

# 一种基于机器学习的自动气象观测站风向异常识别方法

张志坚<sup>1</sup> 张静<sup>2</sup> 伍光胜<sup>1, 3, 4</sup>

1. 广州市突发事件预警信息发布中心, 广州 511430

2. 广州市气象台, 广州 511430

3. 广州市粤港澳大湾区气象智能装备研究中心, 广州 511430

4. 粤港澳大湾区气象研究院, 广州 510641

**摘要:** 针对自动气象观测站资料存在高隐蔽度的异常风向问题, 基于 DBSCAN 算法建立自动气象观测站风向异常识别方法。选取 2016—2022 年影响广州的寒潮、冷空气、台风等 16 个天气过程个例的自动气象观测站历史风向数据和第 2309 号台风苏拉影响广州期间的自动气象观测站实时风向数据进行风向异常识别检测。分析结果表明, 历史个例的风向存疑站点比例介于 0.46%~5.56%, 风向错误的站点比例介于 0.25%~2.05%; 在“苏拉”实时个例中识别出与地面主导风向存在显著偏离的异常风向站点有 13 个, 造成风向异常原因为风向传感器故障和站点观测环境影响。与传统方法相比, 该方法的风向错误识别准确率提高了 20.32%, 为自动气象观测站历史风向数据质量控制提供了新思路, 同时也为自动气象观测站设备运行监控及现场核查提供了有力的参考依据。

**关键词:** 机器学习, DBSCAN, 自动气象观测站, 风向异常识别, 数据质量控制

## A Method for Detecting Abnormal Wind Direction of Automatic Weather Stations Based on Machine Learning

ZHANG Zhijian<sup>1</sup> ZHANG Jing<sup>2</sup> WU Guangsheng<sup>1,3,4</sup>

<sup>1</sup> Guangzhou Emergency Early Warning Release Center, Guangzhou 511430

<sup>2</sup> Guangzhou Meteorological Observatory, Guangzhou 511430

<sup>3</sup> Guangdong-Hong Kong-Macao Bay Area Meteorological Intelligent Equipment Research Center of Guangzhou, Guangzhou 511430

<sup>4</sup> Guangdong-Hong Kong-Macao Greater Bay Area Academy of Meteorological Research, Guangzhou 510641

**Abstract:** To address the issue of high-concealed abnormal wind directions in automatic weather station (AWS) data, this study establishes an abnormal wind direction identification method based on the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm. Historical wind direction data from 16 weather events affecting Guangzhou between 2016 and 2022, including cold waves, cold air masses, and typhoons, as well as real-time wind direction

资助项目: 广州市科技重点研发计划 (2023B04J0704、2023B04J0667) 共同资助。

第一作者: 张志坚, 主要从事气象观测资料分析研究, Email: 790670233@qq.com

通讯作者: 伍光胜, 主要从事气象观测及气象信息技术研究, Email: 596424966@qq.com

---

data from AWSs during the impact of Typhoon Sura (No. 2309), were used to detect abnormal wind directions. The analysis results reveal that the proportion of AWSs with suspicious wind directions in historical cases ranges from 0.46% to 5.56%, while the proportion of AWSs with erroneous wind directions ranges from 0.25% to 2.05%. During the real-time case of Typhoon Sura, the method identified 13 AWSs with significantly deviating wind directions from the dominant ground wind direction, primarily due to wind direction sensor malfunctions and environmental impacts on AWS observations. Compared to the traditional method, the accuracy of wind direction error identification has improved by 20.32%. The new method provides a novel approach for the quality control of historical wind direction data from AWSs and offers an effective reference for the operational monitoring and on-site verification of AWS equipment.

**Key words:** machine learning, DBSCAN, automatic weather station, detect abnormal wind direction, data quality control

## 引 言

测风资料在气象监测预警、气象服务、气候统计分析、城市通风廊道规划、环境污染气象条件分析、港口大风航行管制等行业场景中发挥着至关重要的作用。由于自动气象观测站建设在城市绿地、高楼楼顶、山区、港口码头、海岛、高速公路等多样化的地理环境中，其分布点多面广，多为无人值守站，给仪器的维护带来了不小的挑战。此外，风向传感器安装高度要求、安装不规范等，进一步增加了常规维护工作的难度。这些客观条件导致仪器故障不能被及时发现、诊断和排除，进而使得自动气象观测站日常维护工作相对滞后，观测数据的质量受到影响。因此，部分站点可能存在数据异常或错误记录风向的情况。自动气象观测站数据风向异常检测与处理是气象数据分析处理过程中不可或缺的一环，通过实施有效的异常值检测与处理措施，可以排除异常风向值，提高数据的准确性和可靠性，为气象监测预报预警服务、气候分析和科学研究提供坚实的数据支撑。

风向异常识别检测方面的相关研究可提高风向数据质量。风向异常检测方法研究按识别检测原理的不同可分为两类，基于观测仪器的硬件检测法和基于观测资料的间接异常表征分析法。观测仪器硬件检测法主要依赖于对格雷码盘风向传感器的故障检测装置（黄小静等，2019），将采集到的格雷码信号输出至处理器，检测判断风向传感器故障，该方法简单、直接有效、准确率高，其缺点是需要进行逐设备现场核查，耗费大量人力、物力和时间成本，且在某些情况下可能难以实施，特别是在偏远或难以到达的站点。间接异常表征分析方法是在不检测观测仪器硬件的情况下，对风向资料缺失、内部一致性、样本分布等进行分析识别检测（郭启云等，2020；刘莹等，2015；杨国彬等，2021；王海军等，2007），该方法通过监控风向某一阈值参数、区间统计量的变化来间接推测数据潜在异常或问题，其实施便捷性、实时性好，可以及时发现自动气象观测站的风向异常情况，但无法有效检测出隐蔽性较高的

---

错误风向数据,例如当风向传感器安装时没有正对正北,或者风向在所有区间均可出现但在某一区间表征不足时,这种方法可能无法准确识别此类隐蔽性高的风向错误。Kaya et al (2023) 基于物联网边缘数据处理提出了一种智能的异常检测方法,可实现对温度、压力、湿度等传感器数据缺失和异常的检测,该方法的不足是必须在仪器端部署数据处理模块,需对仪器进行模块升级,实施便利性差。

随着人工智能技术的发展,机器学习和深度学习方法已经被广泛应用于各个领域。基于机器学习方法的数据异常识别在各行业均有成熟的应用案例,如 DBSCAN 聚类方法在风电(封焯文等,2021)、地震(陈利军和王畅,2020)、交通(阮嘉琨等,2019)、环境监测(潘渊洋等,2012)等行业开展数据异常识别,并取得了较好的应用成效。气象行业中,机器学习和深度学习方法在雷达地物回波识别(李巧等,2024;魏鸣等,2019)、雷达数据异常检测及缺失补全(Gong et al, 2023; Li et al, 2019)、掩星反射信号识别(罗文杰等,2023)、灾害性天气识别(兰宇等,2023;李博勇等,2021;闫文辉等,2020)、天气预报误差订正(任萍等,2020)等方面得到应用,有效提高了大气探测资料数据质量、灾害性天气自动识别和天气预报能力。

针对台风、冷空气等典型天气过程,地面自动气象观测站的风向数据通常呈现出较高的一致性,这为风向异常情况的识别提供了有利条件。现有风向异常检测方法在实际应用中仍存在一些不足,特别是在不依赖观测仪器硬件检测的情况下,基于机器学习建立自动气象观测站风向异常识别方法,其利用聚类分析中的 DBSCAN 算法对自动气象观测站的风向数据进行异常检测和分析,旨在及时发现风向异常的自动气象观测站,从而提高自动气象观测站风向数据的质量。

## 1 资料和方法

### 1.1 数据资料

2016—2023 年广州市区域自动气象观测站 10 min 平均风速和平均风向数据,数据时间分辨率为 5 min。业务在用的风向异常识别结果资料来源广东省气象局的实时常规地面观测质量控制系统,资料时间为 2022 年 10 月 8 日至 2023 年 9 月 30 日,用于与本研究方法的识别结果进行对比。

