

基于机器学习的四川盆地雷暴大风格点预警*

罗辉^{1,2} 杨康权^{2,3,4} 向筱铭⁵ 苟阿宁^{6,7} 张武龙¹ 王彬雁¹

- 1 四川省气象台, 成都 610072
- 2 高原与盆地暴雨旱涝灾害四川省重点实验室, 成都 610213
- 3 中国气象科学研究院 青藏高原气象研究院, 北京 100081
- 4 中国气象局成都高原气象研究所, 成都 610213
- 5 四川省气象探测数据中心, 成都 610072
- 6 武汉市气象局, 武汉 430074
- 7 武汉中心气象台, 武汉 430074

提 要: 利用 2018—2022 年四川盆地 3 月 1 日至 9 月 30 日雷暴大风历史个例, 结合雷达三维拼图数据和地面极大风观测, 构建了雷暴大风样本集, 并建立了格点大风预警模型。对 2023 年雷暴大风过程进行独立检验, 评估 4 种模型的预警效果。结果表明, LightGBM 模型具有最高的命中率(POD), 在 15 min 预警时效、10 km 评分半径下达 0.536, 但其空报率(FAR)也最高; 随机森林(RF)模型则展现出最佳的综合性能, 其临界成功指数(CSI)在 30 min 时效、10 km 评分半径下最高为 0.306。CSI 和 POD 均随预警时效延长或评分半径减小而显著下降, 时效从 30 min 延长至 45 min 时 CSI 降幅尤为显著。天气背景显著影响预警效果, 明显冷空气影响下, 回波强度、回波顶高、45 dBz 回波顶高等更易出现高值, 有利于对流强烈发展, 但对流前缘新生雷暴易导致漏报增加; 无强冷空气时, 雷暴大风主要出现在对流主体前沿, 命中率较高。垂直积分液态水含量的时间变化量对模型决策贡献度最高, 其次是垂直积分液态水含量密度、回波顶高及组合反射率, 凸显深对流过程是雷暴大风的核心机制, 无冷空气时, 下沉气流对雷暴大风的预警起主导作用。关键特征值样本及高 SHAP 值分析揭示, 对流回波的时间变化变量是预警的关键, 回波追踪风场大值样本多对应正 SHAP 值, 表明回波移速加快时对流性大风发生概率增大。

关键词: 雷暴大风, 机器学习, 回波特征, 预警

中图分类号: P456,P457 DOI: [10.7519/j.issn.1000-0526.2025.091601](https://doi.org/10.7519/j.issn.1000-0526.2025.091601)

Machine Learning-Based Grid-Point Warning for Thunderstorm Gales in the Sichuan Basin

LUO Hui^{1,2} YANG Kangquan^{2,3,4} XIANG Xiaoming⁵ GOU Aning^{6,7} ZHANG Wu long¹
Wang Binyan¹

1 Sichuan Meteorological Observatory, Chengdu 610072

*高原与盆地暴雨旱涝灾害四川省重点实验室科技发展基金(SCQXJZD202102-09、SCQXKJYJXZD202402)、四川省科技计划重点研发项目(2022YFS0542、2024YFFK0408)、中国气象局创新发展专项(CXFZ2024J013、CXFZ2025J014)、湖北省自然科学基金联合基金重点项目(2024AFD205)、四川省气象局重点创新团队(SCQXZDCXTD202401)共同资助
第一作者: 罗辉, 主要从事短临天气预报与产品研发工作。E-mail: 623885263@qq.com
通讯作者: 杨康权, 主要从事天气预报及相关技术研究。E-mail: yangkangquan@aliyun.com

2 Heavy Rain and Drought – Flood Disaster in Plateau and Basin Key Laboratory of Sichuan Province, Chengdu 610213

3 Institute of Tibetan Plateau Meteorology, Chinese Academy of Meteorological Sciences, Beijing 100081

4 Institute of Tibetan Plateau Meteorology, Chengdu 610213

5 Sichuan Meteorological Observation Data Center, Chengdu 610072

6 Wuhan Meteorological Bureau, Wuhan 430074

7 Wuhan Central Meteorological Observatory, Wuhan 430074

Abstract:Based on historical thunderstorm gales events in the Sichuan Basin from March 1 to September 30 between 2018 and 2022, combined with three-dimensional radar mosaic data and surface maximum wind observations, a thunderstorm wind sample dataset was constructed and a grid-based wind warning model was developed. Independent validation was performed on thunderstorm wind events in 2023 to evaluate the warning performance of four models. The main conclusions are as follows:The LightGBM model achieved the highest probability of detection (POD), reaching 0.536 at a 15-minute lead time with a 10 km evaluation radius, but it also exhibited the highest false alarm rate (FAR). The random forest (RF) model demonstrated the best overall performance, with the highest critical success index (CSI) of 0.306 at a 30-minute lead time and 10 km radius. Both CSI and POD decreased significantly with longer warning lead times or smaller evaluation radii, with a particularly notable decline in CSI when the lead time extended from 30 to 45 minutes. Synoptic conditions significantly influenced warning performance. Under pronounced cold air influence, factors such as echo intensity, echo top height, and 45 dBZ echo top height were more likely to exhibit high values, favoring intense convective development. However, newly initiated storms at convective fronts often led to an increase in missed detections. In the absence of strong cold air, thunderstorm winds mainly occurred at the leading edge of convective systems, resulting in higher POD. The temporal variation of vertically integrated liquid water content contributed the most to the model's decision-making, followed by vertically integrated liquid water content density, echo top height, and composite reflectivity, highlighting the central role of deep convection in thunderstorm wind generation. In scenarios without cold air intrusion, downdrafts played a dominant role in thunderstorm wind warnings. Analysis of key feature values and high SHAP values revealed that temporal changes in convective echoes were critical for effective warnings. Samples with high echo-tracking wind speeds often corresponded to positive SHAP values, indicating an increased probability of convective wind events when echo motion accelerates.

Keywords: thunderstorm gales, machine learning, radar echo characteristic, warning

引言

雷暴大风是由强对流风暴引发的极端天气现象，具有突发性强、破坏力大的特点，常导

致严重灾害。四川盆地虽属雷暴大风低频发生区域（费海燕等，2016；龙柯吉等，2020），但历史案例表明，其致灾性极为显著。例如，2022 年 4 月 11 日，一次强飚线过程横扫四川盆地，多地出现 12 级（ $34.7\text{m} \cdot \text{s}^{-1}$ ）以上的极端大风，导致大量建筑物损毁和基础设施破坏（郭云云等，2024）；2015 年“4.4”、2016 年“6.4”、2017 年“7.28”等强对流过程也都造成了重大人员伤亡和财产损失。这些灾害事件充分暴露出当前雷暴大风预警能力仍存在不足，亟需发展更加精准的预警技术。

在雷暴大风的监测预警领域，学者们围绕雷达回波识别预警特征已取得丰硕成果。研究明确，特定的雷达回波形态学特征和动力场信息是预警的关键指标：弓形回波预示着强烈直线风害（俞小鼎等，2012）；阵风锋作为冷空气出流边界，是其前沿雷暴大风的先兆（王福侠等，2016；杨璐等，2018a）；径向速度图上显著的大值区直接指示强风存在。聚焦四川盆地，龙柯吉等（2020）通过十年特征统计，发现回波质心高度骤降及低层强烈风速大值区的出现，可提供超过 10 分钟的预警提前量；罗辉等（2020a）在一次深入分析中，观测到超级单体风暴内罕见的中反气旋特征，其演变超前于地面大风发生，极具预警价值。在识别预警算法方面，廖玉芳等（2006）综合回波形状、环境场、垂直积分液态水含量、风场特征等多因子，建立了雷暴大风预报预警模型；针对更具爆发性的下击暴流，罗辉等（2015）创新性地从能量学角度设计预警指标，显著提升了大风预测的准确性。周康辉等（2017）、李国翠等（2013）及周金莲等（2001）利用模糊逻辑处理气象要素的模糊性和不确定性，成功实现了雷暴大风的自动识别，验证了该技术路径的可行性。然而，传统模糊逻辑方法在特征深度挖掘与复杂非线性关系学习方面存在局限。

