

山东烟台市短时强降水特征分析

秦 璐^{1,2}, 栾东红^{1,3*}, 孟文童^{1,3}

¹山东省气象防灾减灾重点实验室, 山东 济南

²烟台市牟平区气象局, 山东 烟台

³烟台市气象局, 山东 烟台

收稿日期: 2024年12月7日; 录用日期: 2025年1月8日; 发布日期: 2025年1月17日

摘要

本文利用山东省烟台市2019年1月~2023年12月降水实况资料, 对烟台市的短时强降水发生频次的空间分布、日变化、月变化以及年变化特征进行分析, 并尝试利用CatBoost机器学习算法对短时强降水进行预测, 结论如下: 烟台市短时强降水发生频次及最强降水主要集中在西北以及东南部沿海地区, 其中西北部地区的短时强降水, 从发生频次以及最大强降水来说均较大, 在烟台中部、北部及东部短时强降水发生频次及强度均较小; 烟台市短时强降水总频次日变化、月变化及年变化特征均较为显著, 日变化方面, 傍晚至凌晨短时强降水发生频次最多, 凌晨及午后为发生频次第二多的时段, 而日出前后短时强降水发生频次较低; 其月变化波动幅度较大, 呈先增大后减少的变化特征, 短时强降水较为集中在6~8月; 其年变化波动幅度也较大, 也呈先增大后减少的变化趋势, 在2022年达到最多为499次, 之后逐渐减少, 2019年最少为72次, 短时强降水发生频次较多的时段集中在2021~2023年。用CatBoost模型对烟台市强降水进行初步预测, 模型基本能够很好地捕捉烟台市短时强降水特征, 在测试集上表现出较好的预测能力, 这表明该模型具有很好的泛化能力。

关键词

烟台市, 短时强降水, 机器学习, 相关性, 时空分布

Analysis on Characteristics of Short-Time Heavy Rainfall in Yantai City, Shandong Province

Lu Qin^{1,2}, Donghong Luan^{1,3*}, Wentong Meng^{1,3}

¹Shandong Provincial Key Laboratory of Meteorological Disaster Prevention and Mitigation, Jinan Shandong

²Mouping District Meteorological Bureau of Yantai City, Yantai Shandong

³Yantai Meteorological Bureau, Yantai Shandong

*通讯作者。

Received: Dec. 7th, 2024; accepted: Jan. 8th, 2025; published: Jan. 17th, 2025

Abstract

Based on the actual precipitation data of Yantai City, Shandong Province from January 2019 to December 2023, this paper analyzes the spatial distribution, daily variation, monthly variation and annual variation characteristics of the frequency of short-term heavy precipitation in Yantai City, and tries to use CatBoost machine learning algorithm to predict short-term heavy precipitation. The conclusions are as follows: The frequency of short-term heavy precipitation and the strongest precipitation in Yantai City are mainly concentrated in the northwest and southeast coastal areas. The short-term heavy precipitation in the northwest is larger in terms of frequency and maximum heavy precipitation. The frequency and intensity of short-term heavy precipitation in the central, northern and eastern parts of Yantai are small. The daily variation, monthly variation and annual variation characteristics of the total frequency of short-term heavy precipitation in Yantai City are significant. In terms of daily variation, the frequency of short-term heavy precipitation from the evening to the early morning is the most frequent, and the early morning and afternoon are the second most frequent periods, while the frequency of short-term heavy precipitation before and after sunrise is low. The fluctuation range of monthly variation is large, which increases first and then decreases. The short-term heavy precipitation is more concentrated in June-August. The fluctuation range of its annual change is also large, and it also shows a trend of increasing first and then decreasing. It reaches a maximum of 499 times in 2022, and then gradually decreases. In 2019, it is at least 72 times. The period with more frequency of short-term heavy precipitation is concentrated in 2021~2023. The CatBoost model is used to predict the heavy precipitation in Yantai City. The model can basically capture the characteristics of short-term heavy precipitation in Yantai City, and shows good prediction ability on the test set, which shows that the model has good generalization ability.

