

朱亚妮, 杨溯, 张志强, 等. 中国地面自动气象站小时降水质量控制方法. 应用气象学报, 2024, 35(6): 680-691.
DOI: 10.11898/1001-7313.20240604

中国地面自动气象站小时降水质量控制方法

朱亚妮¹⁾ 杨溯^{1)*} 张志强¹⁾ 仇建华²⁾

¹⁾(国家气象信息中心, 北京 100081)

²⁾(南京信息工程大学计算机学院, 南京 210044)

摘要

高时空分辨率自动气象站降水观测作为重要数据来源, 已被广泛应用于强对流天气监测、模式评估、预报复盘等研究工作。仪器故障、特殊天气条件下观测设备的局限性等因素是自动气象站降水数据不确定性的主要来源, 这些问题在无人值守气象站尤为突出。该研究基于 2021—2023 年中国自动气象站实时观测降水量数据、高时空分辨率雷达数据和高灵敏度降水类天气现象数据, 发展适应于中国自动气象站小时降水数据的多源数据协同质量控制方法(multi-source data collaborative quality control, MDC)。通过综合定量指标与典型个例分析, 对 MDC 的应用效果进行全面评估。结果显示: MDC 判别正确率为 99.92%, 错误数据命中率较现行业务提升 39.3%。基于多源降水观测数据时空一致性, MDC 显著提升了晴空降水、融雪性降水和虚假零值降水等异常数据的甄别能力, 有效弥补了传统方法的不足。

关键词: 自动气象站; 1 h 降水量; 多源数据协同检测; 质量控制

引言

高时空分辨率自动气象站观测降水作为重要数据来源, 已被广泛应用于中小尺度短时效天气预报、遥感观测定量估测降水校准评估、水文气象模型验证和气候变化研究等领域^[1-6]。我国气象部门建立了 2400 个国家级地面自动气象站(简称国家站), 这些站点多数为有人值守站, 数据质量可靠, 但分布稀疏, 难以满足日益高涨的精细化预报和特殊行业天气服务需求。2000 年以来, 中国气象局大力开展自动观测站网建设, 目前已完成 7 万多个区域级地面气象站(简称区域站)的布设^[7]。区域站覆盖范围广、站网密度高, 能更精细地反映中小尺度天气系统和短时效灾害性天气系统的演变特征^[8-9], 为各级政府指挥应对气象灾害提供了重要的决策依据^[10]。

受仪器故障、观测环境和设备性能等因素影响,

自动气象站降水观测数据中不可避免地存在不同形式的随机错误和系统错误^[11-15]。例如受限于仪器灵敏度, 微量降水(小于 0.1 mm)会被遗漏, 导致虚假的零值降水^[16]。此外, 灌溉、仪器壁冷凝水以及附着于雨量计的冰雪融化也可能产生虚假降水记录^[17-19]。相比于国家站, 无人值守区域站的设备维护频次较低, 观测数据质量问题尤为复杂^[20-22]。因此, 研制一套适用于自动气象站降水数据的质量控制方法, 对于提升其应用价值, 具有十分迫切的现实意义。

当前, 降水数据质量控制方法主要分为两大类: 第 1 类是基于降水时空变化规律的传统质量控制方法(简称传统方法)^[14, 23-25]。传统方法包括气候界限值检查、台站或区域主要变化范围检查、僵值检查、单一站点时间一致性检查以及与邻近站的空间一致性检查等, 广泛应用于逐日及以上时间分辨率的降水数据。由于小时降水数据的短时效性和强局地性,

2024-07-04 收到, 2024-09-06 收到再改稿。

资助项目: 中国气象局青年创新团队(CMA2023QN08), 中国气象局创新发展专项(CXFZ2023J049), 国家气象信息中心结余资金项目(NMICJY-202313), 四川省科技计划项目(2023YFS0434)

* 通信作者, 邮箱: yangsu@cma.gov.cn

传统方法质量控制能力明显不足,导致在区域站降水数据的质量控制中,漏判漏检问题尤为突出^[26-28]。第 2 类是与其他数据的协同质量控制方法(简称协同方法),主要与高时空分辨率的雷达数据进行协同^[27-34]。协同方法不再局限于降水本身的时空一致性,有效提升了错误数据判识能力。然而,雷达观测数据可能受到雷达波束遮挡或设备故障等不确定性因素的影响^[35-36]。因此,在地形遮挡严重地区或雨带边缘,仅依赖雷达数据的协同方法可能导致较高的误判风险,进而引发雨雪事件漏报和降水量低估^[29,32]。

