



基于气象大模型的门控时空融合网络闪电预报方法

李怡然^{1,2}, 李清勇^{1,2}, 郑栋^{2,3}, 耿阳李敖^{1,2*}, 郭志清^{1,2}, 徐良韬³, 姚雯³, 吕伟涛³

1. 北京交通大学交通大数据与人工智能教育部重点实验室, 北京 100044

2. 中国气象局雷电重点开放实验室, 北京 100081

3. 中国气象科学研究院灾害天气科学与技术全国重点实验室, 北京 100081

* 通讯作者, E-mail: ylageng@bjtu.edu.cn

收稿日期: 2025-01-01; 收修改稿日期: 2025-06-27; 接受日期: 2025-07-07; 网络版发表日期: 2025-08-08

中国气象局雷电重点开放实验室开放课题(2023KELL-B002、2024KELL-A001)、国家自然科学基金项目(62306028、42075088、U2342215)和深圳市科技计划(KJZD20240903102742055)资助

摘要 闪电属于典型的自然灾害之一, 对人类生活与工业生产有重大影响, 精细化闪电预报对防灾减灾具有重要意义. 传统闪电预报方法主要依赖数值天气模式, 需要通过大量计算资源求解复杂的物理方程. 近年来, 基于人工智能的天气预报模型迅速发展, 尤其是气象大模型技术, 预报准确度可与世界领先的数值天气模式相媲美, 而计算开销仅为其万分之一. 然而, 当前气象大模型无法直接生成闪电预报产品, 且在空间分辨率上也难以满足精细化预报的需求. 为此, 本文研究了基于气象大模型的精细化闪电预报方法, 提出大模型与闪电观测双源数据驱动的预报框架, 融合二者的互补优势提升闪电预报效果. 进一步, 本文设计了一种门控时空融合网络预报模型 gSTFNet 用以解决双源数据在跨时域和跨模态融合中面临的挑战. gSTFNet 首先通过双编码结构对气象大模型预测结果和临近闪电观测数据分别进行特征编码, 减小两类数据在特征空间中的模态差异. 随后利用门控时空融合模块建模双源特征的时空关联, 实现二者的跨时域融合. 最后通过反卷积网络对融合特征进行解码生成预报. 本文利用广东地区 2018~2022 年五年的真实闪电观测数据对方法进行了评估. 实验结果表明: (1) 在 ETS 评分上, 双源融合的预报思路相较于直接基于大模型的闪电预报结果提升约 50%, 比欧洲中期天气预报中心发布的 HRES 闪电预报产品提升约 300%; (2) 与其他基于深度神经网络的预报模型相比, 本文提出的 gSTFNet 模型取得了更优的性能表现.

关键词 闪电预报, 气象大模型, 神经网络, 深度学习, 时空数据融合

1 引言

闪电是一种常见的极端天气现象, 其产生伴随着高电压、大电流和强电磁脉冲, 具有极大的破坏性. 闪电现象多形成于对流发展成熟的积雨云系统内, 不

同冰相物之间的回弹碰撞导致电荷的转移, 使不同的粒子携带不同极性的电荷, 并在动力作用驱动下在云内不同位置形成不同电荷区, 从而导致闪电的发生(郗秀书等, 2014; 郭凤霞等, 2016). 我国雷电灾害造成的人员伤亡和经济损失十分严重, 据统计, 1997~2006 年

中文引用格式: 李怡然, 李清勇, 郑栋, 耿阳李敖, 郭志清, 徐良韬, 姚雯, 吕伟涛. 2025. 基于气象大模型的门控时空融合网络闪电预报方法. 中国科学: 地球科学, 55(9): 3086–3103, doi: 10.1360/SSTe-2025-0001

英文引用格式: Li Y, Li Q, Zheng D, Geng Y, Guo Z, Xu L, Yao W, Lyu W. 2025. A gated spatiotemporal fusion network for lightning forecasting based on weather foundation models. Science China Earth Sciences, 68(9): 2957–2975, <https://doi.org/10.1007/s11430-025-1638-8>

雷灾造成我国4488人死亡、4320人受伤(马明等, 2008). 闪电除了能对人畜造成直接的触电伤害外, 还对电力设施、通信设备、航空航天器、建筑物、树木等有很强的破坏力, 对基础设施和自然环境构成了广泛而严重的威胁(张义军和周秀骥, 2006). 通过对闪电进行及时准确的预报, 可以为相关部门制定应急措施和防范计划提供科学依据, 增强社会的防灾减灾能力. 因此, 发展精准的闪电预报技术具有重要的现实意义.

数值天气预报(numerical weather prediction, NWP)技术是开展闪电活动预报的一种有效手段(Fierro等, 2013; Lopez, 2016). 基于NWP模式输出的动力场、热力场及微物理场信息, 可以建立闪电的诊断参数化方案(McCaul等, 2009; Lynn等, 2012; Wong等, 2013), 利用代用指标开展闪电活动预报. 从云内真实物理过程的角度, 还可以将云内起电物理过程耦合至数值预报模式中, 同时采用放电物理参数化方法对闪电活动进行直接预报(Dafis等, 2018; Xu等, 2022; Mondal等, 2023). 无论是闪电的诊断参数化还是直接的物理参数化, 都依赖于模式对于动力微物理以及闪电参数化方案的合理描述, 而目前对于相关物理过程的认识还存在较大欠缺(Cavaola等, 2024). 此外, NWP模式在运行时需要消耗大量的计算资源, 这也使得NWP在提升预报精细度时面临巨大算力挑战(Bauer等, 2015).

近年来, 以深度学习为代表的人工智能技术得到了快速发展, 催生了基于人工智能的天气预报方法. Shi等(2015)将降水预测问题转化为时空序列预测问题, 并提出基于ConvLSTM的递归时空预测模型, 取得了优于传统方法的预测结果. 基于此, Wang等(2017)提出改进模型PredRNN, 解决了ConvLSTM层与层之间的记忆单元在时域上的相互独立问题, 取得了更优的性能. Tan等(2023b)提出一种基于卷积与注意力机制的非递归时空预测模型, 比递归模型具有更高的预测效率, 在全球12h温度预报上取得了领先性能. Andrychowicz等(2023)提出一种基于UNet结构的天气模型MetNet-3, 通过引入MaxVit模块融合地面气象站、雷达和卫星数据, 实现对未来24h的地表气象要素的准确预报. Guo等(2024)使用ConvLSTM对欧洲中期天气预报中心发布的全球预报系统数据进行校正, 显著提高了4~7天预报时段内中等强度及以上降水的预报准确性. 在闪电预报方面, Geng等(2019)提出基于

ConvLSTM的双编码器模型LightNet, 将NWP数据和历史闪电观测数据编码后进行融合解码, 实现对闪电的短时预报, 相较以往的预报方案取得了更高的精度. 周康辉等(2021)采用SegNet模型将卫星、雷达以及地闪等观测数据和高分辨率数值模式预测数据进行融合, 得到更有效的闪电落区短时预报结果. 张烨方等(2021)采用卷积神经网络为基础架构, 结合多个时间序列的雷达产品与闪电数据, 实现对未来30min的闪电预警. Lu等(2024)提出一种用于闪电临近预报的多模态MCGLN模型, 该模型将生成对抗网络与ConvLSTM集成, 以改进多维时空特征信息的精细化捕捉能力, 从而提高预测准确性. Cavaola等(2024)提出采用一维卷积结构的FlashNet模型, 基于欧洲中期天气预报中心发布的NWP数据实现对未来24h闪电的预报. 这些基于深度学习的天气预报方法能自动从数据中挖掘规律, 具有更灵活的数据融合与非线性建模能力, 在短期预报任务中表现出色. 然而, 这些模型通常侧重于局部区域的单一天气状况预测, 难以实现对大范围、长时间跨度的多气象参数耦合预测(Schultz等, 2021).

最近两年, 随着数据与模型规模的持续增长, 神经网络展现出能力涌现效应, 人工智能进入大模型时代. 在气象预报方面, 不少国内外团队开展了气象大模型的研究. 气象大模型是一个具有大规模参数量的神经网络, 在大量全球再分析数据(Hersbach等, 2020)上进行自回归任务训练而成, 取得了与先进NWP相当的预测精度, 且计算开销仅为NWP的万分之一(Charlton-Perez等, 2024). 典型的气象大模型包括NVIDIA团队的FourCastNet(Kurth等, 2023)、DeepMind团队的GraphCast(Lam等, 2023)、Google Research团队的NeuralGCM(Kochkov等, 2024)以及来自国内华为团队的Pangu-Weather(Bi等, 2023)和复旦大学的FuXi(Chen等, 2023). 表1总结了目前五个典型开源气象大模型的基本信息, 它们的主要预报参量基本一致. 值得注意的是, 多数模型的时间分辨率为6h, 只有Pangu-Weather的时间分辨率可以达到1h, 这使得其在中短期天气预报上更具优势. 然而, 当前气象大模型主要针对基础气象参量, 无法直接预报闪电, 如何将其充分适配闪电预报任务尚无成熟的解决方案. 此外, 气象大模型预报的空间分辨率较粗($0.25^\circ \times 0.25^\circ$), 直接使用参数化方法生成闪电预报在精细程度上难以满足现实需求

表 1 五个典型开源气象大模型的基本信息

模型	时间分辨率	空间分辨率	主要预报参量
FourCastNet	6h	0.25°×0.25°	地表变量: 10m风场的 U 分量及 V 分量、2m温度、表面气压、平均海平面气压、总柱状水汽 高空变量: 温度、风场的 U 分量及 V 分量、相对湿度、位势
GraphCast	6h	0.25°×0.25°	地表变量: 10m风场的 U 分量及 V 分量、2m温度、平均海平面气压、总降水 高空变量: 温度、风场的 U 分量及 V 分量、比湿度、位势、垂直风速
NeuralGCM	6h	0.7°×0.7°	高空变量: 温度、风场的 U 分量及 V 分量、比湿度、位势、比云冰水含量、比云液态水含量
FuXi	6h	0.25°×0.25°	地表变量: 10m风场的 U 分量及 V 分量、2m温度、平均海平面气压、总降水 高空变量: 温度、风场的 U 分量及 V 分量、相对湿度、位势
Pangu-Weather	1h	0.25°×0.25°	地表变量: 10m风场的 U 分量及 V 分量、2m温度、平均海平面气压 高空变量: 温度、风场的 U 分量及 V 分量、比湿度、位势

(Geng等, 2019).

针对上述情况, 本文研究了基于气象大模型(具体采用Pangu-Weather)的精细化闪电预报方法, 提出大模型与临近闪电观测数据双源驱动的预报框架, 旨在融合二者在长期天气趋势预测与短期外推的互补优势提升预报性能. 此外本文设计了一种门控时空融合网络gSTFNet用于解决双源数据在跨时间和跨模态融合中的挑战. gSTFNet模型主要包括双源编码器、门控时空融合模块、预测解码器几个部分. 首先使用双源编码器分别对气象大模型预测结果和临近闪电观测数据进行特征编码, 有效减少潜在特征空间中的模态差异. 随后, 通过门控时空融合模块建模双源特征编码的时空关联, 实现无缝的跨时间融合. 最后, 使用预测解码器提升融合特征的空间分辨率, 最终生成高精度度闪电预报产品. 本文使用广东地区2018~2022年五年真实闪电观测数据对方法进行验证, 预报时长为6h, 时间分辨率为1h, 空间分辨率为 $0.04^{\circ}\times 0.04^{\circ}$.

