龙柯吉，康岚，黄晓龙，等，2025. 四川复杂地形下雷暴大风客观预报方法研究[J]. 气象，51(1)：57-67. Long K J，Kang L，Huang X L，et al，2025. Study on the objective forecasting method of thunderstorm gale under complex terrain in Sichuan Province[J]. Meteor Mon，51(1)：57-67(in Chinese).

四川复杂地形下雷暴大风客观预报方法研究*

龙柯吉[1,3] 康 岚[1,3] 黄晓龙[2,3] 陈朝平[1,3] 但 玻[1,3] 周 威[1,3] 张武龙[1,3]

1 四川省气象台，成都 610072

2 四川省气象探测数据中心，成都 610072

3 高原与盆地暴雨旱涝灾害四川省重点实验室，成都 610072

提 要： 地形复杂的四川地区，虽然雷暴大风发生频次较低，但造成的影响重大，现有的客观预报产品较少且时间分辨率低，为进一步提升四川复杂地形下雷暴大风预报准确率，综合考虑地形因子、模式物理量因子和时间因子，根据海拔高度将四川分为高海拔区和低海拔区，利用2018—2021年数据基于随机森林、自适应提升法、极端随机树三种机器学习方法分区构建雷暴大风预报模型，对2022年进行预报，获得逐3 h雷暴大风潜势预报，再利用气候背景将3 h预报时间降尺度到1 h，形成0～12 h逐小时雷暴大风预报，并检验预报效果。结果表明，逐3 h雷暴大风预报以自适应提升法效果最优，长时间检验和个例检验都表明，基于自适应提升法获得的0～12 h逐小时雷暴大风预报产品优于中央气象台产品，TS评分由0.0104提升至0.0595，空报率由0.988下降至0.808，业务应用价值较高。

关键词： 复杂地形，雷暴大风，机器学习，时间降尺度

中图分类号： P456，P457 **文献标志码：** A **DOI：** 10.7519/j.issn.1000-0526.2024.080501

Study on the Objective Forecasting Method of Thunderstorm Gale Under Complex Terrain in Sichuan Province

LONG Keji[1,3] KANG Lan[1,3] HUANG Xiaolong[2,3] CHEN Chaoping[1,3]
DAN Bo[1,3] ZHOU Wei[1,3] ZHANG Wulong[1,3]

1 Sichuan Meteorological Observatory, Chengdu 610072

2 Sichuan Meteorological Observation and Data Centre, Chengdu 610072

3 Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, Chengdu 610072

Abstract: In the complex terrain of Sichuan Region, although the frequency of thunderstorm gale is relatively low, its impact is significant. There are few objective forecast products of the thunderstorm gale, and even if there are, the time resolution is lower. In order to further improve the accuracy of thunderstorm gale forecasting under complex terrain in Sichuan, in this article we comprehensively consider terrain factors, model physical quantity factors and time factors. According to the altitude, Sichuan is divided into high-altitude and low altitude areas. Based on the data from 2018 to 2021 and three machine learning methods of random forest, adaptive boosting and extreme random tree, we construct a thunderstorm gale prediction model and make a forecast for the 2022 thunderstorm gales obtaining a 3 h thunderstorm gale potential

* 国家重点研发计划（2021YFC3000905）、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项目（SCQXKJYJXZD202201）、四川省气象局智能网格预报创新团队（SCQXCXTD-202201）、中国气象局创新发展专项（CXFZ2023J016）和四川省科学技术厅重点研发项目（2024YFFK0407）共同资助

2024年3月27日收稿； 2024年10月21日收修定稿

第一作者：龙柯吉，主要从事天气学与数值预报研究. E-mail：longkeji945@163.com

通讯作者：康岚，主要从事天气预报及相关研究. E-mail：812700656@qq.com

forecast. Then, based on the climate background, we downscale the 3 h forecast time to 1 h, and make a 0—12 h hourly thunderstorm gale forecast. At the same time, the forecasting effect is tested. The results show that, the adaptive boosting method of 3 h thunderstorm gale forecast has the best effect. The long-time and individual case tests show that the 0—12 h hourly thunderstorm gale forecast product obtained by the adaptive boosting method is superior to the forecasts of National Meteorological Centre with the TS score increased from 0.0104 to 0.0595, and the false alarm rate decreased from 0.988 to 0.808. This indicates that the adaptive boosting method has a higher application value in forecasting operation application value.

Key words: complex terrain, thunderstorm gale, machine learning, downscaling of time

引 言

近年来,强对流天气频发,在全国范围内造成了重大生命财产损失(郑永光等,2016;王婷婷等,2023)。雷暴大风是瞬时风力≥8级并伴有雷暴的强对流天气(刘彬等,2022),是一种强风灾害,具有空间尺度小、突发性强、破坏力大等特点。费海燕等(2016)统计了2004—2013年四川盆地发生雷暴大风次数,四川盆地雷暴大风属于小概率事件,一旦发生,往往会造成严重灾害。例如2022年4月11日晚,四川省安岳县出现罕见雷暴大风天气,极大风速达到了37.4 $m \cdot s^{-1}$(13级),破纪录的狂风裹挟着大雨冰雹给当地40余个乡(镇)带来了不同程度的灾害,多人受灾,农作物大面积受损,直接经济损失巨大。由于雷暴大风等强对流天气局地性强、发展迅速,具有极高的预报难度(郑永光等,2015;2017)。而四川地区由于复杂的地形地貌,雷暴大风的预报预警难度更大。

一直以来,对于雷暴大风的临近预报主要依赖雷达外推,即根据天气雷达探测到的回波数据来确定回波的移动速度和方向(郭瀚阳等,2019),而短时和短期预报则侧重数值模式预报释用(张小玲等,2018)。四川预报雷暴大风依靠预报员的主观经验、传统的统计结果,还有基于物理机理的预报模型,如"配料法"(张小玲等,2010;俞小鼎,2011)。康岚等(2018)指出通过分析对流发展背景条件,确定最有利对流发展的潜势区域,关注回波的生成、形态特点、演变特征,可提前预警大风天气。龙柯吉等(2020)统计给出了四川盆地不同类型雷暴大风发生的环流形势及雷达回波特征。陈永仁等(2018)利用主成分分析确定配料系数及其阈值进行配料,研发了四川省雷暴大风概率预报产品。但仅基于统计或主观得到的特征物理量和阈值范围,难以完全发现数据中有价值的细节信息,尤其是一些中小尺度信息,此外,如果预报员对强对流发生发展规律认识不够全面深刻,也难以捕捉到相关信息。随着智能网格预报业务的快速发展,四川省气象台已构建起一套高分辨率、多时效的格点化预报产品体系,为气象服务提供了有力支撑,但针对强对流天气的预报产品尚显不足,亟待通过研究加以补充完善。

