

项目论文

基于多源城市空间数据融合的 PM2.5 时空预测 与污染暴露风险评估

摘 要

细颗粒物 (PM_{2.5}) 浓度同时受污染排放、气象条件与城市建成环境影响, 准确描述其时间波动和空间差异, 对污染预警与区域风险识别具有重要意义。本文基于 UCI 北京多站点空气质量数据与 OpenStreetMap 公开空间要素, 建立从严格时间切分、多步长预测到空间风险分析的完整研究流程。

(1) 数据构建阶段, 对 12 个监测站的逐小时污染物与气象序列执行因果前向填充和训练期统计量变换, 融合道路、绿地、建筑等静态空间特征; 以 48 小时历史窗口构造 1、6 和 24 小时预测标签, 并将标签限定在同一时间阶段内, 最终得到 8,265 个训练样本、5,833 个验证样本和 4,321 个留出测试样本。

(2) 模型比较阶段, 在相同数据和评价协议下比较持续性基线、24 小时季节朴素基线、LSTM、Transformer 与时空图卷积网络 (STGCN), 每个核心模型使用三个随机种子重复训练。按预先冻结的三步长平均 RMSE 规则, STGCN 为最优模型, 总体 RMSE 为 53.748; 其 1、6 和 24 小时 RMSE 分别为 21.881 ± 0.214 、 53.655 ± 0.735 和 $85.708 \pm 1.683 \mu\text{g}/\text{m}^3$, 较最强朴素基线的三步长平均 RMSE 降低 17.06%。

(3) 风险与解释阶段, 基于留出期预测构建空间网格风险指标, 并使用 Moran's I 、局部空间自相关和替代权重情景评估稳健性。全局 Moran's I 为 0.713, 999 次置换检验 $p = 0.001$; 四组替代权重与基准排序的 Spearman 系数为 0.969–0.989。独立 LightGBM 解释伴随模型的 SHAP 结果显示, 预测原点 PM_{2.5} 贡献居首且整体为正向, 露点温度居第二且整体为正向, 年内周期余弦特征居第三并呈周期性非线性关系; 风速的整体贡献方向为负。

本文将多步长时序预测、分站点误差诊断、空间聚集检验与可解释分析纳入同一证据链, 为城市 PM_{2.5} 预警和区域污染风险筛查提供了可复现的方法参考。

关键词: PM_{2.5}; 时空预测; 时空图卷积网络; 多源空间数据; 空间自相关; SHAP; 暴露风险

ABSTRACT

Fine particulate matter (PM_{2.5}) is jointly affected by emissions, meteorological conditions, and the urban built environment. Accurate characterization of its temporal variability and spatial heterogeneity is important for pollution warning and regional risk screening. Using the UCI Beijing Multi-Site Air Quality data and public OpenStreetMap features, this study develops an integrated workflow covering leakage-controlled temporal splitting, multi-horizon forecasting, and spatial risk analysis.

(1) In the data-construction stage, hourly pollutant and meteorological series from 12 monitoring stations were processed using causal forward filling and training-period statistics, then combined with static road, green-space, and building features. A 48-hour history window was used to construct 1-, 6-, and 24-hour targets, with every target constrained to remain within one temporal split. The resulting data contained 8,265 training, 5,833 validation, and 4,321 held-out test samples.

(2) In the model-comparison stage, persistence, 24-hour seasonal naive, LSTM, Transformer, and spatiotemporal graph convolutional network (STGCN) models were evaluated under the same protocol, with three random seeds for each core model. Under the pre-specified mean-RMSE selection rule, STGCN was the best model, with an overall RMSE of 53.748. Its RMSE values at 1, 6, and 24 hours were 21.881 ± 0.214 , 53.655 ± 0.735 , and $85.708 \pm 1.683 \mu\text{g}/\text{m}^3$, respectively, representing a 17.06% reduction in mean RMSE relative to the strongest naive baseline.

(3) In the risk-and-interpretation stage, held-out predictions were mapped to spatial grids and evaluated using Moran's I , local spatial autocorrelation, and alternative weighting scenarios. Global Moran's I was 0.713 with a permutation p -value of 0.001; rank correlations between the baseline risk index and four alternative scenarios ranged from 0.969 to 0.989. SHAP analysis of an independent LightGBM companion model ranked current PM_{2.5} first with an overall positive contribution, dew-point temperature second with an overall positive contribution, and the annual-cycle cosine term third with a nonlinear periodic pattern; wind speed showed an overall negative contribution.

By integrating multi-horizon forecasting, station-level error diagnosis, spatial clustering tests, and interpretable analysis into one evidence chain, this study provides a reproducible methodological reference for urban PM_{2.5} warning and regional pollution-risk screening.

Keywords: PM_{2.5}; spatiotemporal forecasting; STGCN; multisource spatial data; spatial autocorrelation; SHAP; exposure risk

目 录

摘要	I
ABSTRACT	II
目录	III
图目录	VI
表目录	VII
1 绪论	1
1.1 研究背景及意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 国内外研究现状	2
1.2.1 统计学习与传统机器学习研究	2
1.2.2 深度时序与时空图建模研究	3
1.2.3 污染暴露与空间风险评估研究	3
1.3 研究目标及内容	4
1.3.1 研究目标	4
1.3.2 研究内容	4
1.4 章节安排	4
1.5 技术路线图	5
2 相关研究与理论基础	6
2.1 PM2.5 形成机制与多源影响因素	6
2.1.1 污染物与气象因素	6
2.1.2 时间周期与滞后效应	6
2.1.3 建成环境与空间邻近关系	7
2.2 时序预测理论基础	7
2.2.1 监督学习滑动窗口	7
2.2.2 循环神经网络与长短期记忆	7
2.2.3 自注意力与 Transformer 编码器	8
2.3 图神经网络与时空建模基础	8
2.3.1 空间图与邻接矩阵	8
2.3.2 图卷积与自适应邻接	9
2.3.3 时空特征融合机制	9
2.4 本章小结	9

3	数据集与时序研究设计	10
3.1	数据来源与变量体系	10
3.1.1	北京多站点空气质量数据	10
3.1.2	监测变量与预测目标	10
3.1.3	数据质量与缺失分布	11
3.2	OSM 空间要素与图结构构建	11
3.2.1	道路、建筑、绿地与 POI 要素	11
3.2.2	站点空间特征汇总	11
3.2.3	地理邻接与空间支持边界	12
3.3	时序预处理与空间风险研究设计	12
3.3.1	因果缺失填补与训练期标准化	12
3.3.2	时间切分、多步窗口与测试集锁定	12
3.3.3	风险指标、权重方案与空间统计设计	13
3.4	本章小结	14
4	时序预测方法	15
4.1	朴素基线与 LSTM 模型	15
4.1.1	持续性与 24 小时季节朴素基线	15
4.1.2	LSTM 网络结构	15
4.1.3	多步预测输出	16
4.2	Transformer 编码器模型	16
4.2.1	时间位置编码	16
4.2.2	因果自注意力编码	16
4.2.3	多预测步长解码	17
4.3	STGCN 时空融合模型与训练策略	17
4.3.1	图卷积与时间卷积模块	17
4.3.2	OSM 静态特征与自适应邻接融合	17
4.3.3	模型选择、重复种子与评价指标	18
4.4	本章小结	19
5	实验结果与误差分析	20
5.1	训练收敛与基线比较	20
5.1.1	训练过程与早停结果	20
5.1.2	朴素基线对比	20
5.1.3	三类深度模型总体比较	21
5.2	多步长预测与分站点误差	22
5.2.1	1 小时预测结果	22
5.2.2	6 小时与 24 小时预测结果	22

5.2.3	分站点误差与典型时段诊断	22
5.3	消融、解释与空间风险稳健性	24
5.3.1	OSM 静态特征与邻接结构消融	24
5.3.2	LightGBM 解释伴随模型与 SHAP 结果	24
5.3.3	风险权重敏感性与空间自相关	24
5.4	本章小结	25
6	结论与展望	28
6.1	研究结论	28
6.1.1	时空预测主要结论	28
6.1.2	污染暴露风险主要结论	28
6.2	研究局限	28
6.2.1	数据时段与区域边界	28
6.2.2	模型与评价边界	28
6.2.3	风险指标与因果解释边界	29
6.3	未来展望	29
6.3.1	跨城市外部验证	29
6.3.2	概率预测与极端污染预警	29
6.3.3	动态暴露与多尺度空间建模	29
	参考文献	30

图目录

1.1	技术路线图	5
3.1	时间切分、多步窗口与数据边界	13
5.1	三类核心模型的训练与验证损失	20
5.2	朴素基线与核心预测模型的留出集比较	21
5.3	STGCN 在典型时段的预测一致性与残差诊断	23
5.4	STGCN 的多步长分站点误差与空间信息消融	23
5.5	独立 LightGBM 伴随模型的 SHAP 特征贡献与依赖关系	25
5.6	空间污染风险分布、排序与权重敏感性	26
5.7	空间风险指标的 Moran 散点、置换检验与 LISA 聚集	27

表目录

3.1	动态监测变量与处理方式	10
3.2	原始监测变量缺失比例	11
3.3	时间切分与多步窗口设置	13
4.1	候选模型的结构与信息路径对照	19
5.1	不同模型的 PM2.5 测试集 RMSE 比较 ($\mu\text{g}/\text{m}^3$)	21
5.2	STGCN 的 PM2.5 测试集性能	22
5.3	STGCN 空间信息消融的 PM2.5 测试集 RMSE (随机种子 42)	24

1 绪论

1.1 研究背景及意义

1.1.1 研究背景

细颗粒物（PM_{2.5}）能够长时间悬浮于大气，并通过呼吸系统进入人体。既有流行病学与环境健康研究表明，长期颗粒物暴露与呼吸系统、心血管系统及全因死亡风险密切相关^[1-3]。世界卫生组织发布的全球空气质量指南进一步收紧了 PM_{2.5} 浓度建议值，凸显出细颗粒物污染控制与健康风险管理的长期性^[4-5]。污染源构成、燃料类型和空间尺度还会共同影响颗粒物暴露及其可归因健康损失^[6]，因此仅以单一时间点或单一站点的浓度描述城市污染，难以充分刻画污染过程的动态性和空间异质性。

中国城市 PM_{2.5} 污染具有显著的区域差异、季节变化和污染过程聚集特征^[7]。2013 年以来，能源、产业、交通和末端治理等多项措施推动了全国 PM_{2.5} 水平下降，但改善幅度及其驱动因素在城市间并不一致^[8]。北京空气质量改善的经验也提示，对长期趋势的判断必须区分气象波动、排放变化和观测尺度，不能以少量时段替代完整时间序列^[9]。因此，面向城市精细治理的研究既要回答“未来若干小时浓度如何变化”，也要回答“不同站点和不同空间单元的风险如何分布”。

城市空气污染具有跨站点传播、日内周期、周内规律和季节变化等多重结构。北京多站点公开监测数据为统一分析长期时序与站点差异提供了基础^[10]。短时预测可为监测调度与污染过程跟踪提供提前量，中期预测则更关注污染累积、扩散条件变化和持续性风险。随着预测步长延长，误差通常会逐步累积；北京多站点研究也观察到多小时预测精度随领先时间增加而下降的现象^[11]。若仅对每个站点独立建模，模型难以利用相邻站点之间的同步变化；若只追求复杂网络而忽略朴素基线、时间切分和重复实验，又可能高估模型的实际预测能力。

现有研究已经从传统机器学习、循环神经网络、卷积—循环混合模型逐步发展到图神经网络与 Transformer。相关综述显示，空气质量预测的研究重点正在由单站点、单步长模型转向多源信息、多站点和多步长联合建模^[12-15]。然而，不同研究的数据时段、切分协议、输入变量与评价指标差异较大，模型间结果不宜脱离实验条件直接比较。基于此，本文在统一数据处理和留出测试协议下比较朴素基线、LSTM、Transformer 编码器和时空图卷积模型，并把预测结果与空间风险分析连接起来。