本研究基于高时空分辨率雷达数据和观测灵敏度高的降水类天气现象(简称降水现象),发展并优化传统方法和协同方法,建立适用于我国自动气象站小时降水数据的质量控制方法。这对于解决自动气象站降水数据的复杂质量问题,提升高时空分辨率降水观测数据应用价值具有重要意义。

1 数据

本研究所用数据时段为 2021 年 1 月 1 日 00:00(世界时,下同)—2023 年 12 月 31 日 23:00。该时段内发生了多种类型降水事件,并呈现出复杂的数据质量问题,为质量控制方法应用效果评估提供了足量样本。

采用中国 7.6×10^4 个自动气象站实时观测的逐小时正点前 1 h 降水量作为质量控制对象,开展方法适用性评估。

采用 2400 个国家站正点前 1 h 降水现象、逐小时雷达定量估测降水产品作为协同方法参考数据。降水现象能够区分降水类型(包括雨、雪、冰雹等)并获取较多粒子信息^[37],经过中国气象局气象数据业务系统(Meteorological Data Operation System, MDOS)实时质量控制^[24]。雷达定量估测降水产品由国家气象信息中心研制,由不同雷达实时统计反射率因子-雨强方法反演而成,用于反演的雷达基数据经过了孤立点消除、径向杂波消除、地物杂波消除、海杂波消除、晴空回波抑制、速度退模糊等质量控制处理^[28]。

采用中国地面小时降水数据集(V3.0)^[38]作为质量控制方法正确数据识别能力评估样本。该数据集包含中国 2400 个国家站的 1 h 降水量数据,经过质量控制与人工审核。

采用国家气象信息中心整理的中国自动气象站 1 h 降水量数据更正报,作为质量控制方法错误数据识别能力评估样本。更正报是基于 MDOS 或应用端的告警信息,台站人工更正并重新上传的观测记录。

本文插图中所涉及的中国国界和行政区域界线基于审图号为 GS(2019)3082 号标准地图制作,底图无修改。

2 质量控制及评估方法

2.1 质量控制方法

基于高分辨率雷达、降水现象和邻近国家站降水数据,研制了多源数据协同的自动气象站小时降水量质量控制方法(multi-source data collaborative quality control, MDC)。MDC 包含自检测、多源数据协同检测和动态黑名单 3 个模块。对待检站 1 h 降水量依次执行以上 3 步检查,最终生成相应质量控制码以标识数据质量。质量控制码包含数字 0、1、2、8 和 14,分别表示数据正确、可疑、错误、缺测和黑名单台站。

2.1.1 自检测模块

自检测基于降水自身气候变化特征,识别异常降水记录。自检测模块包含元数据检查、特征值检查^[39]、气候界限值检查、台站界限值检查和僵值检查。

气候界限值检查阈值采用世界气象组织(WMO)最新发布的 1 h 降水量极值,为 305 mm。超出气候界限值的观测值,质量控制码赋值为 2,表示数据错误。

台站界限值检查和僵值检查阈值基于 2400 个国家站 1 h 降水量历史数据统计得到。界限值检查阈值逐站逐月计算,僵值检查阈值则按降水强度逐站确定,并将站点阈值格点化至 $1^\circ \times 1^\circ$ 网格^[40],形成高空间分辨率阈值参数文件。

2.1.2 多源数据协同检测模块

多源数据协同检测利用同时次雷达估测降水和降水现象数据,对自检测模块判识正确的数据开展进一步判识。当待检站 50 km 范围邻近国家站数量不少于 3 时,开展此项检查。此外,为了避免雷达估测数据误差可能带来的误判风险^[6,32],仅在待检站降水情况(有或无)与所有邻近国家站的降水情况均不一致时,启用雷达定量估测降水进行降水量值

的质量判别。具体检验流程如下:

①当待检站小时降水量 $P=0$ 时,检查该站 50 km 范围内的所有邻近国家站降水量。如果所有邻近站的降水量均大于 0,则进行下一步检查;否则,不进行下一步质量判别。

②在待检站 $5\text{ km}\times 5\text{ km}$ 的范围内,查找是否存在有效的雷达定量估测降水数据(非缺测值,下同)。如果存在,计算该范围内的雷达定量估测降水最大值(记为 R_{MAX})。若 $R_{\text{MAX}}>5\text{ mm}$,则判定待检站的降水观测值错误;若 $R_{\text{MAX}}\leq 5\text{ mm}$,则判定观测值正确。

