

袁伟, 蓝俊, 郑玉兰, 等. 福建海洋浮标观测数据接续技术初步研究[J]. 海洋气象学报, 202X, XX(X): 1-10.
YUAN W, LAN J, ZHENG Y L, et al. Preliminary study on data continuation technology for Fujian ocean buoy observations[J].
Journal of Marine Meteorology, 202X, XX(X): 1-10. DOI: 10.19513/j.cnki.hyqxxb.20240513001. (in Chinese)

福建海洋浮标观测数据接续技术初步研究

袁伟^{1,2,3}, 蓝俊^{1,2,3}, 郑玉兰^{1,2,3}, 段卿^{1,2,3}, 高珊^{2,4}

(1. 福建省灾害天气重点实验室 福建 福州 350008; 2. 平潭台湾海峡海洋气象福建省野外科学观测研究站 福建 福州 350008;
3. 福建省气象信息中心 福建 福州 350008; 4. 福建省气象台 福建 福州 350008)

摘要 应用百分位、K-means 聚类等技术, 分析福建海洋浮标的经纬度信息, 尝试解决长序列海洋浮标资料存在位置漂移、摆动情况时的缺测位置估算问题。以此为出发点, 结合中国气象局陆面数据同化系统 CLDAS、欧洲中期天气预报中心 ERA5 等第三方资料, 分别采用反距离权重法、普通克里金插值等方法在给定时段和经纬度位置时进行计算, 得到相应的气温、气压、湿度、风向和风速等要素估测值。在此基础上, 首先, 对不同第三方资料 and 不同插值算法估测结果进行综合评估, 发现数据序列稳定性和时序长度对要素填补接续有很大影响, 反距离权重法相比普通克里金插值法评估效果表现更好。其次, 针对 7 个福建海洋浮标原始观测数据集, 根据结论重新整合得到新数据集, 两者对比评估检验, 各气象要素的皮尔逊相关系数普遍在 0.9 以上, 新数据集不仅完整性得到极大改善, 而且保留了原始观测数据的一致性和稳健性, 局部异常要素值也得到有效控制。最后, 在 2014—2021 年台风个例检验过程中, 对比台风实时路径进行风速要素的评估, 新数据集风速要素相关系数平均提高 0.03、均方根误差平均下降 $0.5 \text{ m} \cdot \text{s}^{-1}$, 实验表明接续方法能提升海洋浮标观测资料的业务适用性。

关键词 海洋浮标; 缺测填补; 要素插值; 评估检验; 相关系数

中图分类号: P413.2 文献标志码: A 文章编号: 2096-3599(2024)04-0000-00

DOI: 10.19513/j.cnki.hyqxxb.20240513001

Preliminary study on data continuation technology for Fujian ocean buoy observations

YUAN Wei^{1,2,3}, LAN Jun^{1,2,3}, ZHENG Yulan^{1,2,3}, DUAN Qing^{1,2,3}, GAO Shan^{2,4}

(1. Fujian Key Laboratory of Severe Weather, Fuzhou 350001, China; 2. Pingtan Taiwan Strait Marine Meteorology Observation and Research Station of Fujian province, Fuzhou 350001, China; 3. Fujian Meteorological Information Center, Fuzhou 350001, China; 4. Fujian Meteorological Observatory, Fuzhou 350001, China)

Abstract The percentile and K-means clustering were applied to analyze the longitude and longitude information of Fujian ocean buoy, and try solve the problem of position estimation of missing observation in the long sequence of ocean buoy data. On this basis, combined with the third-party data such as CLDAS and ERA5, the inverse distance weight method and Ordinary Kriging interpolation algorithms are used to calculate the timing and longitude and latitude positions, and the corresponding temperature,

收稿日期: 2024-05-13; 修回日期: 2024-07-15

基金项目: 中国气象局上海台风研究基金项目 (TFJJ202113)

第一作者简介: 袁伟, 男, 正高级工程师, 主要从事气象信息技术, 38091770@qq.com。

通信作者简介: 高珊, 女, 正高级工程师, 主要从事天气预报研究, 312gs@163.com。

pressure, humidity, wind direction and wind speed are estimated. On the evaluation results of different third-party information and different interpolation algorithms, the stability of data sequence and timing length on the filling of elements are better than Ordinary Kriging interpolation. Subsequently, based on the conclusions, a new data set was reconstituted by integrating the original observation data from seven Fujian marine buoys. A comparative evaluation and validation of the two data sets showed that the Pearson correlation coefficients for various meteorological elements were generally above 0.9. The new data set not only significantly improved its completeness but also preserved the consistency and robustness of the original observation data. Local abnormal element values were also effectively controlled. During the validation process of typhoon cases from 2014 to 2021, the wind speed elements were evaluated by comparing them with the real-time typhoon tracks. The correlation coefficients of the wind speed elements in the new data set is increased by 0.03, and the mean square error decreased by 0.5m/s. The experiment shows that the data continuation technology can improve operational applicability about ocean buoys observation datum.

Keywords Marine buoy; Missing data filling; Element interpolation; Evaluation and Testing; Correlation coefficient

引言

随着福建海洋经济的深入发展,海上气象服务也变得日趋重要,除多普勒天气雷达、风云气象卫星等观测外,海洋观测浮标也是获取海洋气象要素的重要手段。围绕海上风能评估、海洋牧场、航运安全、台风预报等业务应用,国内科研人员在海洋相关数据集研究与应用方面做了大量研究。王春晓等^[1]基于海洋浮标气压、风速、风向和波高等观测要素,结合台风天气过程,通过气压变化截取台风影响期间的海洋浮标观测数据,分析并制作了台风数据集。韩玉康等^[2]基于台湾气象局网站公布的气象水文资料,在计算要素日平均、月平均和其他特征量后,形成了台湾及周边岛屿气象水文观测数据集。一系列的长序列数据集,可为海洋气象服务提供数据支撑,其研究与应用价值不言而喻。在海洋浮标资料的数据质量控制、对比评估检验、缺测填补等方面也有大量科研成果可借鉴,如卢勇夺等^[3]基于极值法则和莱以特法则开展海洋浮标风速、气压、气温、波高要素的异常值检测,对疑似异常要素进行标注。张宇等^[4]基于一定时间窗口内的要素之间的相关性分析,建立要素两两之间的相关系数,以此来判断酸碱度(pH)、溶解氧(DO)浓度等数据值异常与否。张友权^[5]结合美国国家环境预报中心(National Centers for Environmental Prediction, NCEP)的第二代气候预测系统(coupled forecast system model version 2, CFSv2)、欧洲中期天气预报

中心(European Centre for Medium-Range Weather Forecasts, ECMWF)发布的再分析数据 ERA-I、NCEP-DOE 等风场资料,对海洋浮标的风向、风速要素开展均方根误差、相关系数等指标的评估检验,得出不同风场资料的预期应用情况;刘宇^[6]基于全球海洋 Argo 散点资料集,开展海洋表层水温的变化趋势分析与研究,从数据自身出发,采用自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)、长短期记忆网络(Long Short Term Memory, LSTM)等模型进行预测,并通过缺测数据前后时次的有效数据长度进行加权,实现对缺测数据的补全。