Keywords

Yantai, Short-Term Heavy Rainfall, Machine Learning, Correlation, Spatial and Temporal Distribution

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

短时强降水是指在较短时间内出现大量降水的天气现象，通常定义为小时降水量 ≥ 20 毫米的降水事件[1]。短时强降水因其突发性强、强度大，常引发城市内涝、山洪、泥石流等灾害，对人类生命财产安全和经济社会发展构成严重威胁。因此，对短时强降水的研究具有重要的实际意义和科学价值，对于提高气象预报水平、优化城市规划、保障农业生产、管理水资源、应对气候变化以及推动大气科学的发展都具有重要意义。

目前，对于短时强降水的研究进展主要集中在以下几个方面：首先，是地形对短时强降水的影响。已有的研究表明[2]，地形通过调节风的汇聚扰动和对流，导致沿海低洼地区的短时强降水热点。在中国南部沿海地区，短时强降水主要发生在沿海低洼地区而非内陆迎风坡地，这些地区的地形特征对短时强降水的形成起到了重要作用[3][4]。其次，需要考虑气候变化对极端降水的影响：随着全球变暖，大气中

的水汽含量增加, 预期极端降水将加剧[5]。然而, 大气动力学的变化, 如热带哈德来环流的向极地扩张, 可能在区域上减弱或加强这种热力学效应[6]。水资源的可用性在温度-湿度关系中起着重要作用, 不同地区极端降水的变化并不均匀。同时, 不少学者也考虑到人类活动对短时强降水的影响[7]-[9]。人类活动, 特别是气溶胶的排放, 可能会加剧短时强降水。通过使用高分辨率的气候模型, 研究者们发现, 即使在气候变暖的背景下, 气溶胶的增加也可能导致短时强降水的增加。这些进展为我们理解和预测短时强降水提供了重要的科学基础, 并有助于制定应对极端天气事件的策略。

烟台市位于山东半岛东部, 北临渤海, 南临黄海, 西与潍坊市、青岛市接壤, 东与威海市毗邻。了解烟台市短时强降水的分布特征, 并在此基础上初步研究对其进行预警和预测, 对于理解海陆交界地带的气候特征, 服务于当地经济发展建设有重要意义。

2. 资料与方法

2.1. 研究区域概况

山东省烟台市位于我国山东半岛东北部, 与辽东半岛对峙, 地形为低山丘陵地貌为主, 中小河流众多, 烟台市属温带大陆性季风气候, 与同纬度内陆地区相比具有雨水适中、空气湿润、气候温和的特点。本文利用山东省烟台市 2019 年 1 月~2023 年 12 月降水实况资料, 对该区域的短时强降水特征进行分析, 以期进一步提高短时强降水的预报准确率, 在防灾减灾方面, 具有重要意义。

2.2. 资料与方法

2.2.1. 所用资料

- (1) 本文所用山东省烟台市观测数据来自烟台市 143 个气象观测站(国家基准站、基本站、气象观测业务考核区域自动站)。
- (2) 本文用于 CatBoost 模型进行短时强降水预测的物理量数据来自欧洲中期天气预报中心(ECMWF, European Centre for Medium-Range Weather Forecasts)提供的全球气候第五代大气再分析数据集(ERA5), 该数据集分辨率为 $0.25^\circ \times 0.25^\circ$, 可以提供大量大气、陆地和海洋气候变量的每小时估计值。本文用到的变量为 500 hPa 和 850 hPa 比湿、涡度、散度和位势高度以及地面露点温度和平均海平面气压。

2.2.2. 所用方法

利用烟台市短时强降水预报因子, 基于 CatBoost 模型尝试建立短时强降水预警预测模型, 下面介绍 CatBoost 模型的设计方案。

CatBoost 模型是一种基于梯度提升决策树(Gradient Boosting on Decision Trees)的算法, 由俄罗斯的搜索巨头 Yandex 开源。它特别适合处理类别型特征, 并且提供了很好的默认参数, 使得用户在不进行复杂参数调优的情况下也能获得不错的结果。CatBoost 的主要特点包括:

- (1) 对称树(Oblivious Trees): CatBoost 构建对称树, 这意味着在每一层中, 所有叶子节点都使用相同的特征和分割条件进行拆分, 这有助于提高模型的泛化能力并减少过拟合。
- (2) 类别型特征支持: CatBoost 原生支持类别型特征, 无需用户进行额外的预处理, 如独热编码(One-Hot Encoding), 算法会自动处理这些特征。
- (3) 梯度提升算法: CatBoost 使用了一种新颖的梯度提升方案, 称为 Ordered Boosting, 它通过在不同的数据子集上计算梯度和更新模型来减少过拟合。