③当 $P>0$ 时,检查待检站 50 km 范围内是否全部邻近站均无降水和降水类天气现象,如果是,进行下一步检查;否则,不进行下一步质量判别。

④在待检站 $5\text{ km}\times 5\text{ km}$ 的范围内,查找是否存在有效的雷达定量估测降水数据。如果有,则计算与之最接近的雷达定量估测降水(记为 R)和最大值 R_{MAX} 。若 $R_{\text{MAX}}=0$,则判定待检站的降水观测值为错误;若 $R_{\text{MAX}}\neq 0$,则将 P 按照降水强度区间进行分类,并应用雷达定量估测降水与待检站观测值的差值阈值开展质量判别,得到质量控制结果及相应的质量控制码^[28]。若无有效雷达定量估测降水,则判定观测值可疑。

2.1.3 动态黑名单模块

动态黑名单基于数据质量先验性信息,将特定时间段内错误率异常偏高的观测站点临时性停用^[41],依据观测站自检测和多源数据协同模块质量控制结果建立。若在过去 24 h 内,观测站存在 1 次以上错误记录,即被列入黑名单。该站点后续数据仍接受自检测和多源数据协同检测,当站点数据质量改善,不再满足进入黑名单标准时,将恢复正常使用。

2.2 质量控制效果定量评估方法

采用正确率和误剔率,定量评估质量控制方法正确数据识别能力。将中国地面小时降水数据集(V3.0)标记为正确的 1 h 降水量数据设为 B_0 ,对 B_0 进行质量控制后,其被划分为正确数据样本 B_1 和疑误数据样本 B_2 。正确率为 B_1 占 B_0 的百分比,反映质量控制方法对正确数据的判别能力。误剔率为 B_2 占 B_0 的百分比,反映质量控制方法对正确数据的误剔程度。

采用错误数据命中率和漏判率定量评估质量控制方法错误数据识别能力。将台站更正报更正前数据视为错误数据样本 E_0 ,质量控制方法判别数据错误或可疑时,表示命中错误数据,记为 E_1 ;当质量控制方法判别数据正确时,即表示漏判错误数据,记为 E_2 。命中率定义为 E_1 占 E_0 的百分比,反映质量控制方法命中错误数据的能力;漏判率则定义为 E_2 占 E_0 的百分比,表示质量控制方法漏判错误数据程度。

3 质量控制方法应用效果

3.1 质量控制应用效果定量评估结果

3.1.1 正确数据识别能力评估结果

应用 MDC 对 2021—2023 年中国地面小时降水数据集(V3.0)进行质量控制,并对其正确率和误剔率进行评估。2021—2023 年国家站降水观测实有数据共计 6.2005×10^7 站次,其中 4.2×10^4 站次被判为错误。MDC 的正确率达到 99.92%,误剔率为 0.08%。误剔数据的降水量值频率分布显示(图 1),34.64%的误剔数据为零值降水,60.72%为 $[0.1\text{ mm}, 1\text{ mm})$ 的弱降水。随着降水量增加,误剔风险显著下降。不小于 10 mm 小时降水量的误剔

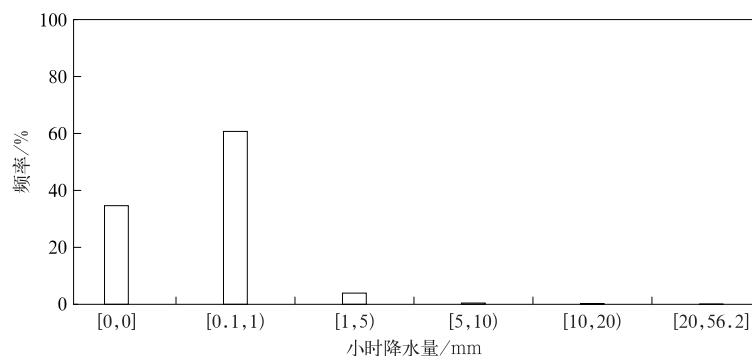


图 1 2021—2023 年误剔数据降水量频率分布

Fig. 1 Frequency distribution of incorrectly excluded data for precipitation values from 2021 to 2023

数据集中分布在西南地区的高山站。这一现象可能归因于高山站降水具有更显著的局地性特征,导致其与邻近站降水及雷达估测降水的时空匹配性较低。

3.1.2 错误数据识别能力评估结果

利用更正报提供的错误数据样本,评估 MDC 错误数据识别能力,并与现行业务 MDOS 实时质量控制方法进行比较。2022—2023 年共收集 1.97×10^5 站次的更正报,包含虚假降水、异常零降水和降水量值错误等 3 类错误数据,分别占总样本量的

85.7%、6.0%和 8.3%。图 2 为更正前与更正后降水量的频率分布。错误数据集中分布在 $[0.1 \text{ mm}, 1 \text{ mm})$,占比达 70.2%, $[1 \text{ mm}, 5 \text{ mm})$ 的小时降水量占比为 15.6%,5 mm 及以上小时降水量的错误数据占比为 8.2%,异常零值降水占比为 6.0%。经台站人工更正后,零值降水量占比提高至 85.7%,5 mm 及以上的小时降水量数据被全部修正,表明原先异常记录的零值降水大多被更正为 0.1~5 mm 范围内降水。