机器学习技术的蓬勃发展为雷暴大风智能预警开辟了新途径，其强大的预测能力在多领域展现出显著优势。杨璐等（2018b）率先应用支持向量机（SVM）构建雷暴大风识别与临近预警模型，有效提升了预警准确率。随机森林（RF）最早是由 Breiman（2001）在决策树算法的基础上发展而来，因其优异的性能受到广泛关注：李文娟等（2018）、罗辉等（2020b）、刘新伟等（2021）基于 RF 的强对流预报预警研究均取得实效，凸显其强大的数据挖掘和预测能力。多项对比研究证实了 RF 的优越性：黄衍和查伟雄（2012）发现 RF 在分类问题泛化能力上优于 SVM 和逻辑回归（Logistic）；白琳等（2017）和 Zhang et al（2017）证明 RF 处理非线性问题及揭示变量关系的能力远超传统 Logistic 模型，也优于神经网络、支持向量回归等方法；在灾害预测领域，Hasanuzzaman et al（2022）、Cracknell and Reading（2014）在 RF、朴素贝叶斯等多种模型效果对比中，RF 评分最高。LightGBM 作为高效梯度提升框架，刘新伟等（2021）基于该模型对雷暴大风、短时强降水、冰雹进行识别，效果优异，张国庆和昌宁（2019）在信用风险预测对比中（Logistic、SVM、RF）发现 LightGBM 表现最优。以上研究表明，LightGBM 与 RF 等机器学习算法在处理高维、非线性复杂系统预测问题上具有巨大潜力和普适优势。

尽管机器学习应用广泛，但在强对流天气精细化预警方面的深入实践仍显不足。大力推

进无缝隙精细化预报业务体系建设，研发更高时空分辨率的气象产品是必然趋势。然而，传统雷暴大风预警多以雷暴整体特征为对象，其空间分辨率取决于雷暴的水平尺度，缺少更为精细和精准的警模技术。为提升四川盆地雷暴大风预警精度，利用雷达三维拼图数据，开展基于 LightGBM、RF 等多种机器学习的雷暴大风 1km 格点预警技术研究。

1 资料和方法

1.1 资料与个例

四川盆地雷暴大风多出现在 3—9 月，样本资料选取了 2018—2022 年 3 月 1 日至 9 月 30 日雷达回波三维拼图数据和地面极大风数据（图 1）。雷达资料采用由广元、绵阳、成都、雅安、达州、南充、乐山、宜宾 8 部 S 波段多普勒新一代天气雷达组成的三维拼图数据，水平分辨率为 $0.01^{\circ} \times 0.01^{\circ}$ ，时间分辨率为 6min；地面风采用四川盆地加密自动站极大风资料（包括极大风速和出现时间），实况极大风时间分辨率为 1min，其中加密站包括国家和区域自动站，自动站在盆地分布不均匀，大部分站点的水平间距在 6~8km，部分超过 10km。

文中所用时间均为北京时

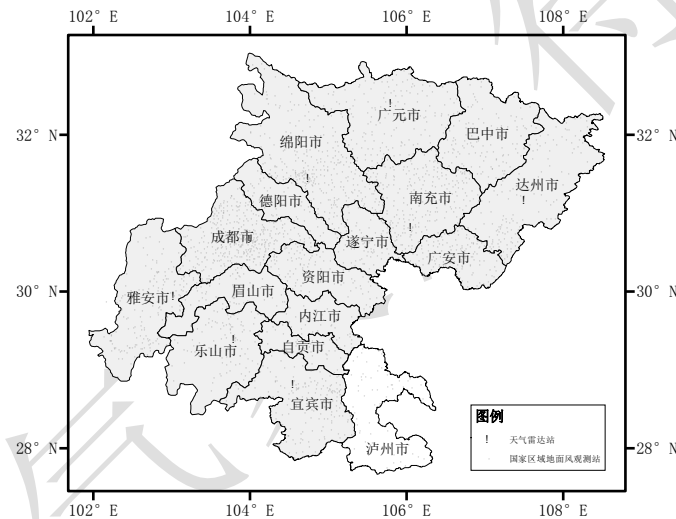


图 1 四川盆地地面观测站和天气雷达站点分布

Fig.1 Distribution of ground observation stations and weather radar sites in Sichuan Basin

四川盆地雷暴大风经常与混合型大风同时出现，大体可分为三种类型：单纯由气压梯度差造成的梯度大风、由热力不稳定造成的雷暴大风、两者都有的混合性大风（龙柯吉等，2020）。盆地大风天气通常在高空系统和冷空气的共同作用下产生，其中混合性大风以层状回波或混合性回波为主，偶尔伴有雷电。然而，这类大风主要是由冷空主导，仅以雷电和大风资料难以准确筛选出雷暴大风。为了能够准确筛选出雷暴大风个例，结合雷达回波，采用主观方式进行个例的挑选。根据四川盆地雷暴大风雷达回波统计结果（龙柯吉等，2020），首先以 1 小时内最大回波强度达到 56dBz 进行初选，再根据 Yang and Sun（2018）的分类方法，主观判断对流组织类型是否属于孤立单体、簇状多单体、线性多单体、非线性对流系统、飢线及弓状回波，确定雷暴大风过程影响时间及区域，形成雷暴大风个例。

1.2 雷暴大风判断

针对雷暴大风个例，首先提取站点实况超过 6 级和 8 级大风发生的时间与位置，结合雷达回波，主观判断大风站否为偏北风（盆地冷空气大风主要为偏北风），剔除由层云或混合性回波覆盖的大风站点，判断大风是否位于强雷暴附近、是否来源于强雷暴区，将剩余大风站点与过去最临近时刻雷达数据进行匹配，确定与大风数据对应的雷达数据时次，记录匹配后的大风和回波信息；最后对于剩余大风站，搜寻站点 5km 范围内是否有超过 40dBz 回波出现，有则计算匹配后雷达回波对应大风站的回波要素值，具体流程见图 2。基于以上大风判别标准，统计近 5 年盆地 8 级大风数量为 249 站次、6 级大风数量为 1433 站次，由于盆地 8 级（ $17.2\text{m} \cdot \text{s}^{-1}$ ）大风多数由冷空气造成的混合性大风，雷暴大风相对较少，结合市县气象台大风的发布标准，确定以 6 级（ $10.8\text{m} \cdot \text{s}^{-1}$ ）雷暴大风为预警目标。

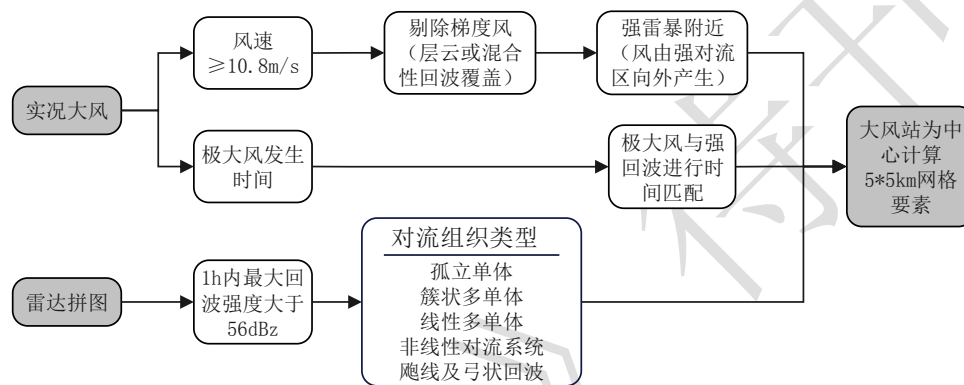


图 2 雷暴大风判断流程

Fig.2 Selection of thunderstorm wind case examples and calculation of sample points