### 1.2 方法介绍

DBSCAN (density-based spatial clustering of applications with noise) 是一种基于密度的空间聚类算法,该算法将具有足够密度的区域划分为簇,并在具有噪声的数据集中划分任意

---

形状的簇，它将簇定义为密度（核心点邻域半径的最少点数MinPts）相连的点的最大集合，把样本空间中距离（邻域半径 $\varepsilon$ ）相近的聚成一簇。算法基本概念定义（图1）如下：

$\varepsilon$ 邻域：对于 $w_i \in D$ ，其中 $\varepsilon$ 邻域包含样本集 $D$ 中与样本 $w_i$ 的距离不大于 $\varepsilon$ 的子样本集，即：

$$N_\varepsilon(w_i) = \{w_j | \text{dist}(w_i, w_j) \leq \varepsilon\} \quad (1)$$

函数 $\text{dist}(x, y)$ 表示两个样本的直线距离，子样本集的样本数量表示为 $|N_\varepsilon(w_i)|$ 。

三种点类别：

- (1) 核心点：对于任一样本 $w_i \in D$ ，若 $|N_\varepsilon(w_i)| \geq \text{MinPts}$ ，则 $w_i$ 是一个核心点；
- (2) 边界点：对于任一样本 $w_i \in D$ ，如果 $|N_\varepsilon(w_i)| < \text{MinPts}$ ，则 $w_i$ 是一个边界点；
- (3) 噪声点：对于任一样本 $w_i \in D$ ， $w_i$ 既不是核心点也不是边界点，则 $w_i$ 是一个噪声点。

四种点关系：

- (1) 密度直达：如果 $w_j$ 处于 $w_i$ 的 $\varepsilon$ 邻域中，且 $w_i$ 是核心点，则称 $w_i$ 密度直达 $w_j$ ；
- (2) 密度可达：对于样本 $w_i$ 和 $w_j$ ，如果存在样本序列 $x_1, x_2, \dots, x_n$ ，满足 $x_1 = w_i$ ， $x_n = w_j$ 且 $x_1, x_2, \dots, x_{n-1}$ 为核心点且 $x_{n-1}$ 到 $x_n$ 密度直达，则称 $w_i$ 密度可达 $w_j$ ；
- (3) 密度相连：对于样本 $w_i$ 和 $w_j$ ，如果存在核心点样本 $w_{\text{core}}$ ， $w_i$ 和 $w_j$ 均从 $w_{\text{core}}$ 密度可达，则称 $w_i$ 和 $w_j$ 密度相连；
- (4) 非密度相连：对于样本 $w_i$ 和 $w_j$ ， $w_i$ 和 $w_j$ 不属于密度相连关系，则称 $w_i$ 和 $w_j$ 非密度相连。

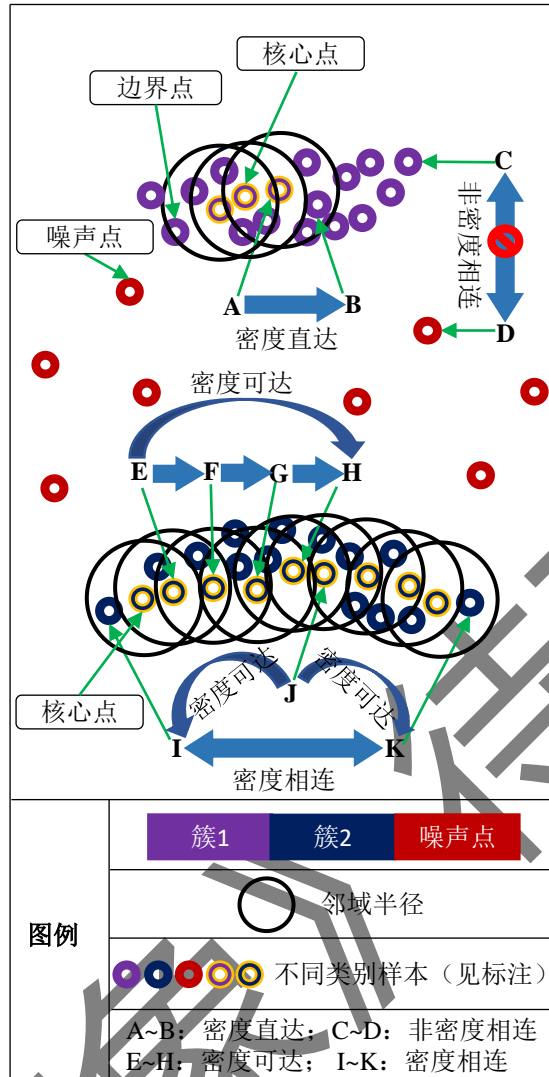


图 1 DBSCAN 算法图解说明

Fig.1 Illustrative explanation of the DBSCAN

### 1.3 技术思路

基于 DBSCAN 算法的聚类分析风向异常识别实现技术思路如图 2 所示。该技术思路主要包含两大核心环节：在地面环境风场风向一致性较好前提下（即有明显主导风向），构建风向异常识别模型和开展风向异常识别。

在构建风向异常识别模型的过程中，首先利用历史台风、冷空气和寒潮等天气过程的风向数据，对 DBSCAN 算法进行细致的参数调整。通过多次试验和效果评估，确定最适合风向异常识别的参数组合。根据调参结果，构建高效的风向异常识别模型。

在开展风向异常识别的过程中，首先启动对典型天气过程的风向异常检测机制，获取自动气象观测站逐时次的风向数据，使用已构建好的 DBSCAN 算法对逐时次风向数据进行聚类分析，通过计算数据点之间的密度和可达性，将风向数据划分为不同的聚类，从而识别出

潜在的异常风向。保存每次聚类分析的结果，以便后续进行对比和分析。接着，对每个自动气象观测站的风向聚类结果进行详细的统计，统计异常聚类的次数和比例。根据统计结果，准确识别出异常次数比例较高的站点，即风向数据存在异常的站点。

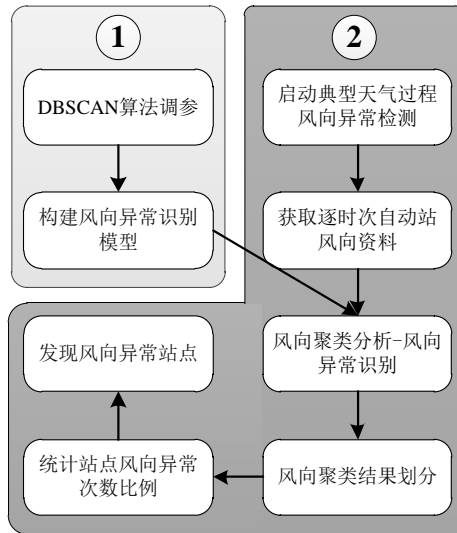
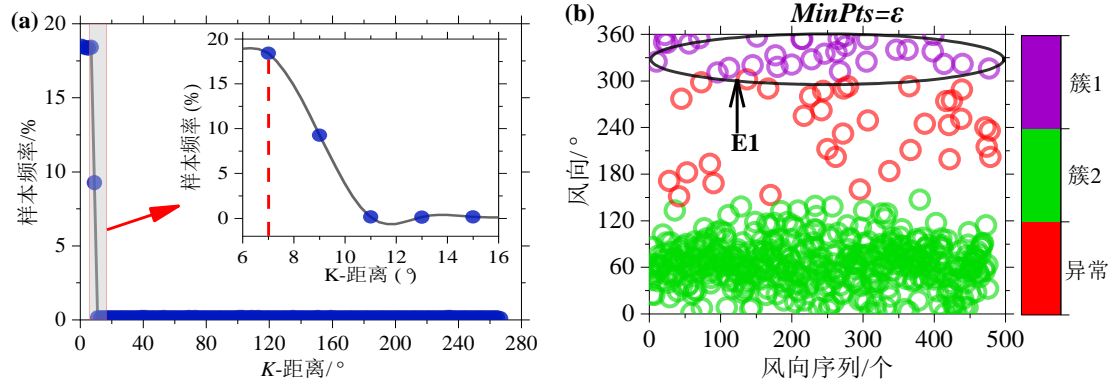