MDC 在中国区域的平均错误数据命中率为

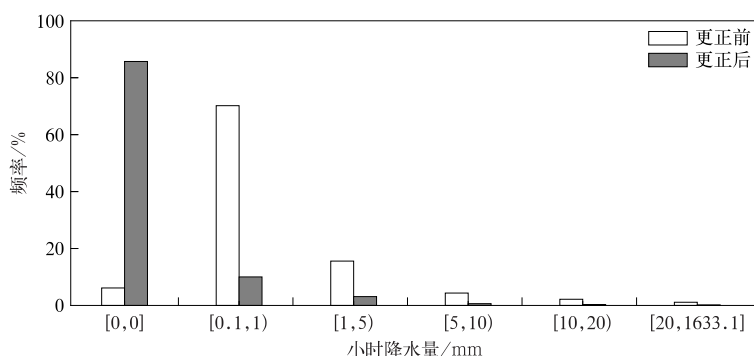


图 2 2022—2023 年更正报更正前后降水量频率分布

Fig. 2 Frequency of precipitation before and after correction report from 2022 to 2023

39.8%,较 MDOS 实时质量控制系统的 0.5%提升 39.3%。由不同降水强度错误数据的命中率(图 3)可知,MDC 对于 $[0 \text{ mm}, 50 \text{ mm})$ 错误数据的判识能力接近,约为 40%。随着降水量增加,命中率显著

提升。降水量不低于 100 mm 时,命中率达到 100%。相比之下,MDOS 实时质量控制几乎无法识别小于 20 mm 的降水错误数据,但对于 20 mm 以上的异常降水具有一定检测能力。

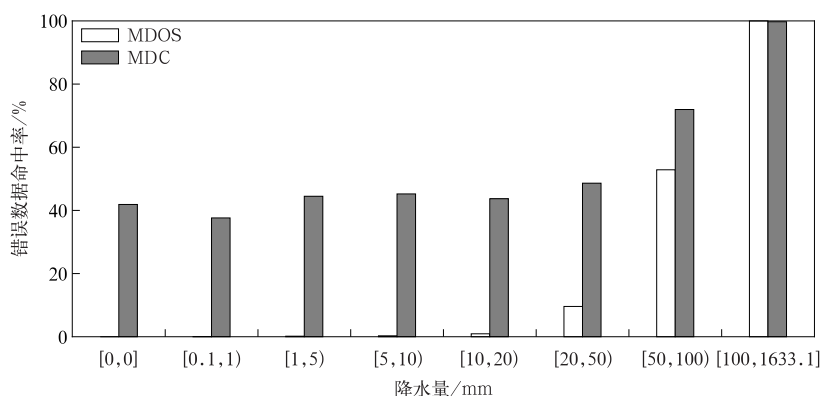


图 3 2022—2023 年 MDOS 和 MDC 不同降水强度错误数据命中率

Fig. 3 Hit rate of erroneous data for different precipitation intensities by MDOS and MDC from 2022 to 2023