1.1.2 研究意义

在理论层面，本文把时间依赖、站点邻接和建成环境静态特征置于同一研究框架，比较纯时间模型与时空融合模型的差异，从而分析空间结构在多站点预测中的增量作用。研究同时设置 1、6 和 24 小时三个预测步长，并把训练期统计量拟合、严格时间切分、重复随机种子和消融实验纳入统一协议，避免以单一预测步长或单次运行概括全部性能。对于解释部分，本文使用独立 LightGBM 伴随模型开展 SHAP 分析，并将其与深度预测模型的内部机制明确区分。

在实践层面，预测结果能够为污染过程识别提供时间维度的信息，而空间风险分析能够补充不同区域的相对筛查线索。本文将北京多站点长期观测与 OpenStreetMap (OSM) 道路、建筑、绿地、水体和兴趣点要素结合，在监测站支持范围内构建探索性风险网格，再通过权重敏感性和空间自相关检验判断风险排序与聚集格局是否稳定。该设计可为监测资源布设、重点区域筛查和后续精细暴露评价提供量化参考，同时不把网格结果外推为个体暴露或因果效应。

1.2 国内外研究现状

1.2.1 统计学习与传统机器学习研究

空气质量预测早期多采用线性回归、支持向量机、树模型以及统计时间序列方法。土地利用回归通过道路、土地覆盖、人口活动和监测浓度之间的统计关系估计空间污染水平；已有研究尝试以遗传算法与支持向量机改进 PM_{2.5} 土地利用回归^[16]，也有研究基于自行车和步行监测构建城市黑碳土地利用回归模型^[17]。这些方法具有结构相对清晰、变量贡献便于分析等优点，但对非线性时间依赖和跨站点传播的表达能力有限。

在评价方面，时间序列模型不能直接照搬随机交叉验证。滚动窗口或顺序留出更符合预测时点的信息可得性^[18]；MAE、RMSE 等指标的选择也会影响模型排序与误差解释^[19]。经典预测工具强调基线模型、残差诊断和可复现的时间序列流程^[20]，Diebold–Mariano 检验则为比较预测误差序列提供了统计检验思路^[21]。这些研究共同说明，规范的评价协议与模型结构本身同等重要。

1.2.2 深度时序与时空图建模研究

LSTM 利用门控结构缓解长序列训练中的梯度消失问题，已广泛用于空气污染预测^[22]。北京 PM_{2.5} 研究比较了不同输入结构和预测时长下的深度模型^[23]，聚合式 LSTM 与结合优化算法的 LSTM 研究则进一步讨论了特征组合与参数搜索的作用^[24-25]。卷积与双向 LSTM、残差网络与领域适配等混合结构也被用于多站点预测^[26-27]。

Transformer 以自注意力替代循环递推，为长程依赖建模提供了另一条路径^[28]。近年来，稀疏注意力、分解—优化—Transformer 和图 Transformer 被应用于 PM_{2.5} 多步预测^[29-31]。与此同时，图卷积网络通过邻接矩阵聚合相邻节点信息^[32]，时空图卷积进一步交替建模空间传播和时间变化^[33]。多视图多任务图网络、时变图、因果卷积和多源图注意力等研究持续扩展空间关系的表达方式^[34-37]。总体看，复杂时空模型具有较强表征能力，但是否稳定优于更简单模型仍需在同一协议下验证。

1.2.3 污染暴露与空间风险评估研究

监测站点提供离散观测，空间风险分析通常需要借助插值或空间统计方法估计站点之间的变化。反距离加权法依据距离衰减进行插值，具有结果受观测值范围约束、实现透明等特点；自适应 IDW 可根据空间分布调整距离参数^[38]。但 IDW 对站点密度、参数和研究边界敏感，不应把站点稀疏区域的插值视为精确浓度^[39-40]。Moran's I 与局部空间关联指标可分别检验总体空间自相关和局部聚集类型^[41-42]，适合用于识别高一高、低—低聚集与空间异常单元。

OSM 为道路网络、建筑和兴趣点等建成环境要素提供了开放数据来源。研究表明，自愿地理信息的完整性和位置精度具有区域差异，使用前必须核查其覆盖范围和数据质量^[43-45]；OSMnx 等工具则为道路网络提取与图结构分析提供了可重复实现^[46]。已有北京空气污染可视分析工作展示了多站点时空数据在探索污染过程中的价值^[47]。因此，本文不把 OSM 要素直接解释为污染成因，而将其作为站点周边建成环境的静态描述，并通过消融和敏感性分析检验其作用。

1.3 研究目标及内容

1.3.1 研究目标

本文以北京多站点逐小时空气质量观测为基础，目标是在统一、可复核的时间边界下建立多步长 PM_{2.5} 时空预测体系，并进一步连接区域污染暴露风险的空间筛查。具体而言，一是比较不同时间与时空模型在 1、6 和 24 小时预测任务中的稳定表现；二是检验 OSM 静态要素和站点邻接结构是否提供增量信息；三是从独立解释伴随模型、误差诊断、权重敏感性和空间自相关四个角度分析结果边界。

1.3.2 研究内容

本文围绕“数据治理—多步预测—空间风险”展开。首先，对北京 12 个监测站 2013 年 3 月至 2017 年 2 月的逐小时空气质量与气象数据进行对齐、缺失处理和周期编码，并构建站点周边 OSM 空间特征。其次，以 48 小时历史窗口预测未来 1、6 和 24 小时的 PM_{2.5} 与 NO₂，在统一时间切分下比较持续性、24 小时季节朴素、LSTM、Transformer 编码器和 STGCN 模型。再次，通过去除 OSM 静态特征、仅保留地理邻接、重复随机种子和分站点误差分析检验结论的稳定性。最后，使用 LightGBM 伴随模型与 SHAP 分析变量贡献^[48-49]，并以 IDW、综合风险指数、权重敏感性和 Moran/LISA 统计描述空间风险格局。

研究采用文献研究、时序监督学习、对比实验、消融分析、模型解释与空间统计相结合的方法。所有预测指标均在锁定测试集上计算；模型优选遵循预先冻结的候选集、主指标和跨随机种子与预测步长聚合规则；空间风险结果限定为研究区内的探索性筛查。

1.4 章节安排

全文共分为六章。第 1 章说明研究背景、意义、国内外研究现状、研究目标与技术路线。第 2 章梳理多步时序预测、空间图建模、模型评价与空间风险分析的理论基础。第 3 章介绍数据来源、变量体系、OSM 空间要素、因果预处理、时间切分与空间风险研究设计。第 4 章给出朴素基线、LSTM、Transformer、STGCN、训练策略和评价指标。第 5 章报告重复随机种子实验、多步长与分站点误差、空间信息消融、独立 SHAP 解

释以及空间风险稳健性。第 6 章总结主要发现，讨论研究局限并提出后续拓展方向。

1.5 技术路线图

本文的技术路线如图1.1所示。图中四个任务带依次对应数据与空间特征构建、时间序列与空间风险研究设计、预测模型与训练评价以及误差解释与空间稳健性分析。每个任务带包含三个模块，每个模块进一步拆分为三个可核验步骤，使论文目录、代码模块和结果证据保持对应。

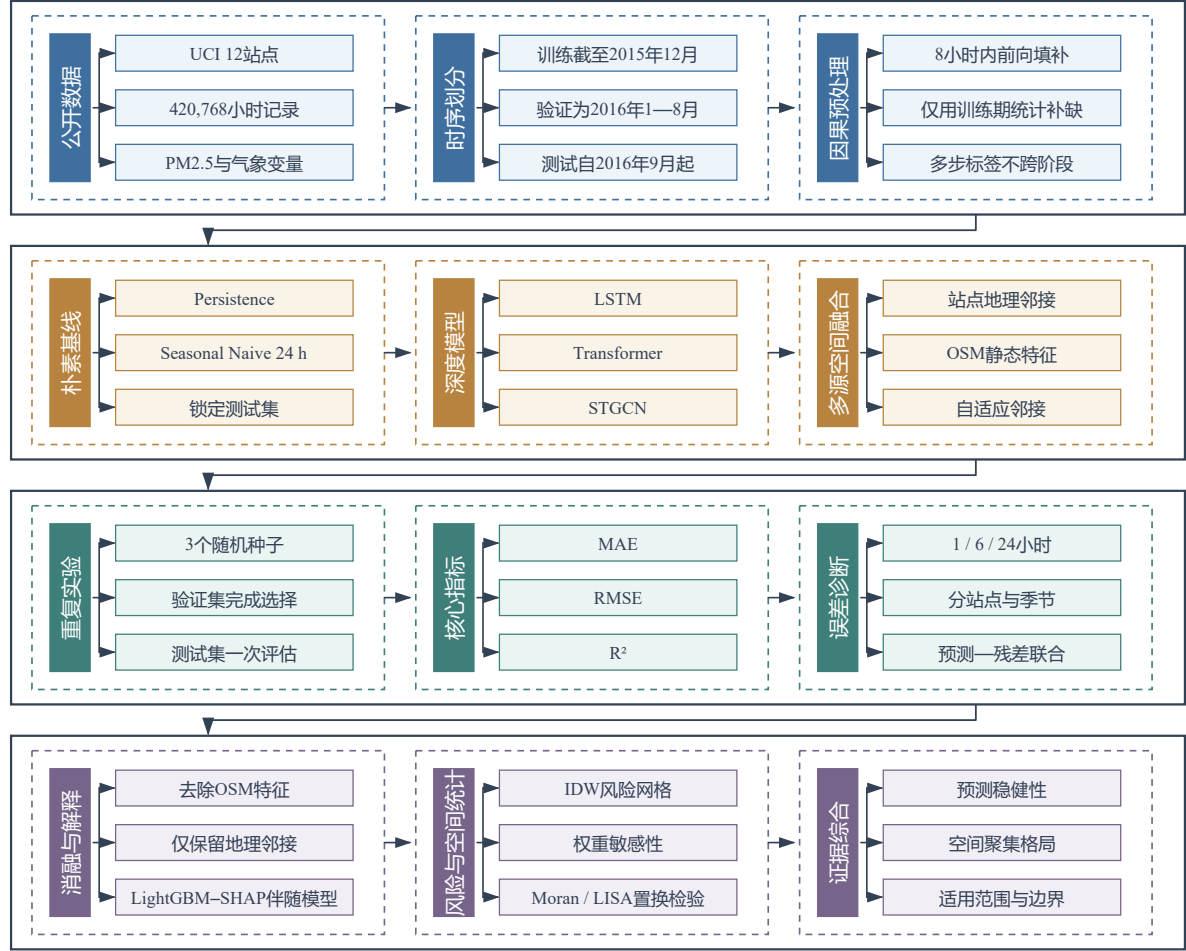


图 1.1 技术路线图

2 相关研究与理论基础

2.1 PM2.5 形成机制与多源影响因素

2.1.1 污染物与气象因素

PM2.5 浓度变化由一次排放、二次生成、区域输送和气象扩散共同决定。共存污染物 PM10、SO₂、NO₂、CO 与 O₃ 能够从不同侧面反映燃烧排放、交通活动与光化学过程；温度、气压、露点、降雨、风速和风向则影响污染物的稀释、沉降和输送。针对气象变量与污染浓度关系的研究指出，相关关系并不天然等同于因果作用，不同污染物和气象条件之间还可能存在非线性与滞后^[50]。因此，本文将污染物与气象变量作为联合预测输入，但不依据预测模型的特征权重作因果判断。