CatBoost 的网络架构主要基于对称决策树, 它使用一种平衡的树结构, 有助于高效的 CPU 实现, 并减少预测时间。在对称决策树中, 只使用一个特征来构建每个树级别上的所有分支。CatBoost 还提供了多种类型的拆分, 包括“FloatFeature”、“OneHotFeature”和“OnlineCtr”。

总的来说, CatBoost 是一个高性能的梯度提升库, 它通过对称树结构、类别型特征的原生支持、Ordered Boosting 梯度提升方案以及 GPU 加速等特性, 为用户提供了一个强大且易于使用的工具, 适用于各种机器学习任务。

3. 烟台市短时强降水分布特征

3.1. 烟台市短时强降水空间分布

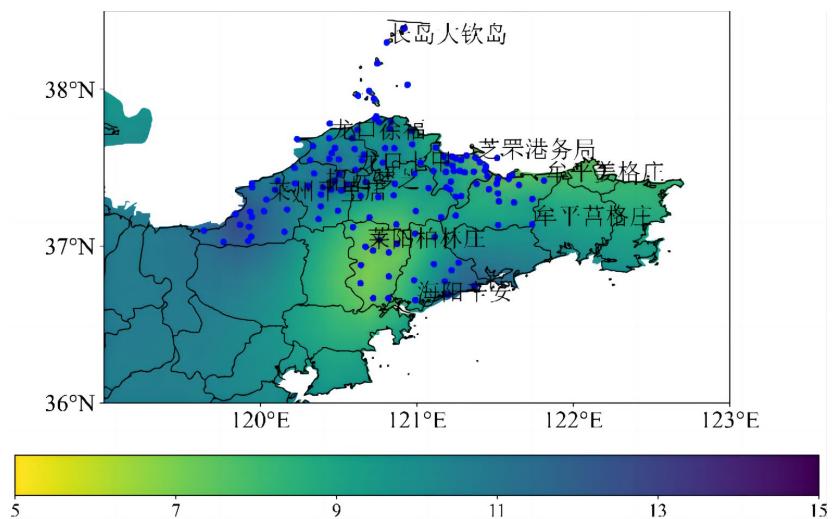


Figure 1. The frequency distribution of short-term heavy precipitation at each station in Yantai City, Shandong Province from January 2019 to December 2023

图 1. 2019 年 1 月~2023 年 12 月山东烟台市各站点短时强降水发生频次分布

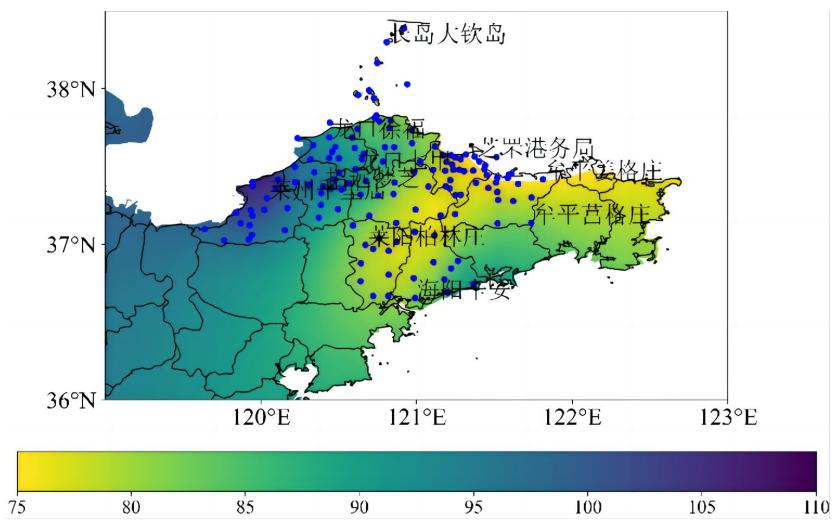


Figure 2. The distribution of maximum heavy precipitation at each station in Yantai City, Shandong Province from January 2019 to December 2023 (unit: mm)

图 2. 2019 年 1 月~2023 年 12 月山东烟台市各站点最大强降水分布(单位: mm)

2019 年 1 月~2023 年 12 月山东烟台市各站点短时强降水发生频次空间分布如图 1 所示, 烟台市短时强降水主要集中在烟台西北以及东南部的沿海地区, 部分站点超过 13 次, 其中莱州平里店及海阳辛安最大, 烟台中部以及东部地区短时强降水发生频次较少, 其中烟台中部莱阳柏辛庄附近最少, 小于 7 次。