与前人研究相比, 本研究的特色和创新包括: (1) 首次探索了基于气象大模型的精细化闪电预报方法, 提出大模型与临近闪电观测数据双源驱动的思路, 融合二者在长期天气趋势预测与短期外推的互补优势, 提升闪电预报的整体性能. (2) 设计了一种门控时空融合网络gSTFNet, 利用双编码结构和门控时空融合机制有效解决闪电预报中跨时间和跨模态数据融合挑战. (3) 实验结果表明, 所提方法相较于直接基于大模型的闪电预报结果在ETS评分上提升约50%, 比欧洲中期天气预报中心发布的HRES闪电预报产品提升约300%.

2 数据与处理

2.1 历史闪电观测数据

本研究采用了2018~2022年广东地区真实闪电观测数据, 采集自粤港澳闪电定位系统(陈绿文等, 2020). 粤港澳闪电定位系统共有19个探测子站, 采用时差-方向综合定位方法, 可测量闪电发生的时间、位置、极性、强度和云地闪识别等参量(张悦等, 2021). 基于广州高建筑雷击观测评估该系统, 表明其对地闪回击和闪电的探测效率都超过90%, 定位误差的均值约为300m(陈绿文等, 2020). 基于广州从化地区的人工触发闪电评估该系统, 其探测到了96%的触发闪电和88%的触发闪电回击事件, 定位误差的平均值约为279m, 对地闪的识别正确率为93%(张悦等, 2022).

本文的研究区域如图1所示, 广东省位于中国大陆东南沿海地区, 地处低纬, 地势总体北高南低, 东北部为中低山地, 中部主要为丘陵盆地, 南部主要为沿海冲积平原. 由于广东省位于太平洋西岸, 是西太平洋热带气旋登陆的主要地区, 热带气旋系统不仅给周边地区带来强烈的大风和暴雨, 也常伴随闪电活动(潘伦湘等, 2010). 根据1995~2018年数据统计显示, 广东省的闪电活动主要集中在4至9月出现, 占全年90%以上(殷启元等, 2021).

选定预报区域范围为 $20^{\circ}\sim 26^{\circ}\text{N}$ 、 $109^{\circ}\sim 118^{\circ}\text{E}$, 覆盖了广东省全境, 选取高发月份(每年4至9月)的闪电数据用于研究, 相关信息总结如表2所示.

数据集具体的处理方式如下: 首先设定空间分辨率为 $0.04^{\circ}\times 0.04^{\circ}$, 将预报区域分为包含 150×225 个格点

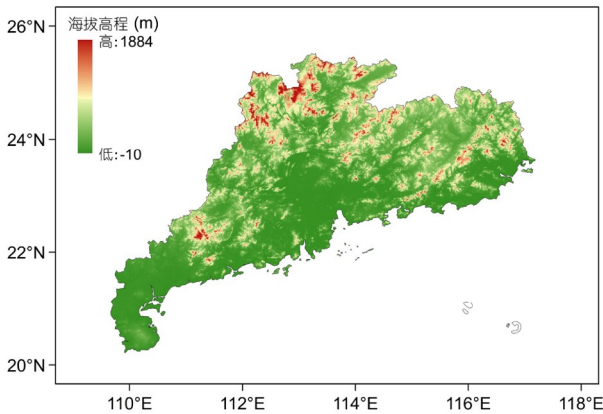


图 1 研究区域示意图

表 2 闪电观测数据基本信息

基本信息	闪电观测数据
时间范围	2018~2022年, 每年4~9月
经纬度范围	20°~26°N、109°~118°E
时间分辨率	逐小时
空间分辨率	0.04°×0.04°

的网格。统计每个格点逐小时的闪电发生次数并处理为二值矩阵, 1代表该小时内该格点发生了闪电, 0代表未发生闪电。由于在一个时刻的网格数据中, 无闪电格点远多于有闪电格点。为避免过量无闪电格点带来的影响, 本文使用如下的数据选取方式: 将连续9h的闪电数据合并为一组样本, 若一组样本中所有小时发生闪电的格点数均不足30, 则过滤掉这组数据。经过过滤后最终得到2597组样本。

2.2 再分析数据

气象大模型需要使用再分析数据作为输入, 为此本研究采用了欧洲中期天气预报中心第五代大气再分析资料(ERA5), 其基于四维变分资料同化方案和综合预测系统生成大气、海陆的各种气象变量。ERA5数据集在垂直方向上提供了从1000hPa到1hPa一共37个不同气压层级的大气数据以及海陆表面数据, 在水平方向上以经纬度网格的方式提供数据, 空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$, 时间分辨率为1h(Hersbach等, 2019)。Pangu-Weather使用了ERA5数据集中四种地表变量以及五种高空变量在13个气压层次的数值, 具体气象要素及相关信息见表3。

2.3 数据集划分

本研究选取2018~2020年三年的数据作为训练集, 2021年的数据为验证集, 2022年的数据为测试集, 数据集具体划分情况见表4。

3 基于Pangu-Weather的门控时空融合网络闪电预报方法

3.1 问题定义

假设当前时刻为 $t = 0$, 首先, 使用Pangu-Weather接收过去一个时刻的ERA5数据作为输入, 预测输出未来 T_f 个时刻的大模型通用气象参量预测数据 $[\hat{E}_t]_{t=1}^{T_f}$, 这些预测结果覆盖全球空间范围。然而, 由于闪电预报任务通常聚焦于特定的局部区域, 因此本文将预测数据 $[\hat{E}_t]_{t=1}^{T_f}$ 裁剪为 $[W_t]_{t=1}^{T_f}$, 接着利用双编码神经网络模型 f 接收未来 T_f 个时刻的Pangu预测数据 $[W_t]_{t=1}^{T_f}$ 以及过去 T_p 个时刻的闪电观测数据 $[O_t]_{t=-T_p}^{-1}$ 作为输入, 预报未来 T_f 个时刻的闪电发生情况 $[\hat{O}_t]_{t=1}^{T_f}$, 问题形式化定义如下:

$$f\left([W_t]_{t=1}^{T_f}, [O_t]_{t=-T_p}^{-1}\right) \rightarrow [\hat{O}_t]_{t=1}^{T_f}. \quad (1)$$

由于数据 $[W_t]_{t=1}^{T_f}$ 和 $[O_t]_{t=-T_p}^{-1}$ 都具有时间和空间结构, 因此需要预测模型能够在这些数据之间建立复杂的时空依赖关系; 此外, 考虑到基础气象参量预测数据和历史观测数据提供了关于闪电发生的互补信息, 本文提出一种门控时空融合网络gSTFNet解决上述预报问题。

3.2 Pangu-Weather

Pangu-Weather是一种使用ERA5数据训练的大规模神经网络模型, 其能够对未来多个气压层的基础气象参量数据(详见表3)进行准确预测, 运行速度比传统NWP提升10000倍以上(Bi等, 2023)。

如图2所示, Pangu-Weather采用了类似Swin Transformer的3D Earth-Specific Transformer架构, 基于Transformer中的注意力机制挖掘不同局部区域的相关性进行特征的聚合与变换。首先将输入数据按照空间关系切分成局部小块, 然后将每个局部块投影至高维

表 3 Pangu-Weather使用的气象要素及相关信息

气象要素	相关信息
地表变量	10mU风(m s^{-1})、10mV风(m s^{-1})、2m温度(K)、平均海平面气压(Pa)
高空变量	U风(m s^{-1})、V风(m s^{-1})、温度(K)、位势($\text{m}^2 \text{s}^{-2}$)、比湿度(g kg^{-1})
气压层	50、100、150、200、250、300、400、500、600、700、850、925、1000hPa
时空分辨率	1h、 $0.25^\circ \times 0.25^\circ$

表 4 数据集划分情况

划分方案	时间范围	样本数量
训练集	2018~2020年, 每年4~9月	1664组
验证集	2021年4~9月	551组
测试集	2022年4~9月	382组

空间中并重组为局部块嵌入矩阵, 接着将其送入具有8个编码器层和8个解码器层的标准编码-解码器结构进行传播. 最后将得到的结果投影到原始空间, 产生所需的输出. 基于上述结构, Pangu-Weather接收某时刻的基础气象参量数据 E_t (初始来自ERA5)作为输入, 预测输出 $\hat{E}_{t+1}$. 递归地将输出作为输入, Pangu-Weather可以实现未来多个时刻的预测: $\hat{E}_{t+1}, \hat{E}_{t+2}, \dots$

3.3 门控时空融合网络模型

图3展示了本文所提出的基于gSTFNet的预测框架. gSTFNet的输入包含两部分: Pangu-Weather输出的广东地区未来 T_f 个小时的基础气象参量预测数据, 以及过去 T_p 个小时的真实闪电观测数据. gSTFNet采用双编码结构(图中模块 I 与模块 II)分别提取双源输入数据的特征编码, 接着引入门控时空融合模块(图中模块III)建模融合特征的时空相关性, 最后通过预测解码器(图中模块IV)对处理后的特征进行解码, 输出最终的预报结果. 下面将对每一个模块进行详细说明.

3.3.1 Pangu预测数据编码器

此模块主要负责对未来 T_f 个小时的Pangu预测数据 $[W_t]_{t=1}^{T_f}$ 进行特征编码, 方便下一步融合处理. 该模块专注于从单帧数据中提取空间特征, 而不涉及时间演化的建模. 首先将大模型预测的四种地表变量以及五种高空变量在4个与闪电现象密切相关的气压层次

(200、500、850、1000hPa)的数据在通道维度进行拼接, 作为该模块的输入数据. 考虑到大模型预测变量的空间分辨率为 $0.25^\circ \times 0.25^\circ$, 为了与闪电网格对齐, 采用线性插值操作将其空间分辨率提升至 $0.04^\circ \times 0.04^\circ$. 编码器采用4层带归一化的二维卷积网络逐小时对数据的空间特征模式进行提取, 得到Pangu预测数据的特征编码. 注意到为了提升模型的运行效率, 第二层与第四层卷积的步长分别设置为2, 因此产生的特征编码的空间尺寸缩减为原始输入的四分之一. 编码过程形式化如下:

$$\Phi^{(i)} = \text{SiLU}(\text{Norm2d}(\text{Conv2d}(\Phi^{(i-1)}))), 1 \leq i \leq 4. \quad (2)$$

其中SiLU表示S型线性激活函数(Elfwing等, 2018), Norm2d表示二维归一化层, 用于加速模型收敛. 具体采用了分组归一化方式(Wu和He, 2018), 其被证明具有对训练批次大小鲁棒的优势, 适合本任务大数据、小批次的特性. Conv2d表示二维卷积层, $\Phi^{(0)} = W$ 为输入的Pangu预测数据, $F^w = \Phi^{(4)}$ 为处理后的特征编码.