一套时序稳定连贯、时间序列长、质量有保障的数据集,有利于海上风能评估、海洋牧场环境评估、气象风险评价等诸多业务的开展。鉴于海洋浮标资料的重要应用价值,考虑福建7个海洋浮标观测存在位置漂移、数据缺测较频繁,完整性、数据质量相对不高等情况,本文从浮标观测原始要素出发,通过浮标位置漂移特性研究、要素接续以及异常值订正等研究,构建要素接续填补技术,制作形成逐时长序列气温、湿度、气压、风速、风向要素数据集。随着在全国范围内多源实况分析产品的应用试验及评估检验,其业务应用场景得到多方位的体现,如李树文等^[7]利用若干多源实况分析产品建立的GBDT模型用于实况气象信息服务,崔园园等^[8]利用中国气象局陆面数据同化系统CLDAS开展气象干旱监测预警服务,文中引入多源实况分析网格产品作为第三

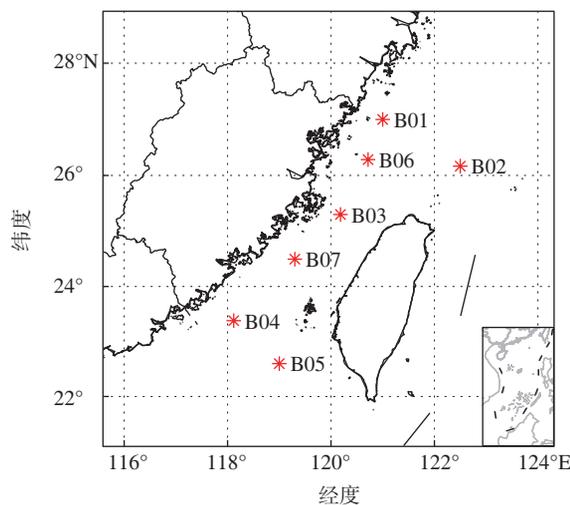
方数据源,开展了经纬度接续、要素接续等相关验证工作,最后结合台风天气过程对数据集进行综合评估和应用检验,以证明借助本文相关技术产生的数据集,在原始观测要素序列的基础上实现数据缺测填补、异常值修订有十分重要的参考意义。

1 资料来源

所使用的数据主要包括:福建 7 个海洋浮标观测数据集、多源融合实况分析产品、ERA5 再分析数据集和中国气象局上海台风研究所 (Shanghai Typhoon Institute of China Meteorological Administration, CMA-STI) 的热带气旋最佳路径数据集。

海洋浮标观测数据集:浮标在台湾海峡区域内布放(图 1),选取 2014—2021 年逐时观测资料,并经过气候阈值、空间一致性、时间一致性等常规的质控检验,过滤各要素的显性异常值,得到 2 m 气温、海平面气压、相对湿度、10 m 风速风向的长序列要

素集。其中,海洋浮标实际数据起止时间、数据完整性(观测时次和理论时次)等情况如表 1 所示。



红色圆点—浮标站位置。

图 1 浮标站位置示意

Fig1 Schematic diagram of the buoy calibration position

表 1 浮标数据基本信息
Table 1 Basic information of buoy data

浮标站	实际数据起止时间	理论时次/h	完整性/%	要素缺测占比					游程统计信息				
				AT/%	BP/%	RH/%	WS/%	WD/%	游程 A /次数	游程 B /次数	游程 C /次数	游程 D /次数	最长游程 /d
B01	2014.4.22—2021.12.31	67 462	80.48	6.09	1.97	5.58	0.002	0.82	1 611	564	77	11	78
B02	2015.9.16—2021.12.31	55 175	90.02	1.24	0.37	1.65	0	2.48	1 506	507	50	5	27
B03	2014.4.21—2021.12.31	67 480	49.87	2.29	0.15	0.39	0.050	2.59	1 324	428	45	9	54
B04	2014.4.2—2021.12.31	70 128	49.70	15.33	0.41	1.03	0.002	0.63	989	342	40	5	92
B05	2017.8.21—2021.12.31	38 254	72.75	10.87	0	0.01	0	3.00	1 337	485	28	6	78
B06	2014.1.1—2021.12.31	70 128	85.95	3.83	0.76	3.18	0.740	2.78	1 393	441	91	10	42
B07	2014.1.1—2021.12.31	70 128	82.49	7.17	0.96	9.22	0.130	2.40	1 092	381	63	11	42

其中,理论时次指具体浮标的有效起止时间内的小时数,完整性指有效时次占比理论时次的比例。要素缺测包括 2 m 气温(AT)、海平面气压(BP)、相对湿度(RH)、10 m 风速(WS)、10 m 风向(WD)的缺测,是指在数据起止时间内观测时次中该要素为空值时的统计占比,0 表示无缺测。

游程统计定义为:统计 2014—2021 年逐时海洋浮标观测数据,若当前时次存在观测要素,则记为 1,否则记为 0,统计连续 1 序列或连续 0 序列,并记录 1 或 0 序列的长度。游程统计用于探测数据完整

性、数据缺测、数据序列稳定性等信息,并根据统计结果,针对数据缺测选取适当的方法来填补处理。表 1 中,游程 A 指缺测 1 个时次的统计量;游程 B 指缺测 2~6 个时次的统计量;游程 C 指缺测 7~144 个时次的统计量;游程 D 指缺测 144 个时次以上的统计量;最长游程指 0 长度最大值,即最长缺测时间,按具体时次换算为缺测天数。

统计表明,气温、相对湿度要素缺测较多。观测资料的连续性缺测方面,连续缺 1 个时次、连续缺 2~6 个时次的统计占比显著,最长缺测时间可达 3 个

月,这也给后期数据接续填补带来挑战。

文中采用多源融合实况分析产品(以下简称“网格产品”)和 ERA5 再分析数据集(以下简称“ERA5”)做为接续的第三方参考数据源,涉及要素有 2 m 气温、海平面气压、相对湿度、10 m 风速风向等。其中网格产品是国家气象信息中心师春香等^[9]研制的多源融合实况分析产品,该产品及相关技术在国内得到了大量应用检验,例如陈玉玮等^[10]就气温、10 m 风速要素开展了站点观测与网格产品之间的对比评估,结果表明逐时、最高、最低气温相关系数在 0.97 以上、平均绝对误差小于 1 °C,10 m 风速平均误差为 $-0.73 \text{ m}\cdot\text{s}^{-1}$ 。诸多分析结论为网格产品在业务中的应用提供了科学依据。ERA5 应用范围也非常广泛,如孔祥宁等^[11]分析山东地区的气温日统计,给出了 ERA5 数据在寒潮过程监测、高温日数分布等方面的局部适用性。

CMA-STI 热带气旋最佳路径数据集(以下简称台风报文)资料取自中国台风网(www.typhoon.org.cn),遴选 2014—2021 年期间覆盖范围在 $21.5^{\circ}\sim 28.5^{\circ}\text{N}$ 、 $113.5^{\circ}\sim 123.5^{\circ}\text{E}$ 的台风路径资料,用于对比评估检验。

2 数据接续方法及评估

2.1 站点位置定位分析

在实际浮标数据中,因浮标漂移、修理等情况导致浮标的经纬度坐标不固定,此外还存在位置信息缺测、异常等情况,因此在数据接续过程中,需解决浮标位置定位问题,为后续数据填补、修订提供保障。

B01 浮标站位置分布如图 2 所示,除错误值外,局部呈很强的聚集情形。考虑浮标存在漂移、摆动、缺测等情况,在一定平滑时间窗口内引入 K -means 聚类算法^[12],可作为一种浮标位置的异常发现、分类分割,进而实现局部预测的手段。