为兼顾数值稳定性与方向连续性，风向不直接作为 16 分类变量输入，而转换为正弦和余弦分量。若风向角为 θ_t ，则编码为

$$w_t^{(s)} = \sin \theta_t, \quad w_t^{(c)} = \cos \theta_t. \quad (1)$$

该表示能够避免 0 度与 360 度在数值空间中被错误地视为相距很远。

2.1.2 时间周期与滞后效应

城市污染序列常同时包含日内周期、工作日与休息日差异、季节变化以及连续污染事件的惯性。多站点、多小时空气质量预测研究表明，季节分组和预测领先时间会影响模型误差^[11]。本文在 48 小时历史窗口中保留两完整日的观测，并增加小时和年内日序的周期编码：

$$h_t^{(s)} = \sin \left(\frac{2\pi h_t}{24} \right), \quad h_t^{(c)} = \cos \left(\frac{2\pi h_t}{24} \right), \quad (2)$$

$$d_t^{(s)} = \sin \left(\frac{2\pi d_t}{365} \right), \quad d_t^{(c)} = \cos \left(\frac{2\pi d_t}{365} \right). \quad (3)$$

其中， h_t 为小时， d_t 为年内日序。周期编码不预设污染变化幅度，只为模型提供相邻周期位置的连续表示。

2.1.3 建成环境与空间邻近关系

道路、建筑、绿地、水体和兴趣点能够描述站点周边的建成环境与潜在人类活动强度。道路和兴趣点密集区域可能对应更高的交通与活动暴露，建筑密度较高、绿地比例较低的区域则可能具有不同的通风与沉降条件。土地利用回归研究说明，建成环境变量能够补充固定监测站之外的空间信息^[16-17]；但 OSM 质量研究同时强调，开放地理数据存在覆盖程度和精度差异^[43-44]。因此，本文在 1 km 和 3 km 缓冲区分别汇总 OSM 要素，并把这些变量定义为静态空间描述，而不是排放清单或个体暴露测量。

2.2 时序预测理论基础

2.2.1 监督学习滑动窗口

设 t 时刻全部站点的动态变量矩阵为 $\mathbf{X}_t \in \mathbb{R}^{N \times F}$ ，其中 N 为站点数， F 为动态特征数；静态空间特征为 $\mathbf{S} \in \mathbb{R}^{N \times P}$ 。给定历史长度 L 和预测步长集合 $\mathcal{H} = \{1, 6, 24\}$ ，单个监督学习样本写为

$$\mathcal{X}_t = [\mathbf{X}_{t-L+1}, \dots, \mathbf{X}_t], \quad \mathcal{Y}_t = [\mathbf{Y}_{t+h} : h \in \mathcal{H}], \quad (4)$$

其中 $L = 48$ ， $\mathbf{Y}$ 包含 PM2.5 与 NO₂ 两个目标。模型学习映射

$$\hat{\mathcal{Y}}_t = f_{\theta}(\mathcal{X}_t, \mathbf{S}, \mathbf{A}), \quad (5)$$

其中 $\mathbf{A}$ 为空间邻接矩阵，纯时间模型不使用该参数。时间序列验证必须保持训练时点早于验证和测试时点，以避免未来观测进入模型选择过程^[18]。

2.2.2 循环神经网络与长短期记忆

普通循环神经网络通过隐藏状态递推表达时间依赖，但在长序列训练中容易出现梯度消失。LSTM 引入输入门、遗忘门和输出门，控制新信息写入、历史记忆保留和当前状

态输出^[22]。其核心更新可写为

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i), \quad (6)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f), \quad (7)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c), \quad (8)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (9)$$

已有空气污染研究表明，LSTM 及其混合结构能够处理多变量非线性序列^[24-25]。本文把 LSTM 作为不显式传播站点信息的深度时序基准，以便判断空间图结构是否带来稳定增益。

2.2.3 自注意力与 Transformer 编码器

Transformer 通过缩放点积注意力直接计算序列中不同时间位置的相关性^[28]：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{M} \right) \mathbf{V}. \quad (10)$$

其中， $\mathbf{M}$ 为因果掩码，未来时间位置取负无穷，从而保证 t 时刻的表示不能访问 t 之后的信息。与循环递推相比，自注意力能够在固定层数内建立远距离时间位置之间的直接联系。本文使用标准 Transformer 编码器，而不将其表述为包含 ProbSparse 注意力或蒸馏结构的 Informer。

2.3 图神经网络与时空建模基础

2.3.1 空间图与邻接矩阵

多站点监测网络可表示为图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ 。节点集合 $\mathcal{V}$ 对应 12 个监测站，边集合 $\mathcal{E}$ 表示站点间空间联系。本文首先依据 Haversine 距离 d_{ij} 构造高斯核邻接：

$$a_{ij} = \exp \left(-\frac{d_{ij}^2}{2\sigma^2} \right) \mathbb{I}(d_{ij} \leq 3\sigma), \quad (11)$$

再按行归一化，使每个节点的邻接权重和为 1。 σ 固定为 25 km，超过 75 km 的联系置零。该矩阵表达空间邻近，而不假定污染一定沿某一方向传播。

2.3.2 图卷积与自适应邻接

图卷积通过邻接矩阵聚合相邻节点特征，是将卷积思想扩展到非欧氏结构的重要方法^[32]。对隐藏表示 $\mathbf{H}$ ，固定邻接分支可简写为

$$\mathbf{H}'_{\text{geo}} = \rho(\mathbf{A}\mathbf{H}\mathbf{W}_{\text{geo}}), \quad (12)$$

其中 $\rho(\cdot)$ 为非线性激活。仅依靠距离可能无法覆盖数据中潜在的非局部同步变化，因此本文同时学习两个节点嵌入 $\mathbf{E}_1$ 和 $\mathbf{E}_2$ ，得到自适应邻接

$$\mathbf{A}_{\text{adp}} = \text{softmax}(\text{ReLU}(\mathbf{E}_1\mathbf{E}_2^T)). \quad (13)$$

固定邻接与自适应邻接分别经线性映射后相加。关闭自适应邻接的消融实验用于判断学习型空间关系是否真正改善留出集性能。

2.3.3 时空特征融合机制

时空图卷积网络通常在时间卷积与图卷积之间交替传递信息^[33]。本文使用扩张率为 1、2 和 4 的三层门控因果卷积提取不同时间尺度特征，每层随后执行固定与自适应图卷积，并通过残差连接和批归一化稳定训练。门控时间卷积形式为

$$\mathbf{Z} = \tanh(\mathbf{W}_f * \mathbf{H}) \odot \sigma(\mathbf{W}_g * \mathbf{H}), \quad (14)$$

其中 $*$ 表示沿时间维的扩张因果卷积。**OSM** 静态特征经两层感知机编码后，与最后一个时间位置的动态隐藏表示拼接，再由统一预测头输出三个预测步长、两个污染物目标。该结构同时保留时间变化、地理邻近、数据驱动邻接和静态建成环境信息。

2.4 本章小结

本章从污染物—气象关系、时间周期、建成环境和空间邻近四个方面说明多源变量进入模型的依据，随后介绍滑动窗口、LSTM、自注意力、图卷积与时空融合的基本原理。上述理论共同支持本文的比较框架：**LSTM** 侧重递归时间依赖，**Transformer** 侧重长程注意力，**STGCN** 进一步引入跨站点空间传播；所有模型均在一致的历史窗口、预测目标和评价协议下比较。

3 数据集与时序研究设计

3.1 数据来源与变量体系

3.1.1 北京多站点空气质量数据

本文使用 UCI Machine Learning Repository 公开的 Beijing Multi-Site Air Quality Data 数据集^[10]。数据覆盖 2013 年 3 月 1 日 0 时至 2017 年 2 月 28 日 23 时，包括北京 12 个国控空气质量监测站的逐小时观测，共 420,768 条站点一小时记录。12 个站点分别为 Aotizhongxin、Changping、Dingling、Dongsi、Guanyuan、Gucheng、Huairou、Nongzhanguan、Shunyi、Tiantan、Wanliu 和 Wanshouxigong。原始数据以站点为单位保存，本文按统一时间索引对齐为 $T \times N \times F$ 三维数组，其中 $T = 35,064$ 个小时， $N = 12$ 个站点。

数据集的时间跨度包含多个采暖季和非采暖季，能够支持对日内变化、季节周期和污染过程持续性的研究。本文保留公开数据的原始监测时段，不把后续年份或其他城市数据混入模型训练，以确保数据来源与结果边界清晰。

3.1.2 监测变量与预测目标

动态输入由 11 个连续监测变量、风向的两个周期分量以及小时和年内日序的四个周期分量组成，共 17 维。连续变量及其作用见表 3.1。模型同时预测 PM2.5 和 NO₂，其中 PM2.5 为本文主要评价目标，NO₂ 作为相关污染物目标，用于检验共享表示能否同时刻画颗粒物与气态污染物变化。

表 3.1 动态监测变量与处理方式

类别	变量	单位	处理方式
污染物	PM2.5、PM10	$\mu\text{g}/\text{m}^3$	连续变量；PM2.5 同时作为主要预测目标
污染物	SO ₂ 、NO ₂ 、O ₃	$\mu\text{g}/\text{m}^3$	连续变量；NO ₂ 同时作为辅助预测目标
污染物	CO	$\mu\text{g}/\text{m}^3$	连续变量
气象	TEMP、DEWP	°C	温度与露点温度
气象	PRES	hPa	地面气压
气象	RAIN	mm	小时降水量
气象	WSPM	m/s	风速
气象	wd	方位	转换为正弦、余弦两个分量
时间	hour、day-of-year	—	分别转换为正弦、余弦周期分量

3.1.3 数据质量与缺失分布

原始数据不存在重复的“站点—时间”组合，但各变量存在不同程度的缺失。表3.2显示，污染物缺失比例总体高于气象变量，其中 CO 缺失比例最高，为 4.92%；气象连续变量缺失比例均不超过 0.10%，风向缺失比例为 0.43%。由于缺失并非集中于单一变量，直接删除包含缺失的站点一小时会破坏多站点时间对齐，并显著减少连续窗口数量。因此，本文采用因果前向填补与训练期统计填补相结合的策略，具体见第 3.3.1 节。

表 3.2 原始监测变量缺失比例

变量	PM2.5	PM10	SO ₂	NO ₂	CO	O ₃
缺失比例 (%)	2.08	1.53	2.14	2.88	4.92	3.16
变量	TEMP	PRES	DEWP	RAIN	WSPM	wd
缺失比例 (%)	0.09	0.09	0.10	0.09	0.08	0.43

3.2 OSM 空间要素与图结构构建

3.2.1 道路、建筑、绿地与 POI 要素

静态空间特征来自 OpenStreetMap 公开地理要素，包括道路、建筑、水体、绿地和兴趣点。OSM 是由用户协作维护的自愿地理信息平台，其开放性适合构建可复核的空间特征，但不同地区的数据完整度并不完全一致^[43,45]。本文使用项目内固定下载结果，保留数据许可和来源说明；空间运算统一在投影坐标系中完成，再转换为 WGS84 经纬度用于制图，避免直接在经纬度坐标系计算长度和面积。

3.2.2 站点空间特征汇总

以每个监测站为中心分别构建 1 km 和 3 km 缓冲区，在两个尺度上汇总道路密度、兴趣点密度、绿地比例、水体比例和建筑密度，共得到 10 个静态特征。1 km 尺度侧重刻画站点近邻建成环境，3 km 尺度用于描述更宽范围的城市结构。全部静态变量按站点形成矩阵 $\mathbf{S} \in \mathbb{R}^{12 \times 10}$ ，在模型中经两层感知机编码为固定长度嵌入。由于这些特征不随时间变化，本文只把它们解释为站点周边空间背景，不用于推断具体时刻的排放变化。

3.2.3 地理邻接与空间支持边界

站点地理邻接由经纬度 Haversine 距离构建。距离通过 25 km 带宽的高斯核转换为权重，超过 75 km 的边被截断，随后按行归一化。该处理使近邻站点获得较高权重，同时保留每个站点的自连接。对于网格风险分析，研究区划分为 208 个 2 km×2 km 单元；仅对距最近监测站不超过 8 km 的 174 个网格计算并解释风险，余下 34 个网格标记为“数据支持不足”。该边界用于避免在监测站稀疏区域对 IDW 结果作过度解释。