3.3.2 闪电观测数据编码器

此模块主要负责对过去 T_p 个小时的闪电观测数据 $[O_t]_{t=-T_p}^{-1}$ 进行特征提取. 与Pangu预测数据编码器类似, 闪电观测数据编码器同样采用4层带归一化的二维卷积网络逐小时对数据的空间特征模式进行提取, 得到闪电观测数据特征编码. 编码过程形式化如下:

$$\Psi^{(i)} = \text{SiLU}(\text{Norm2d}(\text{Conv2d}(\Psi^{(i-1)}))), 1 \leq i \leq 4. \quad (3)$$

其中, $\Psi^{(0)} = O$ 为输入的闪电观测数据, $\Psi^{(4)}$ 为处理后的特征编码.

与逐小时描述未来天气状况的Pangu预测数据不同, 闪电观测数据反映的是历史闪电活动, 其与未来闪电发生的关联性主要体现在时间维度上. 为了捕捉这种时间演化特性, gSTFNet引入了两层分解时空卷积

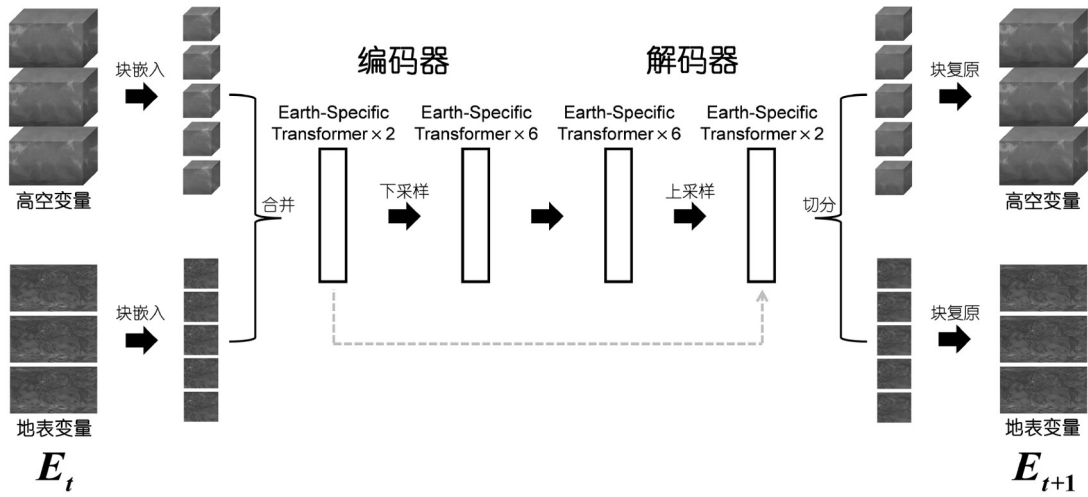


图2 Pangu-Weather气象大模型示意图

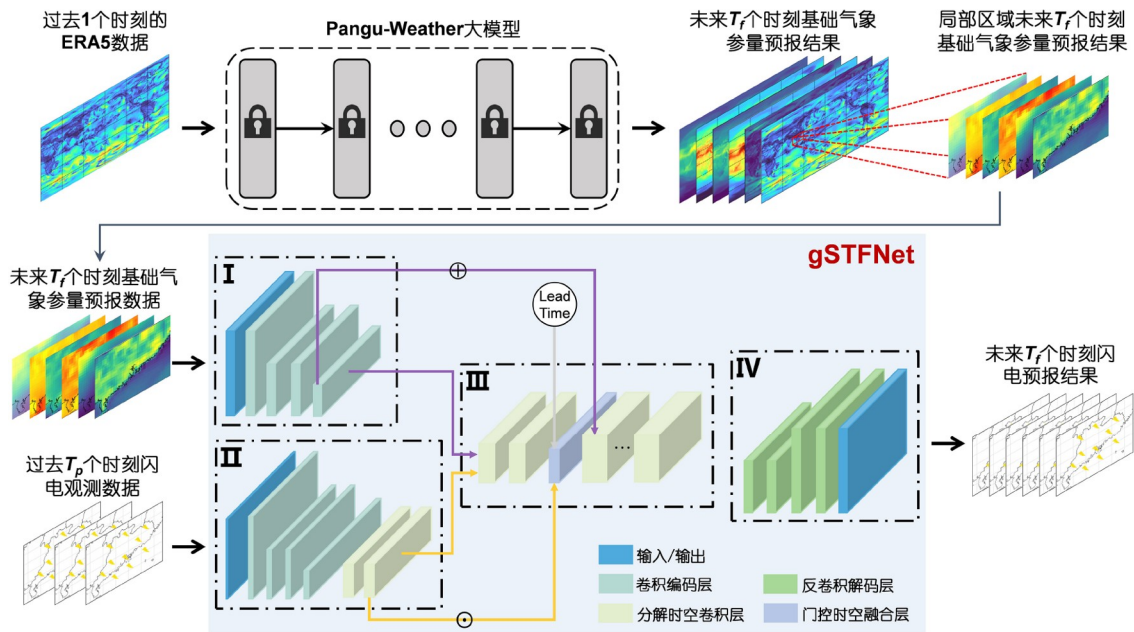


图3 预报框架与gSTFNet模型结构图

(DSTConv)层,用于提取多帧特征,有效建模观测数据的时空依赖性.具体的建模过程描述如下:

$$Z^{(j)} = \text{DSTConv}(Z^{(j-1)}), 1 \leq j \leq 2. \quad (4)$$

其中, $Z^{(0)} = \psi^{(4)}$ 为处理得到的闪电特征编码, $F^l = Z^{(2)}$ 为最终得到的闪电时空特征编码.

DSTConv层的设计参考了Tan等(2023a)提出的方

法,通过堆叠三种参数高效的卷积模块来近似实现大核卷积在时空模式提取上的效果:首先分别通过深度卷积 Conv_{Dw} 与深度空洞卷积 $\text{Conv}_{\text{Dw-d}}$ 挖掘特征的局部与远距离空间相关性,然后基于 1×1 卷积 $\text{Conv}_{2d_{1 \times 1}}$ 实现对时空依赖性建模.此外,还引入了残差连接以增强空间相关特征的提取效果.具体建模过程形式化如下:

$$Z^{(j)} = \text{DSTConv}(Z^{(j-1)}) \\ = Z^{(j-1)} + \text{Conv2d}_{1 \times 1}(\text{Conv}_{\text{Dw-d}}(\text{Conv}_{\text{Dw}}(Z^{(j-1)}))). \quad (5)$$

3.3.3 门控时空融合模块

作为gSTFNet的核心部分, 该模块旨在挖掘双源特征编码的互补信息并实现融合, 其中待融合两种特征源在物理意义和时间范围上均存在差异. 为了实现有效的融合, 模块的设计参考了以下先验观察: 历史闪电数据在未来预测中的重要性会随着预测时长的不同而发生变化, 也即临近的闪电观测与未来短时间内的闪电活动高度相关, 但这种相关性会随着预测周期的延长而逐渐减弱.

上述现象启发本文设计一种动态门控时空融合机制(gated spatio-temporal fusion, gSTF). 如图4所示, 对于每个时刻的大模型编码特征 F_t^w , gSTF利用门控机制将其与闪电观测特征 F^l 进行融合, 其中门控机制可以自适应调节 F^l 对于未来各时刻闪电发生的贡献程度. 具体操作形式化如下:

$$\Gamma_t = \text{sigmoid}(\text{DSTConv}([F_t^w, F^l]) + \text{TimeEmb}(t)) \\ \odot F^l + F_t^w, 1 \leq t \leq T_f. \quad (6)$$

这里门控权重的计算主要包含两部分: 一是利用DSTConv建模 F_t^w 与 F^l 之间的相关性, 二是可学习的时间嵌入TimeEmb(t)(Andrychowicz等, 2023)用于刻画闪

电观测特征在不同预测时长上的时间影响. 具体而言, TimeEmb(t)对每个时间步 t 都创建一个独热向量 $\mathbf{e}_t \in \mathbb{R}^{T_f}$ (其中 $\mathbf{e}_t$ 的第 t 个维度为1, 其他维度为0), 接着通过一个全连接网络将 $\mathbf{e}_t$ 映射至 C 维空间, 其中 C 为DSTConv($[F_t^w, F^l]$)所产生嵌入表示的通道数. 外层的sigmoid为激活函数, 保证门控权重的值位于0~1. 门控权重通过逐元素相乘 $\odot$ 控制 F^l 的影响, 最后与 F_t^w 相加实现融合.

随后, 采用4层DSTConv网络提取逐小时融合特征 $\Gamma = [\Gamma_t]_{t=1}^{T_f}$ 的时空特征模式, 该过程形式化如下:

$$\Theta^{(k)} = \text{DSTConv}(\Theta^{(k-1)}), 1 \leq k \leq 4. \quad (7)$$

其中, $\Theta^{(0)} = \Gamma$ 为输入的融合特征, $F = \Theta^{(4)}$ 表示最终得到的时空融合特征.

3.3.4 预测解码器

此模块将对门控时空融合模块融合后的时空特征 F 进行解码, 输出最终的预报结果. 解码过程采用4层带归一化的二维反卷积网络, 对特征信息逐小时进行解码, 得到最终的闪电预报结果 $\hat{O}$. 解码过程形式化如下:

$$\Omega^{(l)} = \text{SiLU}(\text{Norm2d}(\text{DeConv2d}(\Omega^{(l-1)}))), 1 \leq l \leq 4. \quad (8)$$

其中, $\Omega^{(0)} = F$ 为初始输入特征. 对解码器的输出 $\Omega^{(4)}$ 进一步利用一个 1×1 卷积处理即可得到最终的模型输

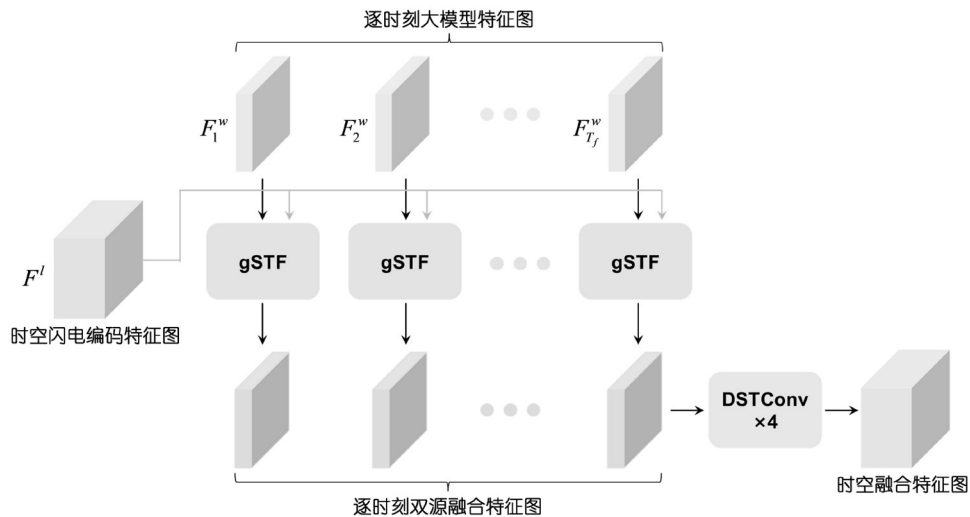


图4 门控时空融合模块结构图

出 $\hat{O}$, 其包含未来 T_f 个时刻逐小时的闪电格点预报, 空间分辨率为 $0.04^\circ \times 0.04^\circ$.

