

韩乐琼,何晓凤,张雪松,等,2023. 强风事件识别及预报订正方法研究[J]. 气象,49(12):1542-1552. Han L Q, He X F, Zhang X S, et al, 2023. Study of approach in identification and modification of gale event forecast[J]. Meteor Mon, 49(12):1542-1552 (in Chinese).

强风事件识别及预报订正方法研究*

韩乐琼^{1,2} 何晓凤² 张雪松³ 肖擎曜^{1,2} 陈 笑^{1,2}

1 北京玖天气象科技有限公司,北京 100081

2 华风气象传媒集团,北京 100081

3 中广核风电有限公司,北京 100070

提 要: 以如东海上风电场升压站激光雷达测风资料为基础,提出了一种强风事件识别方法,设计并比较了三种预报强风事件识别方案。基于决策树和一元线性回归方法,分别开展了针对强风事件的订正方法研究。结果发现:三种预报强风事件识别方案中,等分位阈值方案明显更优,事件命中率达到 76.1%,匹配时长命中率达到 87.6%;采用消偏阈值方案和等分位阈值方案预报的强风事件时长会更接近观测强风事件时长;等分位阈值方案识别的事件基本可以覆盖到各次观测强风事件的全程;两个订正模型相对于模式预报都有一定提升与改进,其中决策树比一元线性回归模型更优,其平均绝对误差、相对误差和均方根误差明显更小。

关键词: 强风事件,数值预报,订正,决策树,一元线性回归

中图分类号: P456, P49

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2023.051501

Study of Approach in Identification and Modification of Gale Event Forecast

HAN Leqiong^{1,2} HE Xiaofeng² ZHANG Xuesong³ XIAO Qingyao^{1,2} CHEN Xiao^{1,2}

1 Beijing Jiutian Meteorology Science and Technology Co., Ltd, Beijing 100081

2 Huafeng Meteorological Media Group Co., Ltd., Beijing 100081

3 China Guangdong Nuclear Wind Power Co., Ltd., Beijing 100070

Abstract: Based on the wind speed observation data of Marine Booster Station in Rudong Wind Farm, this paper proposes a method of identifying gale event. Three identification schemes of gale event forecast are developed and compared through the determination of the crucial parameters. Then, based on the decision tree method and a single linear regression method, the correction methods for gale events are studied. The results show that the gale event forecast of equal cumulative frequency scheme is superior to other schemes, having the hit rate of 76.1% and the hit rate of matching duration of 87.6%. The duration of gale event forecast of eliminating deviation and equal cumulative frequency schemes are more in agreement with the observation data. Besides, the equal cumulative frequency scheme can cover the duration of every observation gale event, so it is good for proposing the beginning and end time of gale warning. The above-mentioned two correction methods can improve the forecast performance to a certain extent. However, the improvement done by the decision tree method is more obvious, for it can significantly reduce the MAE, RE, RMSE.

* 国家重点研发计划(2018YFC1507804)、中广核尖峰计划项目(001-GN-A-2021-SN-0239)共同资助

2022 年 9 月 29 日收稿; 2023 年 10 月 26 日收修定稿

第一作者:韩乐琼,主要从事数值产品的订正应用. E-mail:641822468@qq.com

通讯作者:何晓凤,主要从事数值预报解释应用及专业气象服务. E-mail:hexf@cma.gov.cn

Key words: gale event, numerical weather prediction (NWP), correction, decision tree, a single linear regression

引 言

风是对多行业生产安全有直接影响的基本气象要素,准确的近地层风速预报对企业提质增效保安全具有重要的作用。随着数值天气预报(numerical weather prediction, NWP)技术的快速发展,基于 NWP 的风速预报产品越来越多地应用于各行各业的生产决策、重大活动的气象保障中。在港口气象灾害风险评估方面,一些研究采用模式风场预报产品作为驱动,模拟了风暴潮及其增水过程的水位变化,并基于此开展了灾害性天气下港口的气象灾害风险评估(曹越男等,2014;闫丽凤等,2008)。针对易受大风天气影响的重大室外赛事活动,国内外基于不同数值预报产品开展了较多的赛场风速预报及订正技术研究,致力于提高区域模式的风速预报,为重大赛事提供气象保障(王令等,2008;陈超君等,2012;Kiktev et al,2015;杨璐等,2022;邱贵强等,2023;刘郁钰等,2023)。

随着新能源行业的兴起,NWP 在风能资源精细化评估和气候模拟中也开始发挥重要作用,国内外在该方向均有大量研究成果与工程实践。一些学者基于数值预报产品研究了我国近海风能资源高分辨率模拟方法,分析了近海风能资源分布状况,部分研究侧重于评估 NWP 产品在不同区域的风速预报能力,为区域风能模式预报的业务化应用奠定了基础(周荣卫等,2010;何晓凤等,2014;申华羽等,2020);国外学者也利用 NWP 产品采用动力降尺度方法或着眼于提高某类天气事件的预报能力来评估局地风能资源(Gioli et al,2014;Storm et al,2009);还有许多学者分别基于不同 NWP 产品开展了在各地风能预报中的实践工作(李艳等,2009;Traiteur et al,2012;Cheng and Steenburgh,2015;刘郁钰等,2019;Liu et al,2021)。

虽然现阶段不同的 NWP 风速产品广泛应用于各行各业并起到了一些效果,但是因 NWP 的初值准确性、物理过程参数化方案的近似性、求解过程的近似和截断处理、地形与实际情况的偏差等,均会导致近地层风速的预报结果存在较大偏差,为企业用户做决策带来较大的困扰。鉴于这种情况,多位学

者开展了基于局地观测数据的预报风速订正技术研究,部分学者利用传统线性方法进行统计订正(常蕊等,2013;肖擎曜等,2017),通过多种试验设计提高了局地风速的预报能力;也有部分学者通过对同一风电场运用多种方法做订正试验,研究传统线性方法和机器学习方法的表现差异(江滢等,2013;Bédard et al,2013)。此外,还有许多学者采用谐波分析、向量机方法、K 近邻法、卡尔曼滤波、均生函数法、线性回归法、相似分析法等方法(戚双斌等,2009;程兴宏等,2012;Lin et al,2015;Zjavka,2015;丁煌等,2013;张铁军等,2018;王彬滨等,2016;王在文等,2019;杨程等,2019)开展了风速预报订正方法研究与应用。这些工作显著减小了预报误差中的系统性偏差,对预报产品精度的提升起到了积极作用,但是对于降低强风时段的预报误差效果有限。

此外,不同的企业生产场景对风速有不同的阈值要求,比如风速超过 $10 \text{ m} \cdot \text{s}^{-1}$ 时可能威胁海上施工的作业安全,风速超过 $25 \text{ m} \cdot \text{s}^{-1}$ 时部分风力发电机组需要停机等。对特定企业而言,最关注的是影响其生产安全的风速区间内的风速预报能力,当预报风速值接近该阈值时,由于预报不确定性的存在,企业往往最难决策,因此对超过阈值的事件出现与否的预报准确性提出了很高的要求,而对其他风速段预报准确性的要求反而并不那么高。这种情况下,需要面向特定生产场景定义普适性的强风事件,并将这类事件从预报序列中剥离出来有针对性的进行订正,才能更好地将其应用于企业实际生产中。