3.3 时序预处理与空间风险研究设计

3.3.1 因果缺失填补与训练期标准化

缺失填补按站点、按时间顺序执行。第一步，仅使用过去观测进行最多 8 小时的前向填补，以处理短时连续缺测。第二步，对仍然缺失的连续变量，只使用训练期数据计算“月份—小时”均值，并按对应月份和小时填补。第三步，若某一组合仍无可用统计量，则使用该站点训练期均值。风向同样先执行最多 8 小时的前向填补，再使用训练期众数。验证期和测试期的任何观测均不参与填补参数估计。

动态特征标准化采用训练集均值 μ_f 与标准差 σ_f ：

$$x_{t,n,f}^* = \frac{x_{t,n,f} - \mu_f^{\text{train}}}{\sigma_f^{\text{train}}}. \quad (15)$$

标准化器仅在训练时间段的全部站点观测上拟合，再固定应用于验证和测试数据。静态空间特征在 12 个站点间单独标准化。预测完成后，PM2.5 和 NO₂ 输出还原到原始物理单位后计算指标。

3.3.2 时间切分、多步窗口与测试集锁定

本文按时间顺序划分数据：训练目标不晚于 2015 年 12 月 31 日 23 时；验证目标位于 2016 年 1 月 1 日至 2016 年 8 月 31 日；测试目标从 2016 年 9 月 1 日起至 2017 年 2 月 28 日。每个样本使用过去 48 小时动态序列预测未来 1、6 和 24 小时。只有当一个样本的全部预测目标都位于同一阶段时，该样本才被纳入对应集合，跨越训练—验证或验证—测试边界的窗口全部剔除。

为控制训练成本，训练起点按 3 小时间隔抽样，验证与测试窗口保持逐小时连续。

最终训练、验证和测试窗口数分别为 8,265、5,833 和 4,321。验证集用于早停与模型选择；测试集在训练和选择过程中保持锁定，仅用于最终性能评价。时间划分与样本数见表3.3。

表 3.3 时间切分与多步窗口设置

阶段	目标时间范围	窗口步长	用途	窗口数
训练集	2013-03 至 2015-12	3 小时	参数估计	8,265
验证集	2016-01 至 2016-08	1 小时	早停与模型选择	5,833
测试集	2016-09 至 2017-02	1 小时	最终留出评价	4,321

图3.1把时间阶段、单个监督样本的输入—输出关系与三个数据集的样本规模放在同一证据图中。由图可见，48 小时历史窗口只连接同一阶段内的 1、6 和 24 小时目标，验证集和测试集边界之外的标签不会回流到前一阶段。

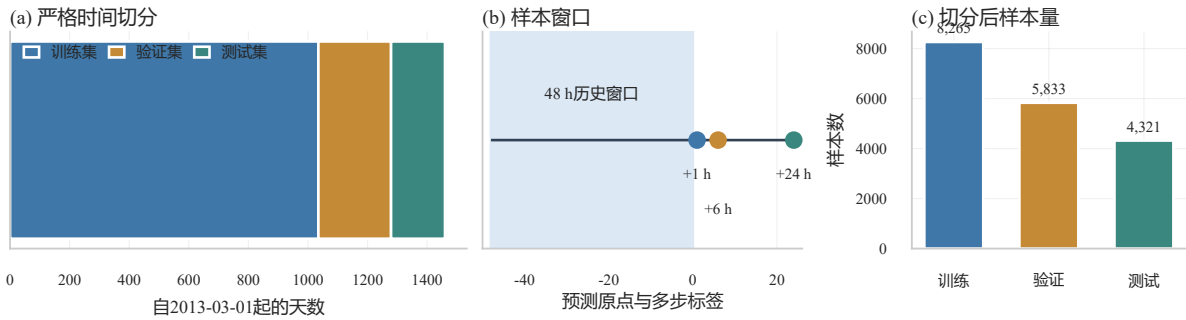


图 3.1 时间切分、多步窗口与数据边界

3.3.3 风险指标、权重方案与空间统计设计

空间风险研究首先以 12 站点 2013—2017 年 PM_{2.5} 观测均值为输入，采用幂指数为 2 的 IDW 插值到网格质心。若网格到最近站点距离大于 8 km，则不参与风险分级。对受支持网格，定义暴露、通风不足和交通敏感三个一级指标：

$$E_i = z(\widehat{PM}_{2.5,i}), \quad (16)$$

$$V_i = z\{z(\text{Building}_i) - z(\text{Green}_i)\}, \quad (17)$$

$$T_i = z\{z(\text{Road}_i) + z(\text{POI}_i)\}. \quad (18)$$

基准综合风险为

$$R_i = 0.50E_i + 0.25V_i + 0.25T_i. \quad (19)$$

为检验权重依赖，另设等权、暴露优先、通风优先和交通优先四种方案，并以 Spearman 等级相关、前 20% 高风险网格 Jaccard 系数及排名位移比较结果。该权重是探索性分析设定，不是官方风险标准。

空间自相关在 174 个受支持网格上计算。全局 Moran's I 采用 Queen 邻接行标准化权重，并通过 999 次随机置换获得双侧伪 p 值；局部 Moran 统计同样采用 999 次置换，将显著网格划分为高一高、低—低、高一低和低—高类型^[41-42]。该设计同时报告空间聚集强度、显著性和局部类型，避免只凭地图颜色判断聚集。

3.4 本章小结

本章说明了公开监测数据、OSM 空间要素、动态与静态变量体系，并冻结了缺失填补、标准化、时间切分和空间支持边界。研究设计的核心原则是：所有随时间变化的统计量只从训练期学习；多步预测标签不得跨越数据阶段；测试集不参与模型选择；风险网格仅在监测支持范围内解释。上述约束为后续模型比较与空间结论提供了统一数据基础。

4 时序预测方法

4.1 朴素基线与 LSTM 模型

4.1.1 持续性与 24 小时季节朴素基线

复杂模型只有在统一协议下超过合理的简单预测，才具有实际比较意义。本文设置两类确定性朴素基线。持续性基线假定未来浓度保持在预测起点：

$$\hat{\mathbf{Y}}_{t+h}^{\text{per}} = \mathbf{Y}_t, \quad h \in \{1, 6, 24\}. \quad (20)$$

24 小时季节朴素基线则使用目标时刻前一日同一小时的观测：

$$\hat{\mathbf{Y}}_{t+h}^{\text{sea}} = \mathbf{Y}_{t+h-24}. \quad (21)$$

持续性基线更适合检验短时惯性，季节朴素基线用于检验日内周期能否解释中期变化。两类基线不需要参数训练，并在与深度模型完全相同的测试窗口上计算指标。

4.1.2 LSTM 网络结构

LSTM 模型对 12 个站点共享参数，但每个站点的时间序列独立进入循环网络，不执行跨站点信息传播。17 维动态特征先经线性层映射至 64 维；随后输入两层 LSTM，提取 48 小时序列最后一个时刻的 64 维隐藏状态。10 维 OSM 静态特征由两层感知机编码为 64 维，并与动态隐藏状态拼接为 128 维表示。该表示经“线性层—ReLU—线性层”预测头输出 3×2 个数值，对应三个步长和两个目标。

训练中使用 0.2 的层间 dropout。LSTM 既能够利用污染物与气象变量的滞后信息，又不显式使用地理邻接，因此其与 STGCN 的差异能够反映空间传播模块的增量作用，而不是输入目标或训练协议差异。

4.1.3 多步预测输出

本文采用直接多输出策略，一次前向传播同时得到 1、6 和 24 小时预测，而不是逐小时递归生成。设最终隐藏表示为 $\mathbf{z}_{t,n}$ ，预测头为 $g_\phi(\cdot)$ ，则

$$[\hat{\mathbf{y}}_{t+1,n}, \hat{\mathbf{y}}_{t+6,n}, \hat{\mathbf{y}}_{t+24,n}] = g_\phi(\mathbf{z}_{t,n}). \quad (22)$$

直接多输出避免递归预测中前一步误差不断作为后一步输入，但不同预测步长共享编码表示，模型仍需同时学习短期惯性和中期变化。

4.2 Transformer 编码器模型

4.2.1 时间位置编码

Transformer 本身不包含时间顺序，本文在 17 维动态特征映射至 64 维后加入正弦位置编码。对位置 pos 和通道 i ，编码为

$$PE_{pos,2i} = \sin(pos/10000^{2i/d}), \quad (23)$$

$$PE_{pos,2i+1} = \cos(pos/10000^{2i/d}). \quad (24)$$

位置编码与输入表示相加，使注意力层能够区分历史窗口中的先后位置。该编码与前述小时、年内日序周期特征作用不同：前者描述窗口内部的相对位置，后者描述真实日历周期。

4.2.2 因果自注意力编码

本文使用两层标准 TransformerEncoder，每层包含 4 头自注意力、维度为 256 的前馈网络和 0.1 的 dropout。注意力矩阵采用上三角因果掩码，使较早位置不能访问其后的信息。各站点共享同一编码器，但在批处理中按“样本一站点”展开，因此 Transformer 与 LSTM 一样不执行站点间直接传播。最后一个历史位置的 64 维输出与 64 维静态空间嵌入拼接，再进入预测头。

需要说明的是，本模型采用完整缩放点积注意力，不含 ProbSparse 稀疏注意力、序列蒸馏或生成式解码器，因此正文统一称为“Transformer 编码器”，不使用 Informer 名

称。稀疏注意力与混合 Transformer 在空气质量预测中的研究可作为后续扩展方向^[29-30], 但不属于本文已实现结构。

4.2.3 多预测步长解码

Transformer 的解码方式与 LSTM 一致: 只取历史窗口最后位置的编码结果, 并通过共享预测头直接输出三个步长。该设计保持模型间输出层一致, 使比较重点落在时间编码方式上。由于 1、6 和 24 小时任务共同参与均方误差训练, 长步长的大误差会在目标尺度标准化后与短步长共同影响参数更新。

4.3 STGCN 时空融合模型与训练策略

4.3.1 图卷积与时间卷积模块

STGCN 首先把 17 维动态特征投影到 32 维通道, 再依次通过三个时空块。第 k 个时空块的时间卷积扩张率为 2^{k-1} , 卷积核宽度为 3; 门控输出随后进入图卷积。图卷积同时聚合固定地理邻接与学习型自适应邻接, 两个分支经线性变换后相加并通过 ReLU。每个时空块使用残差连接和批归一化, 以减少深层训练中的梯度波动。

与经典交通预测 STGCN^[33]相比, 本文结构进一步加入自适应邻接, 其目的不是复制某一现成模型名称, 而是在固定地理关系之外学习数据中的站点联系。与近期多视图、时变图和图注意力空气质量模型^[34-35,37]相比, 本文保持较小网络规模, 以便在重复种子和消融条件下完成一致比较。

4.3.2 OSM 静态特征与自适应邻接融合

STGCN 的动态分支在三个时空块后取最后一个时间位置, 得到每个站点 32 维表示; OSM 静态特征经两层感知机编码为 32 维。二者拼接后进入统一预测头。为区分两类空间信息的作用, 本文设置两项消融: 一是将 OSM 静态特征全部置零但保留图传播, 记为 STGCN-no-OSM; 二是关闭自适应邻接, 仅使用地理高斯核邻接, 记为 STGCN-geo-only。两项消融均采用主实验第一个随机种子, 其余配置保持不变。

4.3.3 模型选择、重复种子与评价指标

三个深度模型均使用 Adam 优化器，初始学习率为 10^{-3} ，权重衰减为 10^{-5} ，批量大小为 128，最多训练 20 轮。验证损失连续 6 轮未改善时提前停止；当验证损失进入平台期时，学习率按 0.5 衰减。梯度范数裁剪阈值为 5。每次训练恢复验证损失最低轮次的参数。主模型使用 42、2026 和 3407 三个随机种子重复训练，并固定数据加载器随机源与确定性计算设置。训练设备为 NVIDIA GeForce RTX 3080 Ti。