3.3.5 模型细节

图5展示了gSTFNet各模块的详细结构. 图中标识符含义如下: Conv表示卷积编码层, DSTConv表示分解时空卷积层, Gated Fusion表示门控时空融合层, DeConv表示反卷积解码层, 最终的Conv 1×1 表示 1×1 卷积层, 用于确保模型输出单通道的闪电预报结果. 表5展示了各层的详细参数配置, 包括卷积核数量、尺寸(为了简洁, 表中省略了卷积核的通道数量, 其与上一层的卷积核个数一致)、步长、填充以及空洞率. 其中, DSTConv层的三种参数配置分别对应其内部三种卷积模块: 深度卷积(Conv_{Dw})、深度空洞卷积(Conv_{Dw-d})、二维 1×1 卷积(Conv_{2d $_{1 \times 1}$}).

3.3.6 损失函数与训练细节

得到模型的预测结果 $\hat{O}$ 后, 需要与真实观测数据 O 对比计算误差损失, 通过最小化误差损失实现模型的训练. 由于正样本(发生闪电)的格点数量远小于负样本(未发生闪电)的格点数量, 为了平衡二者对模型影响, 采用了加权交叉熵损失函数(Weighted Cross Entropy):

$$L_{WCE}(O, \hat{O}) = - \sum_{t=1}^{T_f} \sum_{x,y} [\omega O_t^{x,y} \log(\hat{O}_t^{x,y}) + (1 - O_t^{x,y}) \log(1 - \hat{O}_t^{x,y})]. \quad (9)$$

其中, $O = [O_t]_{t=1}^{T_f}$ 与 $\hat{O} = [\hat{O}_t]_{t=1}^{T_f}$ 分别表示真实观测与模型预测的闪电网格, ω 是正样本的权重, 在实验中设置 $\omega = 10$.

训练过程中, 本文选取SGD优化器(Amari, 1993), 初始学习率设置为0.001, 批次大小为8. 模型共训练70个迭代周期, 在前40个迭代周期之后对学习率进行指数衰减, 衰减系数设置为0.9. 最终选择在验证集上AUC指标(见章节4.1.1)最高的模型用于测试评估.

4 实验与结果分析

4.1 实验设置

4.1.1 评价指标

本研究选取了命中率(POD)、虚警率(FAR)、威胁评分(TS)以及公平威胁评分(ETS)四个指标(Clark等, 2010)用于评估模型的预测效果, 计算公式如表6所示, 其中, TP、FP、TN、FN为闪电预报不同情况格点数统计量, 如表7所示.

在预报任务中, 命中率(POD)和虚警率(FAR)通常存在相关性, 命中率的提升可能会导致虚警率的增加. 因此为了使评估结果具有可比性, 通过调整预测阈值, 分别在命中率为0.3、0.4、0.5的情况下进行模型对比. 此外, 本文还采用了机器学习中常用的AUC(area under curve)指标, 以真阳性率为纵轴和假阳性率为横轴绘制的曲线下方面积, 其优点在于与预测阈值的选择

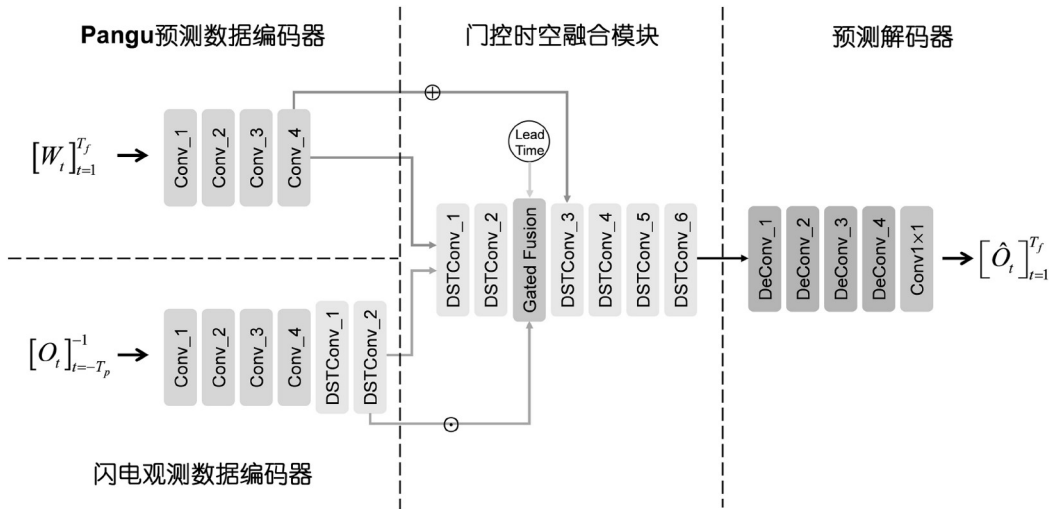


图5 gSTFNet模型详细结构

表 5 gSTFNet网络层详细参数配置

模块	标识符	卷积核个数	卷积核尺寸	步长	填充	空洞率
Pangu预测数据编码器	Conv_1	32	(3×3)	1	1	1
	Conv_2	32	(3×3)	2	1	1
	Conv_3	32	(3×3)	1	1	1
	Conv_4	32	(3×3)	2	1	1
闪电观测数据编码器	Conv_1	16	(3×3)	1	1	1
	Conv_2	16	(3×3)	2	1	1
	Conv_3	16	(3×3)	1	1	1
	Conv_4	16	(3×3)	2	1	1
	DSTConv_1	[48 48 48]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_2	[32 32 32]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
门控时空融合模块	DSTConv_1	[80 80 80]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_2	[256 256 256]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_3	[192 192 192]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_4	[64 64 64]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_5	[64 64 64]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
	DSTConv_6	[64 64 64]	[(5×5) (7×7) (1×1)]	[1 1 1]	[2 9 0]	[1 3 1]
预测解码器	DeConv_1	32	(3×3)	2	1	1
	DeConv_2	32	(3×3)	1	1	1
	DeConv_3	32	(3×3)	2	1	1
	DeConv_4	32	(3×3)	1	1	1
	Conv1×1	1	(1×1)	1	0	1

无关. 当AUC接近于1时, 表示模型在不同阈值下能够高度区分正样本(有闪电)和负样本(无闪电), 保持较高的性能; 相反, 当AUC接近于0.5或更低时, 表示模型的性能与随机猜测相当, 即模型对正负样本没有区分能力.

4.1.2 模型对比基准

模型对比基准主要包括7个现有预报模型以及gSTFNet自身的2个消融模型. (1) HRES(Lopez, 2016; Malardel等, 2016): 欧洲中期天气预报中心发布的HRES数值天气模式闪电预报产品(litota); (2) ConvLSTM(Shi等, 2015): 是一种将卷积神经网络(CNN)与长短期记忆网络(LSTM)相结合的递归时空序列预测模型; (3) PredRNN(Wang等, 2017): ConvLSTM的改进模型, 添加了纵向传播的记忆单元, 具有更强的空间特征提取能力; (4) SimVPv2(Tan等, 2023b): 是一种使用CNN和注意力机制的非递归时空

预测模型; (5) TAU(Tan等, 2023a): 是一种使用注意力机制来并行化地处理时间演变的时空序列预测模型; (6) MetNet-3(Andrychowicz等, 2023): 通过将MaxVit模块加入到UNet网络中, 结合局部和全局的时空信息来实现对未来天气变量的预测; (7) LightNet(Geng等, 2019): 基于ConvLSTM的双时空编码结构的闪电预报模型; (8) gSTFNet-P: 只使用Pangu大模型基础气象参量预测数据训练的gSTFNet消融模型, 其中闪电观测数据输入为零. (9) gSTFNet-L: 只使用闪电观测数据训练的gSTFNet消融模型, 其中Pangu大模型基础气象参量预测数据输入为零.

其中, 六种深度学习模型(ConvLSTM、PredRNN、SimVPv2、TAU、MetNet-3、LightNet)的输入与gSTFNet保持一致, 均采用大模型预测与闪电观测双源输入. 同时, 这些模型的训练参数设置也与gSTFNet保持相同. 对于HRES数据, 采用的是起报时刻向前所能追溯的最新发布的预报产品. 在空间分辨

表 6 评价指标及计算公式

评价指标	公式	取值范围	含义
POD	$POD = \frac{TP}{TP+FN}$	[0,1]	正确预测发生闪电的格点数量与实际发生闪电的格点数量比值
FAR	$FAR = \frac{FP}{TP+FP}$	[0,1]	错误预测发生闪电的格点数量与预测发生闪电的总格点数量比值
TS	$TS = \frac{TP}{TP+FP+FN}$	[0,1]	正确预测发生闪电的格点数量与移除了正确预测未发生闪电的格点数量的比值
ETS	$ETS = \frac{TP-R}{TP+FP+FN-R}$ $\left(R = \frac{(TP+FP)(TP+FN)}{(TP+FP+FN+TN)} \right)$	$\left[-\frac{1}{3}, 1 \right]$	正确预测发生闪电的格点数量与移除了正确预测未发生闪电的格点数量的比值, 但消除了随机预测中偶然命中的贡献(公式中的 R)

表 7 闪电预报不同情况格点数统计量

观测情况\预测结果	预测有闪电	预测无闪电
观测有闪电	TP	FN
观测无闪电	FP	TN

率方面, 所有深度学习模型输出的分辨率均为 $0.04^\circ \times 0.04^\circ$; HRES 预报结果的原始空间分辨率是 $0.1^\circ \times 0.1^\circ$, 为了保证模型结果的一致性和可比性, 本文使用线性插值方法将其提升至 $0.04^\circ \times 0.04^\circ$.

4.1.3 损失函数对比基准

如章节 3.3.6 提到的, 为了解决正负样本不平衡问题, 本文采用了加权交叉熵损失. 为了验证其有效性, 额外引入以下两种经典损失函数作为对比:

(1) Brier Score (Brier, 1950; Cavaola 等, 2024) 是一种均方概率误差, 能够描述预测概率与真实观测概率的偏差程度, 具体定义如下:

$$L_{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2. \quad (10)$$

其中, p 代表预测概率, o 代表真实的标签, N 表示格点数量.