图 2 DBSCAN 算法风向异常识别实现技术思路

Fig.2 The technical approach to identify abnormal wind direction of the DBSCAN

在算法调试训练阶段，使用了 2023 年 1 月 24—26 日广州的冷空气过程以及 2023 年 7 月 16—17 日的台风过程风向数据，样本数据有 1440 时次共 639361 个风向数据，按时次计算风向样本  $K$ -距离（ $K$ -距离是样本点  $w_i$  到除  $w_i$  点外所有点之间距离最近的距离），统计  $K$ -距离的样本分布情况（图 3a），在  $K$ -距离为 7 时样本频率开始出现急剧下降拐点，根据 DBSCAN 算法的聚类原理，明显拐点位置附近对应较好的参数，因此邻域半径参数  $\epsilon$  设置为 8。密度参数  $MinPts$  分别取值为  $\epsilon$  的 1 倍、2 倍和 3 倍，得出算法的异常识别结果如图(3b~3d)所示。 $MinPts$  取值等于  $\epsilon$  时，导致部分异常样本聚成一小簇(E1 标记处)； $MinPts$  取值 3 倍  $\epsilon$  时，导致本应归属簇 1 的部分样本标记为异常样本（E2、E3 标记处）。当  $MinPts$  取值 2 倍  $\epsilon$  时，DBSCAN 算法的风向异常分类相对合理，最终确定算法的密度参数  $MinPts$  设为 16。



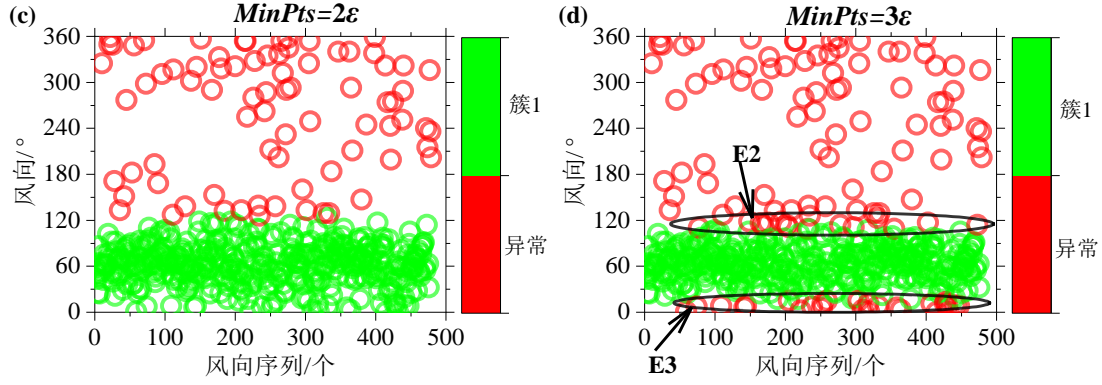


图3 (a) 2023年1月24—26日和2023年7月16—17日样本频率随K-距离分布, (b,c,d) 2023年7月17日12:00不同密度大小的风向异常识别结果

Fig.3 (a) Sample frequency distribution with K-distance from January 24th to 26th, 2023 and from July 16th to 17th, 2023, and classification results of wind direction anomaly identification results with different density levels (b, c, d; data time: 2023-07-17 12:00)

站点风向异常次数比例 ( $P_{AWD}$ ) 计算方法为算法识别出的风向异常总时次 ( $N_{AWD}$ ) 除以天气过程检测时段总时次 ( $N_{TOTAL}$ ), 即:

$$P_{AWD} = \frac{N_{AWD}}{N_{TOTAL}} \times 100\% \quad (2)$$

在个例分析经验基础上, 结合实际识别需求将  $P_{AWD} \geq 85\%$  作为站点风向异常判断标准。

将方法对错误风向站点的识别准确率 ( $P_{AC}$ ) 定义为错误站点数量 ( $N_{ERROR}$ ) 与可疑站点数量 ( $N_{SUSP}$ ) 的比例, 即:

$$P_{AC} = \frac{N_{ERROR}}{N_{SUSP}} \times 100\% \quad (3)$$

## 2 历史风向异常识别

选取2016—2022年广州台风、冷空气和寒潮天气等过程个例16个, 对自动气象观测站历史风向数据进行异常检测(表1)并分析发现, 14个天气过程有明显风向错误的站点, 风向错误站点比例介于0.25%~2.05%, 其中风向错误站点较多的过程是2016年1月22—25日寒潮天气过程和2021年12月25—28日冷空气过程, 风向错误站点数量分别为6个和5个。16个不同天气过程中, 0.46%~5.56%的站点被识别为风向数据异常可疑。

表1 2016—2022年广州天气过程个例风向数据异常识别情况

Table 1 Identification of abnormal wind direction in weather event cases in Guangzhou from 2016 to 2022

天气	开始时间/BT	结束时间/BT	风向错误站点/个	风向可疑站点/个
强冷空气	2022年12月15日00时	2022年12月20日00时	4 (0.92%)	2 (0.46%)

2222 号台风尼格	2022 年 10 月 30 日 00 时	2022 年 11 月 4 日 00 时	3(0.69%)	10 (2.31%)
2203 号台风暹芭	2022 年 7 月 2 日 00 时	2022 年 7 月 5 日 00 时	1 (0.25%)	7 (1.77%)
冷空气	2021 年 12 月 25 日 00 时	2021 年 12 月 28 日 00 时	5 (1.31%)	12 (3.13%)
冷空气	2020 年 12 月 13 日 00 时	2020 年 12 月 18 日 00 时	0	8 (2.14%)
2008 号台风海高斯	2020 年 8 月 18 日 17 时	2020 年 8 月 19 日 17 时	2 (0.53%)	9 (2.39%)
冷空气	2019 年 3 月 22 日 00 时	2019 年 3 月 25 日 00 时	1 (0.28%)	7 (1.99%)
冷空气	2019 年 2 月 8 日 00 时	2019 年 2 月 12 日 00 时	1 (0.29%)	3 (0.86%)
1826 号台风山竹	2018 年 9 月 16 日 00 时	2018 年 9 月 17 日 00 时	0	13 (3.76%)
寒潮	2018 年 1 月 5 日 00 时	2018 年 1 月 13 日 00 时	1 (0.31%)	7 (2.17%)
冷空气	2017 年 11 月 17 日 00 时	2017 年 11 月 25 日 00 时	1 (0.33%)	12 (3.93%)
1715 号台风帕卡	2017 年 8 月 27 日 00 时	2017 年 8 月 28 日 00 时	4 (1.31%)	17 (5.56%)
1714 号台风天鸽	2017 年 8 月 23 日 00 时	2017 年 8 月 24 日 00 时	3 (0.98%)	11 (3.59%)
1624 号台风海马	2016 年 10 月 2 日 00 时	2016 年 10 月 22 日 00 时	3 (1.01%)	12 (4.05%)
1606 号台风妮妲	2016 年 8 月 2 日 00 时	2016 年 8 月 3 日 00 时	1 (0.34%)	13 (4.38%)
寒潮	2016 年 1 月 22 日 00 时	2016 年 1 月 25 日 00 时	6 (2.05%)	15 (5.14%)

注：括号内数据为比例。

为直观分析自动气象观测站的错误风向特征,选取过程个例前后时间长 1~3 个月的逐时风向序列进行分析,错误风向特征表现为:(1)风向长期在两个值区间变化。以 G3162 自动气象观测站为例(图 4a),在 2017 年 8 月 1—30 日,风向始终在 230°~305°波动,8 月 31 日至 9 月 12 日,风向跳转至 0°~14°,风向不变化不符合正常的风向变化规律。(2)风向长期在单个值区间变化。例如自动气象观测站 G1061(图 4b),在 2016 年 10 月 1 日至 11 月 30 日,风向始终局限于 0°~180°,而 180°~359°内则完全无风向分布,这种明显的风向长时间集中现象,与正常的风场变化规律极不相符。(3)风向长期维持恒定值不变。如 G3147 自动气象观测站(图 4c),该站在 2017 年 8 月 1 日至 9 月 30 日的两个月内,风向始终维持