2.1.1 技术方法

考虑站点位置存在错误、缺测等情况,结合缺测游程统计,因连续缺测的时次长短不同,综合当前时序前后可利用的有效信息,设计异常位置信息检测订正、缺测位置信息估测填补处理流程。

设 $P(t) = (x, y)$, $t = t_1, t_2, t_3, \dots, t_m$, 为当前浮标数据所采集到的逐时经纬度位置,其中 $\{t_1, t_2,$

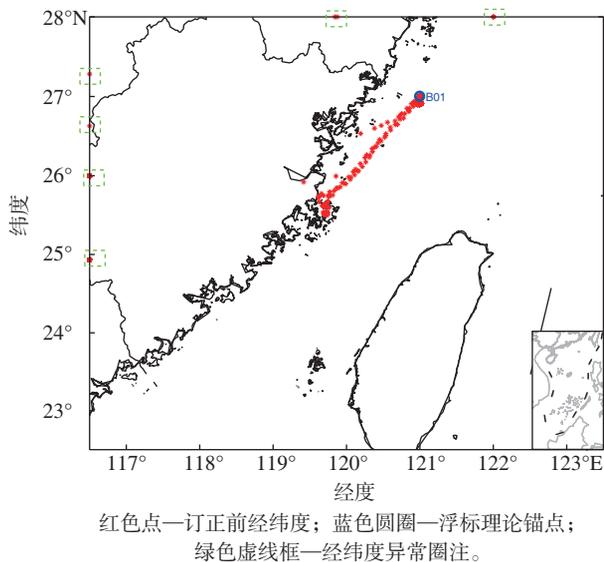


图 2 2014—2021 年 B01 浮标站位置分布
Fig 2 Distribution of B01 buoy positions from 2014 to 2021

t_3, \dots, t_m } 为观测时次时间序列。

记 t'_h 为缺测位置对应时次 ($t_1 < t'_h < t_m$), $P(t'_h)$ 为缺测浮标位置经纬度,步骤如下:

(1) 异常经纬度订正:统计待检测点前后一定时间窗口内(本文取前后各 3 d)的有效观测经纬度信息,采用 K -means 聚类算法,取质心均方差最小的质核作为检测标准,当前经纬度落入该质核,则认为正常,否则认为存在异常,此时用质核来订正当前经纬度坐标值。

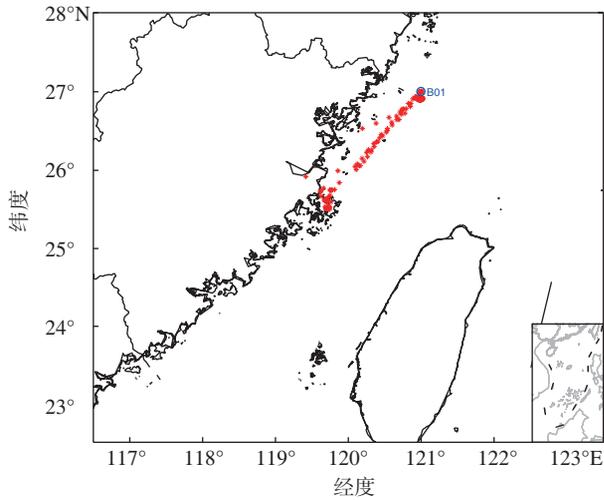
(2) 游程 A、B 类型的缺测时:采用邻近点估算,因缺测时次较少,根据前后有效观测位置取均值作为 $P(t'_h)$ 的估值。

(3) 游程 C 类型的缺测时:用前后各 3 d 的时间窗口来估算数据量,简单采用五分位数的中位数,作为 t'_h 前后的位置,取两者平均值为 $P(t'_h)$ 的估算值。

(4) 游程 D 及其他类型的缺测时:根据 t'_h 的左右实际观测时次,分别计算左右两个窗口聚类族群,相应得到聚类质心。针对两组族群,取质心均方差最小作为前后质心,再取两者平均值作为 $P(t'_h)$ 的估算值。

2.1.2 结果分析

通过上述步骤,完成缺测经纬度信息填补、错误经纬度修订等工作。以 B01 浮标站为例,图 3 是剔除所有可能错误(设备错误、维修时段等)情况下的位置分布。



红色点—订正后经纬度; 蓝色圆圈—浮标理论锚点。

图3 B01 浮标站位置信息订正结果示意

Fig 3 Schematic diagram of correction results for B01 buoy position information

对比缺测或缺测时次的经纬度填补结果,按相邻时次统计差分,对差分序列进行标准差和均方根

误差统计,以 B01 浮标站为例评估结果如表 2 所示。

表 2 相邻时次经纬度差分相关统计

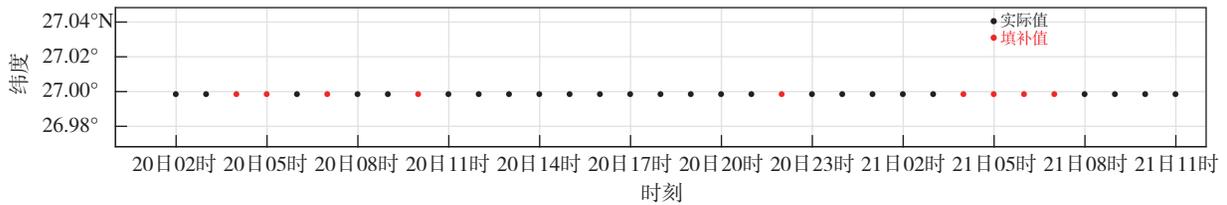
Table 2 Statistics of longitude and latitude differences adjacent time intervals

不同经纬度差分序列	标准差/(°)	均方根误差/(°)
原始数据的经度差分序列	6.449	41.596
原始数据的纬度差分序列	2.016 30	0.000 04
订正填补后的经度差分序列	0.006 348	4.065 610
订正填补后的纬度差分序列	0.007 480 0	0.000 055 9

统计对比表明,经纬度订正填补后,经纬度序列的数据质量和稳定性均得到提升。图 4 是缺测时次较少情形的填补效果,图 5 是缺测时次较多且频繁出现时的填补效果,图 6 为长时段缺测时的填补效果,结果表明,游程 A、B 类型的缺测情形下的填补效果更好些,长时段缺测情形时由于只考虑前后有效位置的聚类值,尽管实际观测要素的持续缺测严重,但填补效果能够保持一定的稳定性,故通过引入第三方数据来填补要素也具有一定参考价值。



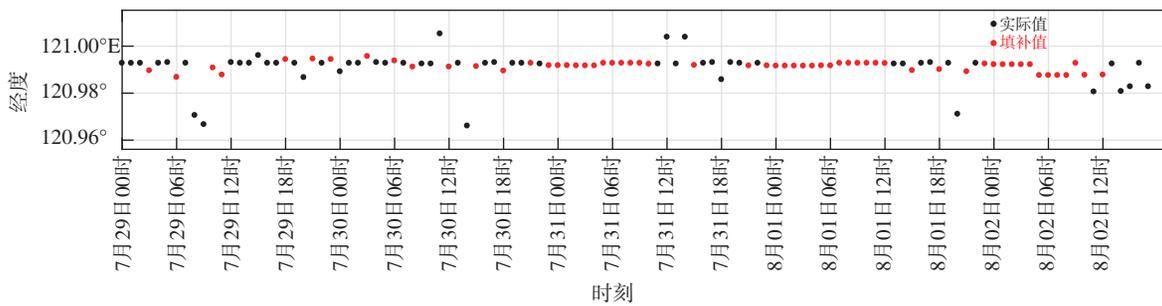
(a) 经度填补序列



(b) 纬度填补序列

图 4 2019 年 4 月缺测时次较少情形下的 B01 浮标站位置填补序列

Fig 4 B01 buoy position filling sequence(less missing)