在人工智能技术的浪潮中,机器学习作为核心动力,正引领各领域变革。杜智涛等(2021)认为机器学习在加速气象观测数据处理、提升数值天气预报质量以及促进地球科学交叉融合方面具有巨大潜力。面对强对流天气这一复杂多变的气象挑战,机器学习技术以其强大的数据处理与模式识别能力,成为了监测与预报预警的重要工具。众多学者已成功运用BP(back propagation)神经网络模型、决策树、随机森林、极端随机树和逻辑回归等算法,实现了对强对流天气的分类与识别,为强对流天气的早期预警提供了有力支持(Gagne Ⅱ et al,2009;Kim et al,2017;杨璐等,2018;周康辉等,2021)。进一步地,通过对基于不同机器学习算法的预测模型的构建并优化,有效提升了降水、雷暴、大风、冰雹等强对流天气及其潜势的预报精度(陈勇伟等,2013;Collins and Tissot,2015;周康辉等,2016;李文娟等,2018;Czernecki et al,2019;顾建峰等,2020),为防灾减灾工作提供了科学依据。鉴于四川地区复杂地形的特殊性及其对气象预报准确性的高要求,本研究将综合利用模式预报和地形数据等(黄晓龙等,2022),对比不同机器学习方法对雷暴大风的预报效果并进行订正,为提升四川地区复杂地形下雷暴大风的预报准确率提供技术支撑。

1　资料与方法

1.1　数据来源

本文使用的实况资料为地面气象观测站逐小时极大风资料,兼顾时间长度和站点数量,从 1924 个有风观测要素的加密站中,剔除缺测时次较多的站点,最终选取站点 1792 个,站点分布如图 1 所示。地闪定位资料为 ADTD 闪电定位仪资料,精度约为 500 m,平均探测效率为 90%。数据来源于四川省气象探测数据中心,已通过业务质量控制,包括内部一致性、持续性、气候极值等方法,具有较好质量保障;模式资料为欧洲中期天气预报中心(ECMWF)模式逐 3 h 预报产品,其中地面和高空要素空间分辨率分别为0.125°×0.125°、0.25°×0.25°;用于描述地表起伏形态特征的数字高程模型(DEM)数据分辨率为90 m。上述资料研究时段均为 2018—2022 年,所有时间均为北京时。

1.2　数据处理

1.2.1　雷暴大风实况

基于地面观测资料及闪电资料判别雷暴大风,即某站点出现≥17.2 m·s^{-1}的大风天气,且伴有闪电时判定该站点出现雷暴大风。具体方法为:将一小时内某网格点周围 0.25°范围内的所有正负闪电次数累计为该格点的闪电次数,最终形成 0.5°×0.5° 的逐小时累计闪电次数的格点数据(周威等,2020),再利用临近插值方法将格点数据插值到观测站点,若站点的闪电次数≥1,且瞬时风速≥17.2 m·s^{-1},则判定该站点出现一次雷暴大风天气。

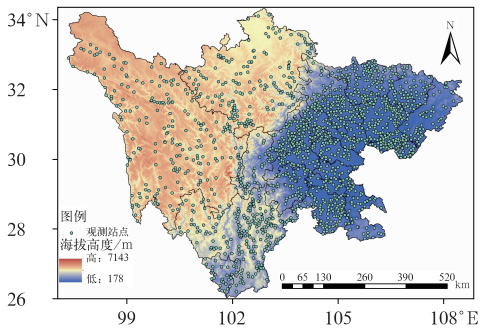

图 1　四川省地形高度及地面气象观测站分布

Fig. 1　Terrain height and distribution of surface meteorological stations in Sichuan Province

1.2.2　模式物理量因子

天气因子对雷暴大风的预报至关重要,强对流天气的形成包含水汽、不稳定层结和动力抬升三个基本条件,垂直风切变对对流风暴的发生、发展也有重要作用。综合考虑这些条件,结合预报经验,初步选取 132 个模式预报因子,包括 ECMWF 模式预报直接输出的物理量和计算获得的物理量,根据变量层次划分为三类:高空变量、单层变量和组合量,其中高空变量层次包括 925、850、700、600、500、400、300 hPa,组合变量包括不同层次垂直温差、假相当位温差,多层次平均风速、平均温度露点差、平均相对湿度,多层次最大温度露点差、最小相对湿度等,具体详见表 1。

1.2.3　地形因子提取

地形与雷暴大风分布密切相关(Jungo et al,2002;杨璐等,2023),因此根据站点或格点的序号(SID)、经度(LON)和纬度(LAT)信息,利用 DEM 数据提取四川区域的地形因子,包括:海拔高度(ALT)、坡度(SLOPE)、坡向(ASPECT)、地表粗糙度(SR)和地表起伏度(RDLS)。其中坡度是指坡面垂直高度与路程之比,坡度值越小,地势越平坦;反之,坡度值越大,地势越陡峭。坡向为线矢量在水平面上的投影与过该点的正北方向的夹角,表征该观测站高程值改变量的最大变化方向。地表粗糙度为特定区域内地球表面积与其投影面积之比,是反映地表形态的宏观指标,能反映地表抗风蚀的能力。地表起伏度为某一特定区域内最高海拔高度和最低海拔高度的差值,是描述一个区域地形特征的宏观性指标。

1.2.4　模型训练数据集

四川地区雷暴大风具有明显日变化和月变化,因此除了天气因子、地形因子,本文还引入时间因子进行建模,包括:月(MONTH)、时(HOUR)。首先,利用双线性插值将 ECMWF 模式预报的物理量因子数据插值到研究站点,然后将模式物理量因子、地形因子、时间因子三类因子与大风实况数据进行匹配处理,建立训练数据集。由于 ECMWF 预报资料的时间分辨率为3 h,并考虑模式具有平衡调整过程,因此将当日06—08、09—11、12—14、15—17 时的大风实况与前一日 20 时起报的 12、15、18、21 h 预报资料进行匹配。而当日 18—20 时、21—23 时和次日 00—02 时、03—05 时的大风实况与当日 08 时起报的 12、15、18、21 h 预报资料进行匹配。

表1 模式物理量因子

Table 1 Model physical quantity factors

高空变量	标识	单层变量	标识	组合量	标识
温度	T	2 m温度	$T_{2\,m}$	温度露点差	$T-T_d$
露点温度	T_d	2 m露点温度	$T_{d\,2\,m}$	700 hPa和500 hPa温差	$T_{700-500}$
假相当位温	θ_{se}	海平面气压	p	850 hPa和500 hPa温差	$T_{850-500}$
经向风	u	3 h海平面变压	Δp_3	700 hPa和500 hPa、850 hPa和500 hPa温差之和	$T_{700-500}+T_{850-500}$
纬向风	v	10 m经向风	$u_{10\,m}$	600 hPa和300 hPa温差	$T_{600-300}$
风速	WS	10 m纬向风	$v_{10\,m}$	500 hPa和300 hPa温差	$T_{500-300}$
位势高度	GH	100 m经向风	$u_{100\,m}$	600 hPa和300 hPa、500 hPa和300 hPa温差之和	$T_{600-300}+T_{500-300}$
散度	Div	100 m纬向风	$v_{100\,m}$	850 hPa和500 hPa假相当位温差	$\theta_{se(850-500)}$
涡度	Vor	对流有效位能	CAPE	700 hPa和500 hPa假相当位温差	$\theta_{se(700-500)}$
垂直速度	w	K指数	K	600 hPa和400 hPa假相当位温差	$\theta_{se(600-400)}$
3 h变温	ΔT_3	大气可降水量	PW	500 hPa和300 hPa假相当位温差	$\theta_{se(500-300)}$
比湿	q	沙氏指数	SI	700～400 hPa平均风速	$WS_{700\sim400}$
相对湿度	RH	2 m温度3 h变温	$\Delta T_{3\,2\,m}$	600～300 hPa平均风速	$WS_{600\sim300}$
		3 h降水	R_3	500～300 hPa平均风速	$WS_{500\sim300}$
		925～700 hPa垂直风切变	$SHR_{925\sim700}$	700～400 hPa平均温度露点差	$(T-T_d)_{700\sim400}$
		925～500 hPa垂直风切变	$SHR_{925\sim500}$	700～400 hPa最大温度露点差	$(T-T_d)^{max}_{700\sim400}$
		600～300 hPa垂直风切变	$SHR_{600\sim300}$	500～300 hPa平均温度露点差	$(T-T_d)_{500\sim300}$
		600～400 hPa垂直风切变	$SHR_{600\sim400}$	500～300 hPa最大温度露点差	$(T-T_d)^{max}_{500\sim300}$
				700～400 hPa平均相对湿度	$RH_{700\sim400}$
				700～400 hPa最小相对湿度	$RH^{min}_{700\sim400}$
				500～300 hPa平均相对湿度	$RH_{500\sim300}$
				500～300 hPa最小相对湿度	$RH^{min}_{500\sim300}$