模型在还原物理单位后的测试预测上计算平均绝对误差 (MAE)、均方根误差 (RMSE) 和决定系数 (R^2):

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |y_i - \hat{y}_i|, \quad (25)$$

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2}, \quad (26)$$

$$R^2 = 1 - \frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{\sum_{i=1}^M (y_i - \bar{y})^2}. \quad (27)$$

MAE 反映平均误差幅度，RMSE 对大误差更敏感， R^2 描述模型相对均值基准解释的方差比例^[19]。本文分别报告每个污染物、每个预测步长和每个站点的指标，并以三个随机种子的均值和标准差衡量训练波动。在三类深度模型之间，以 PM2.5 三个预测步长 RMSE 的等权平均值作为主模型判定依据；该数值先对三个随机种子求均值，再在 1、6 和 24 小时之间等权平均，数值越低越优。该规则在结果汇总前固定。测试指标用于评价，不参与早停或超参数调整。

为补充说明输入变量与短期 PM2.5 预测之间的贡献方向，本文另行训练 LightGBM 解释伴随模型。该模型使用预测原点的气象与污染物滞后特征、时间周期特征以及 OSM 静态特征，预测下一小时 PM2.5；训练期和测试期与主实验保持一致，验证期不参与拟合。随后从测试期固定随机抽取 3000 个样本，使用 TreeSHAP 计算特征贡献^[48-49]。本文报告平均绝对 SHAP 值、蜂群分布和主要变量依赖关系。该结果只解释 LightGBM 伴随模型中的预测关联，不代表 STGCN 或 Transformer 的内部决策，也不作因果效应解释。

4.4 本章小结

表 4.1 候选模型的结构与信息路径对照

模型	时间信息	站点间信息	OSM 静态特征
持续性基线	预测原点观测	不使用	不使用
24 小时季节基线	前一日同时刻	不使用	不使用
LSTM	两层循环编码	共享参数，不传播	多层感知机编码
Transformer	两层四头自注意力	共享参数，不传播	多层感知机编码
STGCN	三个扩张时间卷积块	固定与自适应邻接	与动态表示融合

本章在统一输入、预测目标和训练协议下构建两类朴素基线、LSTM、Transformer 编码器和 STGCN。各模型并非简单追求结构复杂度，而是对应逐层递进的研究问题：朴素基线检验深度模型能否超越短期持续性与日周期规律；LSTM 与 Transformer 比较递归建模和自注意力建模对多步预测的适用性；STGCN 进一步检验显式站点关系能否改善跨站点联合预测。这样的模型分组使性能差异能够回到时间依赖、空间传播和静态城市属性三类信息来源加以解释。

为保证比较具有一致基础，所有学习模型采用相同的 48 小时历史窗口、相同的 1、6 和 24 小时预测目标，以及完全一致的按时间划分训练集、验证集和测试集。训练期均只使用训练段估计标准化参数，验证集用于早停和训练过程选择，锁定测试集只承担最终评价。三个随机种子的均值反映模型的总体水平，标准差用于观察训练波动；分站点指标进一步检查总体优势是否由少数站点驱动。

主模型的确定遵循预先固定的选择规则：先在每个随机种子下分别计算三个预测步长的 PM2.5 RMSE，再对随机种子求均值，最后对三个步长等权平均。该规则既避免只依据某一个短期步长宣布最优，也避免从测试结果中事后挑选有利指标。MAE、 R^2 和逐站点误差作为辅助证据，用于判断误差幅度、解释能力与空间稳定性，不替代主选择指标。

STGCN 的两项空间消融分别移除 OSM 静态特征和自适应邻接，仅用于辨析模型组成部分在当前数据与训练协议下的贡献。LightGBM 及其 SHAP 结果则构成独立的解释证据线，回答主要输入变量与下一小时 PM2.5 预测之间的关联方向，不用于解释 STGCN 内部机制。由此，下一章将按照“总体性能—多步长误差—分站点差异—空间风险—解释分析”的顺序报告结果，形成从预测准确性到应用含义的完整证据链。

5 实验结果与误差分析

5.1 训练收敛与基线比较

5.1.1 训练过程与早停结果

图5.1给出 LSTM、Transformer 和 STGCN 在三个随机种子下的训练损失与验证损失。三类模型的训练损失均随轮次下降，没有出现数值发散。LSTM 的最佳验证轮次集中在 12–13 轮；Transformer 对随机初始化更敏感，三次最佳轮次分别为 4、7 和 12；STGCN 的最佳轮次为 13、17 和 20。验证曲线存在小幅波动，但最佳权重由验证损失单独确定，留出测试集未参与早停。

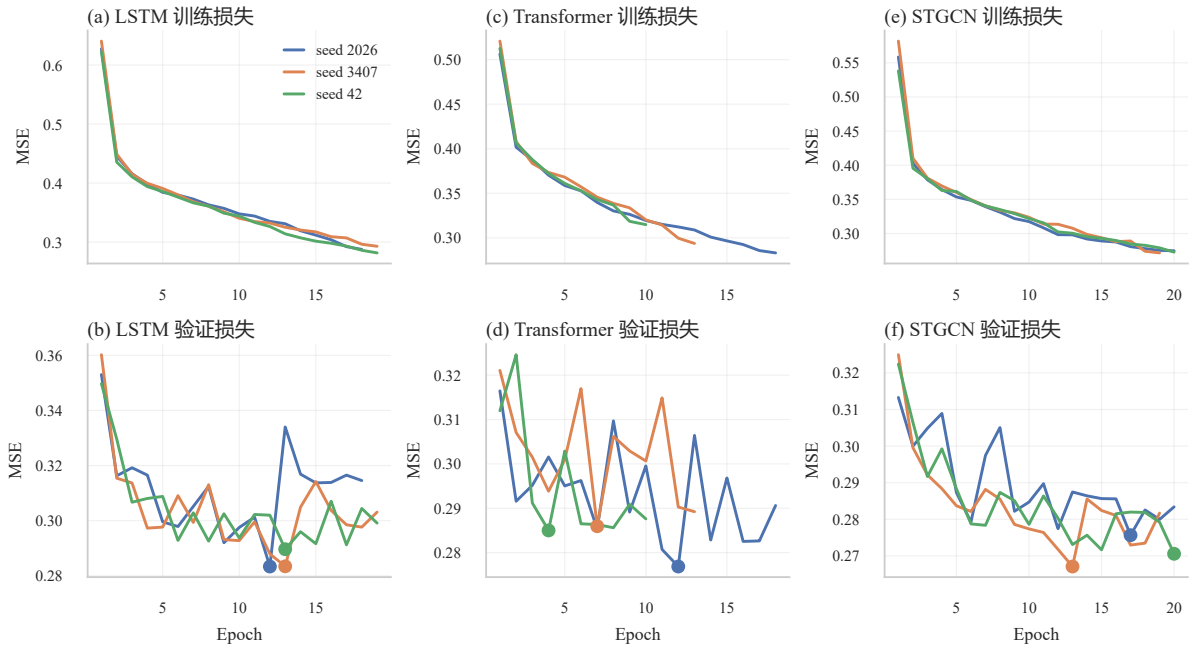


图 5.1 三类核心模型的训练与验证损失

5.1.2 朴素基线对比

表5.1列出各模型的 PM2.5 留出集 RMSE，浓度误差单位均为 $\mu\text{g}/\text{m}^3$ 。持续性基线的 1 小时 RMSE 为 23.037，说明短时污染浓度具有较强惯性；其误差随步长增大，24 小时 RMSE 达 105.358。24 小时季节朴素基线的三步长平均 RMSE 为 105.299，明显高于持续性基线，表明只复制前一日同时刻不足以描述留出期的快速变化。

表 5.1 不同模型的 PM2.5 测试集 RMSE 比较 ($\mu\text{g}/\text{m}^3$)

模型	1 小时	6 小时	24 小时	三步长均值
持续性基线	23.037	66.011	105.358	64.802
24 小时季节朴素基线	105.264	105.274	105.358	105.299
LSTM	24.549 ± 0.861	59.778 ± 1.887	90.027 ± 3.130	58.118
Transformer	23.620 ± 0.540	58.299 ± 0.099	89.453 ± 2.568	57.124
STGCN	21.881 ± 0.214	53.655 ± 0.735	85.708 ± 1.683	53.748

5.1.3 三类深度模型总体比较

按预先冻结的规则，先对各模型三个随机种子求均值，再对 1、6 和 24 小时 RMSE 等权平均。STGCN 以 $53.748 \mu\text{g}/\text{m}^3$ 排名第一，Transformer 和 LSTM 分别为 57.124 和 $58.118 \mu\text{g}/\text{m}^3$ 。STGCN 相对最强朴素基线降低 17.06%，相对排名第二的 Transformer 降低 5.91%。图 5.2 显示，STGCN 在三个步长上的平均 MAE 和 RMSE 均低于其他核心模型，且三次训练的离散程度较小，因此确定为主模型。

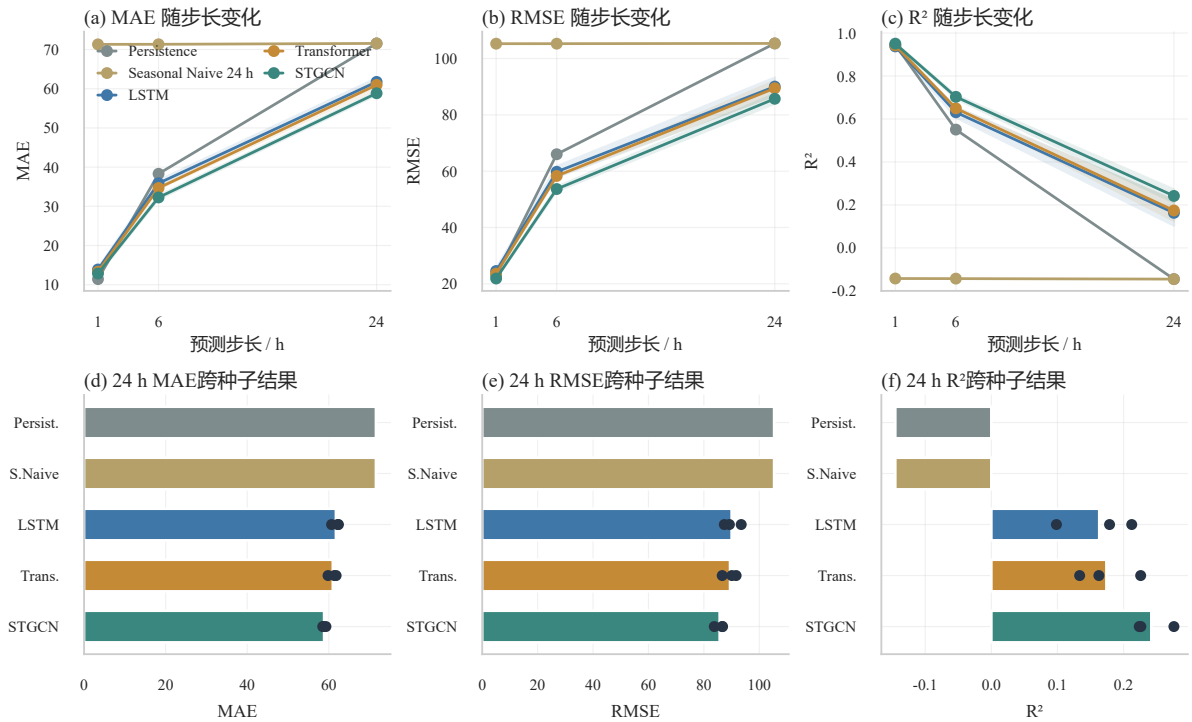


图 5.2 朴素基线与核心预测模型的留出集比较

5.2 多步长预测与分站点误差

5.2.1 1 小时预测结果

表5.2报告 STGCN 的三随机种子均值与标准差。1 小时预测的 MAE 为 $12.854 \pm 0.293 \mu\text{g}/\text{m}^3$, RMSE 为 $21.881 \pm 0.214 \mu\text{g}/\text{m}^3$, R^2 为 0.951 ± 0.001 。较小的跨种子标准差表明近时预测对初始化不敏感,而较高的 R^2 说明历史污染序列和气象条件能够解释留出期大部分短时波动。