这表明,在烟台西北部以及东南部地区短时强降水发生频次较高,在烟台中部短时强降水发生频次较低。

2019年1月~2023年12月烟台市各站点最大强降水分布如图2所示,烟台市最大短时强降水主要集中在西北部以及东南部沿海地区,其中莱州平里店最大,其中心最大值超过110 mm,且逐渐向中部地区递减,烟台中部、北部及东部地区最大强降水较小,其中莱阳柏林庄、芝罘港务局及牟平姜格庄最小,其中心最小值小于75 mm,这也与前文分析的烟台市短时强降水发生频次的分布较为一致,即2019年1月~2023年12月烟台市短时强降水主要集中在西北以及东南部沿海地区,其中西北部地区的短时强降水,从发生频次以及最大强降水来说均较大。

由烟台市短时强降水发生频次以及最大强降水分布可知,其短时强降水主要分布在烟台西北及东南沿海地区,烟台中部、北部及东部短时强降水发生频次及强度均较小。

3.2. 烟台市短时强降水时间分布

由图3可知,烟台市短时强降水总频次的日变化呈波动式变化,0时最多,为160次,6时最少,为13次,其中,夜间21~24时为短时强降水发生频次最多时段,均超过80次,表明烟台市短时强降水主要集中在夜间;0~2时、8~9时及午后13~17时为短时强降水发生的次多时段,其频次均超过50次;5~6时短时强降水发生频次小于15次,表明在该时段内短时强降水发生频次较低。由此可知,烟台市短时强降水总频次日变化特征显著,傍晚至凌晨短时强降水发生频次最多,凌晨及午后为发生频次第二多的时段,而日出前后短时强降水发生频次较低。

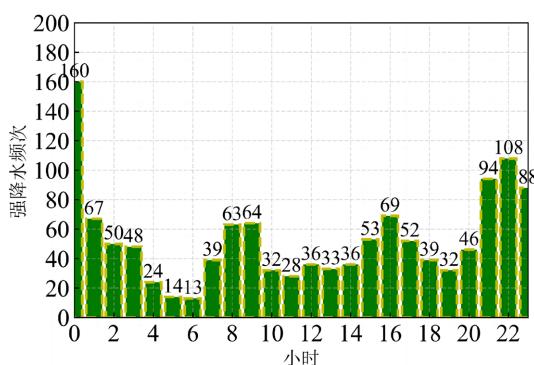


Figure 3. Daily variation of total frequency of short-time heavy precipitation in Yantai City, Shandong Province from January 2019 to December 2023

图3. 2019年1月~2023年12月山东烟台市短时强降水总频次日变化

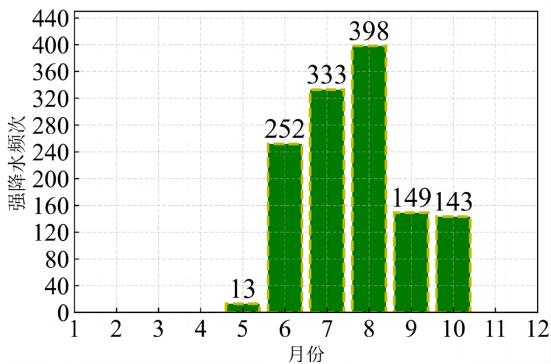


Figure 4. Monthly variation of total frequency of short-time heavy rainfall in Yantai City, Shandong Province from January 2019 to December 2023

图4. 2019年1月~2023年12月山东烟台市短时强降水总频次月变化

进一步分析 2019 年 1 月~2023 年 12 月烟台市短时强降水总频次的月变化(图 4), 烟台市短时强降水总频次月变化总体呈先增大后减小的变化特征, 5~8 月短时强降水总频次呈上升趋势, 且增加的趋势较为显著, 8 月达到最多为 398 次, 之后逐渐减小, 其中 5 月为发生短时强降水频次最少的月份, 为 13 次, 11 月至次年 4 月短时强降水发生频次为 0, 表明烟台市短时强降水月变化特征较为显著, 变化波动幅度较大, 且短时强降水较为集中在 6~8 月。

进一步分析 2019 年 1 月~2023 年 12 月烟台市短时强降水总频次的年变化(图 5)可知, 烟台市短时强降水总频次年变化总体也呈先增大后减小的变化特征, 2019~2022 年短时强降水总频次呈上升趋势, 且增加的趋势较为显著, 2022 年达到最多为 499 次, 之后逐渐减少, 其中 2019 年最少为 72 次。由此可知, 烟台市短时强降水年变化特征较为显著, 变化波动幅度较大, 在 2021~2023 年短时强降水发生频次较高。