结合雷暴大风气候背景，按照海拔高度进行分区：低海拔(≤2000 m)、高海拔(＞2000 m)，其中低海拔站点主要分布在四川盆地，高海拔站点主要分布在川西高原和攀西地区，从而形成2个训练数据集，然后分别建模。

1.3 研究方法

利用2018—2021年数据构建随机森林(RF)、自适应提升法(ADA)、极端随机树(EXT)3种机器学习算法模型，训练集和测试集随机抽取，占比分别为80%、20%，然后利用2022年数据检验模型订正效果，获得最优的逐3 h雷暴大风潜势预报，在此基础上基于气候背景对最优预报进行时间降尺度，形成0～12 h逐小时雷暴大风预报，并与中央气象台下发的基于配料法的雷暴大风产品(SCMOC)进行对比检验。

1.3.1 机器学习方法

集成学习通过组合多个机器学习模型来产生一个优化的模型，从而提高模型的性能，该方法可以减少过拟合，泛化性能好，与单个模型相比具有更高的预测精度。集成算法通常有两种，分别为套袋法(Bagging)和提升法(Boosting)。随机森林属于集成学习中的Bagging算法，由多个决策树集成，在进行随机森林过程中，其输出值是随机森林中所有决策树结果的平均值(孙全德等，2019)，随机森林可以产生高准确度的分类器，处理大量的输入变数，但由于其自身的随机性，可能会导致预测结果出现波动。自适应提升法属于集成学习中的Boosting算法，其核心思想是对某一特定的训练集训练多个弱分类器，然后根据相应策略将这些弱分类器结合起来，构成最终的强分类器，可以使用各种回归分类模型来构建弱学习器，非常灵活，但对异常样本敏感，会影响最终预测结果的准确性。极端随机树是随机森林的变种，都是由多个决策树构成，区别在于随机森林应用的是Bagging模型，极端随机树使用所有的样本，只是随机选取特征，因为分裂存在随机性，在某种程度上可能比随机森林得到更好的结果，但极端随机树的模型解释性较差，对于数据集中类别不平衡的问题处理能力较弱。极端随机树多用于文本分类、图像识别，特别是卫星云图识别等领域(周康辉等，2021)。

1.3.2 特征因子确定

考虑到节省计算资源及业务应用，将3种模型分别迭代训练，获取不同特征因子的重要性，然后将重要性从大到小排序，选取每种模型排名在前30的因子作为最终的特征因子，3种模型雷暴大风在测试集中的整体预测精度均接近1.0，所选取的特征因子及对应的重要性(括号内数字)如表2所示。可

以看出,不同模型在不同分区筛选的特征因子有所不同,但总体而言,大部分因子为模式预报的不同层次风,还包含热力、水汽、动力、垂直风切变相关的物理量以及站点经纬度信息、时次等,和主观预报经验相符。另外,高海拔地区还包含了地形因子,如海拔高度、坡向、起伏度等,印证了地形对大风的预报具有重要作用。

表 2　不同模型分区特征因子及重要性(单位:%)

Table 2　The characteristic factors and importance (unit: %) of different models

序号	高海拔特征因子			低海拔特征因子		
	RF	ADA	EXT	ADA	EXT	RF
1	LAT(12)	LAT(5.9)	LAT(5.7)	LON(10)	LAT(4.6)	LON(4.8)
2	LON(8)	LON(5.6)	LON(5.0)	CAPE(8)	LON(4.5)	LAT(4.6)
3	SID(6)	ALT(3.8)	$RH^{min}_{500\sim300}$(3.8)	R_3(8)	RH_{300}(3.8)	u_{300}(4.2)
4	GH_{700}(6)	$RH^{min}_{500\sim300}$(3.5)	$\theta_{se(600-400)}$(3.6)	v_{100}(6)	CAPE(3.8)	v_{300}(3.8)
5	$\theta_{se\ 600}$(4)	Δp_3(3.4)	u_{600}(3.6)	q_{850}(6)	v_{850}(3.7)	u_{600}(3.8)
6	$T_{2\ m}$(4)	$\theta_{se(500-300)}$(3.4)	$\theta_{se(500-300)}$(3.6)	v_{850}(6)	v_{400}(3.7)	v_{400}(3.7)
7	HOUR(4)	GH_{600}(3.3)	Δp_3(3.5)	$RH^{min}_{500\sim300}$(4)	v_{700}(3.6)	CAPE(3.7)
8	ALT(4)	v_{400}(3.3)	GH_{700}(3.5)	LAT(4)	p(3.6)	v_{850}(3.7)
9	R_3(4)	v_{300}(3.2)	v_{400}(3.5)	u_{925}(4)	u_{850}(3.5)	w_{300}(3.6)
10	$T_{500-300}$(4)	GH_{400}(3.2)	v_{300}(3.5)	$\theta_{se(700-500)}$	$SHR_{925\sim700}$(3.5)	p(3.6)
11	$T_{850-500}$(4)	GH_{500}(3.2)	GH_{600}(3.5)	u_{600}(2)	w_{300}(3.5)	v_{700}(3.6)
12	GH_{400}(4)	u_{600}(3.2)	$T_{500-300}$(3.4)	$T_{850-500}$(2)	$\Delta T_{3\ 850}$(3.5)	Div_{850}(3.3)
13	u_{100}(2)	$\theta_{se(600-400)}$(3.2)	u_{300}(3.3)	w_{300}(2)	v_{925}(3.5)	v_{100}(3.3)
14	$RH^{min}_{700\sim400}$(2)	$T_{500-300}$(3.2)	v_{500}(3.3)	$T_{500-300}$(2)	Δp_3(3.4)	u_{850}(3.3)
15	T_{400}(2)	v_{500}(3.2)	SID(3.3)	T_{850}(2)	$SHR_{925\sim500}$(3.4)	w_{500}(3.3)
16	T_{850}(2)	u_{500}(3.2)	u_{500}(3.3)	T_{400}(2)	v_{100}(3.3)	Δp_3(3.2)
17	u_{600}(2)	$T_{600-300}+T_{500-300}$(3.2)	v_{600}(3.2)	$\Delta T_{3\ 925}$(2)	$\Delta T_{3\ 925}$(3.3)	Div_{925}(3.2)
18	u_{300}(2)	GH_{700}(3.2)	p(3.2)	$\Delta T_{3\ 850}$(2)	w_{400}(3.3)	w_{700}(3.1)
19	u_{400}(2)	u_{300}(3.1)	$SHR_{600\sim300}$(3.2)	w_{700}(2)	w_{500}(3.2)	$\Delta T_{3\ 850}$(3.1)
20	RH_{400}(2)	u_{400}(3.1)	u_{400}(3.1)	$RH_{700\sim400}$(2)	Div_{850}(3.2)	$SHR_{925\sim700}$(3.1)
21	u_{850}(2)	WS_{400}(3.1)	WS_{300}(3.1)	$\theta_{se\ 700}$(2)	u_{925}(3.1)	v_{925}(3.1)
22	w_{300}(2)	HOUR(3.0)	$WS_{500\sim300}$(3.0)	RH_{600}(2)	$T_{d\ 2\ m}$(3.1)	$T_{d\ 2\ m}$(3.1)
23	w_{700}(2)	$SHR_{600\sim300}$(3.0)	w_{300}(3.0)	SID(2)	R_3(3.0)	SID(3.0)
24	WS_{400}(2)	$WS_{500\sim300}$(3.0)	CAPE(3.0)	RH_{925}(2)	HOUR(3.0)	u_{925}(3.0)
25	$RH_{700\sim400}$(2)	WS_{300}(3.0)	w_{400}(3.0)	RH_{700}(2)	$T_{2\ m}$(2.9)	$T_{d\ 2\ m}$(3.0)
26	RDLS(2)	p(2.9)	WS_{400}(3.0)	$(T-T_d)_{300}$(2)	$v_{10\ m}$(2.8)	$\Delta T_{3\ 2\ m}$(2.8)
27	CAPE(2)	GH_{925}(2.9)	WS_{600}(2.9)	v_{500}(2)	$\Delta T_{3\ 2\ m}$(2.7)	R_3(2.7)
28	GH_{600}(2)	w_{300}(2.9)	ALT(2.8)	u_{700}(2)	$u_{100\ m}$(2.5)	$v_{10\ m}$(2.7)
29	GH_{500}(2)	GH_{850}(2.9)	SLOPE(2.1)	Div_{925}(2)	ALT(2.4)	$u_{100\ m}$(2.6)
30	Vor_{600}(2)	w_{400}(2.8)	ASPECT(2.0)	$(T-T_d)_{700}$(2)	$u_{10\ m}$(2.4)	ALT(2.1)