表 5.2 STGCN 的 PM2.5 测试集性能

预测步长	MAE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	R^2
1 小时	12.854 ± 0.293	21.881 ± 0.214	0.951 ± 0.001
6 小时	32.268 ± 0.423	53.655 ± 0.735	0.703 ± 0.008
24 小时	58.846 ± 0.402	85.708 ± 1.683	0.242 ± 0.030

5.2.2 6 小时与 24 小时预测结果

当预测步长增至 6 小时时, STGCN 的 RMSE 上升至 $53.655 \pm 0.735 \mu\text{g}/\text{m}^3$, R^2 降至 0.703 ± 0.008 ; 24 小时 RMSE 进一步上升至 $85.708 \pm 1.683 \mu\text{g}/\text{m}^3$, R^2 为 0.242 ± 0.030 。图5.3的一致性图表明, 1 小时预测点更贴近对角线, 24 小时高浓度区域出现更明显的低估。残差的尾部随步长加宽, 且预测浓度较高时方差增大, 表明极端污染过程仍是中期预测的主要误差来源。

5.2.3 分站点误差与典型时段诊断

图5.4显示站点间的预测难度存在明显差异。1 小时 RMSE 介于 $19.91\text{--}25.04 \mu\text{g}/\text{m}^3$, 6 小时介于 $41.55\text{--}60.55 \mu\text{g}/\text{m}^3$, 24 小时介于 $71.77\text{--}95.24 \mu\text{g}/\text{m}^3$ 。怀柔站在 24 小时预测中的 MAE 最低, 万寿西官站最高。站点间差异说明, 单一总体指标不能完全反映模型在局部区域的表现, 实际应用中应同时检查站点误差排序和典型高污染时段。

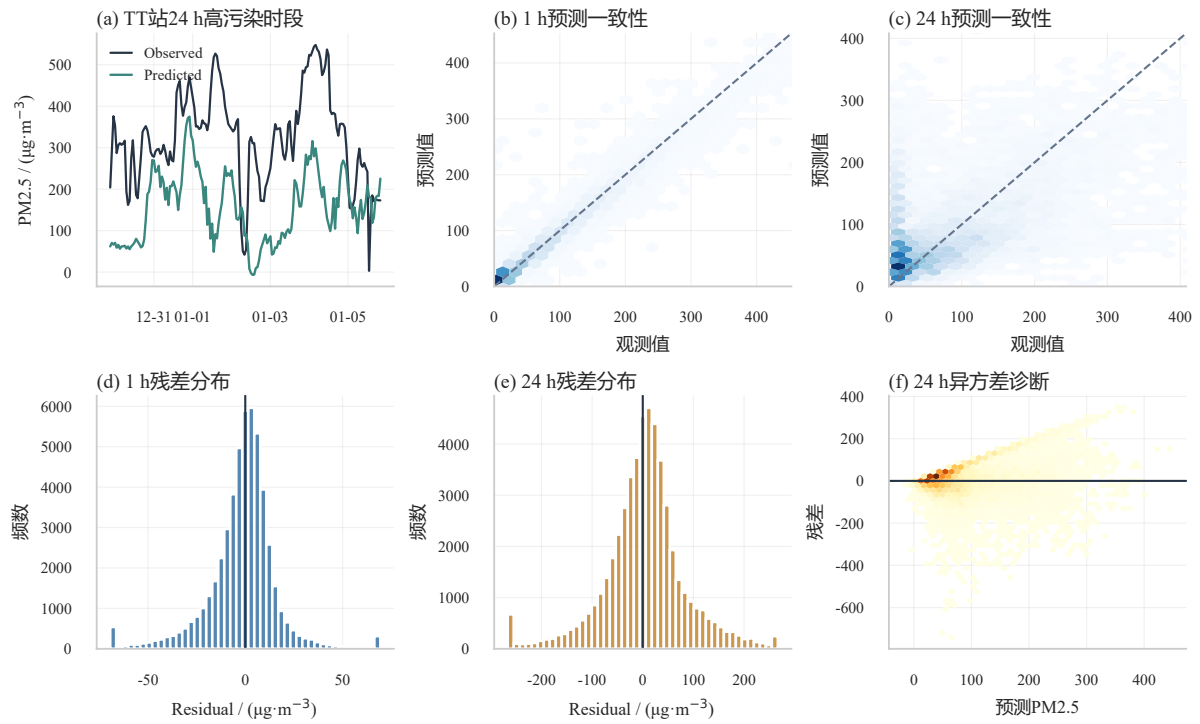


图 5.3 STGCN 在典型时段的预测一致性与残差诊断

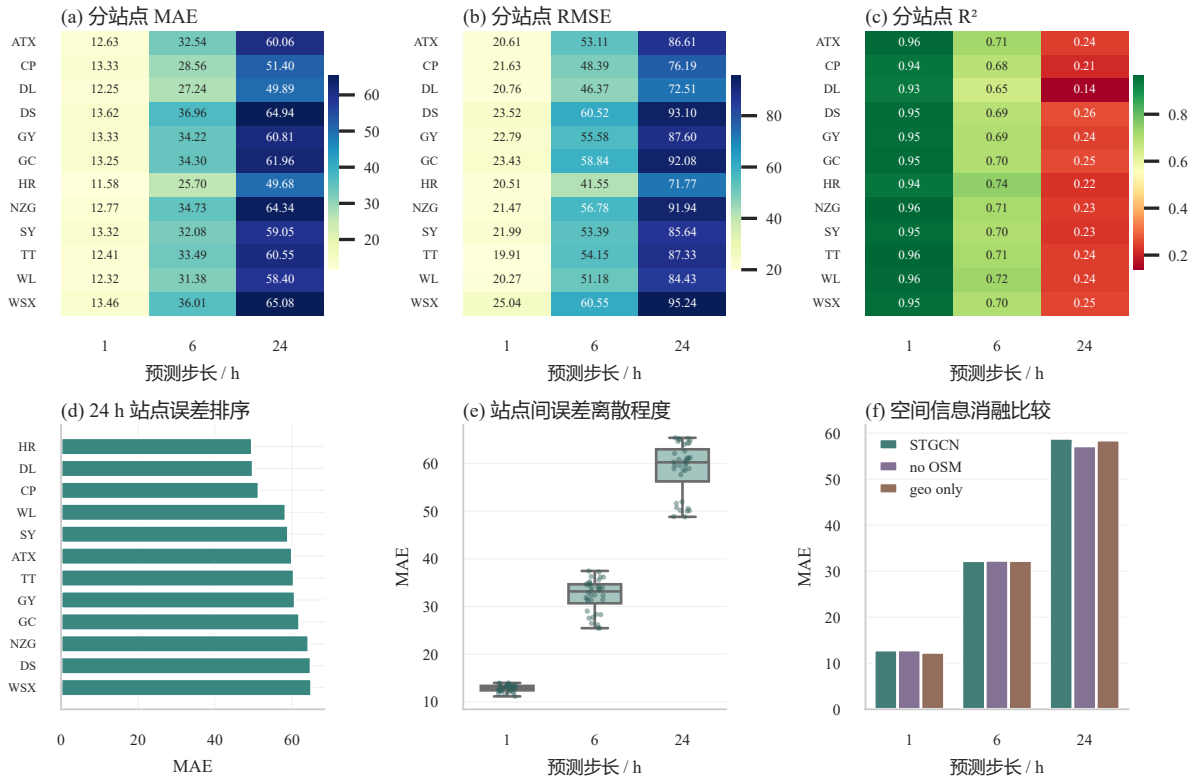


图 5.4 STGCN 的多步长分站点误差与空间信息消融

5.3 消融、解释与空间风险稳健性

5.3.1 OSM 静态特征与邻接结构消融

表5.3在随机种子 42 下比较完整 STGCN、去除 OSM 静态特征的 STGCN-no-OSM 和仅保留固定地理邻接的 STGCN-geo-only。三种配置在 1 和 6 小时上的 RMSE 接近；STGCN-geo-only 的 1 小时 RMSE 最低，STGCN-no-OSM 的 24 小时 RMSE 最低。该结果不支持“OSM 静态特征或自适应邻接在所有步长上均带来增益”的结论。更谨慎的解释是：空间信息的效用随预测步长而变，当前消融只执行一个随机种子，尚不足以将较小差异视为稳定结论。

表 5.3 STGCN 空间信息消融的 PM2.5 测试集 RMSE（随机种子 42）

模型	1 小时	6 小时	24 小时
STGCN	21.675	54.087	86.746
STGCN-no-OSM	21.907	54.442	82.503
STGCN-geo-only	21.387	53.790	84.845

5.3.2 LightGBM 解释伴随模型与 SHAP 结果

独立 LightGBM 伴随模型在按目标时刻切分的测试集上取得 $23.251 \mu\text{g}/\text{m}^3$ 的 RMSE 和 0.9441 的 R^2 。图5.5显示，按平均绝对 SHAP 值排序，预测原点 PM2.5 为第一位，且特征值与 SHAP 值整体呈强正向关联；露点温度为第二位，整体呈正向关联；年内日序余弦特征为第三位，其贡献随季节周期变化，不适合简化为单一正负方向。此外，风速排名第四，整体呈负向关联；24 小时滞后 PM2.5 排名第六，整体呈正向关联。这些排序和方向描述的是 LightGBM 在当前留出数据中的预测依赖，不代表 STGCN 内部机制，也不构成因果关系。

5.3.3 风险权重敏感性与空间自相关

图5.6将留出期 PM2.5 暴露、通风不利和交通强度纳入统一网格。与基准权重相比，等权、暴露优先、通风优先和交通优先四种情景的 Spearman 排序相关系数为 0.9688–0.9890，前 20 个高风险网格的 Jaccard 系数为 0.7500–0.8919。因此，高风险区域的大体格局对权重调整较为稳定，但局部网格的具体名次仍会变动。