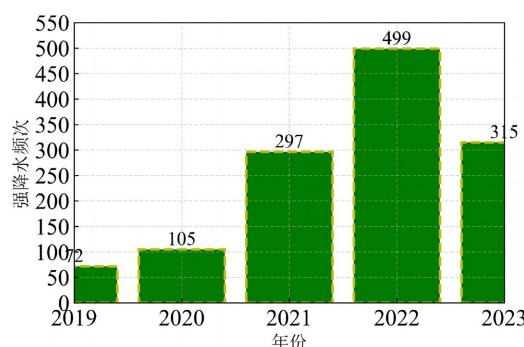


Figure 5. Annual variation of total frequency of short-time heavy precipitation in Yantai City, Shandong Province from January 2019 to December 2023

图 5. 2019 年 1 月~2023 年 12 月山东烟台市短时强降水总频次年变化

3.3. 基于 CatBoost 模型建立的烟台市短时强降水预警预报模型

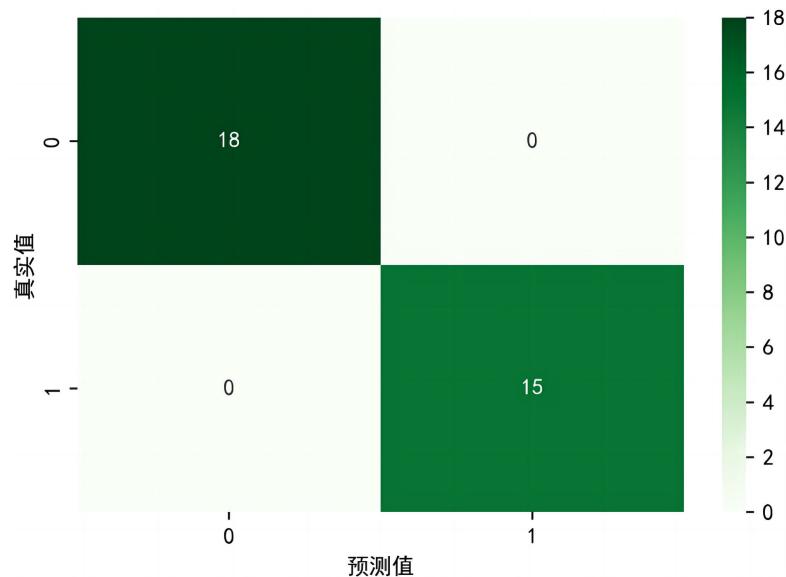


Figure 6. CatBoost short-term heavy rainfall warning and forecasting model prediction confusion matrix (training set)

图 6. CatBoost 短时强降水预警预报模型预测混淆矩阵(训练集)

从图 6 可以看到, CatBoost 短时强降水预警预报模型在训练集上的具体数据划分如下:

真正例(True Positive, TP): 预测为 1 且实际为 1, 共 18 个样本。

真负例(True Negative, TN): 预测为 0 且实际为 0, 共 15 个样本。

假正例(False Positive, FP): 预测为 1 但实际为 0, 没有。

假负例(False Negative, FN): 预测为 0 但实际为 1, 没有。

在图 7(测试集混淆矩阵)中, 短时强降水预警预报模型在测试集上得到的结果如下:

真正例(TP): 3 个。

真负例(TN): 6 个。

假正例(FP): 无。

假负例(FN): 无。

可见, 在测试集上, 模型对于 9 个样本的预测较优, 基本全部预测正确, 这表明模型具有很好的泛化能力。但需要说明的是, 本次结果均基于烟台市短时强降水样本集, 由于整体样本数较少, 得到的结论可能具有局限性。