本文首先基于海上风电安全施工对风速的具体要求,应用滑动平均函数,提出了一种影响海上施工安全的持续性强风事件识别方法,并在充分评估、分析风速预报产品强风预报能力的基础上,结合实况与预报数据,基于所提出的强风识别方法的关键参数,设计了三种预报识别方案,并评估比较了三种预报强风事件,随后基于决策树和一元线性回归方法,分别开展了针对上述强风事件的订正方法研究。

1 资 料

本文观测数据源自江苏省南通市如东县近海某

风电场激光雷达测风资料(以下简称如东站),激光雷达安装于该风电场的海上升压站,其距离海平面 35 m,选用激光雷达 10 m 高度层风速观测序列,故观测点距离平均海面的实际高度约为 45 m。观测资料采集期为 2018 年 1 月 1 日至 12 月 31 日,为了与模式预报输出风速时间对应,提取了每个整点时刻的 10 min 平均风速观测数据,并对观测风速做极值质量控制、僵值质量控制和均一性检验等,最后根据检验结果剔除了异常值和缺测数据。

本文预报资料来自 GDFS(Grid Data Forecast System)(陈笑等,2018)。GDFS 是基于多模式预报产品的集成预报系统(以下简称模式),其空间分辨率为 $0.025^\circ \times 0.025^\circ$,时间分辨率为 1 h,预报时效为 72 h,每天起报 2 次。该模式基于全国 7 万多个自动站实时观测资料和 ECMWF、CMA-MESO、RMAPS_Wind 等多个业务化数值预报系统的预报产品,通过动态误差权重法实现了集成预报。本文直接采用该模式每天 08 时起报的 50 m 高度层的前 24 h 预报风速数据开展研究。鉴于海上的下垫面较光滑,风速的空间变化在一定范围内较小,故未对格点数据做插值,而是直接选用距离观测地点最近的模式格点数据,并将观测资料与该格点的预报数据进行逐时匹配。

2 强风事件的定义与识别方法研究

2.1 强风事件的定义

根据海上施工作业经验, $10 \text{ m} \cdot \text{s}^{-1}$ 是影响施工安全的重要风速阈值,连续超过该阈值的强风过程会导致施工无法进行,因此认为连续几个小时平均风速超过 $10 \text{ m} \cdot \text{s}^{-1}$ 即为一次海上安全施工高影响持续性强风事件,通过以下具体流程识别强风事件。

首先对风速序列做滑动窗口为 5 h 的滑动平均,公式如下:

$$\hat{x}_i = \frac{\sum_{j=-2}^2 x_{i+j}}{5} \quad (1)$$

式中: \hat{x}_i 为经过滑动平均后 i 时刻的风速, i 为时序, $i=3,4,\dots,n-2$, n 为样本数; x_{i+j} 为 $i+j$ 时刻的原始风速, j 为时间增减量, $j=-2,-1,0,1,2$ 。

对滑动平均后得到的新序列做判断,当风速超

过给定风速阈值(针对本文研究的影响海上施工安全的持续性强风事件,观测强风事件阈值为 $10 \text{ m} \cdot \text{s}^{-1}$),定义为一个强风时次。当强风时次连续出现大于等于 3 h,则记为一次强风事件。第一个强风时次为强风事件的开始时间,最后一个强风时次为结束时间。考虑到两次强风事件相距较近时,小风窗口太短无法完成施工,因此将后一次事件开始时间与前一次事件结束时间的差小于等于 3 h 的两次强风事件合并为一次强风事件。

2.2 强风事件识别及特征分析

为开展本文研究工作,首先需在风速的观测和预报序列中识别出强风事件,再开展特征分析和后续订正技术研究。

首先严格根据上文定义,从观测数据中进行海上观测强风事件识别,结果显示 2018 年该海上风电场区域共出现观测强风事件 71 次,总时长为 1374 h,事件平均时长约为 19.35 h。

从预报风速序列中识别强风事件时,首先需要了解模式预报风速的基本情况。统计结果显示,如东站 2018 年风速观测均值为 $7.01 \text{ m} \cdot \text{s}^{-1}$,预报均值为 $6.29 \text{ m} \cdot \text{s}^{-1}$,预报比观测系统性偏小 $0.72 \text{ m} \cdot \text{s}^{-1}$,预报与观测的相关性较高,相关系数达到 0.86,表明预报能较好地反映出实际风速的波动转折。从逐时预报误差的频率和累计频率统计(图 1a)可以看出,如东站预报偏小的频率明显大于预报偏大的频率,预报偏小占 67.4%,频率峰值对应的误差范围为偏小 $0.5 \sim 1 \text{ m} \cdot \text{s}^{-1}$,偏小 $2 \text{ m} \cdot \text{s}^{-1}$ 以上达 20.6%。针对本文研究侧重的强风段,筛选出观测风速大于等于 $10 \text{ m} \cdot \text{s}^{-1}$ 的相对强风时次,提取出对应时次的预报,可计算得到强风段误差分布情况(图 1b),强风段预报偏小频率达 94.1%,频率峰值对应的误差范围偏小 $2 \sim 2.5 \text{ m} \cdot \text{s}^{-1}$,且偏小 $2 \text{ m} \cdot \text{s}^{-1}$ 以上的达到 58.9%,即强风时段预报偏小的情况更为明显。

基于模式预报风速总体偏小且强风时段显著偏小的特征,对预报风速序列进行强风事件识别时,开展了三种敏感性识别方案的对比,以得到与实况强风事件匹配度更高的预报强风事件,这对于后续开展订正非常重要。模式阈值方案(以下简称模式法):基于上文定义的强风事件阈值($10 \text{ m} \cdot \text{s}^{-1}$)和式(1)对预报风速序列进行直接识别,将该方案识别结果称为模式强风事件;消偏阈值方案(以下简称消

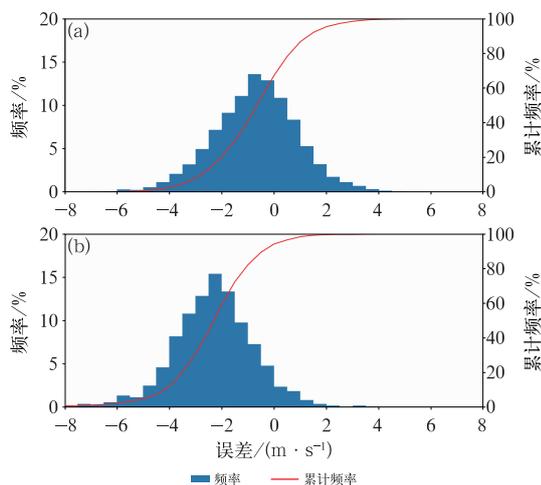


图 1 2018 年如东站 (a) 所有时段和 (b) 观测风速 $\geq 10 \text{ m} \cdot \text{s}^{-1}$ 时段的预报与观测风速的误差频率和累计频率分布