恒定值 240°不变，这种恒定不变的风向数据，显然也与实际风向变化规律不符。(4) 大多数风向长期在单个值区间变化，仅转折天气过程出现极少数风向游离于值区间外。如 G3319 自动气象观测站(图 4d)，在 2021 年 12 月 1—30 日，大多数风向一直处于 180°~240°变化，仅在冷空气天气过程期间出现极少数风向零散分布在 0°~180°和 240°~359°，风向转变连续性特征不明显。

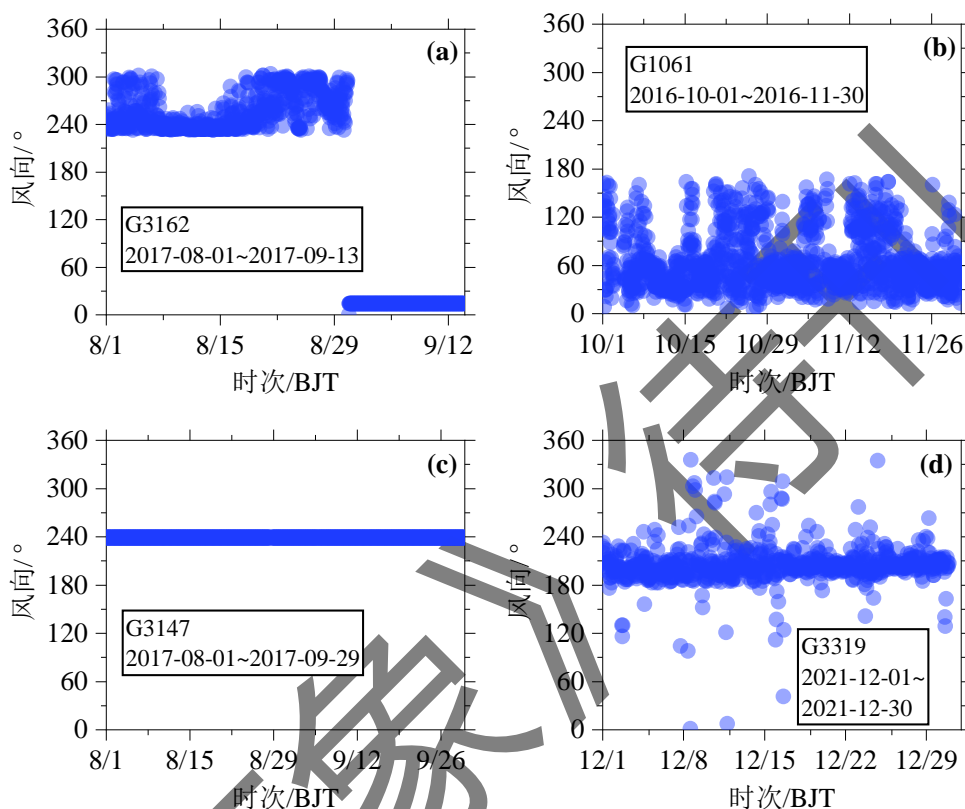


图 4 2016—2022 年广州自动气象观测站风向错误个例

Fig.4 Example of automatic weather stations with incorrect wind direction in Guangzhou from 2016 to 2022

风向可疑自动气象观测站个例如图 5 所示，风向呈现可疑特征表现为站点风向长期与主导风向显著偏离，平均偏差超过 120°，但风向转变连续，分析可能为观测环境影响形成该区域独特的风环境或风向传感器安装时没有指向正北方位等造成。由于历史风向数据无法进行现场核查，无法确定这些站点风向可疑的确切原因，故标志为风向可疑站点。

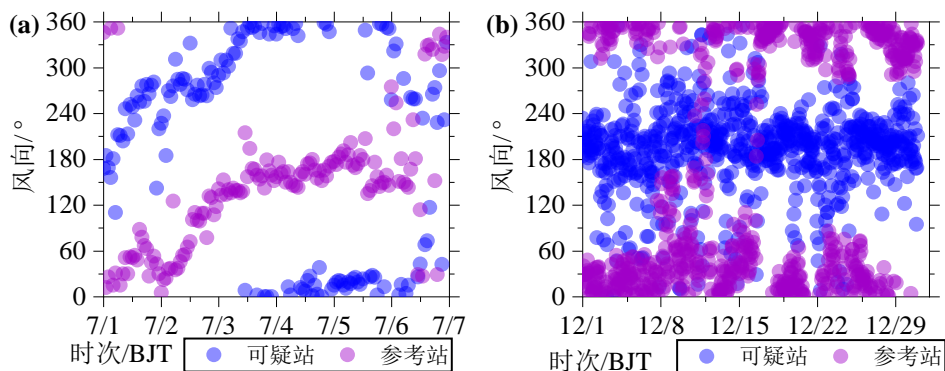


图 5 2016—2022 年广州自动气象观测站风向可疑个例

(a) 可疑站 G3144 和参考站 59294 数据时间均为 2022 年 7 月 1 日 00 时至 6 日 23 时；(b) 可疑站 G3355 和参考站 59285 数据时间均为 2021 年 12 月 1 日 00 时至 30 日 23 时

Fig.5 Suspicious Cases of wind direction at automatic weather stations in Guangzhou from 2016 to 2022.

(a) The data time of the suspicious station G3144 and the reference station 59294 is from 00:00 on July 1, 2022 to 23:00 on July 6, 2022; (b) The data time of the suspicious station G3355 and the reference station 59285 is from 00:00 on December 1, 2021 to 23:00 on December 30, 2021.

### 3 实时风向异常识别

第 2309 号台风苏拉影响期间，广州处于台风外围影响区域，绘制 441 个自动气象观测站逐 6 h 整点时刻风玫瑰图（图 6）。台风影响前期（图 6a~6c）、中期（图 6d, 6e）和后期（图 6f），地面自动气象观测站风向一致性高，超过 83% 站点风向分布在小于 90° 的扇区，符合 DBSCAN 算法聚类分析异常识别条件。在“苏拉”影响前期，地面主导风向为北风，风速逐渐增大，9 月 1 日 15 时（北京时，下同）开始出现  $12 \text{ m s}^{-1}$  以上大风，说明台风对广州的影响开始逐渐加强；影响中期，地面主导风向发生了明显的变化，由北风转为东北风，出现  $12 \text{ m s}^{-1}$  以上大风的站点比例增多；随着台风远离，地面风速开始减小，主导风向仍为东北风。