1.3 评分检验方法

雷暴大风预警效果检验采用气象领域常用的评分指标，以临界成功指数（CSI）、命中率（POD）、空报率（FAR）衡量预警事件的准确性，具体计算方法参考刘新伟等（2021）。

以站点雷暴大风对格点预警结果开展检验，为了说明评分半径对评分结果的影响，结合自动站水平间距，分别设置 5、8、10km 半径进行对比检验，为了验证雷暴大风的最佳预警时效，分别使用未来 15、30、45、60min 的实况开展预警效果检验。

2 回波预警特征与统计分析

2.1 回波预警特征

四川盆地雷暴大风预警回波特征包括：雷达回波强度、回波顶高、垂直液态水含量、中层径向辐合、风暴移动速度、回波质心下降、低仰角风速大值区和辐散等（龙柯吉等，2020），回波特征可归纳为回波强度、回波发展高度、特征量垂直累加和移动速度四类，能够从不同角度预警雷暴大风的发生。以这四类回波特征作为雷暴大风的预警因子，引入其时间变化，计算方法如下：

（1）回波强度反应雷暴中水凝物离子的大小与多少，其值越大代表降水离子在下降过

程中产生大风的可能性越大,其随时间变化间接反映了雷暴中垂直气流的变化,雷暴发展阶段上升气流占主导,回波强度增强;消亡阶段下沉气流占主导,回波强度减弱。回波强度及其时间变化可用于雷暴大风识别与预警(廖玉芳等,2006;周康辉等,2017;吴芳芳等,2013)。

回波强度特征包括:组合反射率(CR)、平均组合反射率因子,其中平均组合反射率因子为

5km×5km 网格内的最大反射率因子。

(2) 回波发展高度能够反应雷暴中上升气流的强弱,不同强度回波的发展高度越高,代表不同浓度水凝物离子被抬升的高度越高,在雷暴减弱出现下沉气流时,降水粒子的拖曳作用更强。回波发展高度及其时间变化能够预警雷暴大风的发生(李国翠等,2014;龙柯吉等,2020)。回波发展高度特征包括:18、35、40 和 50dBz 及最大反射率因子的最大发展高度及其平均值(5km×5km 网格平均)

(3) 垂直累加相关特征。垂直积分液态水含量(VIL)反应单位面积内雷暴中垂直柱体液态水的多少,其作为雷暴大风的预警指标已得到学者们的认可(东高红和吴涛,2007;杨璐等,2018a;龙柯吉等,2020),垂直积分液态水含量密度(VILD)为 VIL 和回波高度的比值,反应单位高度内垂直积分液态水含量的多少,二者的时间变化能够提前预警雷暴大风(肖艳姣等,2009;李国翠等,2013)。基于回波强度的下击暴流预警指标(DWC)能够对单体雷暴产生的大风提前预警(罗辉等,2015),该指标作为预警因子之一引入特征量当中,具体计算方法见公式(1)。

$$DWC = \frac{\sum_{i=1}^n 3.44 \times 10^{-6} \times Z_i^{\frac{4}{7}} \times V_i \times g \times h_i}{Area} \quad (1)$$

式中: Z_i 为第 i 个高度层的反射率因子, h 为回波发展高度, V_i 为 Z_i 回波点所代表的体积, g 为重力加速度, h_i 代表 Z_i 回波所处高度, Area 指计算区域面积,本文取 25km²。

(4) 回波移动速度。雷暴大风除下沉气流、动量下传等影响因子外,雷暴本身的快速移动也会使得地面风速增强,风暴移动速度增大时,雷暴大风出现的概率开始增大(伍志方等,2004;周金莲等,2011;杨璐等,2018a)。陈晓欣等(2022)统计发现,我国大范围雷暴大风都是由移动性对流系统导致的。回波移动速度能够反应雷暴移动速度,可在一定程度体现速度场中的大风区,此处引入光流法反演风场作为雷暴大风的预警因子之一。

2.2 回波统计特征

为了分析冷空气对雷暴大风回波特征的影响,按照地面有无明显冷空气从广元或巴中达州一带进入盆地,将 2018—2022 年雷暴大风个例分为有明显冷空气影响和无明显冷空气影响两类,计算两类雷暴大风预警因子,绘制回波移动速度(V)、组合反射率、回波顶高(TOP)、45dBz 回波发展高度(H_{45})、VIL/VILD6 个因子的概率密度曲线(图 3),可见在明显冷空气影响下,CR 超过 50dBz、TOP 超过 14.2km、 H_{45} 超过 6km、VIL 超过 10kg·m⁻²、VILD 超过 1kg·m⁻³ 出现的概率高于无冷空气影响个例,回波移动速度则无明显差异。