由错误数据命中率的的空间分布(图 4)可知,MDOS 实时质量控制方法命中率基本在 10%以下,

而 MDC 命中率的的空间差异显著。在观测站点密集的东部区域,大多数台站错误数据命中率超过

90%。而在观测台站相对稀疏的西部和东北地区，由于不满足多源协同检测应用条件，MDC 命中率显著下降，与 MDOS 实时质量控制接近。由上文可

知，70.2%的错误数据为[0.1 mm, 1 mm)的虚假弱降水。仅使用 MDC 自检测模块未能有效甄别此类错误数据，质量控制效果与现行业务 MDOS 相当。

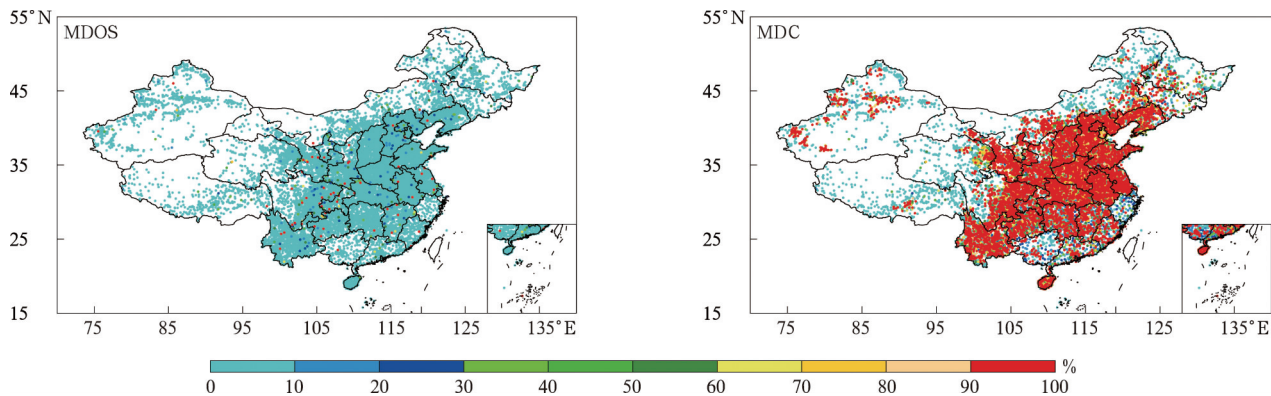


图 4 2022—2023 年中国地区错误数据命中率空间分布

Fig. 4 Spatial distributions of error data hit rate from 2022 to 2023

3.2 个例分析

2021—2023 年我国自动气象站实有小时降水量数据累计达 1.49×10^8 站次，其中 MDC 识别出 1.652×10^6 站次疑误数据，而传统方法仅识别出 1.5×10^5 站次。

3.2.1 异常大降水判识效果

通过逐月逐站精细化设置阈值，有效提升单站异常大降水的质量控制能力。以吉林河南屯站 2021 年 3 月 22 日降水事件(图 5)为例，当日 24 个时次的 1 h 降水量为 19.7 ~ 209.1 mm，均超过该

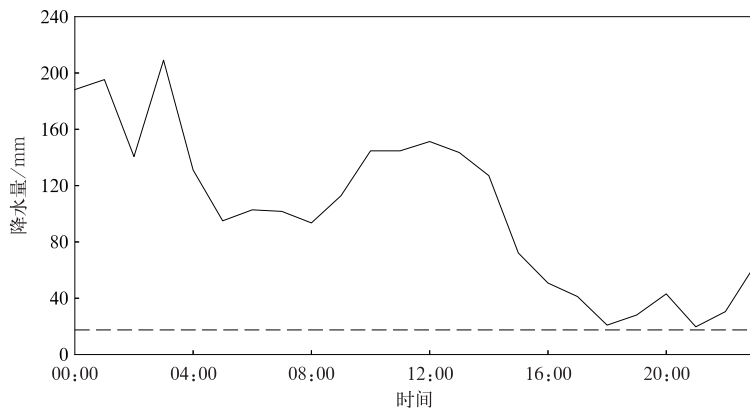


图 5 2021 年 3 月 22 日 00:00—23:00 吉林河南屯站 1 h 降水量(实线)及台站界限值检查的最大阈值(虚线)

Fig. 5 Hourly precipitation(the solid line) at Henantun Station in Jilin from 0000 UTC to 2300 UTC on 22 Mar 2021 and the maximum threshold(the dashed line) for climatic checks

站 3 月的降水最大阈值 17.5 mm。应用 MDC 自检测模块，24 个时次的降水观测值均被判识为错误。人工核查台站更正信息，确认上述观测值错误。而传统方法未能识别上述问题数据，这种错误的单站

大降水可能造成降水结构异常。质量控制前(图 6)当日 04:00 吉林和黑龙江南部地区出现降水，河南屯站小时降水量达到 131.1 mm，而其他台站观测的小时降水量均低于 2 mm。MDC 自检测模块剔