(2) Dice Loss (Li 等, 2019) 是一种用于解决类别不平衡问题的损失函数, 从集合论角度最小化预测与真值的对称差, 具体定义如下:

$$L_{DL} = 1 - \frac{2 \sum_{i=1}^N p_i \times o_i}{\sum_{i=1}^N (p_i + o_i)}. \quad (11)$$

4.2 结果与讨论

4.2.1 实验结果对比

表 8 展示了不同模型未来 6h 闪电预报的累加结果评估. 表格对比了 gSTFNet 模型、欧洲中期天气预报中心发布的 HRES 预报, 以及六种基于深度学习的时空序列预测模型的效果. 表中列出了每种模型在不同 POD 水平下的 FAR、TS 和 ETS 累积评分, 以及 AUC 指标, 其中 $\uparrow/\downarrow$ 符号分别表示该指标值越高/低预报效果越好.

实验结果表明, 七种深度学习方法的预报性能显著优于 HRES, 在 ETS 评分上较 HRES 有 2~3 倍的提升. 这主要由于 HRES 完全依赖于数值天气模式, 其预报精度受初始条件、数值计算误差以及物理参数化的不确定性等多重因素限制. 而深度学习方法通过融合大模型在长期天气趋势预测与闪电观测数据在短期外推的互补优势, 实现了显著的性能提升. 考虑到对 HRES 数据使用线性插值提升分辨率有可能引入一定的误差, 为了保证公平性, 我们补充了一组对比实验: 将 gSTFNet 预测结果的空间分辨率从 $0.04^\circ \times 0.04^\circ$ 插值到 $0.1^\circ \times 0.1^\circ$, 在粗分辨率下与原始 HRES 数据进行对比. 实验结果表明, 虽然降低空间分辨率使得 HRES 与 gSTFNet 的性能差距有所减小, 但原有结论基本成立: gSTFNet 的预报性能依旧显著优于 HRES, 在 ETS 指标上较 HRES 有超过 1 倍的性能提升.

在深度学习模型中, ConvLSTM 和 PredRNN 的表现略逊于其他五种模型. 其原因可能在于这两种方法采用递归时空序列预测架构, 在多步预测过程中, 后一时刻的预报基于前一时刻的输出, 导致累积误差逐

表 8 所有模型的闪电预测实验结果对比^{a)}

预测模型	POD $\approx$ 0.3			POD $\approx$ 0.4			POD $\approx$ 0.5			AUC [†]
	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	
HRES	0.904	0.079	0.057	0.909	0.080	0.057	0.914	0.079	0.055	0.699
ConvLSTM	0.754	0.156	0.141	0.839	0.130	0.110	0.874	0.112	0.090	0.849
PredRNN	0.703	0.176	0.162	0.750	0.181	0.164	0.830	0.145	0.124	0.859
SimVPv2	0.659	0.190	0.178	0.702	0.206	0.190	<u>0.751</u>	<u>0.200</u>	<u>0.181</u>	<u>0.907</u>
TAU	0.650	<u>0.194</u>	<u>0.182</u>	0.717	0.199	0.183	0.778	0.182	0.163	0.891
MetNet-3	<u>0.640</u>	<u>0.194</u>	<u>0.182</u>	<u>0.698</u>	<u>0.207</u>	<u>0.192</u>	0.756	0.195	0.177	0.906
LightNet	0.676	0.184	0.171	0.733	0.190	0.174	0.784	0.178	0.159	0.904
gSTFNet	0.602	0.207	0.195	0.671	0.221	0.207	0.727	0.215	0.197	0.917

a) 表中采用加粗字体表示最优性能, 采用下划线表示次优性能

步增加。相比之下, 其余五种模型的表现较为接近, 但 gSTFNet 的预报效果明显领先于其他模型。具体而言, 在 $\text{POD} \approx 0.3$ 、 0.4 和 0.5 的情况下, gSTFNet 的预报精度比次优模型分别高出 7%、8% 和 9%。这种优势源于 gSTFNet 中的门控时空融合模块, 该模块能够有效地捕捉双源特征的时空关联, 并充分融合双源数据在长期天气趋势预测与短期外推的互补优势, 从而增强了模型的整体预测能力。

表 9 展示了基于不同损失函数训练 gSTFNet 的预报结果对比, 包括 Brier Score (L_{BS})、Dice Loss (L_{DL}) 以及本文使用的加权交叉熵损失 (L_{WCE})。使用加权交叉熵损失函数训练的模型效果显著优于其他两种损失函数, 在 ETS 评分上比使用 Brier Score 训练的模型高出 5~6 倍, 比使用 Dice Loss 训练的模型高出 50% 到 70%。这主要是由于 Brier Score 强调预测概率的准确性, 但闪电数据集中负样本(未发生闪电)的格点数量远大于正样本(发生闪电), 导致模型倾向于预测接近 0 的低概率值以降低整体损失, 影响了模型在闪电事件识别中的有效性。相应地, 虽然 Dice Loss 的设计在一定程度上缓解了正负样本不平衡的问题, 但当正样本数量远少于负样本时, 模型可能会过度关注训练集中的少数正样本, 学习数据中的特定细节而忽略提取更具泛化性的特征, 从而导致过拟合, 影响模型的整体预测性能。相比之下, 加权交叉熵损失通过灵活调整类别权重, 更有效地平衡了正负样本的影响, 使模型取得了较优的性能。

4.2.2 消融实验

为了研究不同数据源对 gSTFNet 模型预报效果的影响, 表 10 展示了分别仅使用大模型预测数据作为输入 (gSTFNet-P) 和仅使用闪电观测数据作为输入 (gSTFNet-L) 的预报累加结果评估, 此外还引入了 HRES 预报结果作为参考基准。

从实验结果来看, gSTFNet-P 的表现明显优于 HRES, 这表明大模型的预测精度能够与当前领先的 NWP 方法相媲美。值得注意的是, gSTFNet-P 在各项指标上的表现均优于 HRES, 这归因于 gSTFNet-P 通过神经网络在广东地区的闪电预报任务中进行了专门训练, 增强了其在该区域的适应性。相比之下, gSTFNet-L 在数值性能指标上优于 gSTFNet-P, 主要因为其输入的历史闪电观测数据与未来闪电短期发展有较强的相关性, 模型通过训练能够有效捕捉和建模这种关联性。通过融合两类数据源, gSTFNet 在整体性能上获得了显著提升, 尤其在 $\text{POD} \approx 0.5$ 的情况下, gSTFNet 的 ETS 指标比 gSTFNet-L 提升了约 11%, 表现出更为优异的综合性能。

为进一步从预报时效角度分析模型的性能表现, 图 6 展示了四个模型在整个预报时长的逐小时对比。从整体趋势来看, gSTFNet-P 与 HRES 的表现相近, 但均不及 gSTFNet-L 和 gSTFNet, 这与表 10 中的数值评估结果一致。就预报时效而言, gSTFNet-L 在短期预测中表现突出, 尤其在早期阶段明显优于 gSTFNet-P 和 HRES, 得益于其输入的历史闪电观测数据与未来短期闪电变

表 9 基于不同损失函数训练gSTFNet的闪电预测实验结果对比^{a)}

损失函数	POD $\approx$ 0.3			POD $\approx$ 0.4			POD $\approx$ 0.5			AUC [†]
	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	
L_{BS}	0.927	0.062	0.039	0.937	0.058	0.033	0.943	0.054	0.028	0.687
L_{DL}	0.780	0.145	0.129	0.816	0.144	0.125	0.847	0.133	0.111	0.849
L_{WCE}	0.602	0.207	0.195	0.671	0.221	0.207	0.727	0.215	0.197	0.917

a) 表中采用加粗字体表示最优性能

表 10 gSTFNet变种模型与HRES的闪电预测实验结果对比^{a)}

预测模型	POD $\approx$ 0.3			POD $\approx$ 0.4			POD $\approx$ 0.5			AUC [†]
	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	FAR [↓]	TS [↑]	ETS [↑]	
HRES	0.904	0.079	0.057	0.909	0.080	0.057	0.915	0.079	0.055	0.699
gSTFNet-P	0.752	0.156	0.141	0.798	0.155	0.137	0.837	0.140	0.119	0.873
gSTFNet-L	0.641	0.197	0.184	0.703	0.207	0.191	0.757	0.195	0.177	0.897
gSTFNet	0.602	0.207	0.195	0.671	0.221	0.207	0.727	0.215	0.197	0.917

a) 表中采用加粗字体表示最优性能

化的紧密关联。然而,随着预报时效的增加,gSTFNet-L的性能开始迅速下降,最终在TS和ETS分数上接近甚至低于gSTFNet-P。通过结合大模型在长期天气趋势预测中的优势以及闪电观测数据在短期预测中的补充作用,gSTFNet在几乎所有时段的表现都优于其他模型,进一步证明了该融合策略的有效性。

综上所述,虽然现有的大模型尚未能直接生成高精度的闪电预报,但通过神经网络对其输出进行适配训练,已展现出一定的预报能力。然而,单独依赖气象大模型难以显著提升闪电预报的准确性,更为有效的策略是将大模型的预测结果与实际观测数据深度融合,充分挖掘两者的互补优势,提高预报性能。

4.2.3 空间分布可视化

本节从空间分布角度对2022年4~9月广东区域闪电发生频次及gSTFNet模型对闪电高发区域的预测能力进行可视化分析。图7a展示了闪电发生次数的空间分布,与图1对比可见,闪电活跃区主要集中在珠三角等地势较低的区域,而北部和东部高海拔地区的闪电活动相对较少。这主要归因于广东省地处东亚季风区,地形呈北高南低、南部沿海的特点,南部地区具备丰富的热量和水汽资源,使雷暴系统在珠三角等低海拔区域更易形成并增强,进而导致闪电活动频繁(徐黄飞

等, 2023)。

图7b展示了gSTFNet在闪电高发区域(闪电发生次数不少于200的格点)的ETS评分表现。整体而言,闪电发生越频繁的区域,gSTFNet的预报性能越突出。特别在珠三角地区,gSTFNet展现出较高的预报技能,大部分区域ETS评分在0.15以上。这表明模型在捕捉和预测高发闪电活动方面具有优势,能够有效刻画未来闪电的发生趋势和空间分布特征。

4.2.4 可视化案例

本节对HRES、gSTFNet-P、gSTFNet-L以及gSTFNet四个模型的四个典型案例进行可视化分析。图8~11分别展示了四个案例的可视化情况,在每张图中,第一行和第二行分别展示了过去三个小时和未来六个小时的闪电真实发生情况。剩余四行从上到下分别对四个模型预测的闪电发生概率进行了可视化。下面是对各个案例的详细分析。

图8展示了一个雷暴单体逐渐发展扩散的过程,且闪电活动频次和范围整体呈现增强趋势。在该案例中,HRES在闪电主要发生的时段和区域上存在一定的偏差。gSTFNet-P虽然能预测出闪电活动区域的扩大趋势,但在前三个小时未能识别出核心闪电区域;相比之下,gSTFNet-L在前四个时刻预测出了闪电发生的