注:括号内数字为重要性。

1.3.3　时间降尺度

四川地区雷暴大风主要发生在 4 月、6—8 月,高频时段主要为午后到凌晨,存在明显的日变化和月变化。因此,根据 2016—2021 年雷暴大风气候背景,获得雷暴大风的总发生频次、各月发生频次和各时次发生频次,在此基础上将逐 3 h 雷暴大风潜势预报进行时间降尺度,具体方案如下:(1)判断高低频格点,月雷暴大风次数≥3 为高频格点,反之为低频格点。(2)判断高低频发生时段,根据雷暴大风日变化特征,14 时至次日 04 时为雷暴大风高发时段,其余为低发时段。(3)当某格点的总雷暴大风次数<1 时,认为雷暴大风发生概率极小,预报值取 0(雷暴大风不发生)。(4)当雷暴大风 3 h 潜势预报出现时,对于高频格点,在高发时段的 3 个时次预报值均取 1(雷暴大风发生),在低发时段仅频次最高时次取 1;对于低频格点,在高发时段的 3 个时次预报值仅频次最高时次取 1,在低发时段均取 0。

1.3.4　检验评估方法

采用点对面(田付友等,2015)的检验方法对雷暴大风预报进行检验,扫描半径为 40 km,检验站点为四川境内 1183 个站点。具体做法为:针对某个检验站点,搜索以该站点为中心,半径为 40 km 的圆圈范围内所有站点,当该范围内有站点发生雷暴大风时,则判定检验站点出现雷暴大风,在此基础上利

用双态分类列联表(表 3)定义的评分进行检验,检验指标包括:TS 评分、Bias 评分、空报率(FAR)和命中率(POD)。

表 3 双态分类列联表

Table 3 Bimodal classification

预报	观测	
	有	无
有	NA(命中)	NB(空报)
无	NC(漏报)	ND(正确否定)

$$\mathrm{TS}=\frac{\mathrm{NA}}{\mathrm{NA}+\mathrm{NB}+\mathrm{NC}} \tag{1}$$

$$\mathrm{Bias}=\frac{\mathrm{NA}+\mathrm{NB}}{\mathrm{NA}+\mathrm{NC}} \tag{2}$$

$$\mathrm{FAR}=\frac{\mathrm{NB}}{\mathrm{NA}+\mathrm{NB}} \tag{3}$$

$$\mathrm{POD}=\frac{\mathrm{NA}}{\mathrm{NA}+\mathrm{NC}} \tag{4}$$

式中:NA、NB 和 NC 分别为雷暴大风的命中站次、空报站次和漏报站次。

2 结果检验

2.1 雷暴大风 3 h 潜势预报检验

从 2022 年全年的评分(图 2)来看,除了 EXT,其余均有 TS 评分,其中以 ADA 的评分最高,为 0.0488;3 种模型的 Bias 评分均在 1 之下,ADA 更接近 1,为 0.30;对于空报率,所有模型均具有较高的空报率,以 ADA 最优,为 0.797;对于命中率,ADA 的评分最高,为 0.0604。对比而言,ADA 预报效果最优。

从不同预报时效检验情况(图 3)来看,对于 TS 评分,所有模型在夜间评分高于白天,这可能与雷暴大风在夜间发生的频次更高有关,其中 ADA 基本在所有预报时效都优于其他模型;对于 Bias 评分,3 种模型均呈现单峰型,ADA 和 EXT 峰值均出现在 24 h 时效,但 EXT 明显偏小,ADA 则接近 1,RF 峰值出现在 30 h 时效;对于空报率,所有模型在 21~36 h 时效的空报率均在 1.0 左右,其余时次则以 ADA 空报率最低,15 h 和 18 h 时效预报均在 0.6 以下;对于命中率,3 种模型夜间命中率更高,以 ADA 模型最优,命中率最高接近 0.18。

综上所述,ADA 模型最优,RF 次之,各模型分时效预报评分存在明显的日变化,普遍夜间评分优于白天,以 ADA 模型最优。虽然 3 种模型在测试集中都展现出了良好的性能,但在实际预测时 ADA 模型表现更优,其原因可能与 3 种模型在每个估计器权重和特征选择的策略有关,RF、EXT 的每个估计器相对独立且权重都相同,而 ADA 的每个估计器都依赖于上一个估计器,分配不同权重。ADA 在每次分类任务中重点关注被错误分类样本,这样逐步可以得到一个比较清晰的分类方向,即一个是否发生雷暴大风的强分类器,使其能够筛选出对预测