从箱线图的分布来看,明显冷空气影响时,CR、 H_{45} 、VIL、VILD 的 50%、75%分位数

和最大值均较高，其中 VIL 高出的幅度最大， H_{45} 次之。对比可见，有明显冷空气参与的雷暴大风的回波更强、发展高度更高，对流发展更加剧烈。

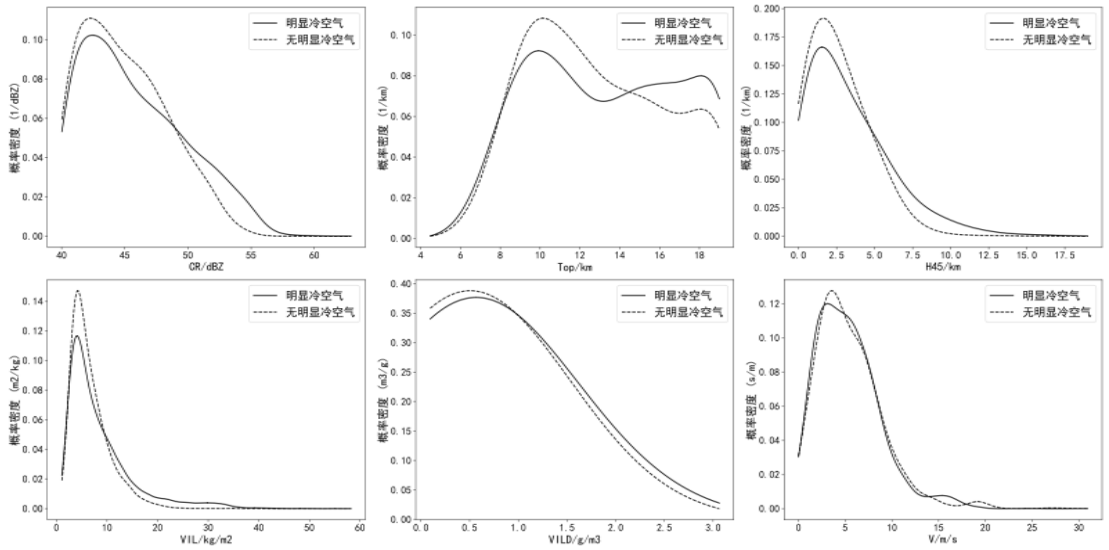


图 3 雷暴大风预警参量概率密度分布

Fig.3 Probability density distribution of thunderstorm gale warning parameters

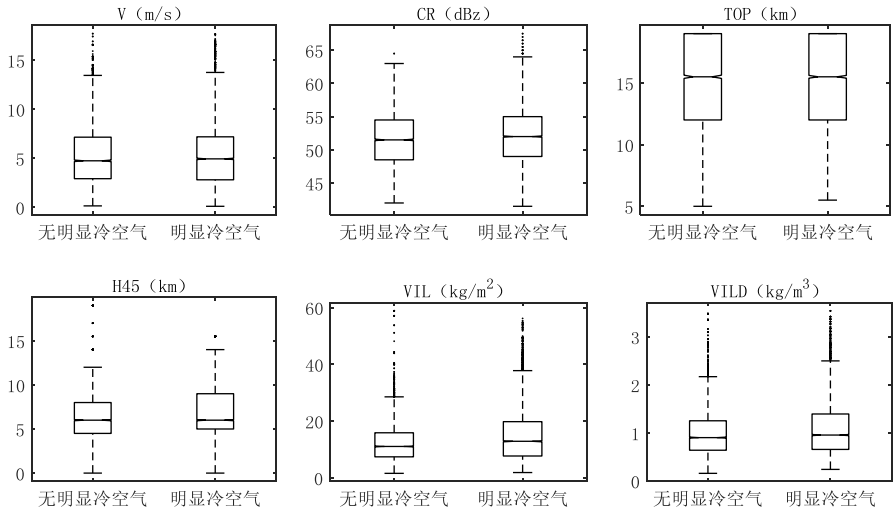


图 4 雷暴大风预警参量箱线图

Fig 4 Box plots of thunderstorm gale warning parameters

3 预警模型与结果检验

3.1 预警模型

随机森林 (RF)、LightGBM、逻辑回归 (LR) 和支持向量机 (SVM) 是机器学习中四种各具特色的算法，适用于不同的场景和需求。随机森林作为一种基于决策树的集成学习方法，通过构建多棵树并综合其结果来提高模型的准确性和鲁棒性，擅长处理高维数据且对

缺失值和异常值不敏感; LightGBM 则是基于梯度提升框架的高效算法, 专为大规模数据设计, 通过直方图算法和单边梯度采样技术显著提升了训练速度和内存效率, 尤其在高维稀疏数据中表现优异; 逻辑回归作为一种经典的线性模型, 以简单高效和良好的可解释性著称, 适合处理二分类问题; 支持向量机则通过间隔最大化和核函数技术能够有效处理非线性问题, 在高维空间中表现优异; 总体而言, 随机森林和 LightGBM 更适合处理复杂的高维数据, 而逻辑回归和 SVM 则在小规模数据集或线性问题中更具优势。

基于以上 4 种算法分别构建大风预警模型的流程 (图 5), 预警特征包括回波强度、回波发展高度、垂直累加相关特征、移动速度四个方面, 共 33 个特征量, 通过时空匹配后计算的特征量与站点大风共同构成数据集, 通过模型训练、检验、调参等步骤构建了雷暴大风格点预警模型。训练使用了自助采样, 训练集是从原始数据中随机抽取, 增加了模型的多样性。数据集共 15976 个, 其中正样本 7815 个, 负样本 8161 个, 为了使正负样本数据均衡, 负样本是从 27246 个非雷暴大风数据中随机取样 30% 得到。

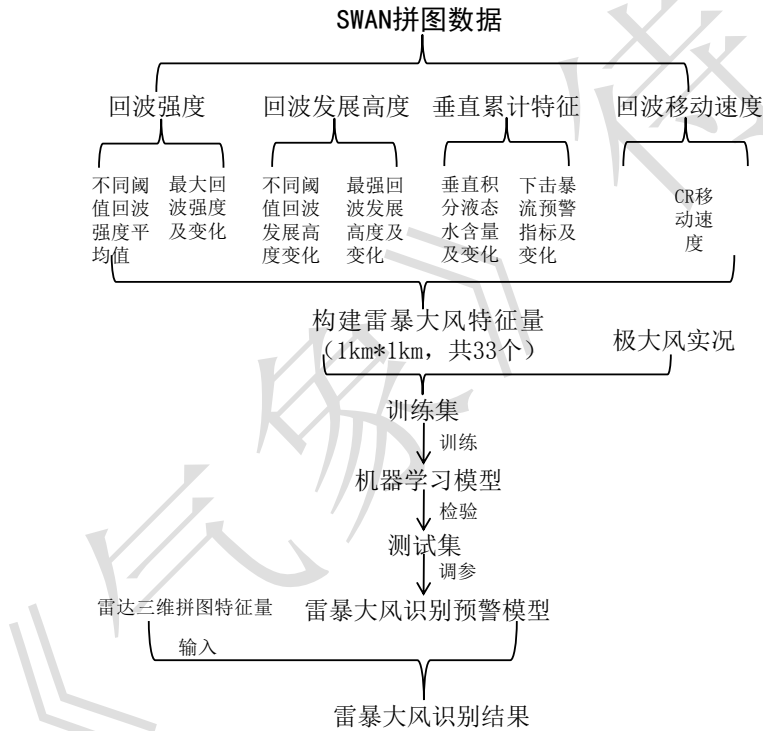


图 5 雷暴大风预警流程
Fig.5 Flowchart of thunderstorm wind identification process

3.2 统计检验结果

为评估雷暴大风预警模型的时效性, 选取 2023 年符合“超过 2 个市出现较集中雷暴大风且持续时间超过 4h”标准的 8 次过程进行检验。检验基于未来 15、30、45、60min 内出现的雷暴大风实况, 并分别设置了 5、8 和 10km 三种评分半径。4 种模型均在 10km 评分半径、15min 预警时效下获得最高 CSI 评分, CSI 评分随着评分半径的减小或预警时效的增加而降低, 预警时效从 30min 延长至 45min 时, CSI 下降尤为显著, 在 8km 和 10km 评分半径下,

15min 与 30min 预警时效的 CSI 评分非常接近。整体上，随机森林（RF）模型的 CSI 评分表现最优，在 10km 半径、30min 时效下达到最高值 0.306，LightGBM 次之。4 种模型同样在 10km 评分半径、15min 预警时效下取得最高 POD 评分，POD 评分也随着评分半径减小或预警时效增加而下降，在不同评分半径和预警时效下，LightGBM 的 POD 评分均最高，RF 次之，在 15min 时效下，LightGBM 的 POD 在 5km、8km、10km 半径下分别高达 0.381、0.479 和 0.536。4 种模型的 FAR 整体维持在 0.5~0.6，在不同评分半径和预警时效下，LightGBM 的 FAR 始终最高（尤其在 15min 时效下达 0.661），RF 次之。模型对比显示，LightGBM 拥有最高的 POD（捕捉能力最强）和最高的 FAR（虚警最多）；随机森林（RF）则拥有最高的 CSI（综合表现最佳），无论是 CSI 还是 POD 评分，均随着预警时效的增加或评分半径的减小而降低。