Fig. 1 Distribution of frequency and cumulative frequency of wind speed error forecast and observed during (a) all time periods and (b) periods with observed wind speed $\geq 10 \text{ m} \cdot \text{s}^{-1}$ at Rudong Station in 2018

偏法):将模式预报的原始风速进行系统性偏差简单订正,即对逐时预报均加上预报与观测的误差均值 ($0.72 \text{ m} \cdot \text{s}^{-1}$)后,再采用 $10 \text{ m} \cdot \text{s}^{-1}$ 阈值和式(1)做强风事件识别,将该方案识别结果称为消偏强风事件;等分位阈值方案(以下简称等分位法):考虑到观测和预报的强风时段在各自的整个时段内占比是近似的,首先对观测的滑动平均序列由小到大排序,计算出观测阈值 $10 \text{ m} \cdot \text{s}^{-1}$ 处的累计频率值(82.4%),再提取模式预报风速的滑动平均序列由小到大排列后累计频率 82.4% 处对应风速值($8.5 \text{ m} \cdot \text{s}^{-1}$),将该值作为预报强风事件阈值(如图 2 所示),再结合式(1)进行强风事件识别,将该方案识别结果称为等分位强风事件。

强风事件评估方法:对持续时间不超过 20 h 的观测强风事件,其时间范围内任意时刻识别到预报强风事件,则记为一次命中,否则记为一次漏报;若该次观测强风事件持续时间超过 20 h,则要求二者的交集时长大于等于 5 h 才记为一次命中,否则记为一次漏报。将与观测强风事件无交集的预报强风事件记为一次空报。事件命中率定义为命中次数除以观测强风事件总数,以百分率表示。各方案预报与观测事件逐时匹配时长:即预报与观测同为大风事件的时刻累计值。其匹配时长命中率定义为匹配

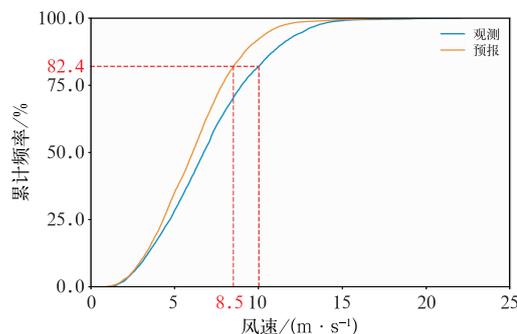


图 2 2018 年如东站观测和预报风速滑动平均序列的累计频率分布

Fig. 2 Cumulative frequency of observed and forecast wind speeds after moving average at Rudong Station in 2018

时长除以总的观测大风事件时长(1374 h)。

基于观测强风事件和三种识别方案得到的预报强风事件开展了效果评估,由表 1 可见:模式法与观测强风事件的匹配较差,命中事件仅为 28 次,命中率为 39.4%,漏报较多;消偏法将命中率提高到了 54.9%;等分位法明显更优,命中事件为 54 次,相比模式法增加命中 26 次,命中率提高到 76.1%。但是随着命中率的提高,消偏法和等分位法的空报也略有增加:模式法空报为 1 次,消偏法和等分位法空报分别增加到了 8 次和 9 次。从事件总时长来看,模式法、消偏法和等分位法的总时长分别为 692、1039 和 1401 h,与 1 年内的观测强风事件总时长相比分别表现为偏少 682 h、偏少 335 h 和偏多 27 h,等分位法识别的强风事件总时长明显更接近观测。

此外,对各方案预报与观测事件逐时匹配时长对比得出:模式法为 625 h,消偏法为 911 h,等分位法为 1203 h,匹配时长命中率分别为 45.5%、66.3%、87.6%,即对观测事件漏报的时长分别为 749、463 和 171 h。而各方案空报事件的时长:模式法为 14 h,消偏法为 47 h,等分位法为 87 h。可见消偏法和等分位法识别的强风事件虽然相比模式法增加了约 2 d 的空报,但是匹配时长分别增加了

表 1 2018 年如东站三种方案识别得到的预报强风事件评估

Table 1 Assessment of forecast gale events identified by three schemes at Rudong Station in 2018

预报强风事件	模式法	消偏法	等分位法
漏报/次	43	32	17
命中/次	28	39	54
命中率/%	39.4	54.9	76.1
空报/次	1	8	9

约 12 d 和 24 d。

针对各月的强风事件,分析后发现各方案空报和漏报情况在同季节的各月间近似,因此从四个季节中分别选取了季节内观测强风事件最多的月份构成了典型月份(分别为 4、7、10、12 月)做进一步分析。图 3 给出了四个典型月份中,观测强风事件及三种方案得到的预报强风事件的各月分布情况。由图 3 可见,虽然消偏法和等分位法识别的预报强风事件会增加少量空报,但是二者能明显减少模式漏报;模式法识别的强风事件时长相对于同期观测普遍偏短,采用消偏法和等分位法识别的预报强风事件时长会更接近观测;消偏法容易漏报影响时间较短的强风事件,而采用等分位法能有效提高这类影响时间较短的强风事件的命中率;并且对比预报和观测的各次事件的起止时间,等分位法基本可以覆盖到各次强风事件的全程。因此等分位法识别的预

报强风事件起止时间可为海上施工作业灾害预警的发布和解除提供有效参考。

在四个典型月份中 7 月和 10 月存在空报事件。图 4b 给出了 6 月 30 日 00 时至 7 月 31 日 23 时观测和预报时序,可看到 7 月 1 日的空报,是因为 6 月 30 日至 7 月 1 日模式预报的大风段滞后于观测大风,观测大风段结束于 6 月 30 日 14 时,而预报风速 14 时才开始高于 $8 \text{ m} \cdot \text{s}^{-1}$,在采用消偏法和等分位法后,等分位法的预报大风事件开始于 12 时,命中了观测大风事件的后半段,而消偏法的预报大风事件开始于 16 时,没有命中大风事件;但是因为预报风速滞后的原因,等分位法预报大风事件的结束时间为 7 月 1 日 01 时,导致了 1 日的短暂空报。7 日的空报则是该段预报与观测反向,表现为预报风速在波峰而观测相对在波谷,预报的相对大风时段提前了一些,导致了预报风速的整体偏大。其中