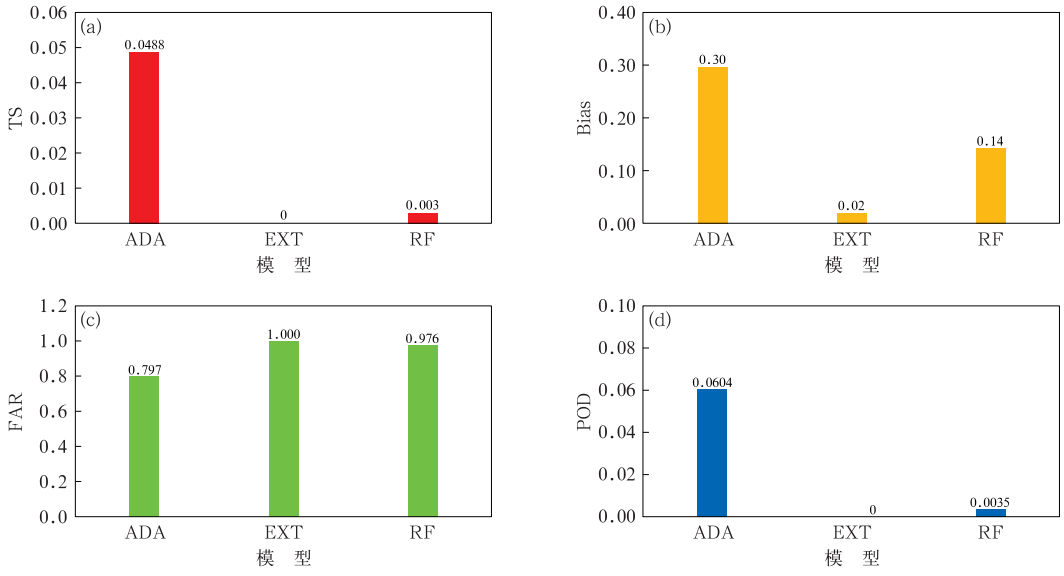

图 2 2022 年 3 种模型的四川省雷暴大风预报评分

Fig. 2 Forecast score of thunderstorm gales of the three models in Sichuan Province in 2022

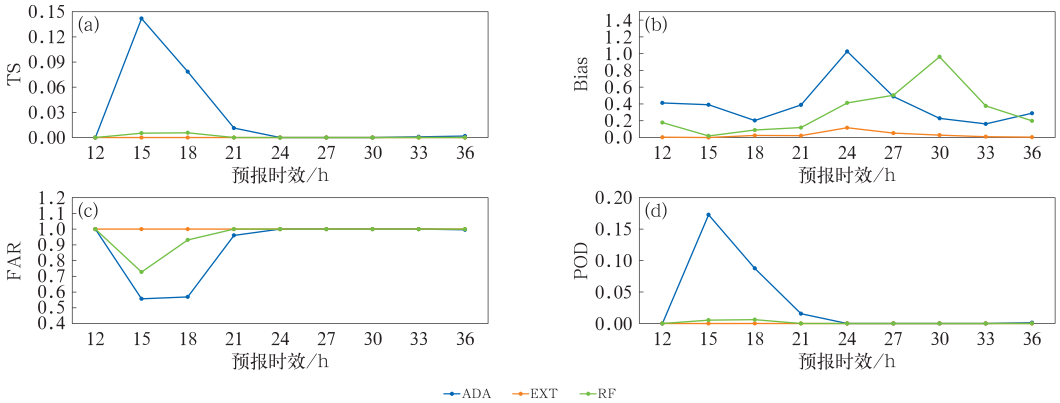

图 3　2022 年 3 种模型四川省雷暴大风不同预报时效评分

Fig. 3　Different forecast lead time score of thunderstorm gales of the three models in Sichuan Province in 2022

结果有重要影响的因子,例如垂直温差,进而在实际预测应用中表现优于 RF 和 EXT。

2.2　雷暴大风 1 h 预报检验

2.2.1　长时间序列预报检验

检验结果表明,雷暴大风 3 h 潜势预报以 ADA 方法最优,将该产品结合气候背景时间降尺度到 1 h,然后与中央气象台 SCMOC 产品进行对比检验,检验时段为 2022 年 4—9 月,主要检验 02 时、08 时、14 时、20 时起报的 0～12 h 逐小时预报。

综合评分来看(表 4),ADA 和 SCMOC 产品的 TS 评分分别为 0.0595、0.0104;Bias 评分分别为 0.4、7.5,表明 ADA 的预报范围较实况偏小,SCMOC 范围明显偏大;空报率分别为 0.808、0.988,ADA 的空报率相对 SCMOC 有明显的下降;命中率分别为 0.0794、0.0870,ADA 略低。总体而言,ADA 优于 SCMOC,TS 评分有较大提升,Bias 评分更接近 1,命中率相当,空报率明显下降。

从不同预报时效检验情况(图 4)来看,对于 TS 评分,所有预报时效均以 ADA 评分更高,2、3、8、9 h

表 4　2022 年 ADA 和 SCMOC 产品四川省雷暴大风逐小时预报评分

Table 4　Hourly forecast score of thunderstorm gales of ADA and SCMOC in Sichuan Province in 2022

模型	TS	Bias	FAR	POD
ADA	0.0595	0.4	0.808	0.0794
SCMOC	0.0104	7.5	0.988	0.0870

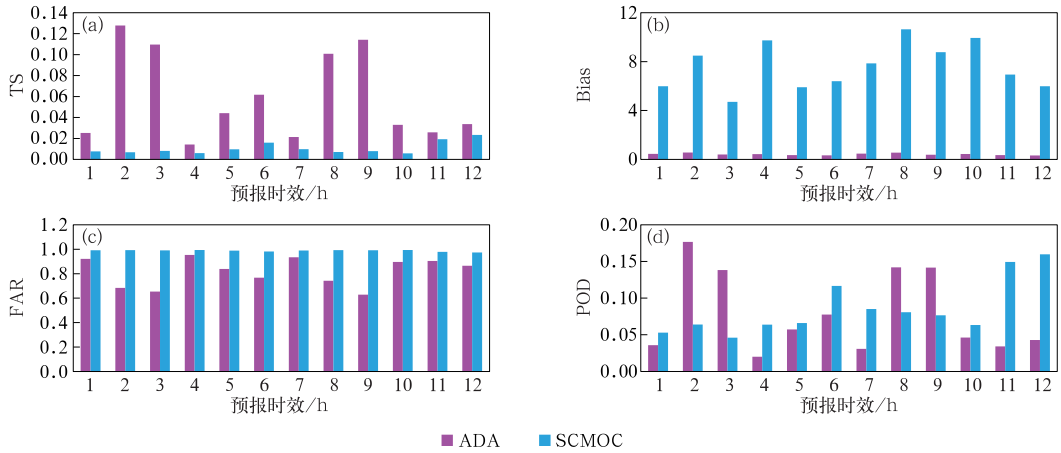

图 4　2022 年 ADA 和 SCMOC 产品四川省雷暴大风不同预报时效评分

Fig. 4　Different forecast lead time score of thunderstorm gales of ADA and SCMOC in Sichuan Province in 2022