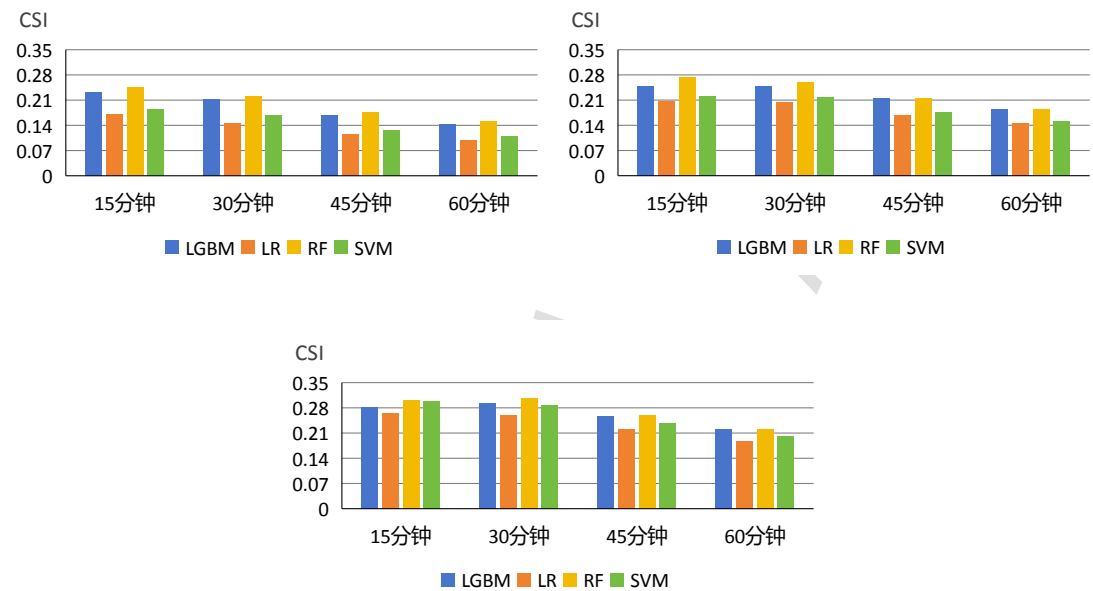
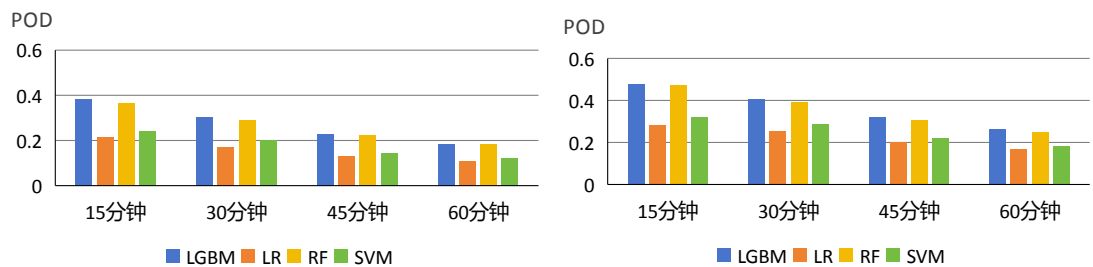


图 6.4 种模型不同评分半径的 CSI 评分
Fig.6 CSI scores of four models at different score radii



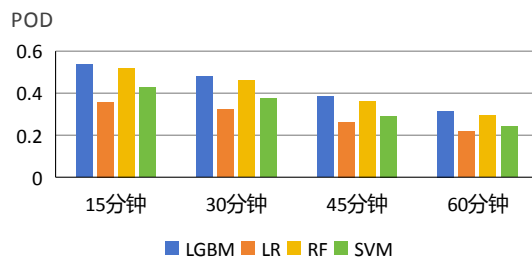


图 7 4 种模型不同评分半径的 POD 评分

Fig.7 POD scores of four models at different score radii

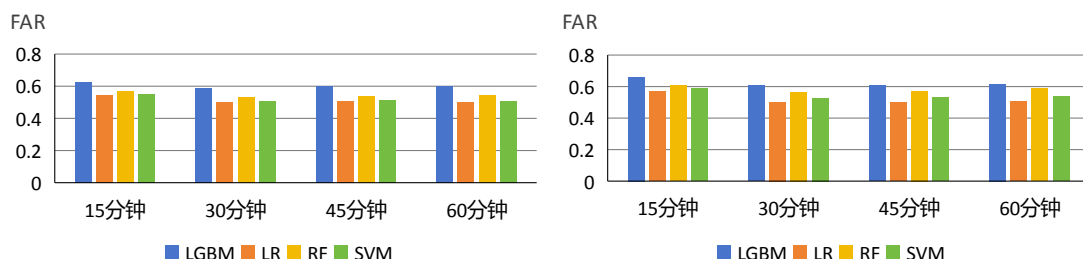


图 8 4 种模型不同评分半径的 FAR 评分

Fig.8 FAR scores of four models at different score radii

3.3 个例预警检验

8 次雷暴大风过程中，低槽（切变）东移型 2 次、副高西侧切变型 4 次、东风扰动型 2 次，其中 3 次伴随明显冷空气、5 次无明显冷空气。选取 2023 年 2 次范围较大的雷暴大风过程开展检验，其中 5 月 5 日有明显冷空气影响，8 月 12 日无明显冷空气影响。

3.2.1 2023 年 8 月 12 日盆地东部雷暴大风个例（无明显冷空气）

2023 年 8 月 11—12 日，受 500hPa 副高外围偏东气流、850hPa 低涡切变及地面弱冷空气共同影响，四川盆地东部出现 6~8 级雷暴大风。过程自 11 日 22: 00 持续至 12 日 13: 00，由南向北影响除盆地西北部和雅安以外的区域。12 日 02: 30，多单体雷暴位于内江、自贡、宜宾、乐山、眉山一带，并向东北方向移动发展；移动前侧回波强度梯度强盛，并伴有新生对流。图 9d 为 02: 30—03: 00 6 级以上大风实况分布，大风主要集中在多单体雷暴前

沿和多个分散的单体雷暴中，图 9a~9c 分别为 LightGBM、逻辑回归和随机森林模型的雷暴大风预警结果，相比而言，LightGBM 预警覆盖范围最广，对应最高的 POD 和最高的 FAR，逻辑回归预警格点最少，POD 和 FAR 最低，随机森林预警范围居中，但综合性能最优，10km 评分半径 CSI 达 0.362，POD 达 0.617。3 种模型均能有效捕捉多单体雷暴前侧及南充、巴中南部的分散雷暴大风，整体预警效果较好。

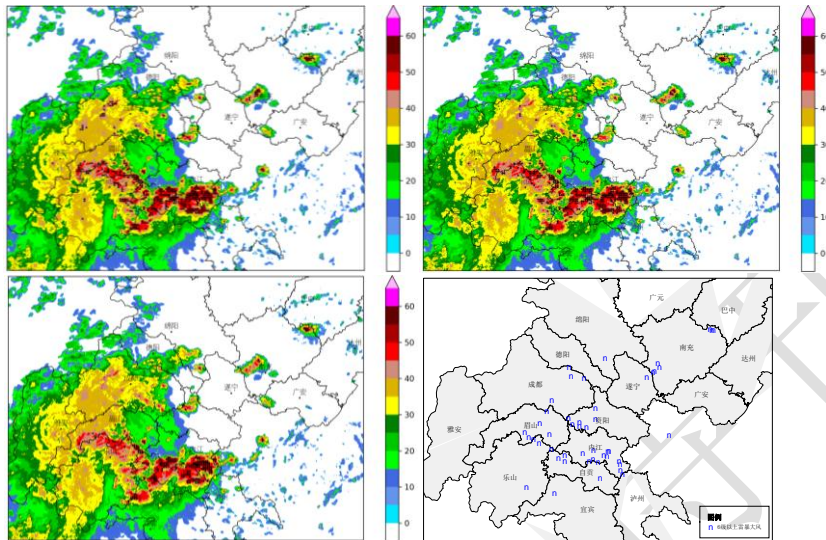


图 9 2023 年 8 月 12 日 (a~c) 02: 30 雷达组合反射率实况 (填色) 和雷暴大风预警结果 (黑色圆点)，
(d) 02: 30—03: 00 实况 6 级以上大风