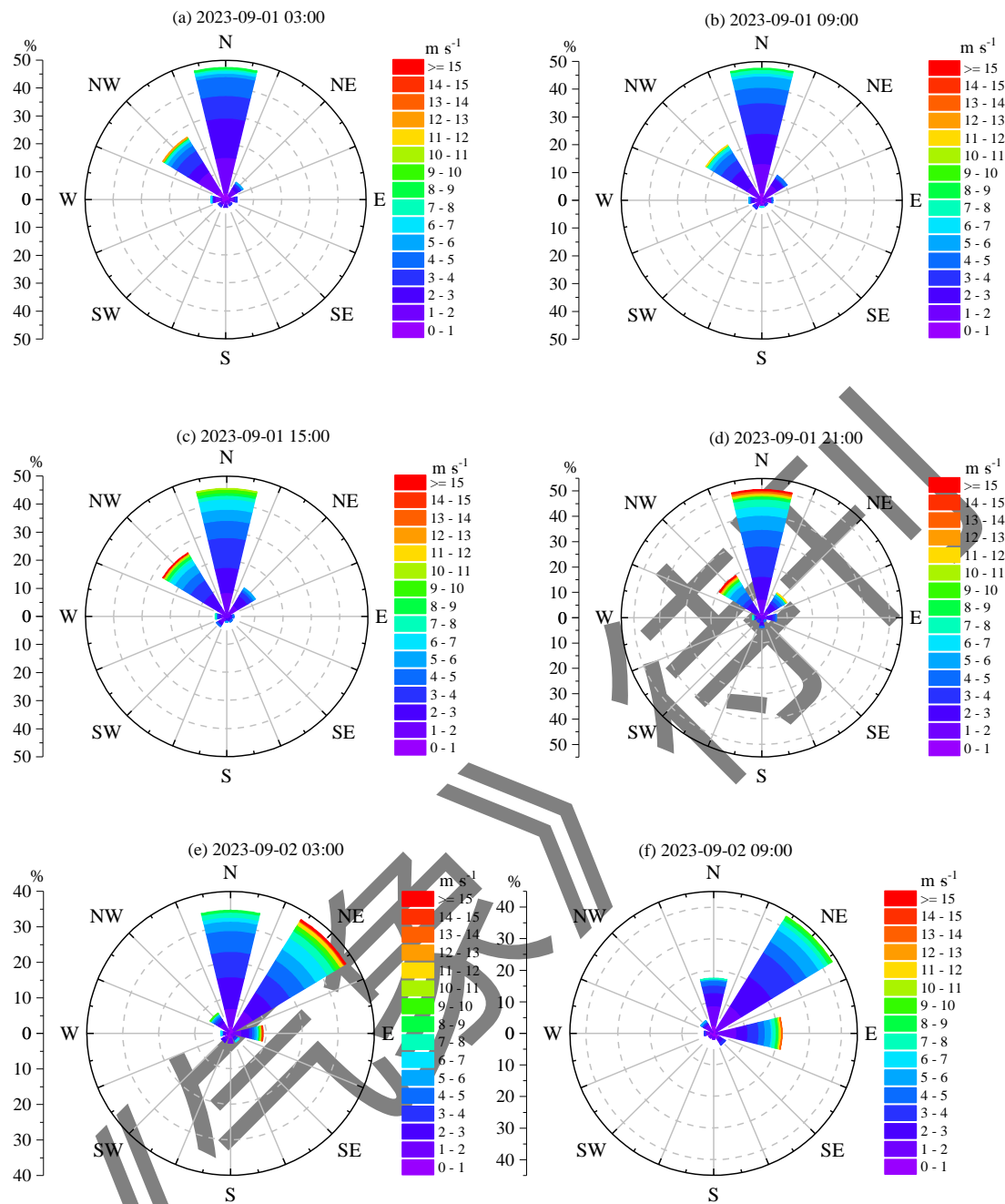


图6 2023年台风苏拉过程期间广州自动气象观测站逐6h风玫瑰图

Fig.6 The wind rose chart of Guangzhou automatic weather stations every 6 hours in the case of typhoon Sula in

2023

与逐6h风玫瑰图对应的风向异常识别结果如图7所示。DBSCAN算法均将9月1日03时、09时、15时和9月2日03时、09时的风向序列划分为2簇，簇内的风向序列呈现出高密度分布的特点，而孤立分布在两簇之间的风向值则被识别为异常值。在9月1日21时，风向分布的高密度区间被算法划分为3簇，同时，那些孤立分布风向值也被识别为异常值。

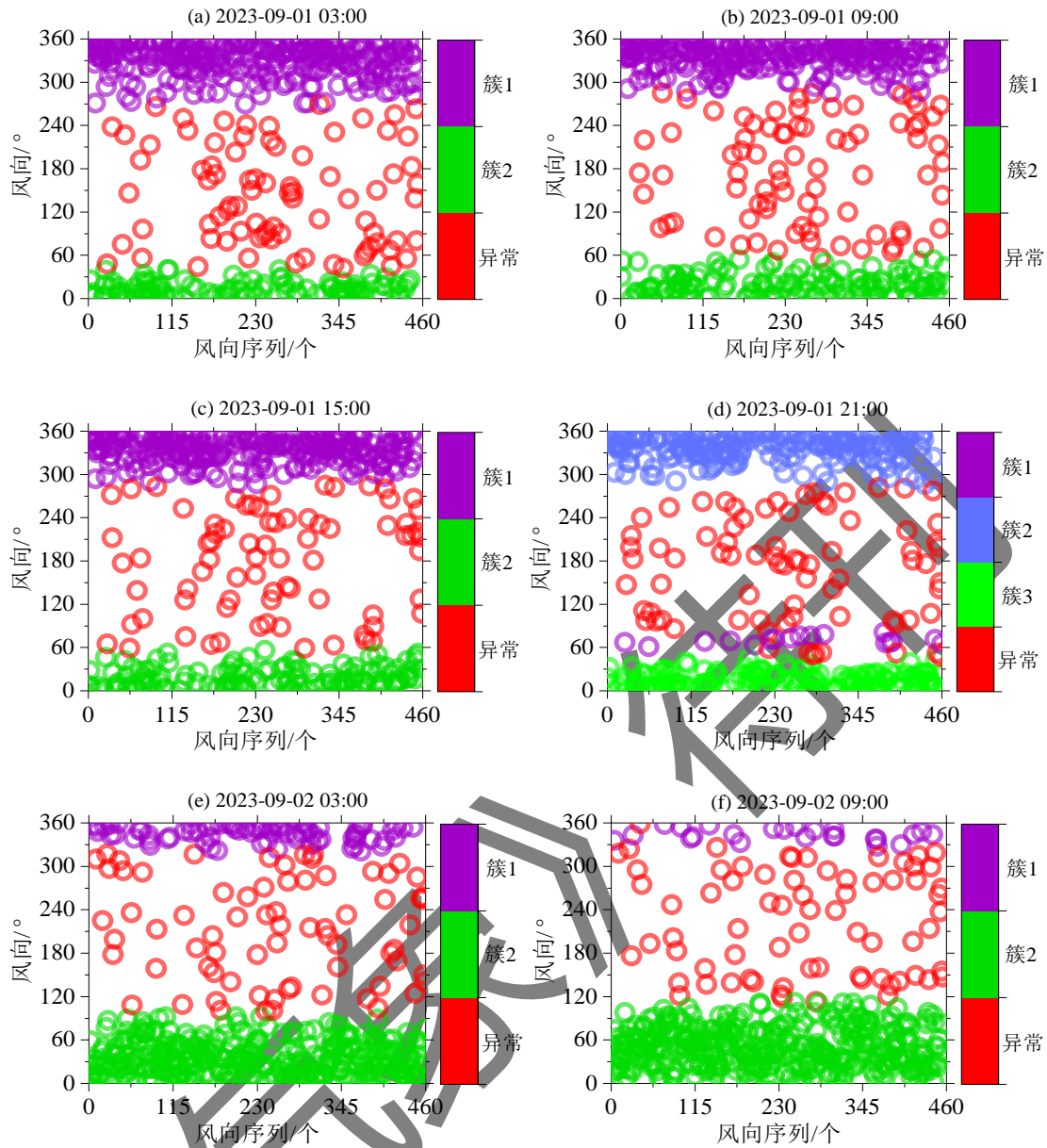


图 7 2023 年台风苏拉过程期间广州自动气象观测站逐 6 h 风向异常识别结果

Fig.7 The results of identifying abnormal wind directions every 6 hours in the case of typhoon Sula in 2023

在台风苏拉过程影响期间，识别出的风向异常自动气象观测站共 15 个（表 2）。为检验识别结果的准确性，采用现场核查方法并对异常原因进行了深入分析。经现场核查，确认风向错误的自动气象观测站有 3 个，由风向传感器故障所致，故障的风向传感器无法准确捕捉风向变化，导致记录的风向数据与实际情况存在较大偏差。另外，受观测环境影响导致该站风向与地面主导风向不一致的站点有 12 个，这些站点位于复杂地形区域、建筑物密集区或受到其他遮挡物的干扰，导致测量站点风向与地面主导风向偏差显著。下文选取风向传感器故障和受站点观测环境影响个例进行详细分析。

表 2 2023 年台风苏拉过程识别出的广州风向异常自动观测气象站

Table 2 Automatic weather stations with identified abnormal wind directions in the case of typhoon Sula in 2023