预报时效评分均在 0.10 以上，最高达到 0.13 左右，SCMOC 则普遍在 0.02 以下；对于 Bias 评分，ADA 所有预报时效均在 1 以下，范围较实况偏小，SCMOC 则普遍在 5 以上，预报范围较实况明显偏大；对于空报率，ADA 的空报率在所有时效都小于 SCMOC，最低达到 0.60 左右，SCMOC 空报率较高，普遍在 0.98 以上；对于命中率，SCMOC 在 1、4～7、10～12 h 时效的命中率更高，但由于其空报率过高，导致 TS 评分较低。

综上所述，ADA 无论从综合评分还是分时效评分来看，均具有较好的效果。

2.2.2 个例检验

选取 2022 年发生在四川的两次雷暴大风过程，对 SCMOC 和 ADA 两种产品的落区预报进行检验。

图 5 为 2022 年 4 月 11 日 21 时至 12 日 00 时四川省雷暴大风实况与预报对比图，可以看出，11 日21 时在广元、成都、乐山、宜宾和甘孜州的个别地方出现了雷暴大风，22 时雷暴大风范围扩大，呈南北向分布，主要位于绵阳东部到自贡一带，23 时雷暴大风落区略有东移北抬，仍然呈南北向分布，12 日00 时，主要位于盆地北部，在广元东部、南充和遂宁一带。从 2 种产品的预报来看，SCMOC 在各个时次均预报甘孜州南部、雅安、乐山一带有成片雷暴大风，空报较为严重，盆地其余地方则为分散雷暴大风，有一定指示意义，但落区总体变化不大，仍然存在空报；ADA 的雷暴大风 21—23 时集中在广元东部到宜宾北部一带，12 日 00 时移至巴中、南充、遂宁一带，虽然仍然存在一定空报，但总体具有较好的参考性。

图 6 为 2022 年 4 月 28 日 01—03 时雷暴大风实况与预报对比图，可以看出，此次雷暴大风主要发生在绵阳、成都东部和遂宁、资阳，雷暴大风落区自北向南缓慢移动，02 时范围最大。从 2 种产品的预报来看，SCMOC 在各个时次仍然存在较大的空报，而在实况雷暴大风区域则漏报较明显；ADA 的雷暴大风分布总体较为集中，仍然存在一定空报，但预报出 01 时和 02 时成都东部、03 时资阳西部的雷暴大风。

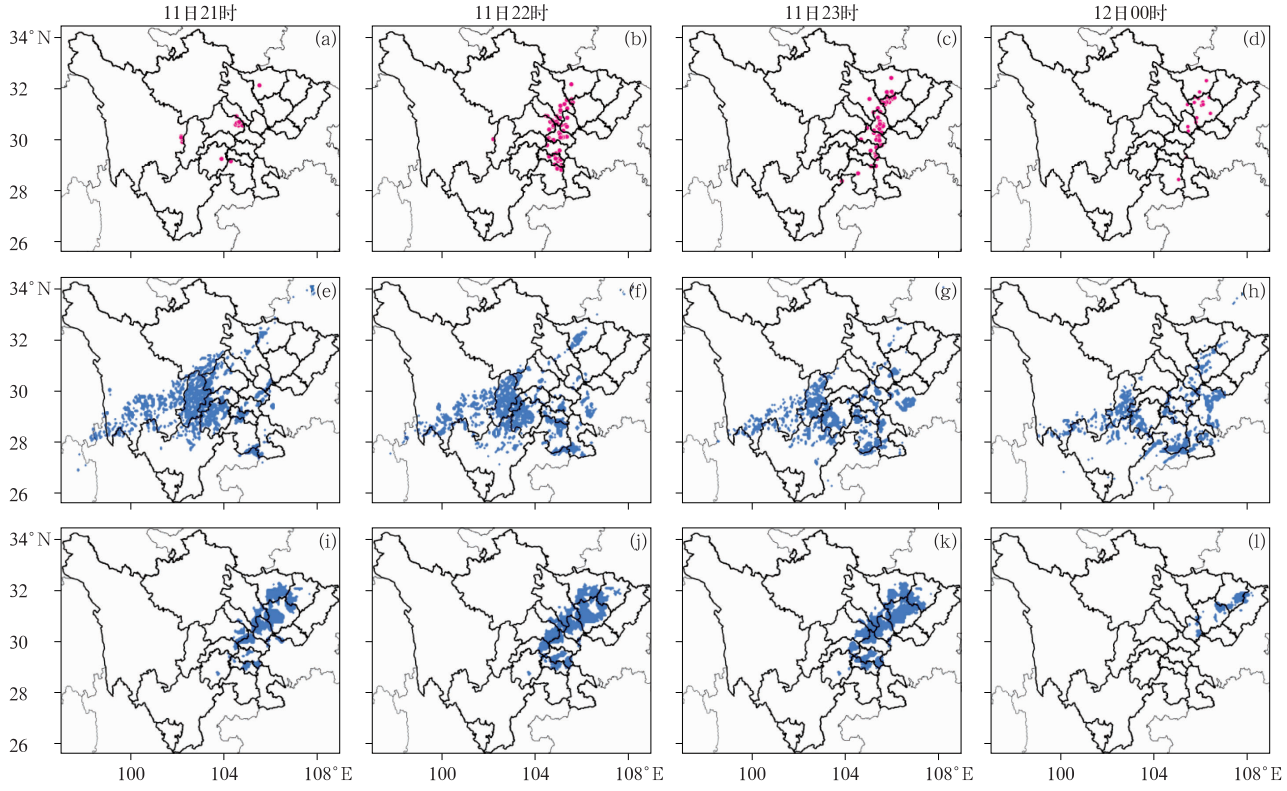

图 5 2022 年 4 月 11 日 21 时至 12 日 00 时四川省雷暴大风(a～d)实况及对应的 11 日 20 时起报的(e～h)SCMOC 和(i～l)ADA 预报

Fig. 5 The thunderstorm gale in Sichuan Province (a—d) observations from 21:00 BT 11 to 00:00 BT 12, and corresponding forecasts of (e—h) SCMOC and (i—l) ADA initiated at 20:00 BT 11 April 2022