(a) 经度填补序列

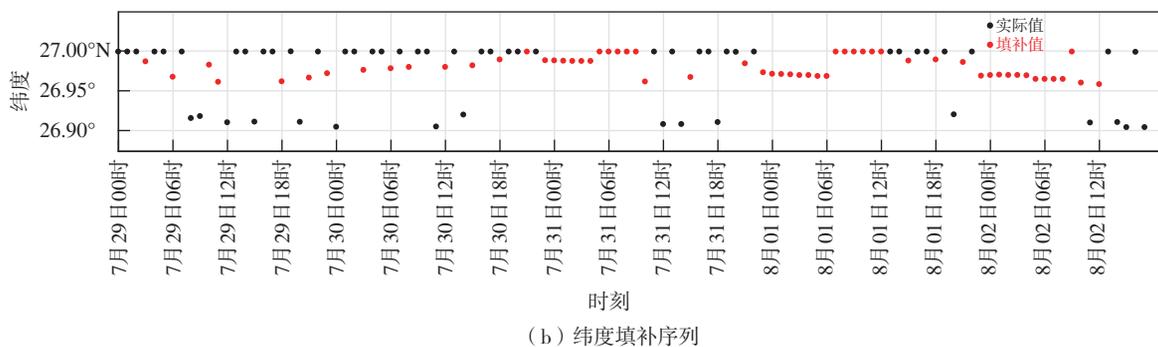


图5 2019年缺测时次较多且缺测频繁情形下的B01浮标站位置填补序列
Fig 5 B01 buoy position filling sequence (more and frequent missing)

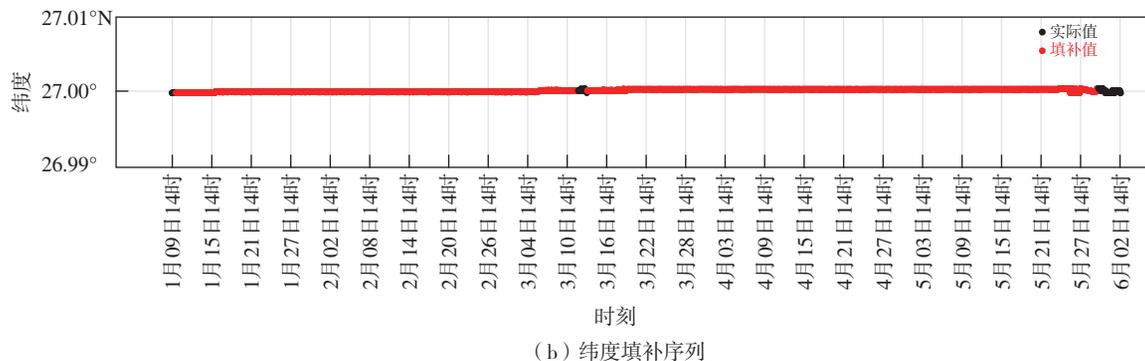
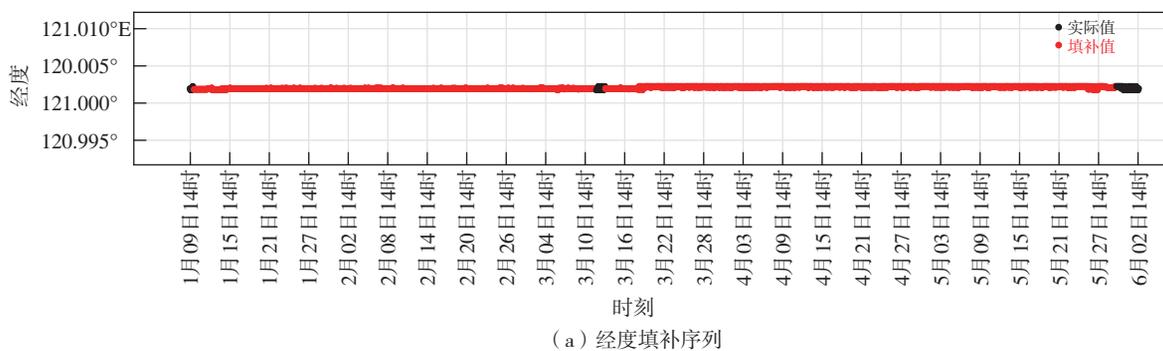


图6 2018年长时段缺测情形下的B01浮标站位置填补序列
Fig 6 B01 buoy position filling sequence (long time missing)

2.2 数据预处理

气象要素质量是开展数据应用的重要前提保障,国内外也开展了大量的技术与应用研究^[3-4,13]。文中研究的海洋浮标数据经过国内气象通信系统(CTS)进行过一次数据质量控制^[14]。然而在数据完整性统计、连续性统计、显见异常值判断等数据预处理过程中,仍然发现要素存在异常情况(图7)。

文中主要以格拉布斯(Grubbs)准则^[15]来发现显见或疑似异常值,对于明显错误的要素按缺测处理,其他疑似异常进行标记,后期将用第三方数据源进行订正。验证观测要素的正态统计分布特性,并计算得到气温、气压、湿度、风速、水温等参数的均值

与标准差,采用Grubbs准则设定要素异常判断门限,通过率均在97%以上(表3)。

表3 不同要素检验统计结果
Table 3 Test statistics of different elements

检验统计指标	要素				
	气温 /°C	气压 /hPa	湿度 /%	风速 /(m·s ⁻¹)	水温 /°C
均值	21.1	1013.6	82.3	7.7	23.2
标准差	6.02	7.9	17.9	4.5	5.7
估测通过率	0.999 3	0.998 9	0.973 0	0.999 3	0.983 6