序号	站号	站名	异常次数/次	异常次数百分比/%	现场核查情况
1	G3369	从化区吕田镇新联村	507	99.41	传感器故障
2	G3224	越秀区建设街黄华路	503	98.63	观测环境影响
3	G3251	花都花山镇东湖村	502	98.43	观测环境影响
4	G3219	白云区景泰街广园中路	500	98.04	观测环境影响
5	G3229	黄埔区新龙镇九龙二小	487	95.49	观测环境影响
6	G3326	从化广外实验小学	477	93.53	观测环境影响
7	G3145	番禺区大石街会江村	476	93.33	观测环境影响
8	G3350	从化区温泉镇中田村	466	91.37	观测环境影响
9	G9507	白云区兴华街白云山风景区	464	90.98	传感器故障
10	G9725	白云区同和街白云山下坑水库	454	89.02	传感器故障
11	G3417	花都梯面镇王子山管理处	448	87.84	观测环境影响
12	G3309	海珠区琶洲街赤沙路	445	87.25	观测环境影响
13	G3243	增城派潭镇榕树下村	440	86.27	观测环境影响
14	G9761	从化太平三百洞气象观测站	437	85.69	观测环境影响
15	G3418	花都花东镇狮前村	434	85.10	观测环境影响

## 2.1 传感器故障个例

对比风向异常自动气象观测站 G9725 与其附近自动气象观测站 G9722 的风速、风向，绘制台风影响期间逐小时风速、风向分布（图 8）。通过对比，发现两者风速变化趋势基本一致，然而 G9725 站与 G9722 站风向存在显著差异。参考自动气象观测站 G9722 的风向为偏北风，与地面自动气象观测站主导风向一致，表明 G9722 站在台风影响期间能够较为准确地反映风向的变化情况。相比之下，G9725 站的风向分布则显得异常，该站风向主要分布在南风至西风区间，与 G9722 站点风向偏离较大甚至相反，故算法识别该站为异常站点。

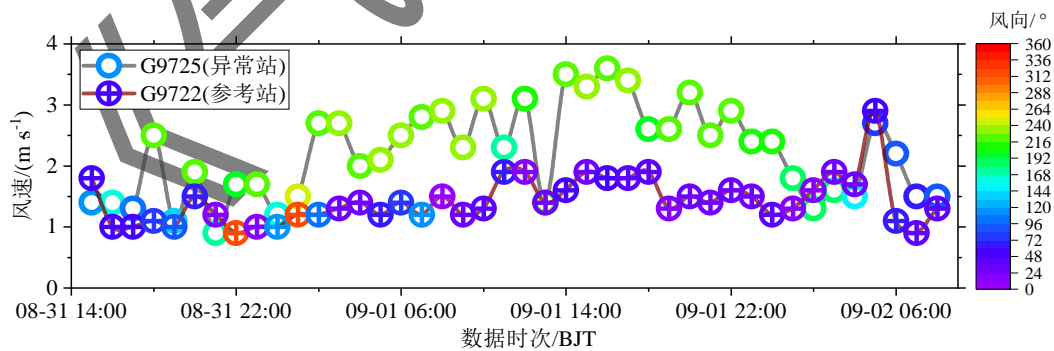


图 8 2023 年 8 月 31 日 15 时至 9 月 2 日 09 时自动气象观测站 G9725、G9722 逐时风速风向序列分布

Fig.8 Hourly wind speed and direction sequence distribution of automatic weather stations G9725 and G9722 from

15:00 on August 31, 2023 to 09:00 on September 2, 2023

现场核查异常站点 G9725 风向传感器，以正北为基准，按固定间隔 10°顺时针旋转风向

标，风向标所处位置为实际风向，记录旋转过程自动气象观测站采集器输出的风向数据，实际风向与输出风向对比如图 9 所示。采集器在 90~359° 输出的风向与实际风向不一致，为此判断风向传感器存在故障。现场更换风向传感器后测试采集器输出风向与实际风向一致。

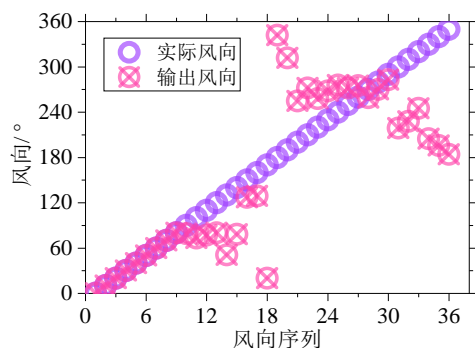


图 9 自动气象观测站 G9725 风向现场核查情况

Fig.9 On-site verification of wind direction at automatic weather stations G9725

## 2.2 观测环境影响个例

分析异常自动气象观测站 G3224、参考站 G1053 风向可知，站点 G1053 风向为偏北风，与地面自动气象观测站主导风向一致，而 G3224 风向一直处于偏南风，两者风向相反（图 10a）。2023 年 9 月 2 日 11 时，现场核查风向标实际方位与采集器输出风向一致（图 10b），风向传感器运行正常。现场核查发现自动气象观测站 G3224 四周有建筑物、大树遮挡（图 10c），南边距离自动气象观测站约 10 m 处为校园十字路口，风从南向街巷吹来，自动气象观测站观测的风向正确。由于周边建筑和植被等因素影响使风受到挤压、阻挡和引流，该处形成独特的街区风环境，导致局地风向与地面主导风向不一致；城市建筑物也迫使城市地面风速变化不均匀，该处风速与参考站相比偏小。

特殊下垫面影响易对识别方法造成误判，由于观测环境一般变化较小，在开展业务应用时可对受下垫面影响较大站点进行剔除标注。街区风场扰动现象提供一个重要启示，为体育赛事等重大活动（刘郁珏等，2022）提供气象服务时，应充分考虑活动现场下垫面对风场的阻挡、拖曳、扰乱和绕流等影响效应（向杰勋等，2019），必要条件下在活动现场（如开幕式体育馆）及附近开展精细化气象监测。

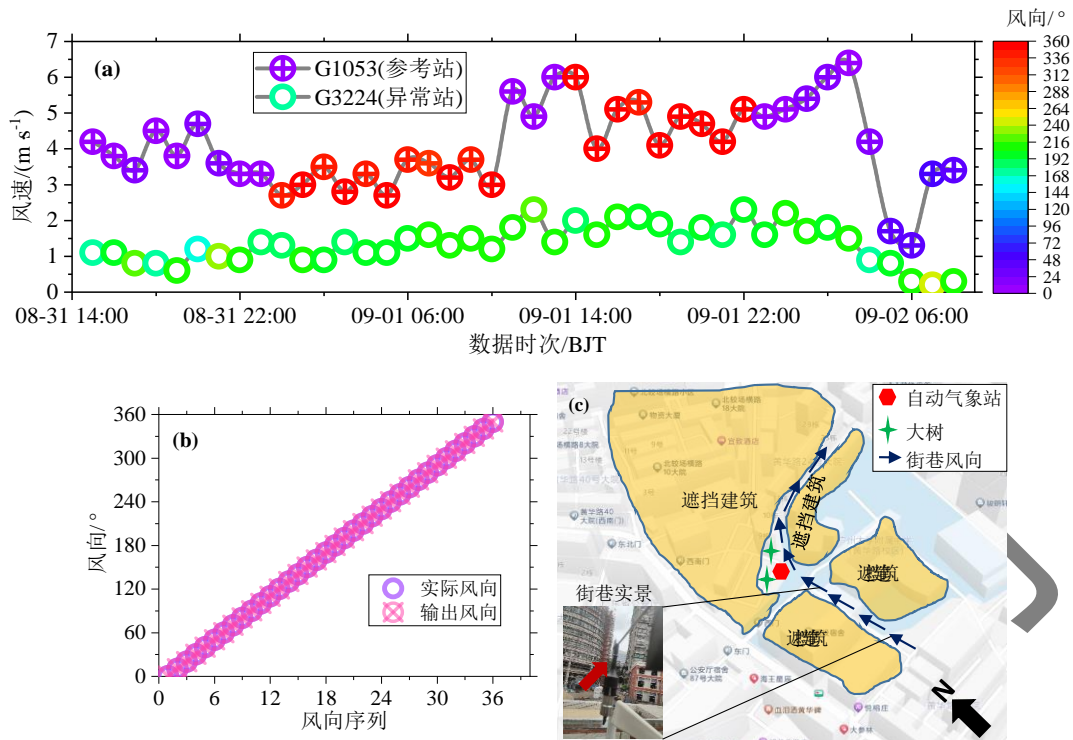


图 10 (a) 2023 年 8 月 31 日 15 时至 9 月 2 日 09 时自动气象观测站 G1053、G3224 逐时风速风向序列分布, (b) 现场风向核查和 (c) 观测环境核查