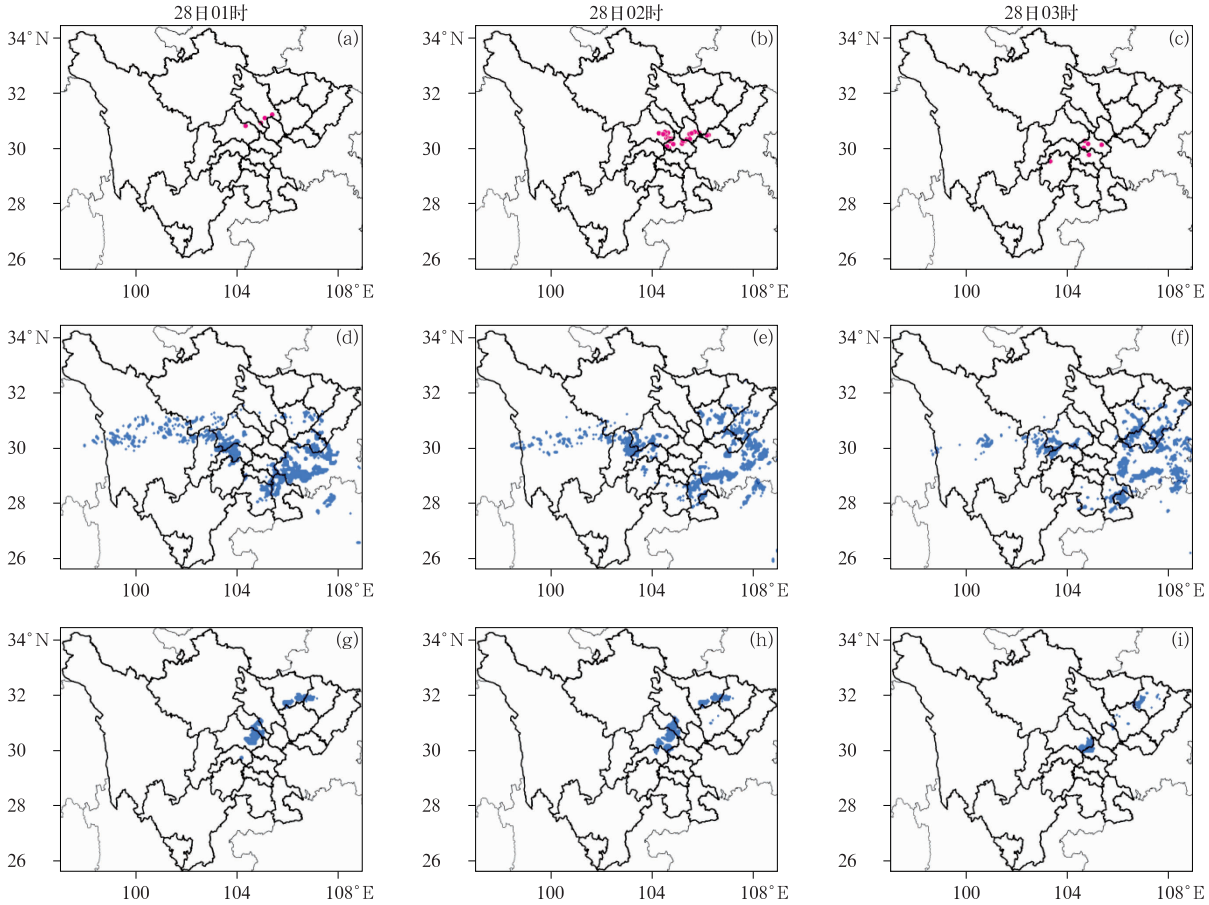

图 6　2022 年 4 月 28 日 01—03 时四川省雷暴大风(a～c)实况及对应的 27 日 20 时起报的(d～f)SCMOC 和(g～i)ADA 预报

Fig. 6　(a－c) The thunderstorm gale in Sichuan Province observations from 01:00 BT to 03:00 BT 28, and corresponding forecasts of (d－f) SCMOC and (g－i) ADA initiated at 20:00 BT 27 April 2022

3　结论和讨论

本文利用 2018—2021 年雷暴大风实况、模式预报、地形因子、时间因子等数据,根据海拔高度分区构建了随机森林、自适应提升法、极端随机树 3 种雷暴大风预报模型,实现逐 3 h 雷暴大风的订正,在此基础上结合气候背景对 3 h 预报进行时间降尺度,获得 0～12 h 逐小时雷暴大风预报,并检验 2022 年预报效果,主要结论如下:

(1)3 种模型在不同分区最终筛选的特征因子略有不同,但都包含不同层次风,站点经纬度信息、时次,以及热力、动力、水汽、垂直风切变相关物理量。高海拔地区还包括海拔高度、坡向、起伏度等地形因子。

(2)逐 3 h 雷暴大风预报以 ADA 效果最优,TS 评分、Bias 评分、空报率和漏报率分别为 0.0488、0.30、0.797、0.0604。分时效检验结果表明,各模型评分存在明显的日变化,普遍夜间评分优于白天,以 ADA 模型最优。

(3)结合 ADA 模型和雷暴大风气候背景形成的 0～12 h 逐小时雷暴大风预报与中央气象台 SCMOC 预报对比检验结果表明,该产品无论综合评分还是分时效评分,均优于 SCMOC,具有较好的订正效果,而 SCMOC 产品因具有较高的空报率,导致 TS 评分偏低。

(4)2022 年 2 次大风个例的订正检验结果表明,SCMOC 产品空报较为明显,ADA 产品相对而言对雷暴大风落区具有更好的指示意义。

虽然本文的雷暴大风预报有一定的预报效果,但仍然存在空报率较高的问题,以后可尝试引入更高时间分辨率模式预报进行消空,并加强四川雷暴

大风的机理分析、提炼和总结消空指标。此外，融合物理理解与模糊逻辑的分类强对流短期客观概率预报产品已纳入国家级智能网格预报产品体系（田付友等，2024a；2024b），可通过检验该产品在四川的预报性能开展融合订正，进一步提升雷暴大风的预报准确率。

参考文献

陈永仁，康岚，曹萍萍，等，2018. 四川雷暴大风的预报因子筛选及预报效果[J]. 高原山地气象研究，38(4)：45-52. Chen Y R，Kang L，Cao P P，et al，2018. Screening of forecasting factors for thunderstorm wind and its forecasting effect in Sichuan[J]. Plateau Mountain Meteor Res，38(4)：45-52(in Chinese).

陈勇伟，郑涛，王汉堃，等，2013. 基于BP神经网络模型的雷电潜势预报[J]. 干旱气象，31(3)：595-601. Chen Y W，Zheng T，Wang H K，et al，2013. Thunderstorm potential prediction based on back propagation neural network[J]. J Arid Meteor，31(3)：595-601(in Chinese).

杜智涛，姜明波，杜晓勇，等，2021. 机器学习在气象领域的应用现状与展望[J]. 气象科技，49(6)：930-941. Du Z T，Jiang M B，Du X Y，et al，2021. Application status and prospect of machine learning in meteorology[J]. Meteor Sci Technol，49(6)：930-941(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).

顾建峰，周国兵，刘伯骏，等，2020. 人工智能技术在重庆临近预报业务中的初步研究与应用[J]. 气象，46(10)：1286-1296. Gu J F，Zhou G B，Liu B J，et al，2020. Study on artificial intelligence technology and its application to Chongqing operational nowcasting[J]. Meteor Mon，46(10)：1286-1296(in Chinese).

郭瀚阳，陈明轩，韩雷，等，2019. 基于深度学习的强对流高分辨率临近预报试验[J]. 气象学报，77(4)：715-727. Guo H Y，Chen M X，Han L，et al，2019. High resolution nowcasting experiment of severe convections based on deep learning[J]. Acta Meteor Sin，77(4)：715-727(in Chinese).

黄晓龙，徐晓莉，吴薇，等，2022. 基于DEM的四川省地面气象台站地形特征分析[J]. 高原山地气象研究，42(1)：135-142. Huang X L，Xu X L，Wu W，et al，2022. Analysis of terrain characteristics of meteorological stations in Sichuan Province based on DEM[J]. Plateau Mountain Meteor Res，42(1)：135-142(in Chinese).