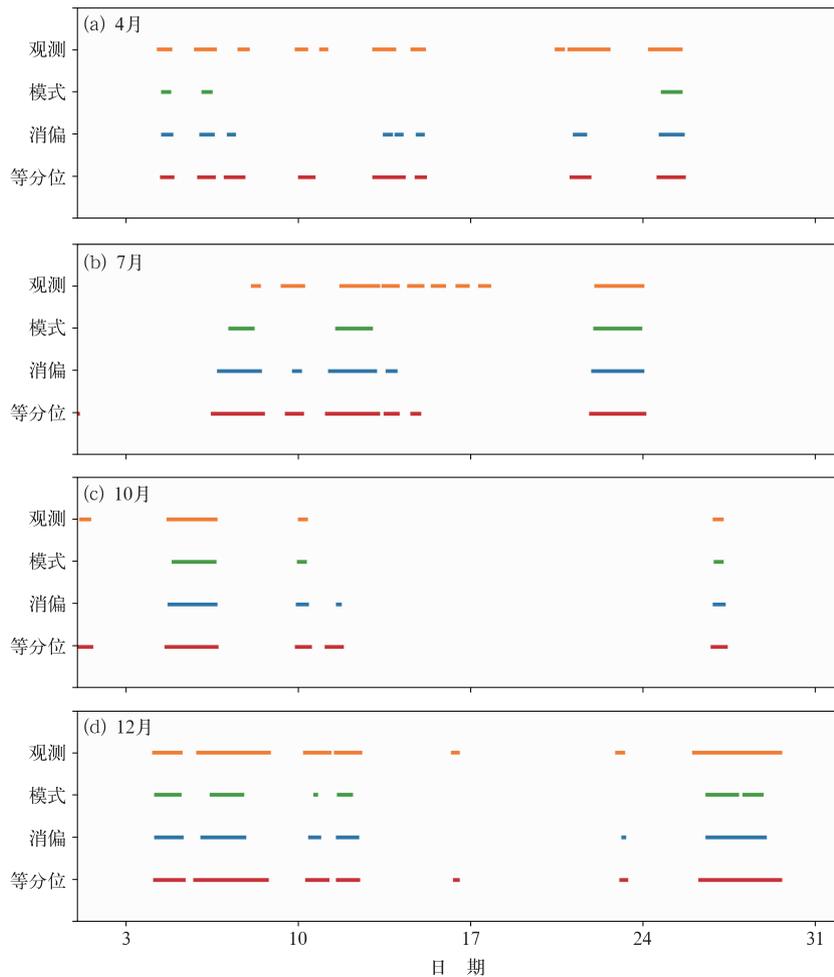


图 3 2018 年四个典型月份如东观测及三种方案预报的强风事件分布

Fig. 3 Distribution of observed and forecast gale events by three schemes at Rudong Station in 2018

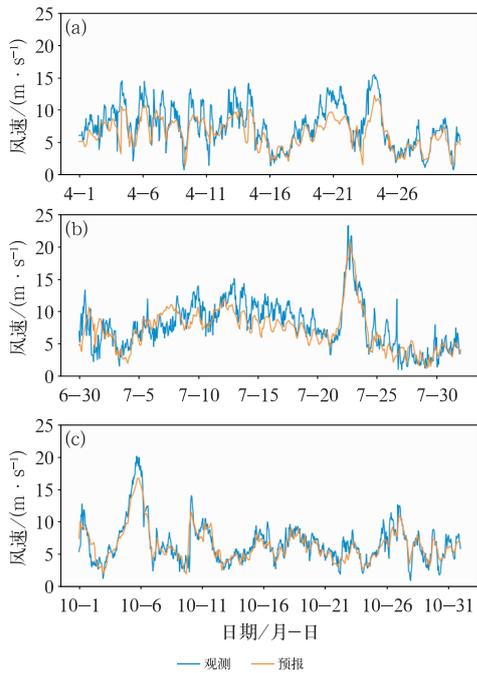


图 4 2018 年(a)4 月、(b)7 月和(c)10 月如东站的观测和预报风速

Fig. 4 The observed and forecast wind speeds at Rudong Station in (a) April, (b) July, (c) October of 2018

7 日的观测平均风速为 $7.19 \text{ m} \cdot \text{s}^{-1}$, 而预报平均为 $10.23 \text{ m} \cdot \text{s}^{-1}$, 预报偏高了 $3.04 \text{ m} \cdot \text{s}^{-1}$, 这导致三个方法均出现了不同时间长度的空报。对 10 月的观测和预报时序图分析(图 4c), 11 日的预报风速非常接近观测风速, 预报和观测风速平均分别为 $8.30 \text{ m} \cdot \text{s}^{-1}$ 和 $8.37 \text{ m} \cdot \text{s}^{-1}$, 预报风速略偏小, 消偏法加上误差均值后风速增加, 从而出现了短暂空报, 而等分位法因为降低了阈值出现了相对较长的空报。上述两种空报的成因共性不强, 且两类空报事件在整个资料时间段内相对较少, 因此难以直接通过总结模式风速预报的特征来消除空报。在实际业务运用时, 针对此类大风事件, 可能需要考虑当前天气型的影响, 做人工修正。

消偏法和等分位法预报的强风事件虽能有效增加命中, 减少漏报, 但是由于模式原始风速预报的固有问题, 仍然存在一些强风事件无法被预报出来(如图 3b, 7 月 15—17 日的观测强风事件), 且存在一些预报强风事件相对观测强风事件时长偏短的现象(如图 3a, 4 月 20—23 日的观测强风事件)。分析对应月份的时序图(图 4a, 4b), 可看到上述提到漏报时段的风速预报波动与观测相似, 但是峰值风速预

报明显偏小, 其中 4 月 20 日、21 日的单日均值分别偏小 $2.6 \text{ m} \cdot \text{s}^{-1}$ 和 $2.9 \text{ m} \cdot \text{s}^{-1}$, 7 月 15 日、16 日、17 日的单日均值分别偏小 2.11 、 2.6 和 $2.12 \text{ m} \cdot \text{s}^{-1}$, 这种显著偏小导致了预报事件的时长偏短或漏报的情况。

3 强风事件强度订正方法研究

基于三种预报强风事件识别方案研究的结果分析, 运用等分位法识别到的强风事件从开始时间、结束时间、持续时长、命中率、时长匹配等各方面都最接近观测强风事件, 其起止时间可为海上施工作业灾害预警的发布和解除提供有效参考。因此, 下文基于等分位法识别的预报强风事件开展预报强风事件的强度订正方法研究。

首先根据等分位强风事件的时间段提取观测及模式预报风速数据, 并按照事件发生的先后顺序依次重构成新的观测强风序列和预报强风序列(以下简称强风序列集)。考虑到平均风速是一个连续性的变量, 某时刻的风速与其之前和之后时刻的风速均有密切关联, 对强风序列集采用如下方法重构样本集:

对强风序列集内某时刻 i 的观测风速 y_i , 找到对应时刻前后 6 h 内($i-6, i-5, \dots, i, i+1, \dots, i+6$)共 13 个预报风速构成预报因子集 $x_i^{(k)}$, 其中 $i=1, 2, \dots, n$ 为时序(n 也为样本总数); $k=1, 2, \dots, 13$ 为因子序号, $k=1$ 时, 对应 $i-6$ 时刻的预报风速, $k=13$ 时, 对应 $i+6$ 时刻的预报风速, 以此类推。将上述变量合并即可构建出样本集 $D = \{(x_1^{(k)}, y_1), (x_2^{(k)}, y_2), \dots, (x_i^{(k)}, y_i), \dots, (x_n^{(k)}, y_n)\}$, 其中 $x_i^{(k)} = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(13)})$ 。随后选取样本集前 80% 时序的数据作为训练样本, 剩下的 20% 时序数据作为独立检验样本, 再分别采用了一元线性回归和决策树两种订正方法基于训练样本段构建了订正模型, 并用独立检验样本段的数据进行订正和检验。