Fig.9 (a-c)Radar composite reflectivity observation (colored)and thunderstorm wind detection results (black dot) at 02:30 BT, (d) observed severe winds of above scale 6 from 02:30 BT to 03:00 BT 12 August 2023

3.2.2 2023 年 5 月 5 日盆地南部雷暴大风个例 (明显冷空气)

2023 年 5 月 5 日，受 500hPa 高空槽、700hPa 切变线及地面强冷空气共同影响，四川盆地大部出现 6~8 级大风，其中雷暴大风于 5 日 20: 00 至 6 日 01: 00 自西北向东南影响自贡、宜宾、泸州，其余地区以冷空气大风为主。5 日 22: 00，宜宾-泸州北部存在多单体雷暴，其前侧持续有新生单体发展。图 10d 为 22: 00—22: 30 6 级以上大风实况分布，大风主要集中在多单体雷暴前沿和多个分散的单体雷暴中，图 10a~10c 分别为 LightGBM、逻辑回归和随机森林模型的雷暴大风预警结果，相比而言，LightGBM 预警范围最广，POD 与 FAR 均最高，逻辑回归预警格点最少，POD 和 FAR 最低，随机森林预警范围居中，但综合性能最优，10km 评分半径 CSI 达 0.323，POD 达 0.382。3 种模型对多单体雷暴前沿及前侧新生单体的预警效果较好，但风暴前侧因大风远离强回波区导致漏报增多。

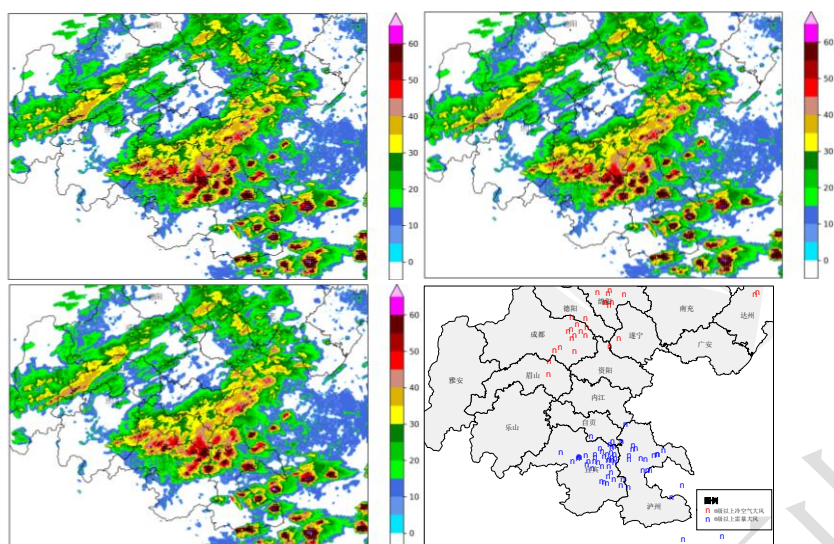


图 10 2023 年 5 月 5 日 (a~c) 22: 00:雷达组合反射率实况 (填色) 和雷暴大风预警结果 (黑色圆点),
(d) 22: 00—22: 30:实况 6 级以上大风

Fig.10 (a—c)Radar composite reflectivity observation (colored)and thunderstorm wind detection results (black dot) at 22:00 BT, (d) observed severe winds of above scale 6 from 22:00 BT to 22:30 BT 5 May 2023

3.4 基于 SHAP 值的模型解释

SHAP 方法 (SHapley Additive exPlanations) 基于合作博弈论中的 Shapley 值框架, 通过量化特征变量的边际贡献增强机器学习模型的可解释性。该方法不仅实现了对影响因素贡献度及交互效应的深度解析, 还能在样本层级精确表征各特征对单次预测的贡献分布。对于每个预测样本, 模型都产生一个预测值, SHAP 值就是该样本中每个特征所分配到的数值 (Lundberg and Lee, 2017; Schoonemann et al, 2024; 周丙锋等, 2025)。图 11 是以散点展示重要性排名前 25 的特征量的 SHAP 值分布图和各特征 SHAP 值绝对平均值重要性柱状图, 即特征值对模型输出的贡献和全局重要性。

预警特征重要性分析表明, VIL 的时间变化量在模型决策中贡献度最高, VILD、回波顶高度及最大反射率因子等特征重要性紧随其后, 这一排序揭示剧烈发展的深对流过程是导致雷暴大风的核心机制。值得注意的是, 无冷空气强迫下, 下击暴流预警因子 (排名第 5、11 位) 与 18 dBz 回波高度 (排名第 3、21 位) 的贡献显著高于冷空气影响, 表明无明显冷空气时, 雷暴下沉气流对地面大风的形成起主导作用。预警特征 SHAP 值分布规律显示, 重要性高的小特征值样本集中于正 SHAP 值区 ($SHAP > 0$), 且以时间变化变量为主, 对模型的预测贡献高, 表明对流回波的动态演变特征是大风预警的关键。部分极端值样本 (如持续高 VILD 或低回波高度) 呈现负贡献 ($SHAP < 0$), 反映特征向量预测的局限性, TREC 风场大值样本普遍集中在正 SHAP 值区, 预测贡献高, 且有无冷空气参与时特征值的作用方向保持一致, 表明当回波移速较快, 出现对流性地面大风的概率开始增大 (李国翠等, 2013)。

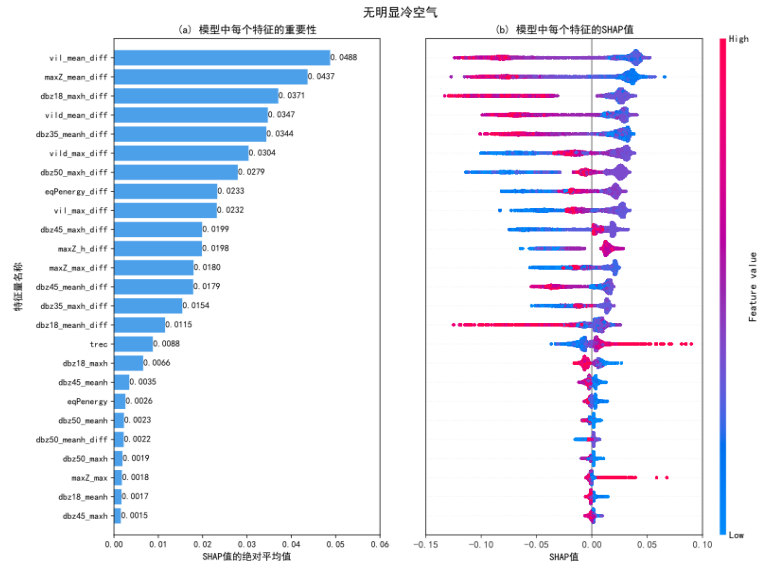


图 11 无明显冷空气影响的雷暴大风预警因子的(a)重要性和 (b) SHAP 值排名

Fig.11 (a) Importance and (b) SHAP value rankings of thunderstorm gale warning factors without strong cold air influence