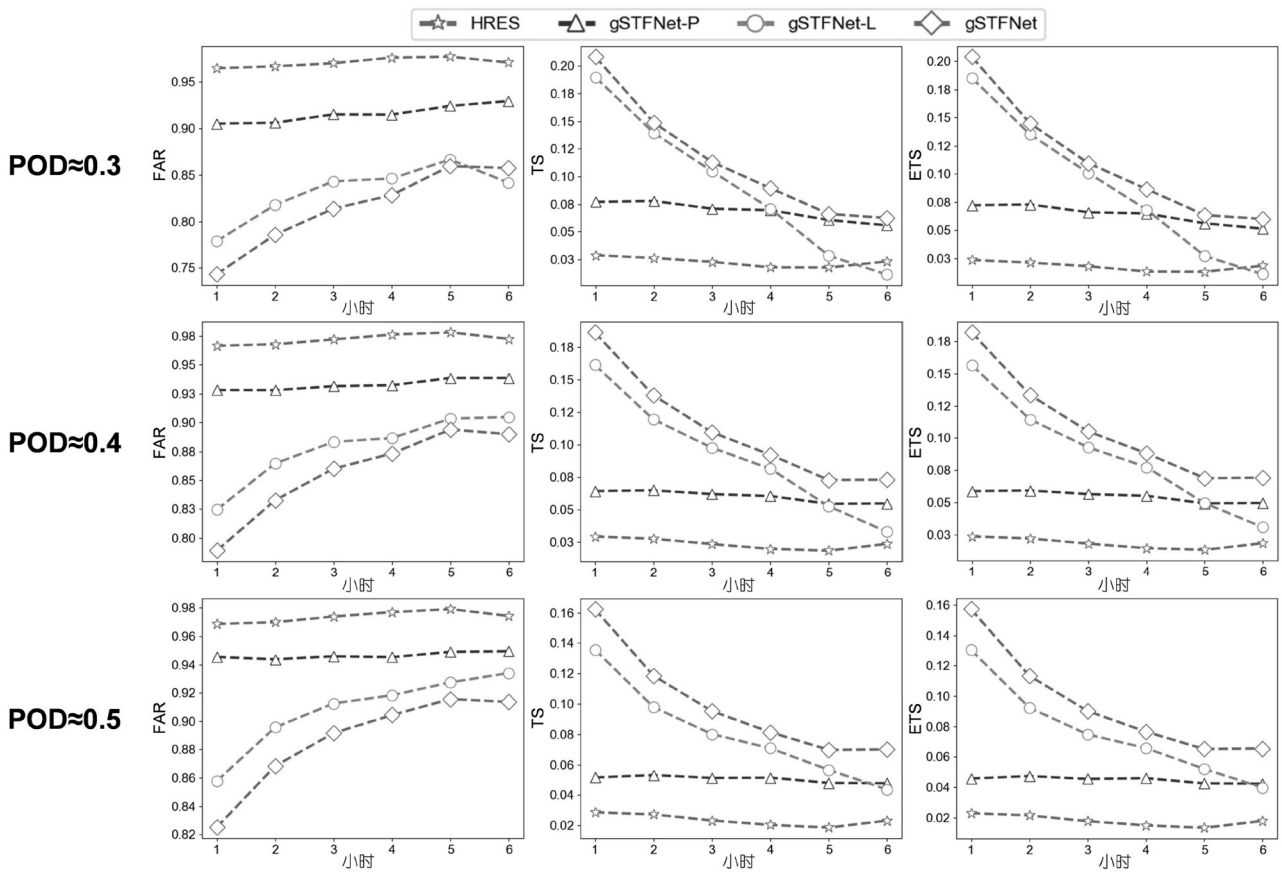


图 6 命中率分别为0.3、0.4、0.5水平下gSTFNet变种模型与HRES逐小时评估曲线

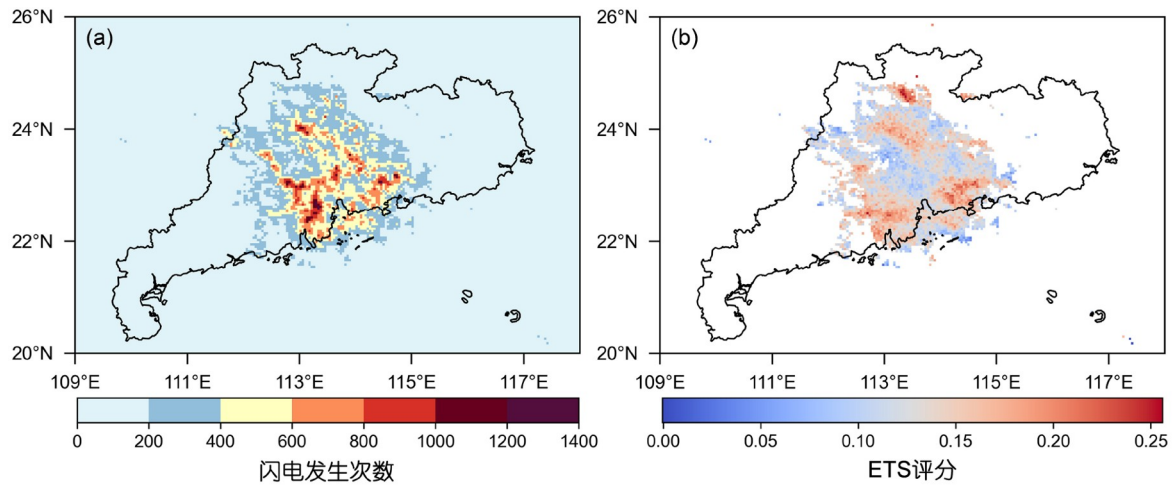


图 7 闪电发生次数空间分布与gSTFNet预测能力评估

(a) 2022年4月至9月闪电发生次数的空间分布; (b) gSTFNet对闪电高发区域的预测能力评估(ETS评分)

核心区域,但在随后的两个小时中仅表现为对这一趋势的简单外推。这些结果揭示出单源数据模型各自的

局限性: 单纯依赖Pangu预测数据的模型容易在空间区域上产生误差,而仅使用闪电观测数据的模型则难

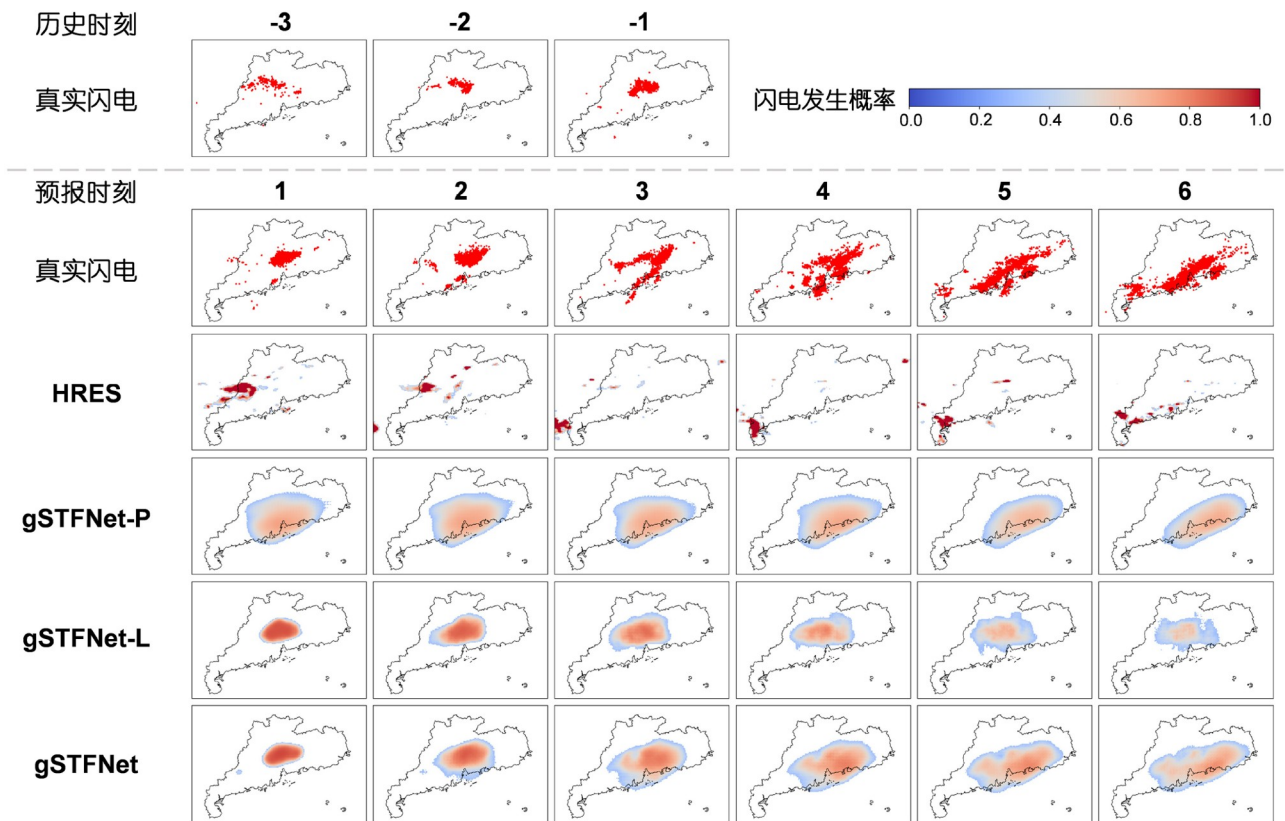


图8 雷暴单体逐渐发展扩散的闪电案例(起报时刻: 2022-06-14 00:00 UTC)可视化效果图

以判断闪电的长期演变趋势。gSTFNet通过利用两者的互补优势取得了更加符合实际观测的预报。

图9展示了一个雷暴逐渐发展并移动的过程案例。与上一案例类似, 基于数值模拟的HRES在预测中仍存在与真实情况的偏差。gSTFNet-P对核心闪电区域的位置预测不准确, 并且倾向于扩大实际闪电区域的扩展范围。gSTFNet-L在前三个时刻对闪电发生区域的预测较为准确, 但在后三个时刻的预测中表现为简单的线性外推。gSTFNet模型在该案例中表现相对较好, 成功捕捉到闪电的运动轨迹, 其预测的闪电区域轮廓与实际情况更为接近。

图10展示了一个雷暴逐渐发展消亡的闪电过程。在此案例中, HRES准确预测了闪电逐渐消失的趋势, 但在闪电发生区域上仍然存在空间误差。gSTFNet-P捕捉到了核心闪电区域逐渐消失的趋势。gSTFNet-L在前两个时刻准确地预测了闪电的发生情况, 但在后几个时刻受到外推趋势的误导, 错误地预测闪电将持