3.1 一元线性回归

一元线性回归是通过寻找自变量与因变量的最小离差平方和来得到数据最优函数解的方法(黄嘉佑, 2000)。在本研究中, 将因变量设置为观测风速 y_i , 通过对前述样本集 $x_i^{(k)}$ 与 y_i 计算线性相关后, 选择相关性最高的 $x_i^{(8)}$ (其相关系数为 0.72) 做为

自变量,假定二者存在下述的一元线性统计关系式[式(2),其为回归理想模型],利用样本集中的数据对该关系式进行估计,得到的估计方程即为一元线性回归方程[式(3)]。

$$y_i = \beta_0 + \beta_1 x_i^{(8)} + e_i \quad (2)$$

$$\hat{y}_i = b_0 + b_1 x_i^{(8)} \quad (3)$$

式中: y_i 为*i*时刻的观测风速, $i=1,2,\dots,n$, n 为建模时段样本数, β_0 为回归常数, β_1 为因子系数, $x_i^{(8)}$ 为预报因子, e_i 为相互独立的随机变量。 \hat{y}_i 为*i*时刻的订正风速, b_0 和 b_1 分别为 β_0 和 β_1 的估计。

式(3)需要满足使得全部的观测风速与订正风速的离差平方和(Q)达到最小,即:

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

将式(3)代入式(4)并展开,根据最小二乘法即可解出方程组的系数。本研究中解出来的一元线性方程如下:

$$\hat{y}_i = 1.119 + 0.98 x_i^{(8)} \quad (5)$$

3.2 决策树

决策树是一种树结构,是近年来机器学习方法中基本的分类与回归方法。回归决策树主要指CART(classification and regression tree)算法,其由Breiman et al(1984)提出。针对连续值,其是通过误差平方和来度量划分叶节点的一种算法。算法本质是将样本时段的特征空间根据当前误差平方和最小的条件来划分成左右两个结点,两个结点分别以同一个特征变量的值作条件划分,依次对划分出的每个结点的结果接着进行划分,即得到若干个单元,每一个划分单元有一个特定的输出。在检验时段的数据应用中,只要按照特征将其归到某个单元,便能得到对应的订正输出值。

本研究针对上文中定义的训练数据集 $D = \{(x_1^{(k)}, y_1), (x_2^{(k)}, y_2), \dots, (x_i^{(k)}, y_i), \dots, (x_n^{(k)}, y_n)\}$,其中 $x_i^{(k)}$ 为不同的特征变量, k 的取值为1,2,3, \dots ,13,分别表示对应观测时刻往前6h至往后6h的预报风速。在决策树CART算法中对上述由所有特征变量组成的特征空间采用遍历当前集合中所有特征变量的所有取值的方法进行划分,根据误差平方和最小化原则选择其中最优的一个作为切分点。如第一轮划分,对训练集中的某一个特征变量 $x_i^{(k)}$ 给出一个*s*值,将 $x_i^{(k)}$ 作为切分变量,*s*作为切分点,即可将*D*根据*s*划分为两个区域 $R_1(k, s)$ 和

$R_2(k, s)$,即 $R_1(k, s) = \{(x_i^{(k)}, y_i) | x_i^{(k)} \leq s\}$, $R_2(k, s) = \{(x_i^{(k)}, y_i) | x_i^{(k)} > s\}$ 。该*s*的取值要使得当前两个区域的误差平方和最小,即满足式(6):

$$SE_k = \min_{k,s} \left[\sum_{R_1(k,s)} (y_i - \bar{c}_1)^2 + \sum_{R_2(k,s)} (y_i - \bar{c}_2)^2 \right] \quad (6)$$

式中: \bar{c}_1 和 \bar{c}_2 分别为 $R_1(k, s)$ 和 $R_2(k, s)$ 内 y_i 的均值, SE_k 为以特征变量 $x_i^{(k)}$ 为分区时得到的最小误差平方和。对每个特征变量计算 SE_k ,并记录下对应的切分点*s*的值,再从所有的 SE_k 中找出最小的值,即能求得本次划分过程中使式(6)结果最小时的*s*值和对应的特征变量 $x_i^{(k)}$,做为当前的最优切分点*s*和最优切分变量 $x_i^{(k)}$,同时能根据 $x_i^{(k)}$ 和*s*的取值计算得到确定的 \bar{c}_1 和 \bar{c}_2 。此时可将训练集*D*看作父节点,而 $R_1(k, s)$ 和 $R_2(k, s)$ 为其叶子节点,树的深度为一层。随后以 $R_1(k, s)$ 和 $R_2(k, s)$ 作为新的训练集分别继续采用上式进行遍历划分求解,不断生成新的叶子节点,直到所有样本切割完或达到给定的树的最大深度。其中,样本切割完的定义为当前叶子节点训练集内的样本个数小于等于给定的最小样本个数,或当前叶子节点的最小误差平方和与进行划分后的最小误差平方和相差不大。这样就生成了一颗回归树,其属于不同特征变量区间范围内(即最后一层叶子节点的样本集内)所对应的均值 \bar{C} ,即为决策树模型的订正结果。

在本研究中,最终生成的决策树模型深度为8,最小叶子节点样本个数为1个,第一个根节点以特征变量 $x_i^{(8)}$ (订正时刻往后1h的预报风速)为切分变量,切分点*s*为 $13.9 \text{ m} \cdot \text{s}^{-1}$,当 $x_i^{(8)} > s$ 时,就进入右节点,接着以 $x_i^{(11)}$ (即订正时刻往后4h的预报风速)做切分变量,即以切分点*s*为 $15.1 \text{ m} \cdot \text{s}^{-1}$ 做新一轮判断;若 $x_i^{(8)} \leq s$,进入左节点,以 $x_i^{(7)}$ (即订正时刻的预报风速)为切分变量,即以切分点*s*为 $10.4 \text{ m} \cdot \text{s}^{-1}$ 做新一轮判断。每次判断比较会得到左右两个分支节点,根据当前的判断结果选择对应分支的叶子节点,再根据叶子节点上新的切分变量和切分点继续判断,一直走到底层叶子节点,即可得到该节点样本均值 \bar{C} 作为其订正结果。

本文的决策树回归模型其本质可看作是模型根据输入预报的多特征变量的风速与对应的观测风速,总结每一次输入的13个风速预报点的变化曲线与对应观测的特征,通过对各个特征变量的阈值划分将预报风速划分为不同的天气型,并给出了判断