通过基本数据处理后,可剔除较为明显的要素异常值,结果如下图8所示。

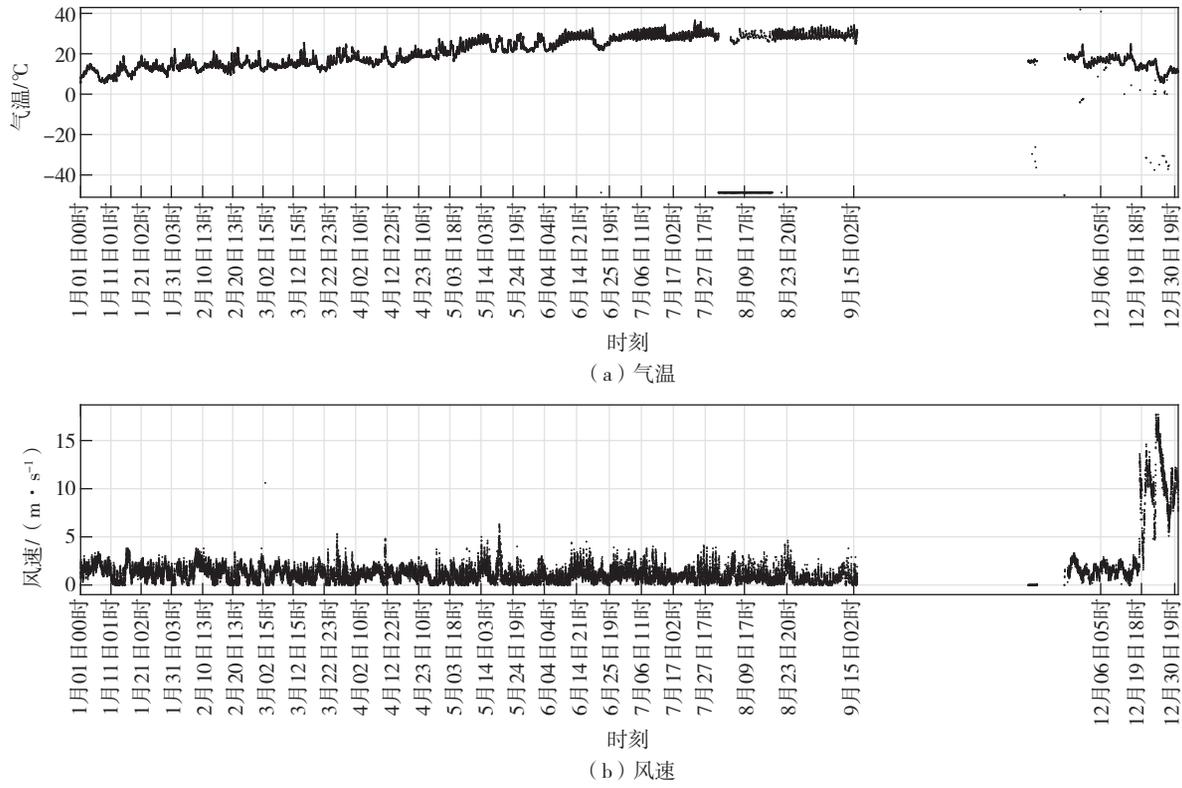


图7 2021年B01浮标站原始观测要素序列
Fig 7 Original observation element sequence of B01 buoy

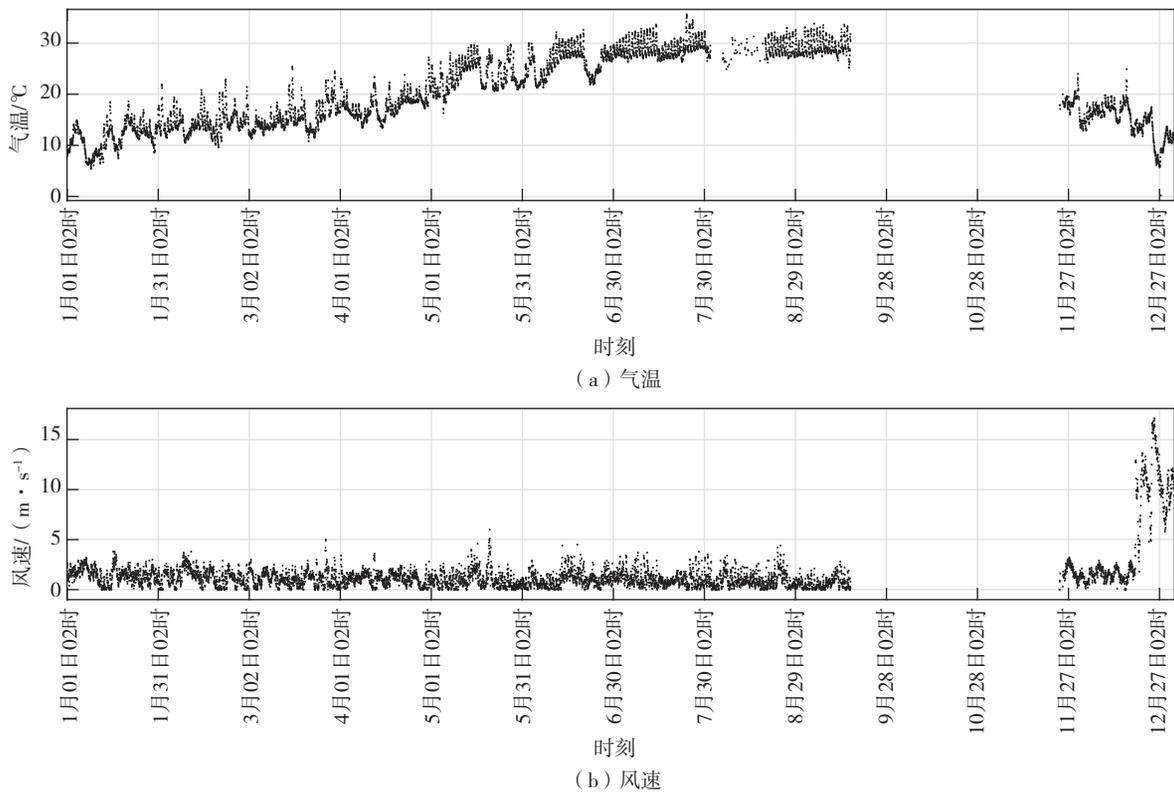


图8 2021年B01浮标站要素订正后序列
Fig 8 Corrected element sequence of B01 buoy

2.3 气象要素接续方法

利用网格产品,提供对缺测要素的填补,综合当前一些关于插值算法对比评估方面的研究成果^[16-20],选取最近邻插值法(Nearest Neighbor Interpolation, NNI)、反距离权重法(Inverse Distance Weighted, IDW)、普通克里金(Ordinary Kriging, Krg)插值法^[21-22]并进行比较,同时将浮标数据分别与CLDAS和ERA5对比,采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)、拟合优度(R^2)、皮尔逊相关系数(r)、斯皮尔曼等级相关系数(ρ)等进行综合评估,综合结论后选取合适的网格数据和插值方法来实现要素接续。

在3.1节中,对任意待预测时间 t'_h ,得到对应时次的浮标估测位置 $P(t'_h)$ 。通过时次 t'_h 的实况网格资料,取 $P(t'_h)$ 位置周边 M 个(本文取 $M=25$)网格点,得到网格经纬度和格点要素值:

(1)对最近邻插值法而言,取网格点中距离 $P(t'_h)$ 最近的网格点,直接作为 $P(t'_h)$ 处要素的估算值,记为 $V_{NNI}(t'_h)$ 。

(2)对反距离权重法而言,对 M 个网格点进行计算,得到结果作为 $P(t'_h)$ 处要素的估算值,记为 $V_{IDW}(t'_h)$ 。

(3)对普通克里金插值法而言,对 M 个网格点

进行计算,得到结果作为 $P(t'_h)$ 处要素的估算值,记为 $V_{Krig}(t'_h)$ 。

为更好地实现对缺测值的填补,通过对比评估插值算法和两种不同网格资料的接近程度,选择相似度和误差精度最佳插值算法及对应的网格资料作为缺测时次要素估算值。

利用2014—2021年期间的浮标资料,结合网格产品进行要素评估与接续。主要通过最近邻插值法、反距离权重法、普通克里金插值法等,在网格资料(CLDAS、ERA5)中计算给定浮标位置的数值,再与实测值开展皮尔逊相关系数、均方根误差、平均绝对值误差等指标检验分析。