续发生。gSTFNet的整体预报趋势与真实情况较为一致, 但在后三小时中, 其预测结果与实际闪电区域之间存在一定差异。

图11展示了一个发展新生的闪电过程。在该案例中, 一个新的雷暴单体逐渐在广东省东部地区形成。由于过去几个小时内该区域的闪电活动较弱, 这意味着仅从闪电观测数据难以挖掘充分的指导信息。因此, gSTFNet-L对后四个小时的闪电几乎没有预测能力。相比之下, HRES、gSTFNet-P和gSTFNet成功地预测了新生闪电的趋势, 其中gSTFNet模型的预测结果在空间范围上与实际情况最为接近。

通过上述四个案例的分析可以看出, 各模型在不同类型的闪电过程预测中展现出不同的适用性。HRES在捕捉闪电趋势方面有一定优势, 但在闪电发生主要区域预报上表现不佳; gSTFNet-P和gSTFNet-L在预报时效方面各有所长, 而gSTFNet通过双源数据融合能够在大多数情况下取得更加准确和可靠的预报。

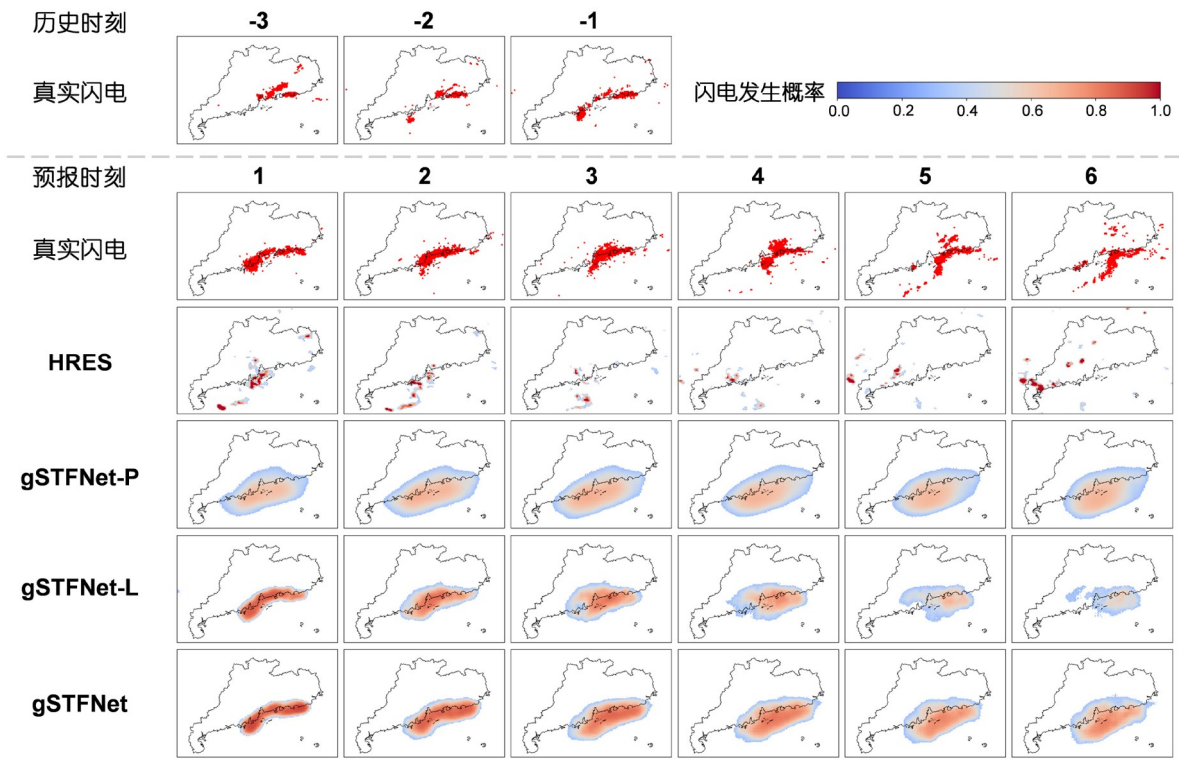


图 9 雷暴逐渐发展移动的闪电案例(起报时刻: 2022-06-06 18:00 UTC)可视化效果图

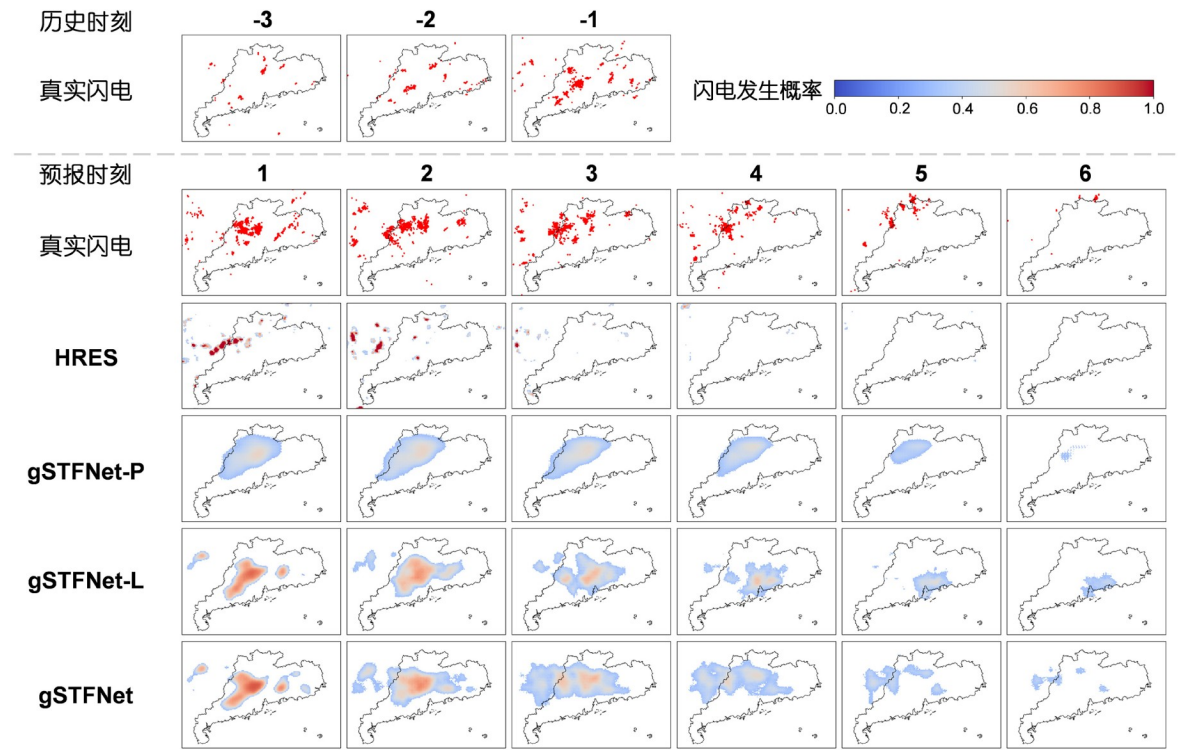


图 10 雷暴逐渐发展消亡的闪电案例(起报时刻: 2022-05-28 6:00 UTC)可视化效果图

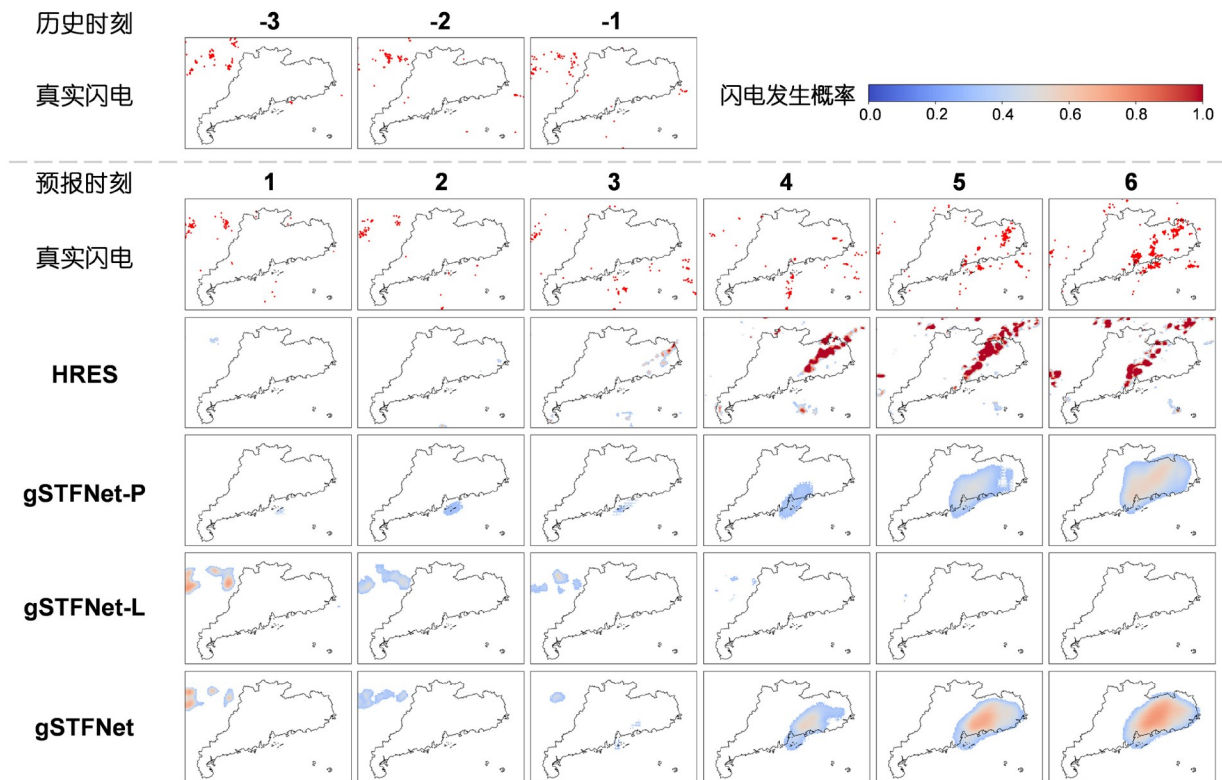


图 11 发展新生的闪电案例(起报时刻: 2022-06-29 00:00 UTC)可视化效果图

5 结论与展望

本研究探讨了基于气象大模型的闪电预报方法, 提出了一种气象大模型输出与历史闪电观测双源数据驱动预报框架, 并设计了双编码门控时空融合神经网络预报模型gSTFNet, 通过广东地区2018~2022年五年的闪电观测数据对该方法进行了验证. 主要结论包括:

(1) 气象大模型的闪电预报能力. 虽然现有气象大模型尚不能直接生成闪电预报, 但通过对其输出进行神经网络的适配训练, 展示出一定的闪电预报能力, 预报精度优于领先的NWP预报 (详见图6中gSTFNet-P与HRES的对比).

(2) 双源数据融合的优势. 单纯依赖气象大模型难以进一步提升闪电预报的准确性, 更为有效的策略是将大模型的预测结果与实际观测数据深度融合, 挖掘二者之间的互补优势, 能够实现预报性能的显著提升 (如表10与图6结果所示). 该双源数据融合策略可进一步扩展至降水、风速等其他天气要素的预报中.