康岚，刘炜桦，肖递祥，等，2018. 四川盆地一次极端大风天气过程成因及预报着眼点分析[J]. 气象，44(11)：1414-1423. Kang L，Liu W H，Xiao D X，et al，2018. Analysis on formation reason and forecast of an extreme gale in Sichuan Basin[J]. Meteor Mon，44(11)：1414-1423(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).

刘彬，邹灵宇，李晓鹏，等，2022. 云南雷暴大风天气的环境条件特征分析[J]. 气象，48(11)：1402-1417. Liu B，Zou L Y，Li X P，et al，2022. Analysis on environmental characteristics of thunderstorm gales in Yunnan[J]. Meteor Mon，48(11)：1402-1417(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).

孙全德，焦瑞莉，夏江江，等，2019. 基于机器学习的数值天气预报风速订正研究[J]. 气象，45(3)：426-436. Sun Q D，Jiao R L，Xia J J，et al，2019. Adjusting wind speed prediction of numerical weather forecast model based on machine learning methods[J]. Meteor Mon，45(3)：426-436(in Chinese).

田付友，郑永光，坚参扎西，等，2024a. 融合物理理解与模糊逻辑的分类强对流客观短期预报系统：(2)表现评估[J]. 气象，50(6)：649-660. Tian F Y，Zheng Y G，Jiancan Z X，et al，2024a. Forecasting system for short-term multi-category convective phenomena combining physical understanding and fuzzy logic Part Ⅱ：performance evaluation[J]. Meteor Mon，50(6)：649-660(in Chinese).

田付友，郑永光，孙建华，等，2024b. 融合物理理解与模糊逻辑的分类强对流客观短期预报系统：(1)系统构成[J]. 气象，50(5)：521-531. Tian F Y，Zheng Y G，Sun J H，et al，2024b. Forecasting system for short-term multi-category convective phenomena combining physical understanding and fuzzy logic Part Ⅰ：system construction[J]. Meteor Mon，50(5)：521-531(in Chinese).

田付友，郑永光，张涛，等，2015. 短时强降水诊断物理量敏感性的点对面检验[J]. 应用气象学报，26(4)：385-396. Tian F Y，Zheng Y G，Zhang T，et al，2015. Sensitivity analysis of short-duration heavy rainfall related diagnostic parameters with point-area verification[J]. J Appl Meteor Sci，26(4)：385-396(in Chinese).

王婷婷，朱文剑，姚凯，等，2023. 2021年吉林中部一次极端雷暴大风的中尺度成因分析[J]. 气象，49(10)：1215-1226. Wang T T，Zhu W J，Yao K，et al，2023. Mesoscale analysis of an extreme thunderstorm gale in central Jilin Province in 2021[J]. Meteor Mon，49(10)：1215-1226(in Chinese).

杨璐，韩丰，陈明轩，等，2018. 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，2018. Thunderstorm gale identification method based on support vector machine[J]. J Appl Meteor Sci，29(6)：680-689(in Chinese).

杨璐，王晓丽，宋林烨，等，2023. 基于阵风系数模型的百米级阵风客观预报算法研究[J]. 气象学报，81(1)：94-109. Yang L，Wang X

L，Song L Y，et al，2023. An algorithm for objective forecasting of gust winds at 100 m horizontal resolution based on a gust coefficient model [J]. Acta Meteor Sin，81 (1)：94-109 (in Chinese).

俞小鼎，2011. 基于构成要素的预报方法——配料法[J]. 气象，37(8)：913-918. Yu X D，2011. Ingredients based forecasting methodology[J]. Meteor Mon，37(8)：913-918(in Chinese).

张小玲，陶诗言，孙建华，2010. 基于“配料”的暴雨预报[J]. 大气科学，34(4)：754-766. Zhang X L，Tao S Y，Sun J H，2010. Ingredients-based heavy rainfall forecasting[J]. Chin J Atmos Sci，34(4)：754-766(in Chinese).

张小玲，杨波，盛杰，等，2018. 中国强对流天气预报业务发展[J]. 气象科技进展，8(3)：8-18. Zhang X L，Yang B，Sheng J，et al，2018. Development of operations on forecasting severe convective weather in China[J]. Adv Meteor Sci Technol，8(3)：8-18 (in Chinese).

郑永光，陶祖钰，俞小鼎，2017. 强对流天气预报的一些基本问题[J]. 气象，43(6)：641-652. Zheng Y G，Tao Z Y，Yu X D，2017. Some essential issues of severe convective weather forecasting[J]. Meteor Mon，43(6)：641-652(in Chinese).

郑永光，田付友，孟智勇，等，2016. “东方之星”客轮翻沉事件周边区域风灾现场调查与多尺度特征分析[J]. 气象，42(1)：1-13. Zheng Y G，Tian F Y，Meng Z Y，et al，2016. Survey and multi scale characteristics of wind damage caused by convective storms in the surrounding area of the capsizing accident of cruise ship “Dongfangzhixing”[J]. Meteor Mon，42(1)：1-13(in Chinese).

郑永光，周康辉，盛杰，等，2015. 强对流天气监测预报预警技术进展[J]. 应用气象学报，26(6)：641-657. Zheng Y G，Zhou K H，Sheng J，et al，2015. Advances in techniques of monitoring，forecasting and warning of severe convective weather[J]. J Appl Meteor Sci，26(6)：641-657(in Chinese).

周康辉，郑永光，韩雷，等，2021. 机器学习在强对流监测预报中的应用进展[J]. 气象，47(3)：274-289. Zhou K H，Zheng Y G，Han L，et al，2021. Advances in application of machine learning to severe convective weather monitoring and forecasting[J]. Meteor Mon，47(3)：274-289(in Chinese).

周康辉，郑永光，蓝渝，2016. 基于闪电数据的雷暴识别、追踪与外推方法[J]. 应用气象学报，27(2)：173-181. Zhou K H，Zheng Y G，Lan Y，2016. Flash cell identification，tracking and nowcasting with lightning data[J]. J Appl Meteor Sci，27(2)：173-181 (in Chinese).

周威，张武龙，康岚，等，2020. 基于 EC 模式闪电格点概率预报模型及应用[J]. 气象科技，48(6)：862-870. Zhou W，Zhang W L，Kang L，et al，2020. Application of lightning grid probability prediction products based on EC model[J]. Meteor Sci Technol，48(6)：862-870(in Chinese).

Collins W，Tissot P，2015. An artificial neural network model to predict thunderstorms within 400 km^2 South Texas domains[J]. Meteor Appl，22(3)：650-665.

Czernecki B，Taszarek M，Marosz M，et al，2019. Application of machine learning to large hail prediction-the importance of radar reflectivity，lightning occurrence and convective parameters derived from ERA5[J]. Atmos Res，227：249-262.

Gagne Ⅱ D J，McGovern A，Brotzge J，2009. Classification of convective areas using decision trees[J]. J Atmos Oceanic Technol，26(7)：1341-1353.

Jungo P，Goyette S，Beniston M，2002. Daily wind gust speed probabilities over Switzerland according to three types of synoptic circulation[J]. Int J Climatol，22(4)：485-499.

Kim M，Im J，Park H，et al，2017. Detection of tropical overshooting cloud tops using Himawari-8 imagery[J]. Remote Sens，9(7)：685.

（本文责编：张芳）