARCÍA M, AZNARTE J L. Shapley additive explanations for NO_2 forecasting [J]. Ecological Informatics, 2020, 56: 101039.
- [35] XU J H, LI W T, XIE H M, et al. Long-term trends and spatiotemporal variations in atmospheric XCH_4 over China utilizing satellite observations [J]. Atmosphere, 2022, 13(4): 525.
- [36] LI K, BAI K X, JIAO P L, et al. Developing unbiased estimation of atmospheric methane via machine learning and multiobjective programming based on TROPOMI and GOSAT data [J]. Remote Sensing of Environment, 2024, 304: 114039.
- [37] SCHNEISING O, BUCHWITZ M, REUTER M, et al. A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor [J]. Atmospheric Measurement Techniques, 2019, 12(12): 6771-6802.
- [38] LINDQVIST H, KIVIMÄKI E, HÄKKILÄ T, et al. Evaluation of sentinel-5P TROPOMI methane observations at northern high latitudes [J]. Remote Sensing, 2024, 16(16): 2979.
- [39] CRIPPA M, GUIZZARDI D, PAGANI F, et al. GHG emissions of all world countries [J]. Publications Office of the European Union, 2023, 10: 953322.
- [40] 浦静姣, 徐宏辉, 顾骏强, 等. 气象因素对长三角背景地区甲烷浓度的影响分析 [J]. 环境科学, 2013, 34(3): 835-841.
PU J J, XU H H, GU J Q, et al. Impacts of meteorological factors on atmospheric methane mole fractions in the background area of Yangtze River Delta [J]. Environmental Science, 2013, 34(3): 835-841.
- [41] 张世勃, 曹闪闪, 胡礼庭, 等. 亚洲季风区大气 CH_4 浓度时空变化特征与影响因素 [J]. 应用生态学报, 2021, 32(4): 1406-1416.
ZHANG S Q, CAO S S, HU L T, et al. Spatio-temporal variation of atmospheric CH_4 concentration and its driving factors in monsoon Asia [J]. Chinese Journal of Applied Ecology, 2021, 32(4): 1406-1416.
- [42] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system

- [C].San Francisco California USA:ACM,2016:785-794.
- [43] CHEN T,HE T,BENESTY M,et al.Xgboost:extreme gradient boosting[EB/OL].Washington DC:CRAN,(2025-03-05)[2026-03-04].<https://cran.r-project.org/package=xgboost>.
- [44] LUNDBERG S M,LEE S I.A unified approach to interpreting model predictions[J].Advances in neural information processing systems,2017,30:627-637.
- [45] LUNDBERG S M,ERION G G,LEE S I.Consistent individualized feature attribution for tree ensembles[EB/OL].Ithaca,New York:arXiv.org,(2025-03-05)[2026-03-04].<https://doi.org/10.48550/arXiv.1802.03888>.
- [46] HELSEL D R,HIRSCH R M.Statistical methods in water resources[M].Amsterdam:Elsevier,1993.
- [47] BOX G E,JENKINS G M,REINSEL G C,et al.Time series analysis:forecasting and control[M].New York:John Wiley & Sons,2015.
- [48] XU W H,LI Q X,WANG X L,et al.Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices[J].Journal of Geophysical Research: Atmospheres,2013,118(17):9708-9720.
- [49] HOOKER G,MENTCH L,ZHOU S Y.Unrestricted permutation forces extrapolation:variable importance requires at least one more model,or there is no free variable importance[J].Statistics and Computing,2021,31(6):82.
- [50] DAOUD J I.Multicollinearity and regression analysis[C].Kuala Lumpur,Malaysia:IOP Publishing,2017:012009.
- [51] HASELBECK F,JOHN M,GRIMM D G.easyPheno:an easy-to-use and easy-to-extend Python framework for phenotype prediction using Bayesian optimization[J].Bioinformatics Advances,2023,3(1):vbad035.
- [52] SEABOLD S,PERKTOLD J.Statsmodels:econometric and statistical modeling with Python[C].Texas:SciPy,2010:92-96.
- [53] OLDS R H,SAGE B H,LACEY W N.Phase equilibria in hydrocarbon systems.composition of the dew-point gas of the methane-water system[J].Industrial & Engineering Chemistry, 1942,34(10):1223-1227.
- [54] PETROPOULOS G P,KALAITZIDISZ C.Multispectral vegetation indices in remote sensing:an overview[J].Ecological Modelling,2012,2:15-39.
- [55] WANG G F,LIU P,HU J M,et al.Spatiotemporal patterns and influencing factors of agriculture methane emissions in China[J]. Agriculture,2022,12(10):1573.
- [56] JACOB D J,TURNER A J,MAASAKKERS J D,et al.Satellite observations of atmospheric methane and their value for quantifying methane emissions[J].Atmospheric Chemistry and Physics,2016,16(22):14371-14396.

(责任编辑:刘 方)