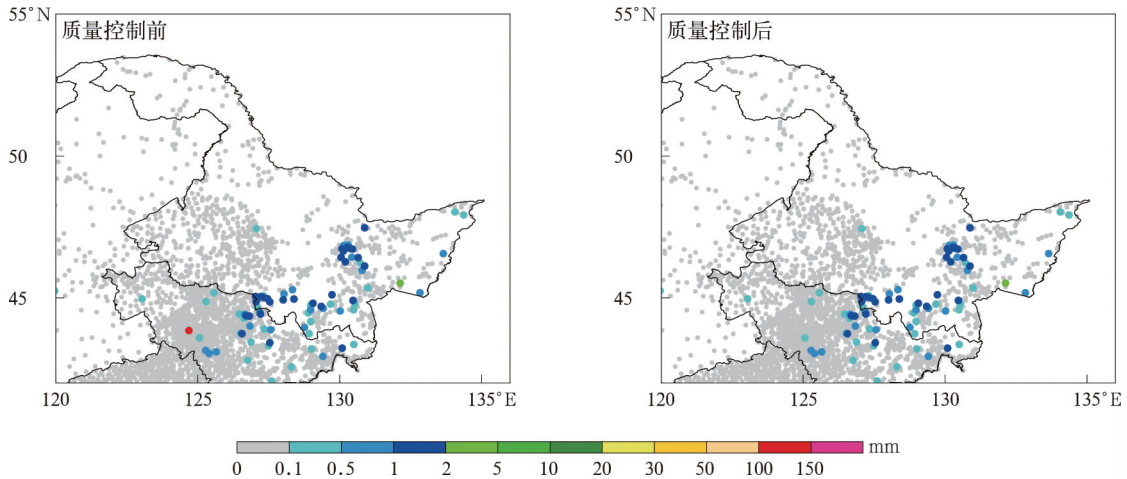


图 6 2021 年 3 月 22 日 04:00 质量控制前后 1 h 降水量空间分布
 Fig. 6 Spatial distribution of hourly precipitation before and after quality control at 0400 UTC 22 Mar 2021

除了河南屯站的异常大降水,使降水结构更加合理。
 3.2.2 虚假降水判别效果

MDC 多源数据协同检测模块能够有效识别晴空虚假降水。2023 年 7 月 28 日第 31 届世界大学生夏季运动会(简称大运会)在成都开幕。当日 02:00 FY-4A 气象卫星云图、雷达估测降水以及天气现象数据(图略)均显示赛区为晴空区,但质量控

制前(图 7)区域内 S1761 站等 10 个区域站出现虚假弱降水。传统方法未能识别此类错误数据,导致区域站降水观测与实际天气不一致。应用 MDC 多源数据协同检测模块,上述错误数据均被准确识别剔除,降水结构与当日晴好天气一致。与台站核查后,确认上述弱降水为虚假降水,由蚊虫空翻、草坪浇水和仪器维护残留降水量等原因产生。

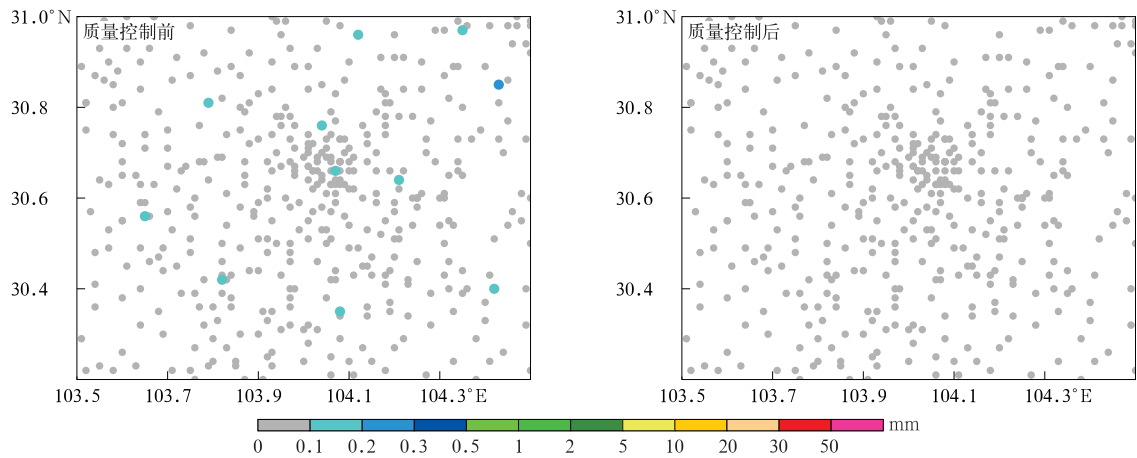


图 7 2023 年 7 月 28 日 02:00 成都大运会赛区质量控制前、后 1 h 降水量空间分布
 Fig. 7 Spatial distribution of hourly precipitation before and after quality control in Chengdu Universiade Competition Area at 0200 UTC 28 Jul 2023

对于空间连续分布的融雪性虚假降水,MDC 多源数据协同检测模块亦具备判别能力。2022 年 2 月 5—9 日华中地区(30°~35°N,108°~116°E)出现一轮大范围雨雪过程。冬半年,国家站采用称重式雨量计进行降水观测,具备不同形态降水观测能

力^[17-19]。因此选取国家站降水观测数据分析此次雨雪过程演变特征。图 8 为发生液态降水和固态降水的国家级台站数量的时间序列。液态降水包括雨、毛毛雨等,固态降水包括雪、雨夹雪和冰粒等。由图 8 可见,降水主要集中在 2 月 5 日 18:00—7 日