图9—11分别给出气温、U风和V风(将风向与风速转换为U风、V风)所对应的不同算法、不同第三方数据源时的评估指标,对比发现B01、B02、B03、B05浮标站与CLDAS之间各项评估指标更优,而B04、B06、B07浮标站与ERA5的各项评估指标更有优势,结合第2节图1、表1信息,数据序列稳定性和时序长度对要素填补接续有很大影响,数据越稳定,引入CLDAS的接续效果要好,反之,引入ERA5的效果要好。

图12是不同插值方法对数据填补接续的效果分析(平均绝对误差越小越好),整体而言IDW插值相比另外两种表现更佳。

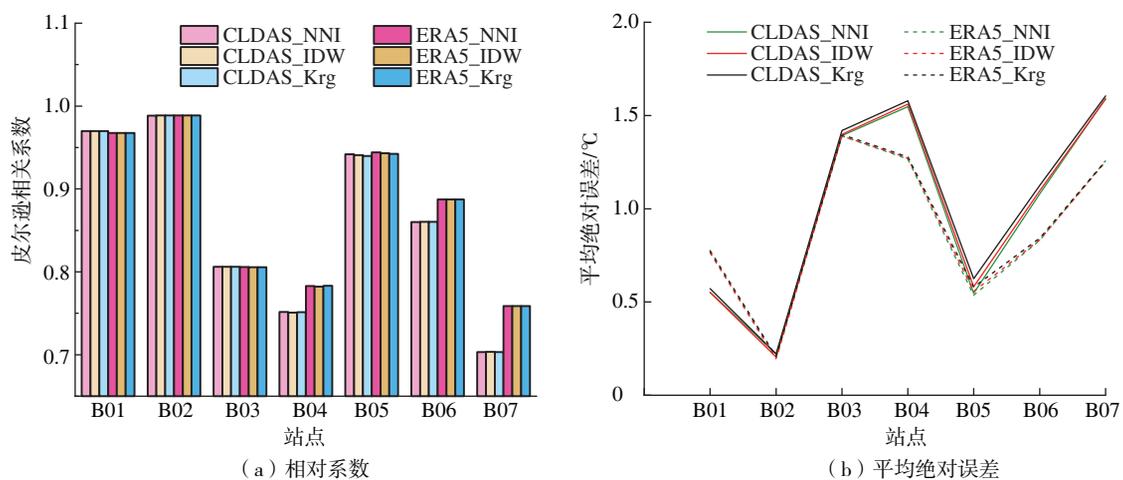


图9 不同浮标站、不同插值方法的气温要素相关系数和平均绝对误差
Fig 9 Relationship between correlation coefficients and absolute errors of temperature elements for different buoy stations and interpolation methods

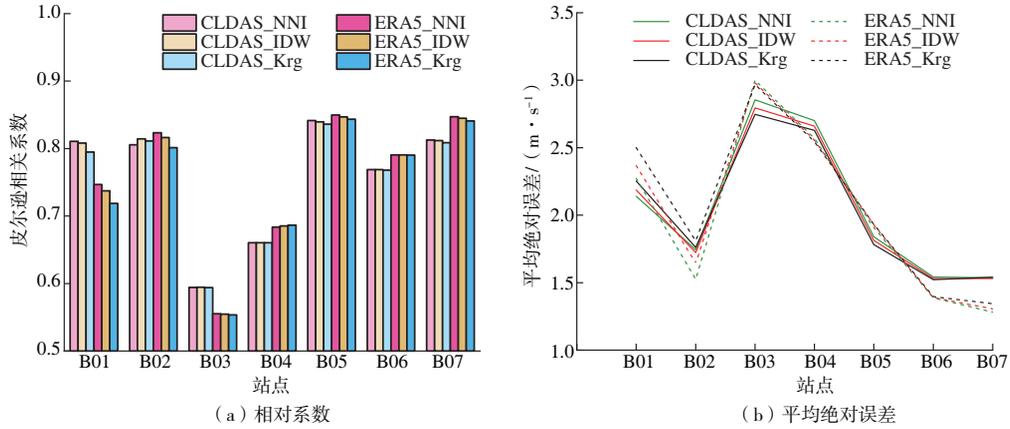


图 10 不同浮标站、不同插值方法的 U 风要素相关系数和平均绝对误差
 Fig 10 Relationship between correlation coefficients and absolute errors of U-wind elements for different buoy stations and interpolation methods

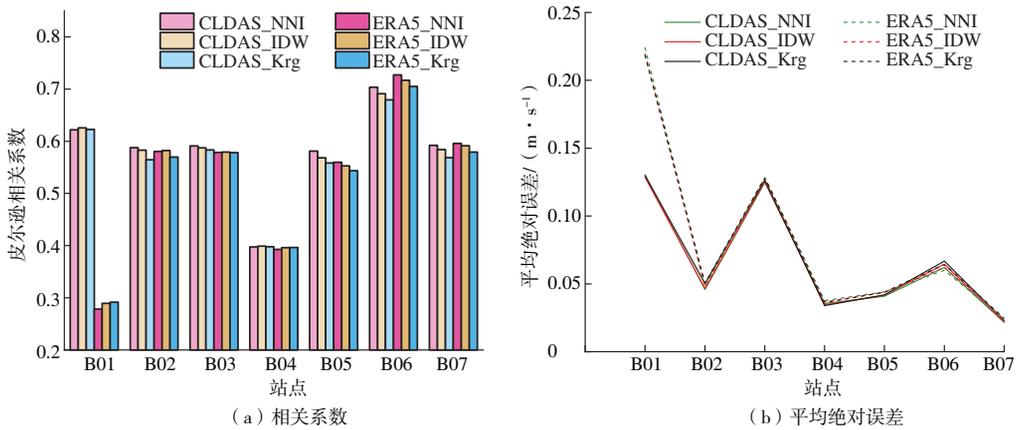


图 11 不同浮标站、不同插值方法的 V 风要素相关系数和平均绝对误差
 Fig 11 Relationship between correlation coefficients and absolute errors of V-wind elements for different buoy stations and interpolation methods

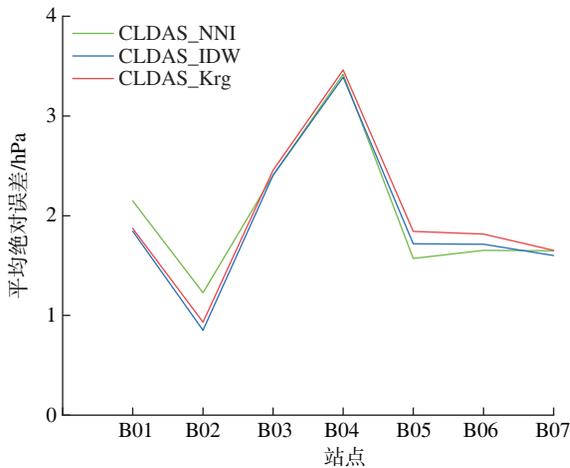


图 12 不同插值方法下 CLDAS 与原始观测气压要素平均绝对误差的对比评估

Fig 12 Comparison and evaluation of absolute error relationships of air pressure between CLDAS and original observation using different interpolation methods