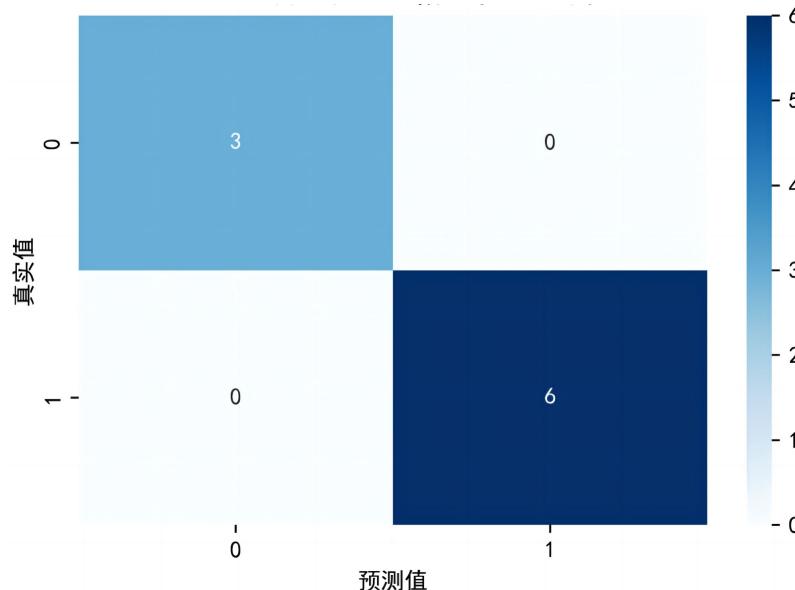


Figure 7. CatBoost short-term heavy rainfall warning and forecasting model prediction confusion matrix (test set)
图 7. CatBoost 短时强降水预警预报模型预测混淆矩阵(测试集)

3.4. 特征重要性及相关性分析

图 8 展示了各个特征在模型预测中的重要性, 可以看出, 不同特征对模型的贡献是不均衡的。500 hPa 散度(500_d)贡献度最高超过 0.14, 它是模型做出预测时最关键的特征, 说明这个变量在样本分类过程中起到了决定性的作用。在数据中, 它与目标变量 y 即短时强降水的相关性非常高, 使该特征的分布与类别之间的界限更加显著。850 hPa 比湿(850_q)贡献度最低接近 0.07, 几乎对模型的预测没有影响(接近 0)。其中 500 hPa 涡度(500_vo)、500 hPa 散度(500_d)、850 hPa 散度(850_d)与地面露点温度(d2m)贡献度均超过 0.1, 表明这些物理量对于短时强降水的贡献度相对较高, 对于短时强降水的发生发展影响较大。从特征重要性分析来看, 500 hPa 涡度(500_vo)、500 hPa 散度(500_d)、850 hPa 散度(850_d)与地面露点温度(d2m)是该模型做出预测时的主要依据, 而其他特征对分类任务的帮助较为有限。这也意味着, 如果需要简化模型或进行特征选择, 可以考虑减少对其他物理量的依赖。

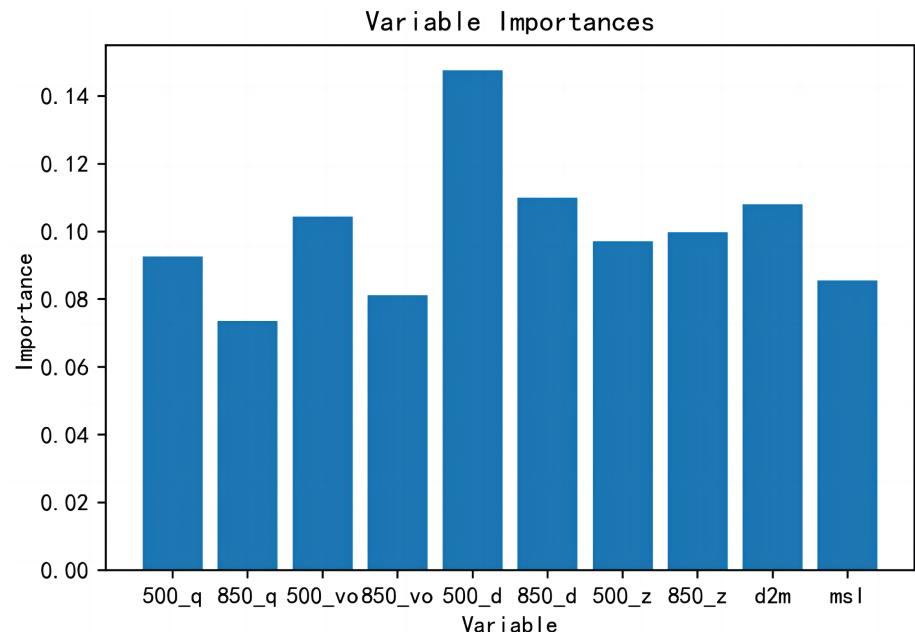


Figure 8. Contribution of each characteristic parameter of CatBoost short-term heavy rainfall warning and forecasting model
图 8. CatBoost 短时强降水预警预报模型各特征参数贡献度