不同天气型的逐级关键切分变量和切分点,因此在读入新的预报风速序列时,可根据预报风速值确定对应天气型,最终得到该类天气型下的订正结果。

3.3 两种订正方法性能评估与分析

3.3.1 评估指标

针对两种订正方法在强风事件中订正性能的评估,整体定量评估方面采用平均绝对误差(MAE)、相对误差(RE)、均方根误差(RMSE)三个方法定义检验指标,如式(7)~式(9):

$$MAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (7)$$

$$RE = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{\sum_{i=1}^n y_i} \times 100\% \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (9)$$

式中: y_i 和 \hat{y}_i 分别为观测风速、模式预报(或订正)风速, i 为序号, n 为样本总数。

3.3.2 综合评价

上述独立检验样本时段包含 300 个样本,其中观测风速均值为 $11.91 \text{ m} \cdot \text{s}^{-1}$,模式预报风速均值为 $10.15 \text{ m} \cdot \text{s}^{-1}$ 。经过一元线性回归和决策树订正后的风速均值分别为 $10.98 \text{ m} \cdot \text{s}^{-1}$ 和 $11.23 \text{ m} \cdot \text{s}^{-1}$ 。可见两种订正方法均能有效减少均值误差,使得订正风速均值更接近观测均值。由表 2 的模式预报及两种订正方法的风速检验结果可见,MAE 从模式预报的 $1.97 \text{ m} \cdot \text{s}^{-1}$ 变为 $1.43 \text{ m} \cdot \text{s}^{-1}$ (一元线性回归) 和 $1.31 \text{ m} \cdot \text{s}^{-1}$ (决策树),分别减小了 $0.54 \text{ m} \cdot \text{s}^{-1}$ 和 $0.66 \text{ m} \cdot \text{s}^{-1}$;订正后 RE 虽然仍然偏小,但是偏小程度明显降低,从模式预报的 -14.7% 变为 -7.81% (一元线性回归) 和 -5.71% (决策树);此外, RMSE 也明显减小。无论是从 MAE、RE 或 RMSE 来看,经过两种订正方法订正后的风速预报效果相对于模式风速预报效果都有一定提升与改进,其中决策树模型比一元线性回归模型更优。

从整个检验样本时段的时序图来看(图 5),在每个强风事件的波峰处,模式预报的风速均是明显偏小于观测的,经过两种方法订正后,风速均有一定增加,且基本在每个强风事件的波峰处一元线性回

归法订正风速的误差均高于决策树订正风速的误差,比观测风速偏小相对较多。

部分波峰处(如序号 1~15 和 94~100)决策树订正后风速偏大于观测,是因为建模样本相对较少,供模型学习分类的历史事件不足,导致一些天气型未学习到,将输入的新时序作为类似于 7 月 12—14 日观测出现较长的强风事件中,对预报风速明显偏小的天气型进行订正,导致订正相对实况偏大。

3.3.3 最强强风事件评价

从检验时段中,以观测风速的最大值为条件筛选出了最强强风事件(2018 年 11 月 16 日 05 时至 17 日 05 时的强风事件)。图 6 给出了这次强风事件的时序图,该次强风事件因为东亚大槽东移,地面冷锋南压,带来一股股强冷空气形成大风降温事件。可以看到 16 日 05—10 时风速较大,为 $12 \sim 18 \text{ m} \cdot \text{s}^{-1}$,此为该次事件的第一个波峰,随后风速逐渐回落到 $10 \text{ m} \cdot \text{s}^{-1}$ 左右再次增大,18—22 时为第二个波峰,风速约为 $12 \sim 14 \text{ m} \cdot \text{s}^{-1}$,此后回落到最低 $7 \text{ m} \cdot \text{s}^{-1}$ 左右。模式预报出了第一个波峰,但是模

表 2 2018 年如东站的模式原始预报和两种订正方法的效果评估

Table 2 The effect comparison of forecasted by model and corrected wind speeds by two methods at Rudong Station in 2018

检验指标	模式预报	一元线性回归	决策树
MAE/($\text{m} \cdot \text{s}^{-1}$)	1.97	1.43	1.31
RE/%	-14.7	-7.81	-5.71
RMSE/($\text{m} \cdot \text{s}^{-1}$)	2.36	1.79	1.72

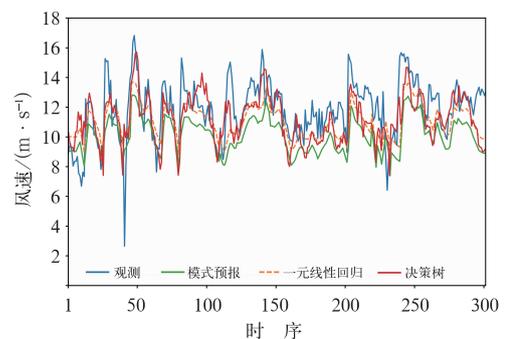


图 5 2018 年如东强风序列集检验时段观测、模式预报、两种订正方法的风速时序变化
Fig. 5 Temporal variation of observed, forecasted by model and corrected by two methods wind speeds in gale event test periods at Rudong Station in 2018

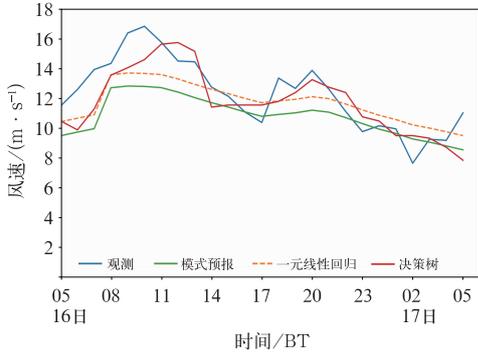


图 6 2018 年 11 月 16 日 05 时至 17 日 05 时如东站的观测、模式预报两种订正方法的风速对比

Fig. 6 Comparison of observed, forecasted by model and corrected by two methods wind speeds at Rudong Station from 05:00 BT 16 to 05:00 BT 17 November 2018

式预报的波峰风速明显小于实况,可能是因为该观测站位于近海,从陆地到海洋的下垫面变化使得风力增强,而模式对地形变化的刻画能力不足导致了误差,且模式对第二个波峰几乎没有体现,一元线性回归订正后也仅仅是增加了风速值,整体形态和模式预报一致,没有预报出第二个波峰,而决策树订正后的风速序列能明显看出两个波峰,如上文所述,决策树可看作是通过建模时段输入的预报风速序列学习划分不同天气型,从而根据新的预报风速序列判断对应天气型的过程,此处决策树模型学习到了出现双波峰的天气型,因此增大了第二个波峰处的订正风速,从而订正出第二个波峰。