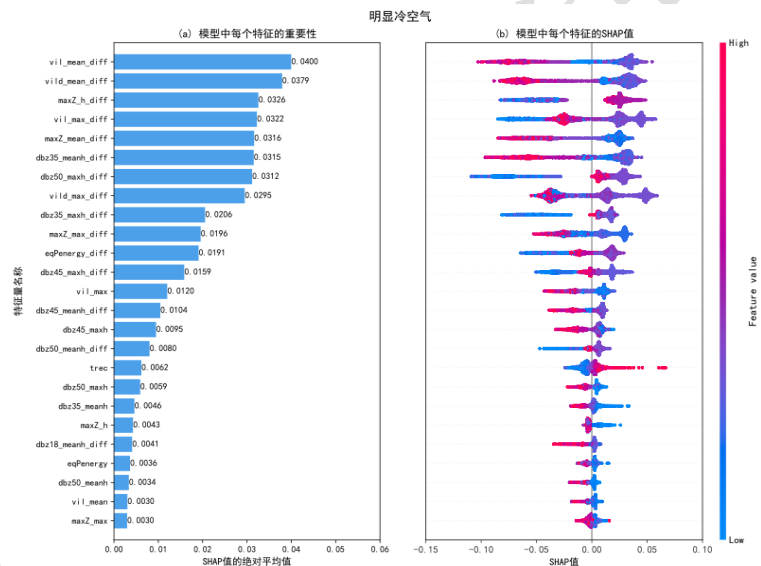


图 12 有明显冷空气影响的雷暴大风预警因子的 (a) 重要性和 (b) SHAP 值排名

Fig.12 (a) Importance and (b) SHAP value rankings of thunderstorm gale warning factors in strong cold air influence

4.结论与讨论

选取 2018—2022 年四川盆地 3 月 1 日至 9 月 30 日雷暴大风个例,使用雷达回波三维拼图和地面极大风数据,构建雷暴大风数据样本集和格点大风预警模型,利用 2023 年雷暴大风天气过程对 4 种模型预警效果进行评估,得到以下结论:

(1) LightGBM 模型的 POD 最高, 15min 预警时效下, 10km 半径达 0.536, 表明捕捉能力最强, 但 FAR 也最高, 达 0.661, 随机森林 (RF) 模型的 CSI 最高, 30min 预警时效下, 10km 半径达 0.306, 综合表现最佳。在不同评分半径和预警时效下, CSI 和 POD 评分

均随预警时效增加或评分半径减小而显著降低, 时效从 30min 延至 45min 时 CSI 下降尤为明显。

(2) 在明显冷空气影响下, 表征预警关键因子(如组合反射率 CR、回波顶高 TOP、45dBZ 回波顶高 H45)更易呈现高值, 表明冷空气的侵入显著促进了对流的发展。个例分析表明, 模型的预警效果和时效会受天气背景影响, 有明显冷空气参与时, 对流主体前侧可能有较多新生雷暴, 导致漏报较多, CSI 和 POD 分别达 0.325 和 0.382; 而无强冷空气时, 雷暴大风主要出现在对流主体前沿部分, CSI 和 POD 分别达到 0.362 和 0.617, 高于有明显冷空气参与时。

(3) 预警特征重要性分析表明, 垂直积分液态水含量 VIL 的时间变化量在模型决策中贡献度最高, 垂直积分液态水含量密度、回波顶高度及最大反射率因子等特征重要性紧随其后, 表明深对流过程是导致雷暴大风的核心机制, 无明显冷空气时, 雷暴下沉气流对地面大风的形成起主导作用。预警特征重要性高的小特征值样本集中于正 SHAP 值区, 且以时间变化变量为主, 对模型的预测贡献高, 表明对流回波的动态演变特征是大风预警的关键, TREC 风场大值样本普遍集中在正 SHAP 值区, 当回波移速较快, 出现对流性地面大风的概率开始增大。

本文的雷暴大风预警模型基于雷达三维拼图数据, 对于已经远离对流主体的大风站点, 存在一定程度的误评价, 可能将一些雷暴大风站点处理为冷空气大风。另外, 基于 TREC 风进行预警因子的空间匹配, 计算的时间变化量存在一定偏差, 对预警效果也会有一定影响。由于训练和检验的样本量有限, 对不同类型的雷暴大风回波特征的理解不够, 预警因子需要进一步优化, 对于 8 级或 10 级以上雷暴大风, 需要后期增加更多雷暴大风过程数据, 扩充样本量, 实现不同强度雷暴大风的预警。

参考文献

- 白琳, 徐永明, 何苗, 等, 2017. 基于随机森林算法的近地表气温遥感反演研究[J]. 地球信息科学学报, 19(3): 390-397. Bai L, Xu Y M, He M, et al, 2017. Remote sensing inversion of near surface air temperature based on random forest[J]. J Geo-inform Sci, 19(3): 390-397 (in Chinese).
- 陈晓欣, 俞小鼎, 王秀明, 2022. 中国大范围雷暴大风事件(Derechos)研究: 时空分布、环境背景和对流系统形态特征[J]. 气象学报, 80(1): 67-81. Chen X X, Yu X D, Wang X M, 2022. Investigation of derechos in China: spatiotemporal distribution, environmental characteristics, and morphology of derechos producing convective systems[J]. Acta Meteor Sin, 80(1): 67-81 (in Chinese).
- 东高红, 吴涛, 2007. 垂直积分液态水含量在地面大风预报中的应用[J]. 气象科技, 35(6): 877-881. Dong G H, Wu T, 2007. Application of vertically integrated liquid (VIL) water in disastrous wind nowcasting[J]. Meteor Sci Technol, 35(6): 877-881 (in Chinese).
- 费海燕, 王秀明, 周小刚, 等, 2016. 中国强雷暴大风的气候特征和环境参数分析[J]. 气象, 42(12): 1513-1521. Fei H Y, Wang X M, Zhou X G, et al, 2016. Climatic characteristics and environmental parameters of severe thunderstorm gales in China[J]. Meteor Mon, 42(12): 1513-1521 (in Chinese).
- 郭云云, 罗辉, 青泉, 等, 2024. 2022 年春季四川盆地一次飑线的环境条件与回波特征[J]. 气象, 50(12): 1480-1494. Guo Y Y, Luo H, Qing Q, et al, 2024. Environmental conditions and radar echo characteristics of a squall line in Sichuan Basin in spring 2022[J]. Meteor Mon, 50(12): 1480-1494 (in Chinese).
- 黄衍, 查伟雄, 2012. 随机森林与支持向量机分类性能比较[J]. 软件, 33(6): 107-110. Huang Y, Zha X W, 2012. Comparison on

classification performance between random forests and support vector machine[J]. *Software*, 33(6): 107-110 (in Chinese).

李国翠, 刘黎平, 张秉祥, 等, 2013. 基于雷达三维组网数据的对流性地面大风自动识别[J]. *气象学报*, 71(6): 1160-1171. Li G C, Liu L P, Zhang B X, et al, 2013. Automatic identification of ground thunderstorm gale based on the radar mosaic 3D data[J]. *Acta Meteor Sin*, 71(6): 1160-1171 (in Chinese).

李国翠, 刘黎平, 连志鸾, 等, 2014. 利用雷达回波三维拼图资料识别雷暴大风统计研究[J]. *气象学报*, 72(1): 168-181. Li G C, Liu L P, Lian Z L, et al, 2014. Statistical study of the identification of thunderstorm gale based on the radar 3D mosaic data[J]. *Acta Meteor Sin*, 72(1): 168-181 (in Chinese).

李文娟, 赵放, 郦敏杰, 等, 2018. 基于数值预报和随机森林算法的强对流天气分类预报技术[J]. *气象*, 44(12): 1555-1564. Li W J, Zhao F, Li M J, et al, 2018. Forecasting and classification of severe convective weather based on numerical forecast and random forest algorithm[J]. *Meteor Mon*, 44(12): 1555-1564 (in Chinese).

廖玉芳, 潘志祥, 郭庆, 2006. 基于单多普勒天气雷达产品的强对流天气预报预警方法[J]. *气象科学*, 26(5): 564-571. Liao Y F, Pan Z Q, Guo Q, 2006. Technique for forecasting and warning of severe convective weather based on single Doppler weather radar products[J]. *Sci Meteor Sinica*, 26(5): 564-571 (in Chinese).