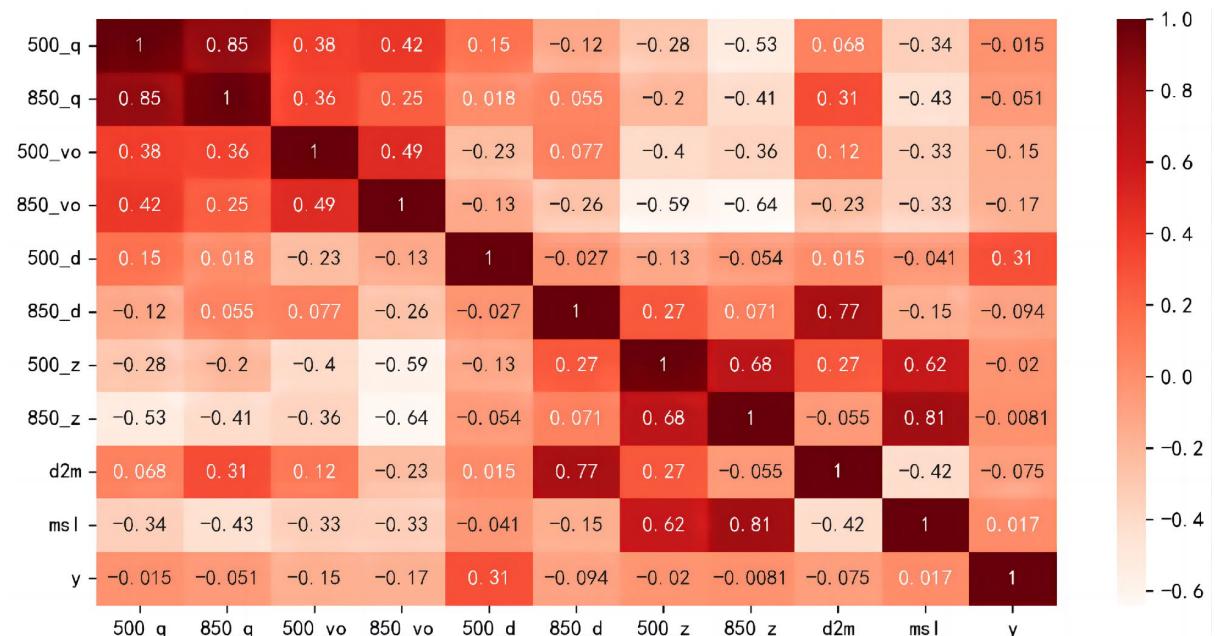


Figure 9. Correlation analysis of CatBoost short-term heavy rainfall warning and forecasting model parameters
图 9. CatBoost 短时强降水预警预报模型各参数相关性分析

图 9 展示了不同特征之间的相关性情况，以及这些特征与目标变量短时强降水(y)的关系。在相关性矩阵中，数值的范围从-1 到 1，数值越接近 1 表示两个特征正相关，接近-1 表示负相关，而接近 0 则表示不相关。由短时强降水各参数相关性分析可知(图 3)，500 hPa 散度(500_d)与短时强降水(y)相关系数最大为 0.31，且为正相关，表明 500 hPa 散度与短时强降水为同向变化，且相关性较强，即 500 hPa 散度(500_d)增大时目标变量 y 更有可能是正类；500 hPa 涡度(500_vo)、850 hPa 涡度(850_vo)与短时强降水(y)

的负相关性最大, 相关系数为-0.15、-0.17; 平均海平面气压(msl)、850 hPa 比湿(850_z)、500 hPa 比湿(500_q)、500 hPa 位势高度(500_z)与短时强降水(y)相关系数为 0.017、-0.0081、-0.015、-0.02, 表明上述物理量与短时强降水(y)的相关性较小, 独立性较高, 说明这些特征对目标分类的帮助不大, 这与特征重要性分析的结果一致。由此可知, 500 hPa 散度(500_d)与短时强降水(y)正相关性最高, 500 hPa (500_vo) 涡度、850 hPa 涡度(850_vo)与短时强降水的负相关性最高。