此外,从这次事件的风速最大值来看:观测风速最大值为 $16.81 \text{ m} \cdot \text{s}^{-1}$ (7 级),模式预报风速为 $12.82 \text{ m} \cdot \text{s}^{-1}$ (6 级),一元线性回归订正风速为 $13.68 \text{ m} \cdot \text{s}^{-1}$ (6 级),决策树订正风速为 $15.62 \text{ m} \cdot \text{s}^{-1}$ (7 级),模式预报和一元线性回归订正的风速均明显低于观测风速,误差分别为 $3.99 \text{ m} \cdot \text{s}^{-1}$ 和 $3.13 \text{ m} \cdot \text{s}^{-1}$,而决策树订正的风速将误差缩小到 $1.19 \text{ m} \cdot \text{s}^{-1}$ 。从最大强度和序列形态上看,决策树订正后的风速都更接近观测,适合用作海上施工作业中的强风预警。

在业务化应用时,可以首先对预报风速采用等分位法识别强风事件,若识别到强风事件,则对强风事件对应时间段的预报风速采用上述决策树模型做订正,并替换模式预报的该时间段风速,其余非强风

事件时段的预报风速则采用原始模式预报风速。

4 结论与讨论

针对海上施工作业对气象条件的要求,本文定义了一种持续性强风事件识别方法。利用 GDFS 模式资料,采用实况与预报相结合的方法,设计并比较了三种预报强风识别方案,根据评估结果选出了最优识别方案。在其基础上,采用一元线性回归和决策树方法对预报强风事件的风速序列进行订正,并开展了整体订正效果评估和典型过程订正效果分析,得到了以下结论:

(1) 与原始模式预报相比,消偏法和等分位法均能够有效提高预报强风事件的识别命中率,而等分位阈值方案更优,其识别命中率达到 76.1%,匹配时长命中率达 87.6%。

(2) 采用消偏法和等分位法均能增加识别强风事件的时长、减少漏报,其中等分位法识别的预报强风时长与观测非常接近,且能明显增加对影响时间较短的观测强风事件的命中率。

(3) 采用等分位法识别的强风事件基本可以覆盖到各次观测强风事件的全程,有利于精准给出强风预警的开始和结束时间。

(4) 一元线性回归和决策树模型都能一定程度上减少强风事件中风速预报的误差,其中决策树相比一元线性回归模型更优。

(5) 从最大强度和序列形态上看,由决策树订正后的风速比一元线性回归订正后的风速更接近实况。

本文研究了海上观测的持续性强风事件特征并分析了模式的预报表现,给出了影响海上施工安全的持续性强风事件的识别方法及基于敏感性参数的预报强风识别方案和预报序列强度的订正方法,有效改进了强风事件预报能力,但仍然存在一些强风事件漏报或预报事件相对观测事件的时长偏短的现象。此外,因为资料时长较短,做决策树订正建模时,数据相对较少且部分天气型未学习到,会导致部分订正偏大,未来可考虑引入风向因子并增加资料时长,根据观测强风事件进行分类,从而归纳不同风向的强风事件,再构建订正模型,尝试进一步降低漏报和减少空报。另外,本文只采用了一个模式对一个海上观测点的数据做研究,得到的预报强风事件识别方案仍有一些局限性。对于其他模式预报,若

预报本身与观测风速的相关性太差,可能难以通过上述方法来提高预报强风事件的预报能力。此外,在中国各海域的强风特征也不尽相同,所以识别方法和订正方法有待进一步研究。

参考文献

- 曹越男,张增海,原野,等,2014. 南海台风风暴潮数值模拟及其对气象强迫场的敏感性试验[J]. 热带气象学报,30(6):1119-1126.
- Cao Y N, Zhang Z H, Yuan Y, et al, 2014. Sensitivity tests of surface wind fields in typhoon storm surge modeling: cases study in the South China Sea[J]. J Trop Meteor, 30(6):1119-1126(in Chinese).
- 常蕊,朱蓉,柳艳香,等,2013. 基于均生函数的风电场风速短临预报模型[J]. 气象,39(2):226-233. Chang R, Zhu R, Liu Y X, et al, 2013. Nowcasting model of wind speed based on mean generating function for wind farms[J]. Meteor Mon, 39(2):226-233(in Chinese).
- 陈超君,王东海,李国平,等,2012. 冬季高海拔复杂地形下 GRAPES-Meso 要素预报的检验评估[J]. 气象,38(6):657-668. Chen C J, Wang D H, Li G P, et al, 2012. A study of the GRAPES-Meso prediction verification for high altitude and complex terrain during winter time[J]. Meteor Mon, 38(6):657-668(in Chinese).
- 陈笑,赵东,何晓风,等,2018. 基于 MODE 对模式预报强风风场的检验分析[J]. 气象,44(8):1009-1019. Chen X, Zhao D, He X F, et al, 2018. Evaluation and analysis of model forecast performance of high wind based on MODE method[J]. Meteor Mon, 44(8):1009-1019(in Chinese).
- 程兴宏,陶树旺,魏磊,等,2012. 基于 WRF 模式和自适应偏最小二乘回归法的风能预报试验研究[J]. 高原气象,31(5):1461-1469. Cheng X H, Tao S W, Wei L, et al, 2012. Short-term wind power forecasting experiment based on WRF model and adapting partial least square regression method[J]. Plateau Meteor, 31(5):1461-1469(in Chinese).
- 丁煌,陶树旺,肖子牛,等,2013. 基于 WRF 和 SVM 方法的风电场功率预报技术研究[J]. 高原气象,32(2):581-587. Ding H, Tao S W, Xiao Z N, et al, 2013. Study on wind power forecasting of wind farm based on WRF and SVM[J]. Plateau Meteor, 32(2):581-587(in Chinese).
- 何晓风,周荣卫,孙逸涵,2014. 3 个全球模式对近地层风场预报能力的对比检验[J]. 高原气象,33(5):1315-1322. He X F, Zhou R W, Sun Y H, 2014. Verification on surface wind speed of three global circulation models in China[J]. Plateau Meteor, 33(5):1315-1322(in Chinese).
- 黄嘉佑,2000. 气象统计分析与预报方法[M]. 北京:气象出版社. Huang J Y, 2000. Statistic Analysis and Forecast Methods in Meteorology[M]. Beijing: China Meteorological Press(in Chinese).
- 江滢,宋丽莉,程兴宏,2013. 风电场风速预报集合订正方法的尝试性研究[J]. 资源科学,35(3):673-680. Jiang Y, Song L L, Cheng X H, 2013. An integrated and revised method of forecasting wind speed for wind farms[J]. Resour Sci, 35(3):673-680(in Chinese).
- 李艳,汤剑平,王元,等,2009. 区域风能资源评价分析的动力降尺度研究[J]. 气候与环境研究,14(2):192-200. Li Y, Tang J P, Wang Y, et al, 2009. Application of dynamical downscaling method for assessment of wind energy resources[J]. Climatic Environ Res, 14(2):192-200(in Chinese).
- 刘郁珏,苗世光,黄倩倩,等,2023. 睿图-大涡系统支撑北京冬奥会气象服务保障的评估分析[J]. 气象,49(6):733-744. Liu Y J, Miao S G, Huang Q Q, et al, 2023. Evaluation and analysis of meteorological service for Beijing Winter Olympic Games supported by RMAPS-LES prediction system[J]. Meteor Mon, 49(6):733-744(in Chinese).
- 刘郁珏,苗世光,刘磊,等,2019. 修正 WRF 次网格地形方案及其对风速模拟的影响[J]. 应用气象学报,30(1):70-81. Liu Y J, Miao S G, Liu L, et al, 2019. Effects of a modified sub-grid-scale terrain parameterization scheme on the simulation of low-layer wind over complex terrain[J]. J Appl Meteor Sci, 30(1):70-81(in Chinese).
- 戚双斌,王维庆,张新燕,2009. 基于支持向量机的风速与风功率预测方法研究[J]. 华东电力,37(9):1600-1603. Qi S B, Wang W Q, Zhang X Y, 2009. Wind speed and wind power prediction based on SVM[J]. East China Electric Power, 37(9):1600-1603(in Chinese).
- 邱贵强,于波,陶亦为,等,2023. 基于集成学习算法的冬奥会延庆赛区极大风速预报[J]. 气象,49(6):721-732. Qiu G Q, Yu B, Tao Y W, et al, 2023. Forecasting of extreme wind speed in Yanqing competition zone of the Winter Olympic Games based on ensemble learning algorithm[J]. Meteor Mon, 49(6):721-732(in Chinese).
- 申华羽,方艳莹,涂小萍,等,2020. ECMWF 细网格 10 m 风预报在浙江沿海的评估与订正[J]. 气象,46(11):1485-1494. Shen H Y, Fang Y Y, Tu X P, et al, 2020. Verification of ECMWF 10 m wind forecast for coastal Zhejiang Province[J]. Meteor Mon, 46(11):1485-1494(in Chinese).
- 王彬滨,吴息,余江,等,2016. 谐波分析方法在沿海风速数值预报订正的应用[J]. 热带气象学报,32(5):752-758. Wang B B, Wu X, Yu J, et al, 2016. The application of harmonic analysis method in correction of numerically forecast wind speed in coastal areas[J]. J Trop Meteor, 32(5):752-758(in Chinese).
- 王令,丁青兰,卞素芬,等,2008. 奥运气象服务中的短时预报及预警[J]. 气象,34(S1):263-268. Wang L, Ding Q L, Bian S F, et al, 2008. The short-term prediction and warning in Olympic Games weather service[J]. Meteor Mon, 34(S1):263-268(in Chinese).
- 王在文,陈敏,Monache L D,等,2019. 相似集合预报方法在北京区域地面气温和风速预报中的应用[J]. 气象学报,77(5):869-884. Wang Z W, Chen M, Monache L D, et al, 2019. Application of analog ensemble method to surface temperature and wind speed prediction in Beijing Area[J]. Acta Meteor Sin, 77(5):869-884(in Chinese).
- 肖擎曜,胡非,范绍佳,2017. 风能数值预报的模式输出统计(MOS)