3 评估检验

3.1 要素评估检验

为更好地评估接续效果,按浮标要素数据理论观测时次,结合实际数据的完整性情况及上述分析结果,针对不同的浮标观测站和不同的要素选取最佳评估指标所对应的实况网格资料,采用反距离权重插值技术取样,得到一套完整序列的新数据集。并对新数据集开展适用性检验,包括与原始观测数据之间的对比评估、结合台风报文的检验等,用于检验站点位置定位前后、数据预处理前后的对比情况。为利用新数据集对要素的接续应用提供参考依据。

以气温、风速等要素为例,计算新数据集值与原始观测值要素完整性差值、相关系数、均方根误差等评价指标,如表 4、5 所示。

表4 气温要素性能对比

浮标站	完整性提升/%	皮尔逊相关系数	均方根误差/°C
B01	25.2	0.999 06	0.171
B02	18.2	0.999 32	0.080
B03	52.2	0.913 99	0.308
B04	67.3	0.999 08	0.014
B05	41.4	0.994 86	0.096
B06	40.4	0.995 74	0.052
B07	44.1	0.998 05	0.032

针对海洋浮标要素,结果表明接续后的数据集完整性得到极大的提升,普遍超过 0.9 的相关系数和可接受的均方根误差也表明提升后的数据集具有很好

的原始观测信息保持性。图 13 是 B01 浮标站接续前后数据趋势分析,拟合优度为 0.999,误差在 0.3 °C 以内,也表明数据具有很高的一致性和稳健性。

表5 风速要素性能对比

浮标站	完整性提升/%	皮尔逊相关系数	均方根误差/(m·s ⁻¹)
B01	20.3	0.893 68	1.809
B02	16.9	0.872 99	1.493
B03	50.7	0.934 56	1.264
B04	62.0	0.987 95	1.150
B05	34.0	0.982 20	1.059
B06	38.6	0.918 05	0.752
B07	40.0	0.983 55	0.585

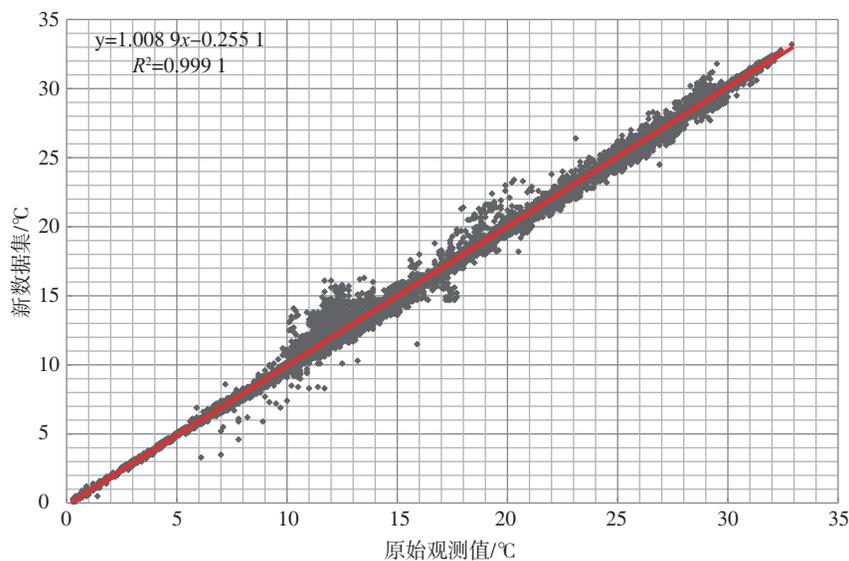


图 13 B01 浮标站新数据集与原始观测数据气温要素趋势

Fig 13 The trend distribution of air temperature elements between the new dataset and the original observational data of B01

3.2 在台风天气过程中的应用检验

原始观测信息和新数据集分别应用在 2014—2021 年台风个例中进行评估检验,针对台风实时路径,按照 7、10、12 级台风风圈(7 级风圈,风速为 13.9~24.4 m·s⁻¹;10 级风圈,风速为 24.5~32.6 m·s⁻¹;12 级风圈,风速超过 32.7 m·s⁻¹),匹配台风中心点与浮标位置关系(图 14)得到相应样本点,按风力等级换算后与台风风圈进行对比评估。

统计结果表明新数据集对比原始观测数据在台风个例中的检验,其相关系数、均方根误差等评估指标均得到了改善(表 6),其中 B05 浮标站进入台风 7 级风圈的有效风速观测值绝大部分在 13.9~24.4 m·s⁻¹内。

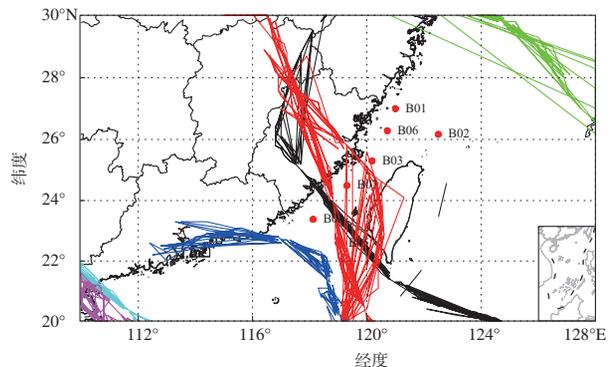


图 14 台风路径与浮标相应关系示意

Fig 14 Schematic diagram of the correspondence between typhoon paths and buoys

图 15、16 分别以 B01 浮标站和 B05 浮标站为

例,结合台风个例期间原始观测与新数据集之间的变化,表明新数据集与台风风圈风速变化更加吻合,

可见缺测时用新数据集中的要素值来填补原始观测要素具有很强的业务适用性。

表 6 数据在台风个例中的检验对比情况

Table 6 Comparison and testing of data in typhoon cases

浮标站	样本数	对比新数据集相关系数	对比新数据集均方根误差	对比原始观测相关系数	对比原始观测均方根误差
B01	51	0.564	1.810	0.535	2.110
B02	79	0.787	1.765	0.739	2.601
B03	34	0.553	1.855	0.514	2.808
B04	37	0.455	1.755	0.339	1.952
B05	30	1.000	0	0.985	0
B06	58	0.764	1.026	0.453	1.468
B07	131	0.585	2.822	0.541	3.019

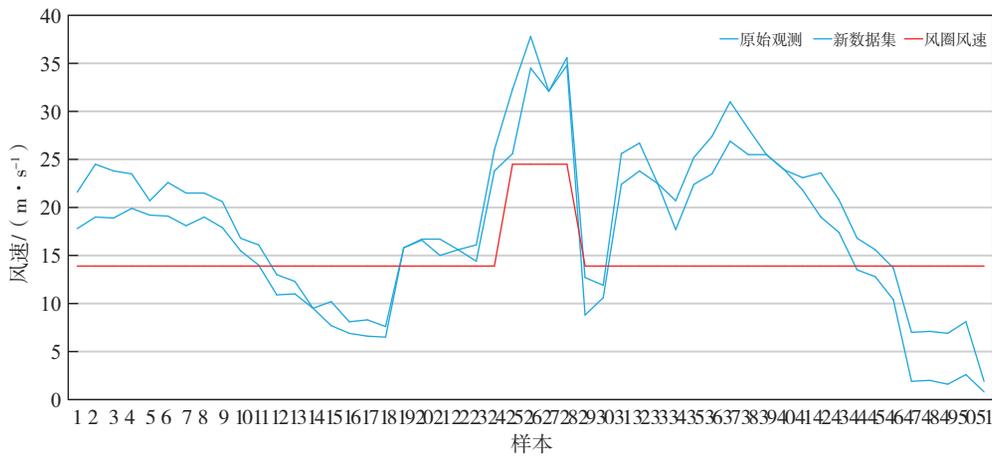


图 15 B01 浮标站原始观测与新数据集风速要素在台风个例中的综合对比
Fig 15 Comprehensive comparison of wind speed elements between the original observations of B01 and the new dataset in typhoon cases