11:00,且以固态降水为主;2月7日 12:00—22:00 降水过程结束;2月7日 23:00—2月9日 04:00 出

现一次小范围降水,以液态降水为主;此后,该轮雨雪天气过程结束。

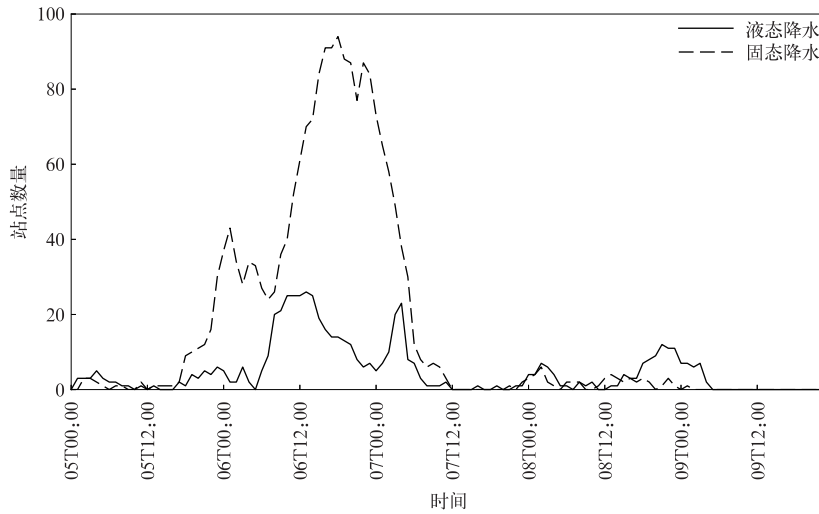


图 8 2022 年 2 月 5 日 00:00—10 日 23:00 华中地区发生液态降水和固态降水的国家站数量

Fig. 8 Number of national stations in the Central China with liquid precipitation and solid precipitation from 0000 UTC 5 Feb to 2300 UTC 9 Feb in 2022

在此次雨雪过程,区域站降水结构与国家站差异显著。以 2 月 9 日 06:00 为例,此时降雪过程结束,使用称重式雨量计的国家站降水量为零值;但质量控制前(图 9)大量区域站出现降水。经多源数据协同检测,鄂西北地区的降水被剔除,降水结构与国家站基本一致。区域站异常降水归因于翻斗雨量计

对固态降水观测的不适应性。在降雪期间,雨量筒被冻结,积雪堆积;随着气温上升,积雪融化,导致融雪性虚假降水。对于这种同期发生的大范围虚假降水,传统方法难以识别。MDC 多源数据协同检测,有效提升此类错误数据检测能力,避免大范围降水事件的误报。

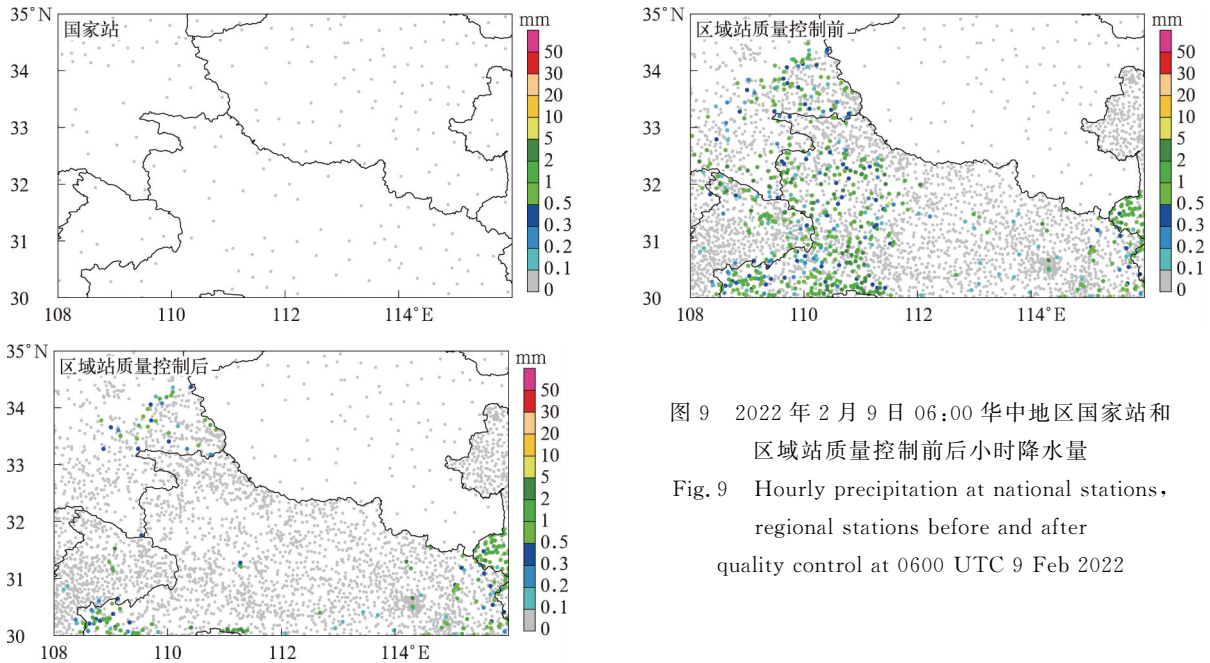


图 9 2022 年 2 月 9 日 06:00 华中地区国家站和区域站质量控制前后小时降水量
Fig. 9 Hourly precipitation at national stations, regional stations before and after quality control at 0600 UTC 9 Feb 2022