其他特征之间的相关性, 500 hPa 比湿(500_q)与 850 hPa 比湿(850_q)正相关性最高, 相关系数为 0.85; 850 hPa 涡度(850_vo)与 850 hPa 位势高度(850_z)负相关性最高, 相关系数为-0.64; 地面露点温度(d2m)与 850 hPa 散度(850_d)相关系数为 0.015, 500 hPa 散度(500_d)与 850 hPa 散度(850_d)相关系数为-0.027, 其相关性较小, 独立性较高。从相关性热力图来看, 特征之间有部分高度相关的变量, 这意味着数据的多重共线性问题较大, 模型的结果会受到某一组高度相关变量的较多影响。地面露点温度(d2m)、500 hPa 散度(500_d)与 850 hPa 散度(850_d)的特征重要性较低, 而且与目标变量 y 的相关性也不强。这表明这些特征对于模型的预测贡献有限, 可以考虑在未来的数据处理中进行特征简化, 减少对这些变量的依赖。

4. 结论

本文利用山东省烟台市 2019 年 1 月~2023 年 12 月降水实况资料, 对烟台市的短时强降水发生频次的空间分布、日变化、月变化以及年变化特征进行分析, 结论如下:

(1) 2019 年 1 月~2023 年 12 月烟台市短时强降水发生频次及最强降水主要集中在西北以及东南部沿海地区, 其中西北部地区的短时强降水, 从发生频次以及最大强降水来说均较大, 在烟台中部、北部及东部短时强降水发生频次及强度均较小。

(2) 烟台市短时强降水总频次日变化、月变化及年变化特征均较为显著, 其中日变化, 傍晚至凌晨短时强降水发生频次最多, 凌晨及午后为发生频次第二多的时段, 而日出前后短时强降水发生频次较少; 其月变化波动幅度较大, 呈先增大后减少的变化特征, 短时强降水较为集中在 6~8 月; 其年变化波动幅度也较大, 也呈先增大后减少的变化趋势, 在 2022 年达到最多为 499 次, 之后逐渐减少, 2019 年最少为 72 次, 短时强降水发生频次较多的时段集中在 2021~2023 年。

(3) 用 CatBoost 模型对烟台市强降水进行初步预测, 模型基本能够很好地捕捉烟台市短时强降水特征, 在测试集上表现出较好的预测能力, 这表明该模型具有很好的泛化能力。基于 CatBoost 模型从特征重要性分析来看, 500 hPa 涡度(500_vo)、500 hPa 散度(500_d)、850 hPa 散度(850_d)与地面露点温度(d2m)是该模型做出预测时的主要依据。500 hPa 散度(500_d)与短时强降水(y)正相关性最高, 500 hPa (500_vo) 涡度、850 hPa 涡度(850_vo)与短时强降水的负相关性最高。同时从相关性热力图来看, 特征之间有部分高度相关的变量, 这意味着数据的多重共线性问题较大, 模型的结果会受到某一组高度相关变量的较多影响。

参考文献

- [1] 陈军. 短时强降水个例的特征分析[D]: [硕士学位论文]. 南京: 南京大学, 2016.
- [2] 樊李苗. 短时强降水发生的环境条件和个例分析[D]: [硕士学位论文]. 北京: 中国气象科学研究院, 2012.
- [3] 孙晓巍. 辽宁短时强降水分布特征及预报技术研究[D]: [硕士学位论文]. 兰州: 兰州大学, 2015.
- [4] 丁丽佳, 谢汉良, 余元标, 等. 潮州极端短时强降水时空分布特征及地形作用[J]. 广东气象, 2022, 44(6): 16-20.
- [5] 孙颖, 邓运超, 赵程程. 青岛市短时强降水及其大尺度环流背景特征分析[J/OL]. 海洋湖沼通报(中英文), 1-16. <http://kns.cnki.net/kcms/detail/37.1141.P.20240903.1144.002.html>, 2025-01-09.
- [6] 唐永兰, 徐桂荣, 万蓉. 2020 年主汛期长江流域短时强降水时空分布特征[J]. 大气科学学报, 2022, 45(2): 212-224.

- [7] 王子月. 长江中下游夏季极端降水事件的归因和未来预估[D]: [博士学位论文]. 南京: 南京信息工程大学, 2023.
- [8] 王雅芳. 人为和自然强迫对全球降水变化的影响及预估[D]: [博士学位论文]. 兰州: 兰州大学, 2023.
- [9] 刘婕妤. 不同空间尺度极端降水变化的人类活动检测归因[D]: [硕士学位论文]. 扬州: 扬州大学, 2022.