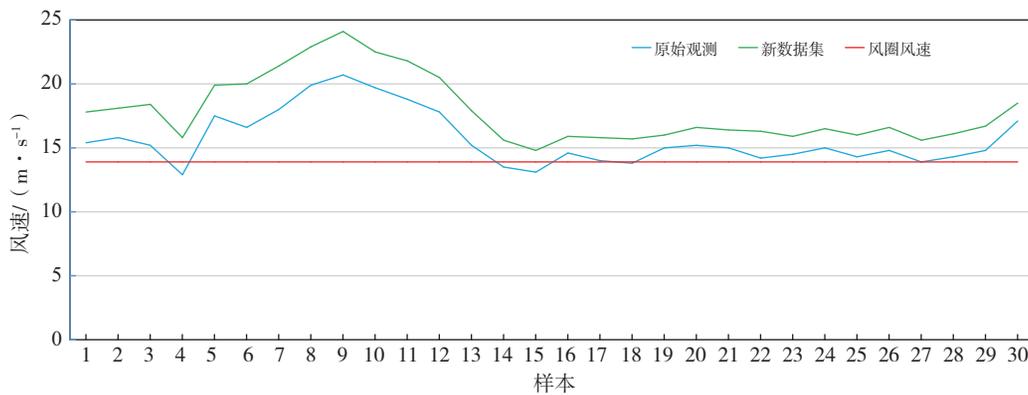


图 16 B05 浮标站原始观测与新数据集风速要素在台风个例中的综合对比
Fig 16 Comprehensive comparison of wind speed elements between the original observations of B05 and the new dataset in typhoon cases

4 结论

在分析 2014—2021 年期间福建海洋浮标观测资料信息的基础上,以气象要素补缺接续为研究对

象,结合 CLDAS、ERA5 等网格资料,分析了浮标位置摆动漂移特点、不同插值方法的应用效果、对比评估检验指标等,得出如下结论:

(1) 通过 K-means 聚类算法,对于游程较短的

经纬度缺测或错误时的位置修订,普遍修正比例在90%以上,且能够保留较好的位置连贯性。但针对缺测时次较多且缺测频繁的情况,应用效果不佳,难以反映真实海洋浮标位置状态。

(2)综合评估表明,根据浮标整体数据情况,选取不同的网格资料作为要素补缺接续的数据来源,能够较好的保障要素接续后的稳定性。针对浮标的高海拔为0 m的特点,IDW 算法稍有技术优势。

(3)浮标要素接续前后的数据对比评估表明,除完整性得到提高外,新数据集与原始观测数据的相关系数普遍在0.9以上,均方根误差也在可控范围内,在时序、空间上保留了更好的要素连贯性,数据质量也进一步得到改善。尤其是对比台风个例,相关系数平均提高了0.03、均方根误差平均下降 $0.5\text{ m}\cdot\text{s}^{-1}$,用新数据集中的要素值来填补原始观测要素具有很好的业务适用性。此外,用新数据集作为原始观测数据的接续来源,目前通过 Grubbs 准则来标识疑似异常,容易出现误判,从而导致过度订正,在尽可能保留原始观测要素信息方面还有很大研究空间。

参考文献:

- [1] 王春晓,王旭,刘长华,等.2010—2018年黄海、东海海域台风数据集[J/OL].中国科学数据,2020,5(4):120-130 [2024-05-13]. <http://www.sciencedb.cn/dataSet/handle/924>.
- [2] 韩玉康,赵艳玲,关吉平,等.台湾及周边岛屿气象水文观测数据集[J/OL].中国科学数据,2020,5(4):152-158 [2024-05-13]. <https://www.scidb.cn/en/detail?dataSetId=720626421566799872>.
- [3] 卢勇夺,王朝阳,王豹,等.我国海洋锚系浮标数据异常值检测方法研究:以 QF110 和 QF306 为例[J].海洋预报,2019,36(6):37-43.
- [4] 张宇,周燕,陶邦一,等.基于时序相关性分析方法的浮标异常数据识别[J].海洋学报,2020,42(11):131-141.
- [5] 张友权.台湾海峡三种再分析海面风场资料的评估[J].渔业研究,2020,42(6):556-571.
- [6] 刘宇.基于海洋时序数据的温度预测与补全方法研究[D].长春:吉林大学,2020.
- [7] 李树文,赵桂香,王一颀,等.基于机器学习的多源实况分析产品和观测数据融合应用试验[J].海洋气象学报,2024,44(1):108-117.
- [8] 崔园园,张强,李威,等.CLDAS 融合土壤相对湿度产品适用性评估及在气象干旱监测中的应用[J].海洋气象学报,2020,40(4):105-113.
- [9] 师春香,潘昉,谷军霞,等.多源气象数据融合格点实况产品研制进展[J].气象学报,2019,77(4):774-783.
- [10] 陈玉玮,黄晓莹.广东区域智能网格实况分析产品的检验评估[J].广东气象,2021,43(1):42-44.
- [11] 孔祥宁,刘焕彬,曹洁,等.HRCLDAS 与 ERA5-Land 气温日统计数据在山东地区的对比评估[J].海洋气象学报,2023,43(2):88-99.
- [12] 周颖,张贺,张珂玮.基于 K-均值聚类方法的大气环流模式 IAP AGCM4.1 对西北太平洋热带气旋的模拟评估[J].大气科学,2020,44(5):1141-1154.
- [13] 黄少平,赖亮,刘海燕,等.基于格点场的江西省气象和水文降水资料对比分析[J].气象与减灾研究,2020,43(2):130-136.
- [14] 熊安元,赵芳,王颖,等.全国综合气象信息共享系统的设计与实现[J].应用气象学报,2015,26(4):500-512.
- [15] 刘首华,陈满春,董明媚,等.一种实用海洋浮标数据异常值质控方法[J].海洋通报,2016,35(3):264-270.
- [16] 杨春华,郑莉,黄河清,等.重庆山地区域气象要素空间插值方法对比[J].气象与环境学报,2022,38(4):57-66.
- [17] 许雯,董美莹,陈锋.基于逐时降水站点资料空间插值方法对比研究[J].气象与环境学报,2017,33(1):34-43.
- [18] 金君,彭思岭,刘启亮,等.中国陆地区域气象要素空间插值方法比较研究[J].工程勘察,2010,38(11):48-51.
- [19] 狄天一.气象领域时间序列配准与缺失值处理方法研究与应用[D].长春:长春工业大学,2019.
- [20] 王艳萍,李新庆,刘垚,等.不同插值算法对气温空间插值效果评估分析[J].信息技术,2020,44(6):31-35.
- [21] 王靖波,潘懋,张绪定.基于 Kriging 方法的空间散乱点插值[J].计算机辅助设计与图形学学报,1999,11(6):525-529.
- [22] 杜宇健,萧德云.Delaunay-固定距离滑动邻域 Kriging 算法[J].工程图学学报,2005,26(2):64-68.