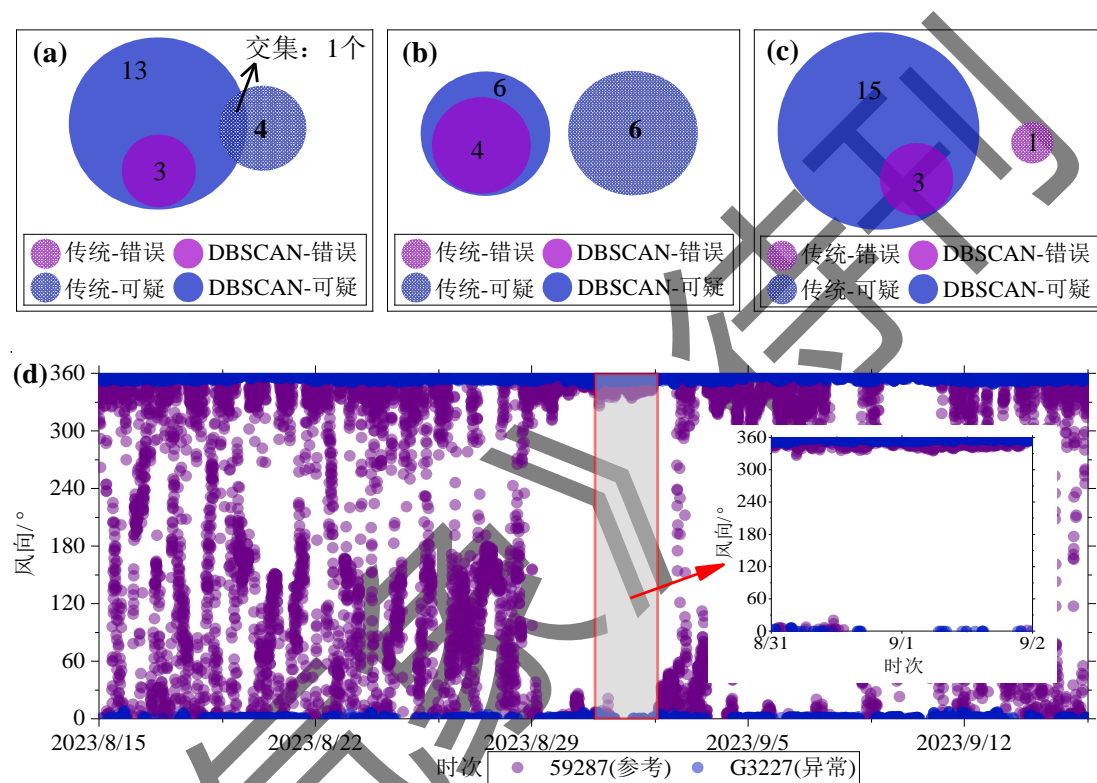
Fig.10 (a) Distribution of hourly wind speed and wind direction sequences of automatic meteorological observation stations G1053 and G3224 from 15:00 on August 31, 2023 to 09:00 on September 2, 2023, (b) On-site wind direction verification and (c) Observation environment verification.

## 4 对比分析

对比当前业务在用的风向质量控制方法（以下称传统方法），风向异常识别结果如图 11 所示。在台风尼格（图 11a）过程中，传统方法识别出 4 个可疑站点，而基于 DBSCAN 算法则识别出 13 个，两种方法识别结果均为可疑的有 1 个站；DBSCAN 算法识别结果中有 3 个站点风向存在明显错误特征，而传统方法并未发现任何错误风向站点。2022 年 12 月 15—19 日强冷空气（图 11b）过程中，现用质量控制方法识别出 6 个可疑站点，但分析这些站点并未表现出明显的错误特征；DBSCAN 算法同样识别出 6 个可疑站点，但其中有 4 个站点的风向存在明显的错误特征。此次过程中，两种方法识别的可疑站点并无交集。台风苏拉（图 11c）过程中，传统方法成功识别出 1 个风向有明显错误特征的站点，但 DBSCAN 算法却未能识别出这一错误；DBSCAN 算法识别出 15 个可疑站点，其中 3 个站点的风向存在明显错误。

分析传统方法未能识别的错误站点风向数据，其特征与前文分析得出的错误风向（1）、

(2)、(4) 特征相符合，站点风向存在变动，但这种变化并不符合正常的风向变化规律。分析 DBSCAN 方法未识别出的错误站点原因（图 11d 所示），在 8 月 31 日至 9 月 2 日的短期内，异常站与参考站的风向分布呈现出高度的一致性；而在 8 月 15 日至 9 月 15 日的长期序列中，与参考站相比，异常站的风向一直维持在正北风向附近，其错误特征明显。由此可见，站点风向资料的时间窗口期对 DBSCAN 算法识别结果有一定影响。另外，DBSCAN 算法误识别的站点均受到了站点观测环境的影响，其风向与主导风向的不一致性是导致误识别的主要原因。



注：图 (d) 红色标注框为 DBSCAN 算法识别的时间窗口期（8 月 31 日至 9 月 1 日），局部放大图为时间窗口期参考站 59287

和异常站 G3227 的逐 5 min 风向序列对比，两个站的风向在时间窗口期内均为北向

图 11 (a~c) 传统方法与 DBSCAN 方法的风向异常识别结果对比 (a: 台风尼格, 2022 年 10 月 30 日至 11 月 3 日; b: 强冷空气, 2022 年 12 月 15—19 日; c: 台风苏拉, 2023 年 8 月 31 日至 9 月 2 日) 及 (d) 时间窗口期对 DBSCAN 算法识别结果影响分析

Fig.11 (a~c) Comparison of wind direction anomaly detection results between traditional methods and DBSCAN method (a: typhoon Nigog, October 30, 2022 to November 3, 2022; b: strong cold air, December 15, 2022 to December 19, 2022; c: typhoon Sula, August 31, 2023 to September 2, 2023) and (d) analysis of the impact of the time window period on the identification results of the DBSCAN algorithm

传统方法在这三次过程中共识别出风向可疑站点 11 个，风向错误站点 1 个，计算得出



的识别准确率为 9.09%。相比之下，DBSCAN 算法表现更为优异，共识别出风向可疑站点 34 个，其中确认的风向错误站点达到 10 个，识别准确率提升至 29.41%。DBSCAN 算法相较于传统方法在识别风向异常方面的显著提升，其识别准确率提高了 20.32%。在典型天气个例中，DBSCAN 算法在识别错误风向中相较于传统方法具有更高的准确性。

表 3 识别准确率统计对比

Table 3 Statistical comparison of recognition accuracy

类别	站点数量/个	准确率/%
传统方法	可疑站点	11
	错误站点	1
DBSCAN 方法	可疑站点	34
	错误站点	10

## 5 结论

基于 DBSCAN 聚类分析算法提出了一种自动气象观测站风向异常识别方法，并在自动气象观测站历史风向数据和实时风向数据异常识别两个场景进行应用分析，并与业务现用的方法相对比，得出结论如下：

(1) 在第 2309 号台风苏拉影响过程期间，方法成功识别出与主导风向偏离大的异常风向站点 15 个，错误风向站点识别准确率为 20%，识别结果为自动气象观测站设备运行监控和现场核查提供有力的参考依据，有助于及时发现并解决仪器问题，提升气象数据的准确性和可靠性。

(2) 历史风向数据异常识别应用表明，方法在历史台风、冷空气和寒潮过程中可有效识别出错误风向的自动气象观测站，16 个过程个例中，有 87.5% 的个例识别出错误风向的自动气象观测站，不同个例的风向错误站点比例介于 0.25%~2.05%。该方法为自动气象观测站历史风向数据质量控制提供了一种新的有效思路和方法。

(3) 与现用的方法相比较，该方法在风向异常识别方面表现更为出色，风向错误识别准确率提高了 20.32%。

值得注意的是，本方法的使用具有一定的条件限制，它要求地面自动气象观测站风向一致性程度高。此外，选取时间窗长短也对识别结果有影响。虽然该方法在广州区域的自动气象观测站风向异常识别研究中取得了良好效果，但当区域面积扩大后，不同区域主导风向可能不一致，风向异常识别方法能否正确、有效划分风向簇，还需要进一步研究和分析。