- 研究[J]. 资源科学, 39(1): 116-124. Xiao Q Y, Hu F, Fan S J, 2017. Model output statistics and wind power numerical prediction[J]. Resour Sci, 39(1): 116-124(in Chinese).
- 闫丽凤, 江文胜, 周淑玲, 等, 2008. 0703 温带气旋特大风暴潮数值模拟对比分析[J]. 应用气象学报, 19(5): 595-601. Yan L F, Jiang W S, Zhou S L, et al, 2008. Contrast analysis on numerical simulation of 0703 cyclone strong storm surge[J]. J Appl Meteor Sci, 19(5): 595-601(in Chinese).
- 杨程, 姜瑜君, 余贞寿, 等, 2019. 基于偏最小二乘回归的区域换式风速预报订正技术研究[J]. 气象, 45(5): 676-684. Yang C, Jiang Y J, Yu Z S, et al, 2019. Correction technology of regional wind speed forecasting based on partial least square regression[J]. Meteor Mon, 45(5): 676-684(in Chinese).
- 杨璐, 宋林焯, 荆浩, 等, 2022. 复杂地形下高精度风场融合预报订正技术在冬奥会赛区风速预报中的应用研究[J]. 气象, 48(2): 162-176. Yang L, Song L Y, Jing H, et al, 2022. Fusion prediction and correction technique for high-resolution wind field in Winter Olympic Games area under complex terrain[J]. Meteor Mon, 48(2): 162-176(in Chinese).
- 张铁军, 颜鹏程, 张正英, 等, 2018. 多种订正技术在风电场风速预报订正中的应用[J]. 干旱气象, 36(5): 835-844. Zhang T J, Yan P C, Zhang Z Y, et al, 2018. Application of various technologies in modification of wind speed forecast in wind farms[J]. J Arid Meteor, 36(5): 835-844(in Chinese).
- 周荣卫, 何晓风, 朱蓉, 等, 2010. 中国近海风能资源开发潜力数值模拟[J]. 资源科学, 32(8): 1434-1443. Zhou R W, He X F, Zhu R, et al, 2010. Numerical simulation of the development potential of wind energy resources over China's offshore areas[J]. Resour Sci, 32(8): 1434-1443(in Chinese).
- Bédard J, Yu W, Gagnon Y et al, 2013. Development of a geophysics model output statistics module for improving short-term numerical wind predictions over complex sites[J]. Wind Energy, 16(8): 1131-1147.
- Breiman L, Friedman J, Olshen R, et al, 1984. Classification and Regression Trees [M]. New York: Chapman & Hall (Wadsworth, Inc.): 368.
- Cheng W Y Y, Steenburgh W J, 2005. Evaluation of surface sensible weather forecasts by the WRF and the Eta models over the Western United States[J]. Wea Forecasting, 20(5): 812-821.
- Gioli B, Gualtieri G, Busillo C, et al, 2014. Aircraft wind measurements to assess a coupled WRF-CALMET mesoscale system [J]. Meteor Appl, 21(1): 117-128.
- Kiktev D B, Astakhova E D, Zaripov R B, et al, 2015. FROST-2014 project and meteorological support of the Sochi-2014 Olympics [J]. Russ Meteorol Hydrol, 40(8): 504-512.
- Lin Y J, Kruger U, Zhang J P, et al, 2015. Seasonal analysis and prediction of wind energy using random forests and ARX model structures[J]. IEEE Trans Control Syst Technol, 23(5): 1994-2002.
- Liu Z Y, Hara R, Kita H, 2021. Hybrid forecasting system based on data area division and deep learning neural network for short-term wind speed forecasting[J]. Energy Convers Manage, 238(1): 114136.
- Storm B, Dudhia J, Basu S, et al, 2009. Evaluation of the weather research and forecasting model on forecasting low-level jets: implications for wind energy[J]. Wind Energy, 12(1): 81-90.
- Traiteur J J, Callicutt D J, Smith M, et al, 2012. A short-term ensemble wind speed forecasting system for wind power applications [J]. J Appl Meteor Climatol, 51(10): 1763-1774.
- Zjavka L, 2015. Wind speed forecast correction models using polynomial neural networks[J]. Renewable Energy, 83: 998-1006.

(本文责编: 王婷波 何晓欢)