(3) 预报时长对模型表现的影响. 基于气象大模型

的预报模型(gSTFNet-P)在短期内的表现不及基于历史观测数据的预报模型(gSTFNet-L), 但随着预报时长的增加, gSTFNet-P的预报精度更加稳定. 而gSTFNet-L的预报性能则快速衰退, 最终不及gSTFNet-P. 通过有效融合两类数据, gSTFNet在短期和长期的预报精度上均表现出色(详见图6结果).

(4) gSTFNet在双源时空数据融合中的有效性. 本研究中所使用的时空数据具有各自的特性, 在物理意义、时间范围以及与预测任务的相关性上均存在差异. 在针对这些特殊条件下的时空数据融合挑战时, gSTFNet展现出优于当前最先进模型的性能(如表8所示). 这一结果表明了gSTFNet在双源时空数据融合领域中的巨大潜力.

未来研究将从以下几个方面展开:

(1) 更丰富的多源观测数据输入. 实验表明观测数据对提升短期闪电预报精度有显著作用. 未来可引入更多与闪电密切相关的数据源, 如雷达回波和卫星遥感数据, 进一步提高预报精度.

(2) 预报时长的延长. 随着预报时长的增加, 模型

需应对更复杂的非线性天气系统演变。气象大模型在捕捉大气系统长期趋势上的潜力将在这一情境下尤为关键。为此, 将研究更有效的融合技术提升大模型对于长时效预报能力的提升作用。

(3) 更先进的气象大模型知识提取与迁移方法。当前研究主要聚焦于从气象大模型输出中挖掘闪电预报信息, 然而气象大模型内部蕴含的大量复杂气象规律的先验知识尚未得到充分利用。未来可探索更先进的知识提取与迁移技术, 如增量微调和模型蒸馏, 以进一步提升气象大模型在闪电预报中的应用效果。

致谢 感谢欧洲中期天气预报中心提供的ERA5与HRES数据, 本研究中所用数据遵循CC BY 4.0国际许可协议(<https://creativecommons.org/licenses/by/4.0>)。感谢中国气象科学研究院提供的广东地区闪电观测数据, 感谢中国科学院计算机网络信息中心地理空间数据云平台提供的广东省海拔高程数据(<http://www.gscloud.cn>)。感谢编委与两位匿名审稿人提出的建设性意见, 对改善本文有很大的帮助。

参考文献

陈绿文, 吕伟涛, 张义军, 马颖, 齐奇, 武斌. 2020. 粤港澳闪电定位系统对高建筑物雷电的探测. 应用气象学报, 31: 165–174

郭凤霞, 陆干沂, 吴鑫, 王昊亮, 刘祖培, 鲍敏, 李雅雯. 2016. 强雷暴中正地闪发生的条件. 中国科学: 地球科学, 46: 730–742

马明, 吕伟涛, 张义军, 孟青, 杨晶. 2008. 1997–2006年我国雷电灾情特征. 应用气象学报, 19: 393–400

潘伦湘, 鄒秀书, 刘冬霞, 王东方, 杨静. 2010. 西北太平洋地区强台风的闪电活动特征. 中国科学: 地球科学, 40: 252–260

鄒秀书, 刘冬霞, 孙竹玲. 2014. 闪电气象学研究进展. 气象学报, 72: 1054–1068

徐黄飞, 李建勇, 周嘉健, 朱家杉, 邹庆彪, 杜赛, 谢宸浩. 2023. 基于DDW1全闪三维闪电定位系统的闪电活动探测分析. 热带气象学报, 39: 643–652

殷启元, 郭泽勇, 张义军, 庄燕洵, 陈绿文. 2021. 1995–2018年广东地区雷灾伤亡特征. 热带气象学报, 37: 512–520

张焯方, 冯真祯, 刘冰. 2021. 基于卷积神经网络的雷电临近预警模型. 气象, 47: 373–380

张义军, 周秀骥. 2006. 雷电研究的回顾和进展. 应用气象学报, 17: 829–834

张悦, 吕伟涛, 陈绿文, 武斌, 马颖, 齐奇. 2021. 粤港澳大湾区两套闪电定位系统地闪探测性能的对比分析. 热带气象学报, 37: 409–418

张悦, 吕伟涛, 陈绿文, 武斌, 齐奇, 马颖, 张阳, 郑栋, 颜旭, 孟青.

2022. 基于人工引雷的粤港澳闪电定位系统性能评估. 应用气象学报, 33: 329–340

周康辉, 郑永光, 王婷波. 2021. 利用深度学习融合NWP和多源观测数据的闪电落区短时预报方法. 气象学报, 79: 1–14

Amari S. 1993. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5: 185–196

Andrychowicz M, Espeholt L, Li D, Merchant S, Meroze A, Zyda F, Agrawal S, Kalchbrenner N. 2023. Deep learning for day forecasts from sparse observations. arXiv preprint arXiv, 2306.06079, doi: 10.48550/arXiv.2306.06079

Bauer P, Thorpe A, Brunet G. 2015. The quiet revolution of numerical weather prediction. *Nature*, 525: 47–55

Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619: 533–538

Brier G W. 1950. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*, 78: 1–3

Cavaiola M, Cassola F, Sacchetti D, Ferrari F, Mazzino A. 2024. Hybrid AI-enhanced lightning flash prediction in the medium-range forecast horizon. *Nat Commun*, 15: 1188

Charlton-Perez A J, Dacre H F, Driscoll S, Gray S L, Harvey B, Harvey N J, Hunt K M R, Lee R W, Swaminathan R, Vandeale R, Volonté A. 2024. Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán. *NPJ Clim Atmos Sci*, 7: 93

Chen L, Zhong X, Zhang F, Cheng Y, Xu Y, Qi Y, Li H. 2023. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *NPJ Clim Atmos Sci*, 6: 190

Clark A J, Gallus Jr. W A, Weisman M L. 2010. Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Weather Forecast*, 25: 1495–1509

Dafis S, Fierro A, Giannaros T M, Kotroni V, Lagouvardos K, Mansell E. 2018. Performance evaluation of an explicit lightning forecasting system. *J Geophys Res-Atmos*, 123: 5130–5148

Elfving S, Uchibe E, Doya K. 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netws*, 107: 3–11

Fierro A O, Mansell E R, MacGorman D R, Ziegler C L. 2013. The implementation of an explicit charging and discharge lightning scheme within the wrf-arw model: Benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. *Mon Weather Rev*, 141: 2390–2415

Geng Y, Li Q, Lin T, Jiang L, Xu L, Zheng D, Yao W, Lü W, Zhang Y. 2019. Lightnet: A dual spatiotemporal encoder network model for lightning prediction. In: Proceedings of the 25th ACM SIGKDD

- International Conference on Knowledge Discovery & Data Mining. Anchorage. 2439–2447
- Guo S, Yang Y, Liu P. 2024. Assimilation of the deep learning-corrected global forecast system fields into the regional model for improving medium-range persistent precipitation forecasts. *Atmos Res*, 303: 107318
- Hersbach H, Bell B, Berrisford P, Horányi A, Muñoz Sabater J, Nicolas J, Radu R, Schepers D, Simmons A, Soci C, Dee D. 2019. Global reanalysis: Goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, 159: 17–24
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan R J, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J. 2020. The ERA5 global reanalysis. *Q J R Meteorol Soc*, 146: 1999–2049
- Kochkov D, Yuval J, Langmore I, Norgaard P, Smith J, Mooers G, Klöwer M, Lottes J, Rasp S, Düben P, Hatfield S, Battaglia P, Sanchez-Gonzalez A, Willson M, Brenner M P, Hoyer S. 2024. Neural general circulation models for weather and climate. *Nature*, 632: 1060–1066
- Kurth T, Subramanian S, Harrington P, Pathak J, Mardani M, Hall D, Miele A, Kashinath K, Anandkumar A. 2023. FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. In: *Proceedings of the Platform for Advanced Scientific Computing Conference*. Davos. 13: 1–11
- Lam R, Sanchez-Gonzalez A, Willson M, Wirsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W, Merose A, Hoyer S, Holland G, Vinyals O, Stott J, Pritzel A, Mohamed S, Battaglia P. 2023. Learning skillful medium-range global weather forecasting. *Science*, 382: 1416–1421
- Li X, Sun X, Meng Y, Liang J, Wu F, Li J. 2019. Dice loss for data-imbalanced NLP tasks. *arXiv preprint*, doi: 10.18653/v1/2020.acl-main.45
- Lopez P. 2016. A lightning parameterization for the ECMWF integrated forecasting system. *Mon Weather Rev*, 144: 3057–3075
- Lu M, Jin C, Yu M, Zhang Q, Liu H, Huang Z, Dong T. 2024. MCGLN: A multimodal ConvLSTM-GAN framework for lightning nowcasting utilizing multi-source spatiotemporal data. *Atmos Res*, 297: 107093
- Lynn B H, Yair Y, Price C, Kelman G, Clark A J. 2012. Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Weather Forecast*, 27: 1470–1488
- Malardel S, Wedi N, Deconinck W, Diamantakis M, Kühnlein C, Mozdzyński G, Hamrud M, Smolarkiewicz P. 2016. A new grid for the IFS. *ECMWF Newsletter*, 146: 23–28
- McCaul E W, Goodman S J, LaCasse K M, Cecil D J. 2009. Forecasting lightning threat using cloud-resolving model simulations. *Weather Forecast*, 24: 709–729
- Mondal U, Panda S K, Banerjee B K, Kumar A, Sharma D. 2023. Performance evaluation of lightning potential index and flash count using WRF microphysical parameters over Rajasthan and West Bengal, India. *Dyn Atmos Oceans*, 104: 101404
- Schultz M G, Betancourt C, Gong B, Kleinert F, Langguth M, Leufen L H, Mozaffari A, Stadler S. 2021. Can deep learning beat numerical weather prediction? *Phil Trans R Soc A*, 379: 20200097
- Shi X, Chen Z, Wang H, Yeung D Y, Wong W K, Woo W C. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal. 802–810
- Tan C, Gao Z, Wu L, Xu Y, Xia J, Li S, Li S Z. 2023a. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver. 18770–18782
- Tan C, Li S, Gao Z, Guan W, Wang Z, Liu Z, Wu L, Li S Z. 2023b. OpenSTL: A comprehensive benchmark of spatio-temporal predictive learning. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. New Orleans. 69819–69831
- Wang Y, Long M, Wang J, Gao Z, Yu P S. 2017. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach California. 879–888
- Wong J, Barth M C, Noone D. 2013. Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations. *Geosci Model Dev*, 6: 429–443
- Wu Y, He K. 2018. Group normalization. In: *Proceedings of the European Conference on Computer Vision*. Munich. 3–19
- Xu L, Chen S, Yao W, Karacostas T. 2022. Evaluation of Lightning prediction by an electrification and discharge model in long-term forecasting experiments. *Adv Meteorol*, doi: 10.1155/2022/4583030

(责任编辑: 黄建平)