全 174 个有效网格的全局 Moran's I 为 0.7131，999 次置换检验的 p 值为 0.001，说

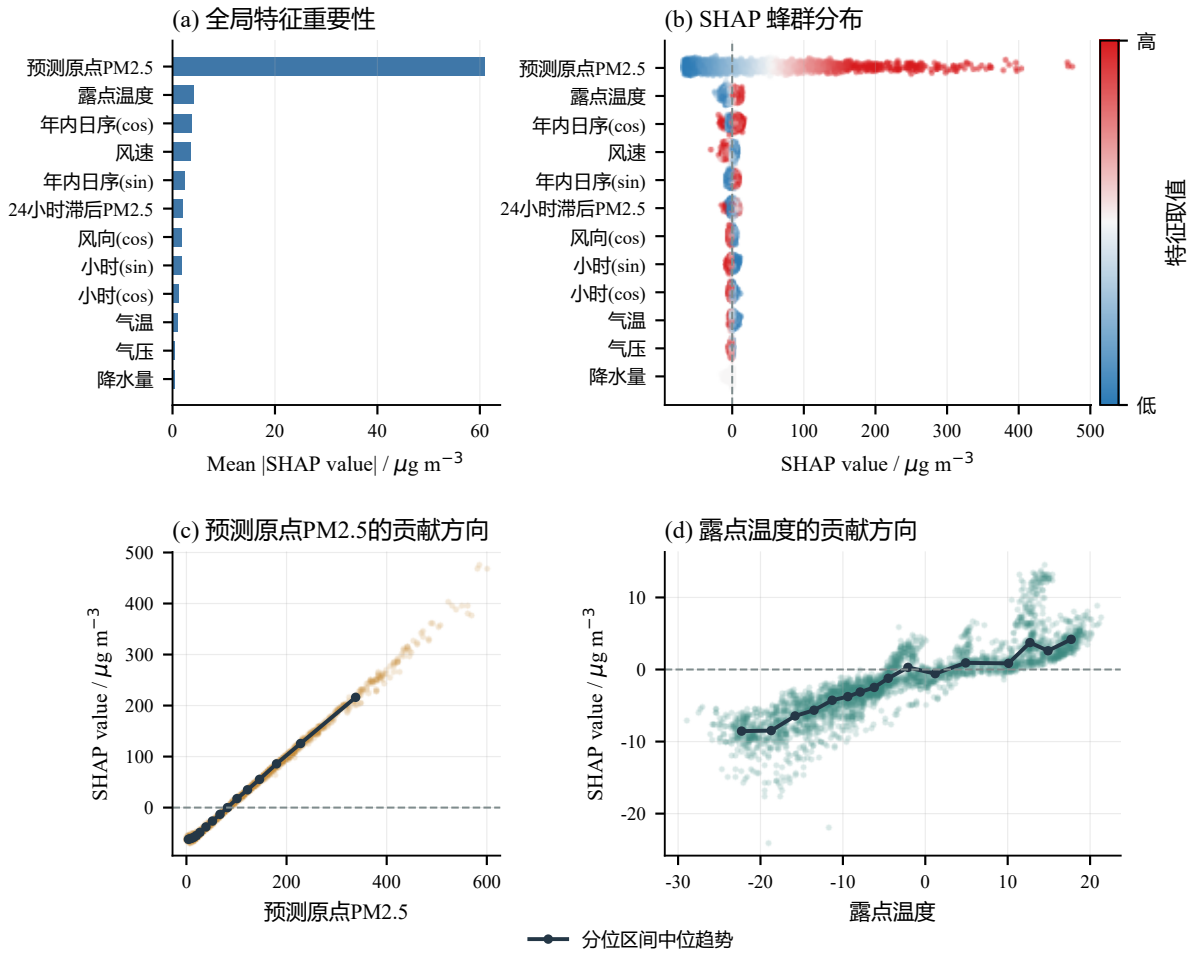


图 5.5 独立 LightGBM 伴随模型的 SHAP 特征贡献与依赖关系

明当前风险指标存在显著正空间自相关。局部置换检验识别出 62 个 $p < 0.05$ 的网格。图5.7中的高一高和低一低聚集说明风险并非随机散布，而是在相邻区域形成连续格局。这一结果支持将空间聚集用于区域筛查，但不能将网格指数解释为个体实际吸入剂量。

5.4 本章小结

本章在锁定留出集上完成模型选择、多步长评价、分站点诊断、空间消融和风险稳健性分析。STGCN 按事先冻结的三步长平均 RMSE 规则取得最优总体结果，但预测性能随步长延长而下降，不同站点和高污染时段的误差差异明显。消融实验没有显示空间模块在所有步长上的一致增益，因此本文保留该不一致证据。空间风险排序在替代权重下较稳定，并呈显著正空间自相关；LightGBM-SHAP 则从独立伴随模型角度补充了主要特征的排序和关联方向。

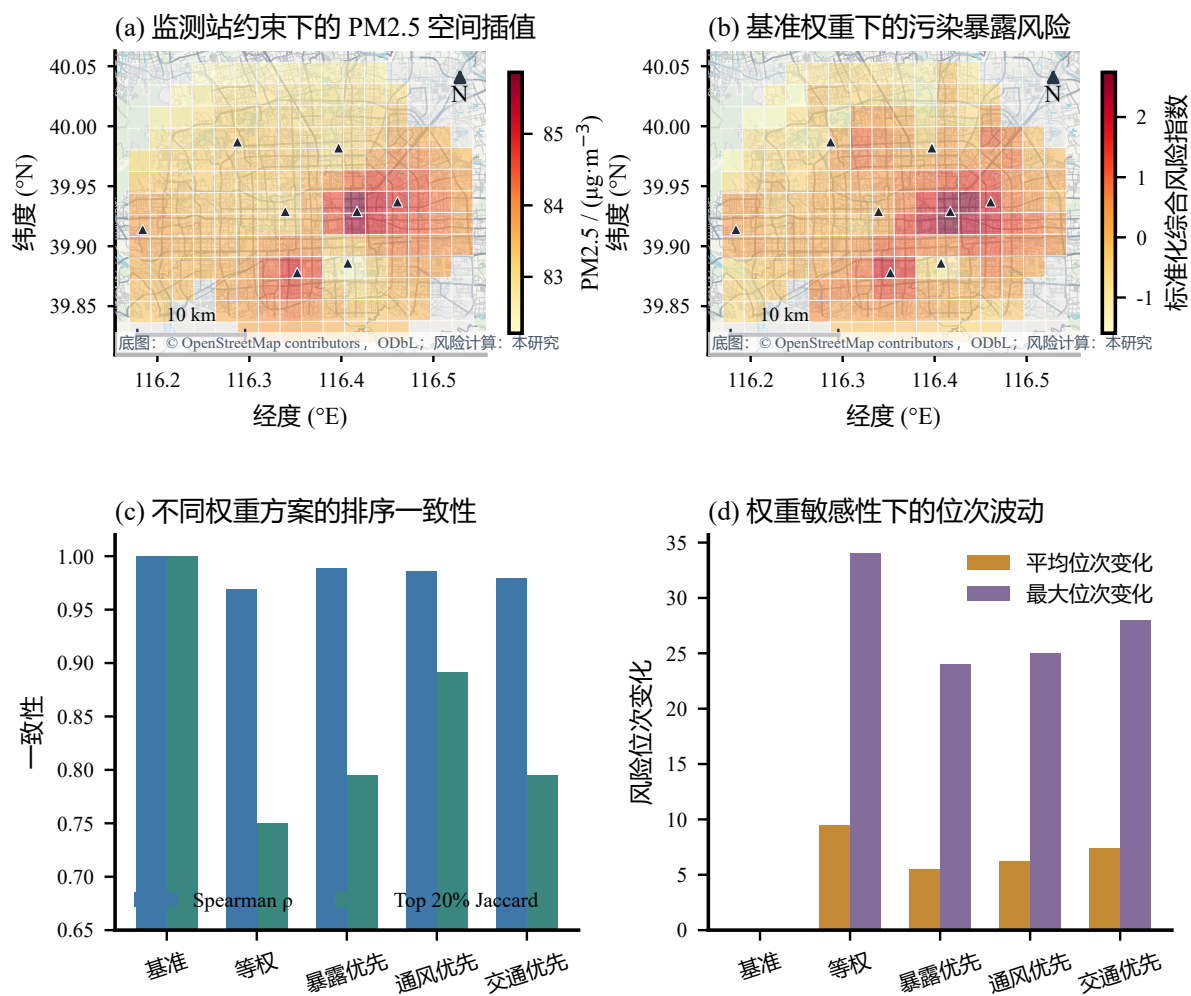


图 5.6 空间污染风险分布、排序与权重敏感性

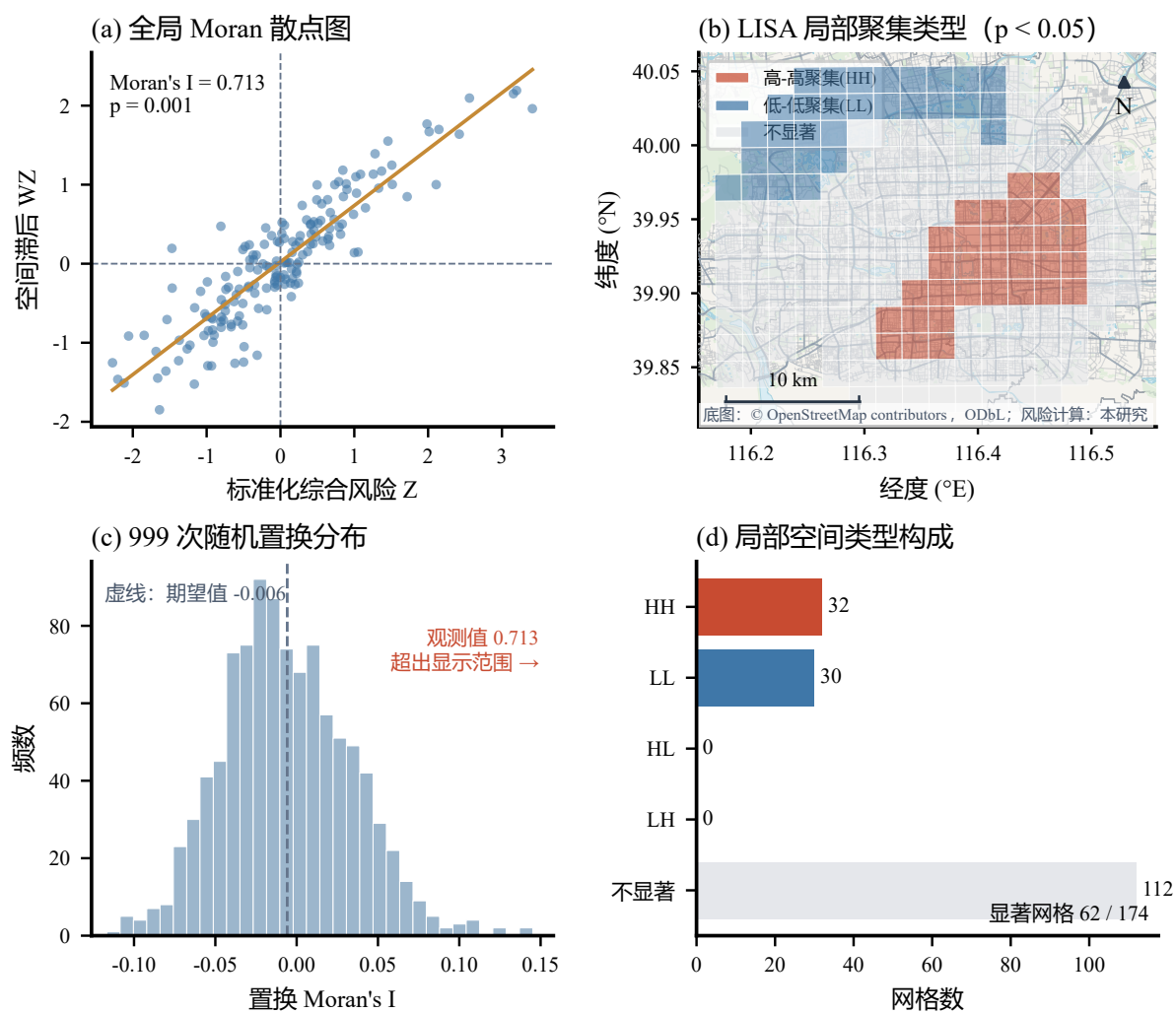


图 5.7 空间风险指标的 Moran 散点、置换检验与 LISA 聚集

6 结论与展望

6.1 研究结论

6.1.1 时空预测主要结论

本文基于北京 12 个监测站的逐小时公开数据，在严格时间切分和统一训练协议下比较两类朴素基线、LSTM、Transformer 和 STGCN。按三随机种子平均、三预测步长等权的 RMSE 规则，STGCN 以 $53.748 \mu\text{g}/\text{m}^3$ 的均值成为最优模型，较最强朴素基线降低 17.06%。其 1 小时结果最稳定，24 小时的误差和跨种子波动明显增大。分站点结果进一步表明，模型在不同区域的预测难度不同，应将总体性能与局部误差共同报告。

6.1.2 污染暴露风险主要结论

留出期预测与 OSM 空间要素共同支撑了网格风险分析。综合风险指标的全局 Moran's I 为 0.7131，置换检验 $p = 0.001$ ，显示高风险和低风险网格均存在明显聚集。四类替代权重情景的 Spearman 系数均高于 0.96，说明区域风险的总体排序对适度权重变化具有一定稳健性。独立 LightGBM-SHAP 分析则表明，当前 PM_{2.5}、露点温度、季节周期与风速是伴随模型中较重要的预测依赖，其中前两者整体呈正向关联，风速整体呈负向关联。