## 参考文献:

- 陈利军, 王畅, 2020. 基于 DBSCAN 的地震电离层扰动异常数据检测方法[J]. 地震工程学报, 42(2): 410-415. Chen L J, Wang C, 2020. Detection method for seismic ionospheric disturbance anomaly data based on DBSCAN[J]. China Earthq Eng J, 42(2): 410-415 (in Chinese).
- 封焯文, 朱世平, 赵志华, 等, 2021. 风功率异常数据检测方法对比研究[J]. 电工电能新技术, 40(7): 55-61. Feng Z W, Zhu S P, Zhao Z H, et al, 2021. Comparative study on detection methods of wind power abnormal data[J]. Adv Technol Electr Eng Energy, 40(7): 55-61 (in Chinese).
- 郭启云, 钱媛, 杨荣康, 等, 2020. L 波段探空雷达测风质量控制方法研究[J]. 大气科学学报, 43(5): 845-855. Guo Q Y, Qian Y, Yang R K, et al, 2020. Study on the quality control method of wind measurement of L-band sounding radar[J]. Trans Atmos Sci, 43(5): 845-855 (in Chinese).
- 黄小静, 陈涛, 李晓红, 等, 2019. EL15-2C 型风向传感器现场故障排除实例分析[J]. 气象水文海洋仪器, 36(1): 68-71. Huang X J, Chen T, Li X H, et al, 2019. Case analysis for field troubleshooting of EL15-2C wind direction sensor[J]. Meteor Hydrol Marine Instrum, 36(1): 68-71 (in Chinese).
- 兰宇, 罗聪, 伍志方, 等, 2023. 三种机器学习方法在广东雷暴大风自动识别的应用效果评估[J]. 热带气象学报, 39(2): 256-266. Lan Y, Luo C, Wu Z F, et al, 2023. The assessment of application effectiveness of three machine learning methods in automatic identification of thunderstorm gale in Guangdong[J]. J Trop Meteor, 39(2): 256-266 (in Chinese).
- 李博勇, 胡志群, 郑佳峰, 等, 2021. 利用贝叶斯方法改进华南地区冰雹识别效果[J]. 热带气象学报, 37(1): 112-125. Li B Y, Hu Z Q, Zheng J F, et al, 2021. Using Bayesian method to improve hail identification in South China[J]. J Trop Meteor, 37(1): 112-125 (in Chinese).
- 李巧, 戚友存, 张哲, 等, 2024. 基于贝叶斯分类器和回波物理特征的 C 波段雷达非气象回波识别方法和性能分析[J]. 大气科学, 48(3): 823-836. Li Q, Qi Y C, Zhang Z, et al, 2024. Nonmeteorological echoes identification method based on bayesian classifier and echo physical characteristics using C-band radar and its performance[J]. Chin J Atmos Sci, 48(3): 823-836 (in Chinese).
- 刘莹, 王海军, 李中华, 2015. 基于观测数据的风向传感器故障检测方法设计与应用[J]. 气象, 41(11): 1408-1416. Liu Y, Wang H J, Li Z H, 2015. Design and application of fault detection method for wind direction sensors based on observation data[J]. Meteor Mon, 41(11): 1408-1416 (in Chinese).
- 刘郁珏, 黄倩倩, 张涵斌, 等, 2022. 基于大涡模拟的冬奥赛区风环境精细化评估[J]. 应用气象学报, 33(2): 129-141. Liu Y J, Huang Q Q, Zhang H B, et al, 2022. Refined assessment of wind environment over Winter Olympic competition zone based on large eddy simulation[J]. J Appl Meteor Sci, 33(2): 129-141 (in Chinese).
- 罗文杰, 项杰, 杜华栋, 2023. 利用改进的 GoogLeNet 深度学习模型识别 COSMIC-2 掩星信号中的反射信号[J]. 大气科学, 47(3): 631-641. Luo W J, Xiang J, Du H D, 2023. Identification of the reflected signal in the COSMIC-2 occultation signal using the improved GoogLeNet deep learning model[J]. Chin J Atmos Sci, 47(3): 631-641 (in Chinese).
- 潘渊洋, 李光辉, 徐勇军, 2012. 基于 DBSCAN 的环境传感器网络异常数据检测方法[J]. 计算机应用与软件, 29(11): 69-72, 111. Pan Y Y, Li G H, Xu Y J, 2012. Abnormal data detection method for environment wireless sensor networks based on DBSCAN[J]. Comput Appl Soft, 29(11): 69-72, 111 (in Chinese).
- 任萍, 陈明轩, 曹伟华, 等, 2020. 基于机器学习的复杂地形下短期数值天气预报误差分析与订正[J]. 气象学报, 78(6): 1002-1020. Ren P, Chen M X, Cao W H, et al, 2020. Error analysis and correction of short-term numerical weather prediction under complex terrain based on machine learning[J]. Acta Meteor Sin, 78(6): 1002-1020 (in Chinese).
- 阮嘉琨, 蔡延光, 乐冰, 2019. 基于 DBSCAN 密度聚类算法的高速公路交通流异常数据检测[J]. 工业控制计算机, 32(7): 92-94. Ruan J K, Cai Y G, Le B, 2019. Highway traffic flow anomaly data detection based on DBSCAN density clustering algorithm[J]. Indust Control Comput, 32(7): 92-94 (in Chinese).
- 王海军, 杨志彪, 杨代才, 等, 2007. 自动气象观测站实时资料自动质量控制方法及其应用[J]. 气象, 33(10): 102-109. Wang H J, Yang Z B, Yang D C, et al, 2007. The method and application of automatic quality control for real time data from automatic weather stations[J]. Meteor Mon, 33(10): 102-109 (in Chinese).

- 
- 魏鸣, 管理, 梁学伟, 等, 2019. 基于支持向量机的雷达地物回波识别研究[J]. 大气科学学报, 42(4): 631-640. Wei M, Guan L, Liang X W, et al, 2019. Ground clutter identification based on the support vector machine method with Doppler weather radar data[J]. *Trans Atmos Sci*, 42(4): 631-640 (in Chinese).
- 向杰勋, 陈桂兴, 姜平, 等, 2019. 冷空气强风在大型城市中的精细结构和形成机制[J]. 大气科学, 43(3): 577-597. Xiang J X, Chen G X, Jiang P, et al, 2019. Fine-scale structures and formation of strong winds over a megacity during a cold surge process[J]. *Chin J Atmos Sci*, 43(3): 577-597 (in Chinese).
- 闫文辉, 黄兴友, 赵钰锦, 等, 2020. 基于改进 DBSCAN 聚类算法的雷暴单体三维结构识别技术介绍[J]. 热带气象学报, 36(4): 542-551. Yan W H, Huang X Y, Zhao Y J, et al, 2020. Introduction of 3D structure detection technology of thunderstorm cell based on improved DBSCAN clustering algorithm[J]. *J Trop Meteor*, 36(4): 542-551 (in Chinese).
- 杨国彬, 郭启云, 舒康宁, 等, 2021. 基于名单控制方法的探空测风数据质量分析[J]. 气象, 47(6): 727-736. Yang G B, Guo Q Y, Shu K N, et al, 2021. Quality analysis of the radiosonde wind observation data based on the list control method[J]. *Meteor Mon*, 47(6): 727-736 (in Chinese).
- Gong A F, Chen H N, Ni G H, 2023. Improving the completion of weather radar missing data with deep learning[J]. *Remote Sens*, 15(18): 4568.
- Kaya S M, İşler B, Abu-Mahfouz A M, et al, 2023. An intelligent anomaly detection approach for accurate and reliable weather forecasting at IoT Edges: a case study[J]. *Sensors*, 23(5): 2426.
- Li W, Zhou W, Wang Y M, et al, 2019. Meteorological radar fault diagnosis based on deep learning[C]//2019 International Conference on Meteorology Observations (ICMO). Chengdu, China: IEEE: 1-4.