刘新伟, 黄武斌, 蒋盈沙, 等, 2021. 基于 LightGBM 算法的强对流天气分类识别研究[J]. *高原气象*, 40(4): 909-918. Liu X W, Huang W B, Jiang Y S, et al, 2021. Study of the classified identification of the strong convective weathers based on the LightGBM algorithm[J]. *Plateau Meteor*, 40(4): 909-918 (in Chinese).

龙柯吉, 康岚, 罗辉, 等, 2020. 四川盆地雷暴大风雷达回波特征统计分析[J]. *气象*, 46(2): 212-222. Long K J, Kang L, Luo H, et al, 2020. Statistical analysis of radar echo characteristics of thunderstorm gales in Sichuan Basin[J]. *Meteor Mon*, 46(2): 212-222 (in Chinese).

罗辉, 张杰, 朱克云, 等, 2015. 下击暴流的雷达预警量化指标研究[J]. *气象学报*, 73(5): 853-867. Luo H, Zhang J, Zhu K Y, et al, 2015. Study of the radar quantitative index of forewarning downburst[J]. *Acta Meteor Sin*, 73(5): 853-867 (in Chinese).

罗辉, 苟阿宁, 康岚, 等, 2020a. 四川盆地一次中反气旋超级单体的雷达回波特征研究[J]. *气象*, 46(10): 1362-1374. Luo H, Gou A N, Kang L, et al, 2020a. Radar echo characteristics of an meso-anticyclonic supercell of Sichuan in August 2016[J]. *Meteor Mon*, 46(10): 1362-1374 (in Chinese).

罗辉, 肖递祥, 匡秋明, 等, 2020b. 四川盆地暖区暴雨的雷达回波特征及分类识别[J]. *应用气象学报*, 31(4): 460-470. Luo H, Xiao D X, Kuang Q M, et al, 2020b. Radar echo characteristics and recognition of warm-sector torrential rain in Sichuan Basin[J]. *J Appl Meteor Sci*, 31(4): 460-470 (in Chinese).

王福侠, 俞小鼎, 裴宇杰, 等, 2016. 河北省雷暴大风的雷达回波特征及预报关键点[J]. *应用气象学报*, 27(3): 342-351. Wang F X, Yu X D, Pei Y J, et al, 2016. Radar echo characteristics of thunderstorm gales and forecast key points in Hebei Province[J]. *J Appl Meteor Sci*, 27(3): 342-351 (in Chinese).

吴芳芳, 俞小鼎, 张志刚, 等, 2013. 苏北地区超级单体风暴环境条件与雷达回波特征[J]. *气象学报*, 71(2): 209-227. Wu F F, Yu X D, Zhang Z G, et al, 2013. A study of the environmental conditions and radar echo characteristics of the supercell-storms in northern Jiangsu[J]. *Acta Meteor Sin*, 71(2): 209-227 (in Chinese).

伍志方, 叶爱芬, 胡胜, 等, 2004. 中小尺度天气系统的多普勒统计特征[J]. *热带气象学报*, 20(4): 391-400. Wu Z F, Ye A F, Hu S, et al, 2004. The statistic characteristics of mesoscale and microscale systems with the new generation weather radar[J]. *J Trop Meteor*, 20(4): 391-400 (in Chinese).

肖艳姣, 马中元, 李中华, 2009. 改进的雷达回波顶高、垂直积分液态水含量及其密度算法[J]. *暴雨灾害*, 28(3): 210-214. Xiao Y J, Ma Z Y, Li Z H, 2009. Advanced algorithms of ET, VIL and VIL density for the CINRAD[J]. *Torrential Rain Dis*, 28(3): 210-214 (in Chinese).

杨璐, 陈明轩, 孟金平, 等, 2018a. 北京地区雷暴大风不同生命期内的雷达统计特征及预警提前量分析[J]. *气象*, 44(6): 802-813. Yang L, Chen M X, Meng J P, et al, 2018a. Radar statistical characteristics and warning lead analysis of thunderstorm gales in different life periods in Beijing[J]. *Meteor Mon*, 44(6): 802-813 (in Chinese).

杨璐, 韩丰, 陈明轩, 等, 2018b. 基于支持向量机的雷暴大风识别方法[J]. *应用气象学报*, 29(6): 680-689. Yang L, Han F, Chen M X, et al, 2018b. Thunderstorm gale identification method based on support vector machine[J]. *J Appl Meteor Sci*, 29(6): 680-689 (in

Chinese).

俞小鼎, 周小刚, 王秀明, 2012. 雷暴与强对流临近天气预报技术进展[J]. 气象学报, 70(3): 311-337. Yu X D, Zhou X G, Wang X M, 2012. The advances in the nowcasting techniques on thunderstorms and severe convection[J]. Acta Meteor Sin, 70(3): 311-337 (in Chinese).

张国庆, 昌宁, 2019. 基于 LightGBM 的银行信用卡违约研究[J]. 科技资讯, 17(12): 8-9. Zhang G Q, Chang N, 2019. Study on bank credit card default based on LightGBM[J]. Science & Technology Information, 17(12), 8-9.

周丙锋, 史静, 谢新水, 等, 2025. 大运河文化带新质生产力的时空分异特征与影响因素——基于机器学习 XGBoost-SHAP 模型[J]. 地域研究与开发, 44(1): 14-22. Zhou B F, Shi J, Xie X S, et al. 2025. Exploration of characteristics and influencing factors of spatial and temporal differentiation of neoplastic productivity in the grand canal cultural belt: based on the machine learning XGBoost-SHAP model[J]. Areal Res Dev, 44(1): 14-22 (in Chinese).

周金莲, 魏鸣, 吴涛, 等, 2001. 对流性大风天气的多普勒雷达资料识别方法研究[C]//第 28 届中国气象学会年会——S3 天气预报灾害天气研究与预报. 厦门: 中国气象学会. Zhou J L, Wei M, Wu T, et al, 2001. Research on Doppler radar data identification method for convective wind weather[C]//Proceedings of the 28th Chinese Meteorological Society Annual Meeting. Xiamen: Chinese Meteorological Society (in Chinese).

周康辉, 郑永光, 王婷波, 等, 2017. 基于模糊逻辑的雷暴大风和非雷暴大风区分方法[J]. 气象, 43(7): 781-791. Zhou K H, Zheng Y G, Wang T B, et al, 2017. Fuzzy logic algorithm of thunderstorm gale identification using multisource data[J]. Meteor Mon, 43(7): 781-791 (in Chinese).

Breiman L, 2001. Random forests[J]. Mach Learn, 45(1): 5-32.

Cracknell M J, Reading A M, 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information[J]. Comput Geosci, 63: 22-33.

Hasanuzzaman M, Islam A, Bera B, et al, 2022. A comparison of performance measures of three machine learning algorithms for flood susceptibility mapping of river Silabati (tropical river, India)[J]. Physics Chem Earth, Parts A/B/C, 127: 103198.

Lundberg S M, Lee S I, 2017. A unified approach to interpreting model predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc: 4768-4777.

Schoonemann J, Nagelkerke J, Seuntjens T G, et al, 2024. Applying XGBoost and SHAP to open source data to identify key drivers and predict likelihood of wolf pair presence[J]. Environ Manage, 73(5): 1072-1087.

Yang X L, Sun J H, 2018. Organizational modes of severe wind-producing convective systems over North China[J]. Adv Atmos Sci, 35(5): 540-549.

Zhang H, Wu P B, Yin A J, et al, 2017. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: a comparison of multiple linear regressions and the random forest model[J]. Sci Total Environ, 592: 704-713.