6.2 研究局限

6.2.1 数据时段与区域边界

本文数据范围为 2013 年 3 月至 2017 年 2 月的北京 12 个固定监测站。严格时间留出能够检验同一城市在后续时段的预测性能，但不等同于跨城市、跨气候带或近年排放结构下的外部验证。站点观测反映的是固定位置的环境浓度，未直接包含人群活动轨迹、室内外转换和个体时间活动模式。

6.2.2 模型与评价边界

本文比较的深度模型范围为 LSTM、标准 Transformer 编码器和轻量 STGCN，未纳入大规模预训练时序模型、神经微分方程或动态图结构搜索。主结果使用确定性点预测

指标，尚未给出预测区间的覆盖率与宽度。空间消融只在随机种子 42 下执行，且各配置在不同预测步长上的优势不一致，因此不能将较小差异解释为稳定组件贡献。

6.2.3 风险指标与因果解释边界

网格风险指标由污染暴露、通风不利和交通强度组成，属于区域探索性筛查量，不是个体吸入暴露、健康损害概率或因果效应估计。 $\text{Moran's } I$ 和 LISA 识别的是指标值在空间上的相似聚集，不能自动推出具体污染源或传播路径。同样， SHAP 仅描述 LightGBM 伴随模型中的特征依赖，既不代表 STGCN 内部机制，也不证明特征对 $\text{PM}_{2.5}$ 的因果作用。

6.3 未来展望

6.3.1 跨城市外部验证

后续可引入不同气候区、城市形态与排放结构的公开监测数据，在不重新调参的条件下评估现有模型的直接迁移能力，并比较少量标签微调、域适应和站点嵌入校准。外部验证应冻结源城市模型与预处理参数，并单独报告目标城市的站点覆盖、污染分布变化和失败模式。

6.3.2 概率预测与极端污染预警

针对 24 小时预测中高浓度低估和残差异方差，可引入分位数损失、深度集成或分布预测，同时报告预测区间覆盖率、平均宽度和极端污染召回率。在预警场景中，评价应从平均误差扩展到阈值跨越的提前量、漏报与误报，使模型性能与实际预警目标对齐。

6.3.3 动态暴露与多尺度空间建模

未来可在合法合规的前提下融合更细粒度道路流量、遥感反演、边界层高度与人群活动时间信息，区分环境浓度、活动空间暴露与个体暴露。空间建模可从固定网格扩展为站点、道路与社区多尺度图，并在风险排序中报告尺度敏感性和空间不确定性，从而为更精细的城市污染管理提供证据基础。

参考文献

- [1] POPE C A. Epidemiology of fine particulate air pollution and human health: biologic mechanisms and who's at risk?[J]. *Environmental Health Perspectives*, 2000, 108(suppl 4): 713-723.
- [2] MANISALIDIS I, STAVROPOULOU E, STAVROPOULOS A, et al. Environmental and Health Impacts of Air Pollution: A Review[J]. *Frontiers in Public Health*, 2020, 8: 14.
- [3] ORELLANO P, KASDAGLI M I, PÉREZ VELASCO R, et al. Long-Term Exposure to Particulate Matter and Mortality: An Update of the WHO Global Air Quality Guidelines Systematic Review and Meta-Analysis[J]. *International Journal of Public Health*, 2024, 69: 1607683.
- [4] World Health Organization. WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide [EB/OL]. World Health Organization. 2021 [2026-08-20]. <https://www.who.int/publications/i/item/9789240034228>.
- [5] LIU J, HE C, SI Y, et al. Toward Better and Healthier Air Quality: Global PM_{2.5} and O₃ Pollution Status and Risk Assessment Based on the New WHO Air Quality Guidelines for 2021[J]. *Global Challenges*, 2024, 8(4): 2300258.
- [6] MCDUFFIE E E, MARTIN R V, SPADARO J V, et al. Source sector and fuel contributions to ambient PM_{2.5} and attributable mortality across multiple spatial scales[J]. *Nature Communications*, 2021, 12(1): 3594.
- [7] ZHANG Y L, CAO F. Fine particulate matter (PM_{2.5}) in China at a city level[J]. *Scientific Reports*, 2015, 5(1): 14884.
- [8] ZHANG Q, ZHENG Y, TONG D, et al. Drivers of improved PM 2.5 air quality in China from 2013 to 2017[J]. *Proceedings of the National Academy of Sciences*, 2019, 116(49): 24463-24469.
- [9] ZHANG S, GUO B, DONG A, et al. Cautionary tales on air-quality improvement in Beijing[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2017, 473(2205): 20170457.
- [10] ZHANG S, GUO B, DONG A, et al. Beijing Multi-Site Air Quality Data[EB/OL]. UCI Machine Learning Repository. 2017 [2026-08-20]. <https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data>.
- [11] YAN R, LIAO J, YANG J, et al. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering[J]. *Expert Systems with Applications*, 2021, 169: 114513.
- [12] MÉNDEZ M, MERAYO M G, NÚÑEZ M. Machine learning algorithms to forecast air quality: a survey[J]. *Artificial Intelligence Review*, 2023, 56(9): 10031-10066.

- [13] ZAININ, EAN L W, AHMED A N, et al. A systematic literature review of deep learning neural network for time series air quality forecasting[J]. *Environmental Science and Pollution Research*, 2022, 29(4): 4958-4990.
- [14] AGBEHADJIE, OBAGBUWA I C. Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction[J]. *Atmosphere*, 2024, 15(11): 1352.
- [15] KALANTARI E, GHOLAMI H, MALAKOOTI H, et al. Machine learning for air quality index (AQI) forecasting: shallow learning or deep learning?[J]. *Environmental Science and Pollution Research*, 2024, 31(54): 62962-62982.
- [16] ZHANG P, MA W, WEN F, et al. Estimating PM_{2.5} concentration using the machine learning GA-SVM method to improve the land use regression model in Shaanxi, China [J]. *Ecotoxicology and Environmental Safety*, 2021, 225: 112772.
- [17] VAN DEN BOSSCHE J, DE BAETS B, VERWAEREN J, et al. Development and evaluation of land use regression models for black carbon based on bicycle and pedestrian measurements in the urban environment[J]. *Environmental Modelling & Software*, 2018, 99: 58-69.
- [18] BERGMEIR C, HYNDMAN R J, KOO B. A note on the validity of cross-validation for evaluating autoregressive time series prediction[J]. *Computational Statistics & Data Analysis*, 2018, 120: 70-83.
- [19] HYNDMAN R J, KOEHLER A B. Another look at measures of forecast accuracy[J]. *International Journal of Forecasting*, 2006, 22(4): 679-688.
- [20] HYNDMAN R J, KHANDAKAR Y. Automatic Time Series Forecasting: The forecast Package for R[J]. *Journal of Statistical Software*, 2008, 27(3).
- [21] DIEBOLD F X, MARIANO R S. Comparing Predictive Accuracy[J]. *Journal of Business & Economic Statistics*, 1995, 13(3): 253-263.
- [22] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [23] YANG J, YAN R, NONG M, et al. PM_{2.5} concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time[J]. *Atmospheric Pollution Research*, 2021, 12(9): 101168.
- [24] CHANG Y S, CHIAO H T, ABIMANNAN S, et al. An LSTM-based aggregated model for air pollution forecasting[J]. *Atmospheric Pollution Research*, 2020, 11(8): 1451-1463.
- [25] DREWIL G I, AL-BAHADILI R J. Air pollution prediction using LSTM deep learning and metaheuristics algorithms[J]. *Measurement: Sensors*, 2022, 24: 100546.
- [26] LI D, LIU J, ZHAO Y. Prediction of Multi-Site PM_{2.5} Concentrations in Beijing Using CNN-Bi LSTM with CBAM[J]. *Atmosphere*, 2022, 13(10): 1719.

- [27] CHENG X, ZHANG W, WENZEL A, et al. Stacked ResNet-LSTM and CORAL model for multi-site air quality prediction[J]. *Neural Computing and Applications*, 2022, 34(16): 13849-13866.
- [28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[C]// *Advances in Neural Information Processing Systems* 30. 2017: 5998-6008.
- [29] ZHANG Z, ZHANG S. Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks[J]. *International Journal of Environmental Science and Technology*, 2023, 20(12): 13535-13550.
- [30] TANG B, LI Q, HUANG T. Development of a decomposition-optimization-transformer hybrid model for spatiotemporal forecasting of PM2.5 air pollution in Chinese cities: A case study[J]. *Ecotoxicology and Environmental Safety*, 2025, 305: 119209.
- [31] SHI Y, YANG J, QIN D. An Adaptive Spatiotemporal Graph Transformer for Multi-Site PM2.5 Multi-Step Forecasting with Non-stationary and Sparse-Aware Method[J]. *Journal of Environmental Management*, 2026, 414: 130515.
- [32] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks[C]// *International Conference on Learning Representations*. 2017.
- [33] YU B, YIN H, ZHU Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting[C]// *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. 2018: 3634-3640.
- [34] SUI S, HAN Q. Multi-view multi-task spatiotemporal graph convolutional network for air quality prediction[J]. *Science of The Total Environment*, 2023, 893: 164699.
- [35] WANG H, ZHANG L, WU R, et al. Spatio-temporal fusion of meteorological factors for multi-site PM2.5 prediction: A deep learning and time-variant graph approach[J]. *Environmental Research*, 2023, 239: 117286.
- [36] WANG Y, TIAN S, ZHANG P. Novel spatio-temporal attention causal convolutional neural network for multi-site PM2.5 prediction[J]. *Frontiers in Environmental Science*, 2024, 12: 1408370.
- [37] WU J, XU L, FAN S, et al. A Spatiotemporal graph attention network for PM2.5 forecasting using multi-source data[J]. *Air Quality, Atmosphere & Health*, 2025, 18(10): 3037-3052.
- [38] LU G Y, WONG D W. An adaptive inverse-distance weighting spatial interpolation technique[J]. *Computers & Geosciences*, 2008, 34(9): 1044-1055.
- [39] De MESNARD L. Pollution models and inverse distance weighting: Some critical remarks[J]. *Computers & Geosciences*, 2013, 52: 459-469.
- [40] MASROOR K, FANAEI F, YOUSEFI S, et al. Spatial modelling of PM2.5 concentrations in Tehran using Kriging and inverse distance weighting (IDW) methods[J]. *Journal of Air Pollution and Health*, 2020.

- [41] MORAN P A P. Notes on Continuous Stochastic Phenomena[J]. *Biometrika*, 1950, 37(1/2): 17.
- [42] ANSELIN L. Local Indicators of Spatial Association—LISA[J]. *Geographical Analysis*, 1995, 27(2): 93-115.
- [43] HAKLAY M. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets[J]. *Environment and Planning B: Planning and Design*, 2010, 37(4): 682-703.
- [44] GIRRES J F, TOUYA G. Quality Assessment of the French OpenStreetMap Dataset[J]. *Transactions in GIS*, 2010, 14(4): 435-459.
- [45] NEIS P, ZIELSTRA D. Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap[J]. *Future Internet*, 2014, 6(1): 76-106.
- [46] BOEING G. OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks[J]. *The Journal of Open Source Software*, 2017, 2(12): 215.
- [47] LI H, FAN H, MAO F. A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM_{2.5}) in Beijing, China[J]. *Atmosphere*, 2016, 7(3): 35.
- [48] KE G, MENG Q, FINLEY T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree[C]//*Advances in Neural Information Processing Systems* 30. 2017: 3146-3154.
- [49] LUNDBERG S M, LEE S I. A Unified Approach to Interpreting Model Predictions[C]//*Advances in Neural Information Processing Systems* 30. 2017: 4765-4774.
- [50] Yuval, LEVI Y, BRODAY D M. Revealing causality in the associations between meteorological variables and air pollutant concentrations[J]. *Environmental Pollution*, 2024, 345: 123526.