3.2.3 虚假零值降水判识效果

雨量计漏斗堵塞、机械故障或信号传输问题可能导致虚假零值数据^[32]。本文以“23·7”(2023 年 7 月 29 日—8 月 2 日)华北特大暴雨事件为例,分析虚假零值数据判识效果。该个例中北京西部、河北中部和西南部等多地遭遇特大暴雨,累积降水量超过 400 mm,局地达 1000 mm 以上^[42]。其中, A1361 站因仪器故障,小时降水量持续出现零值,严重影响了真实降水结构。质量控制前(图 10),7 月 31 日 04:00 A1361 站观测的零降水量与其邻近站点 50 mm 以上降水量存在明显差异。经 MDC 多源数据协同检测,可剔除此类错误数据,从而更准确

表征强降水中心及其覆盖范围。

MDC 多源数据协同检测对弱降水带中大面积虚假零值降水也具备甄别能力。2023 年 4 月 2 日 00:00—4 日 23:00 我国黄淮以及华北地区出现一次大范围降水过程。图 11 为 72 h 累积降水量的空间分布。由图 11 可见,山西省自动气象站观测显示该省出现大范围无降水区,与周边地区呈现的连续雨带存在显著差异;而从国家站降水空间分布看,雨带覆盖整个山西,与邻近区域降水结构一致。

图 12 为 2023 年 4 月 3 日 05:00 质量控制前后山西 1 h 降水量分布。质量控制前,山西大量区域站出现零值降水,与国家站的降水结构存在明显差

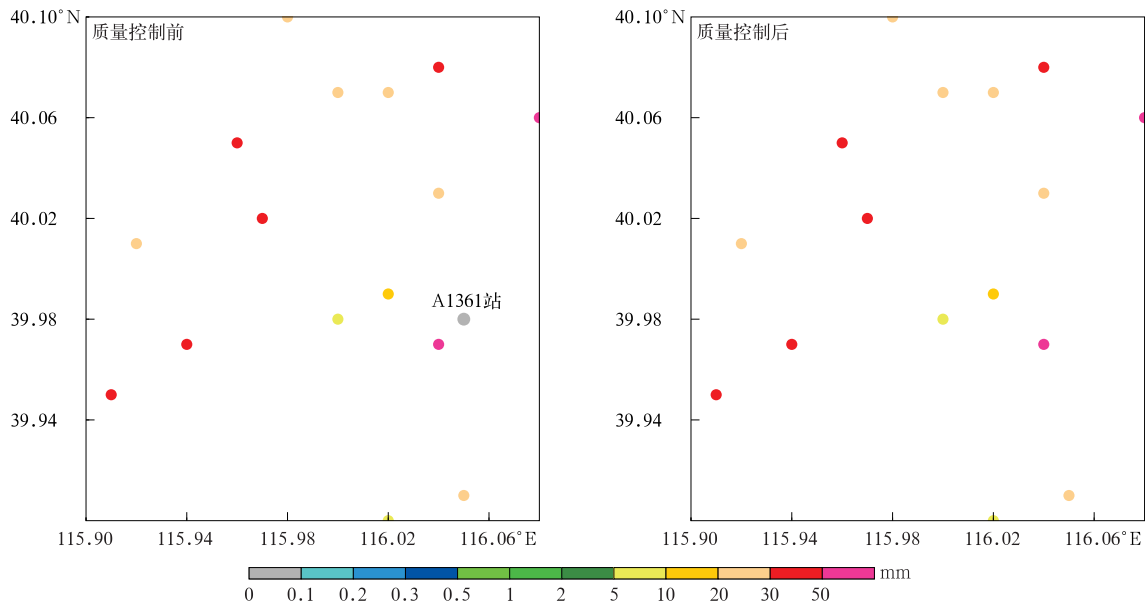


图 10 2023 年 7 月 31 日 04:00 质量控制前后北京西部自动气象站 1 h 降水量空间分布
 Fig. 10 Spatial distribution of hourly precipitation at automatic stations before and after quality control in the west of Beijing at 0400 UTC 31 Jul 2023

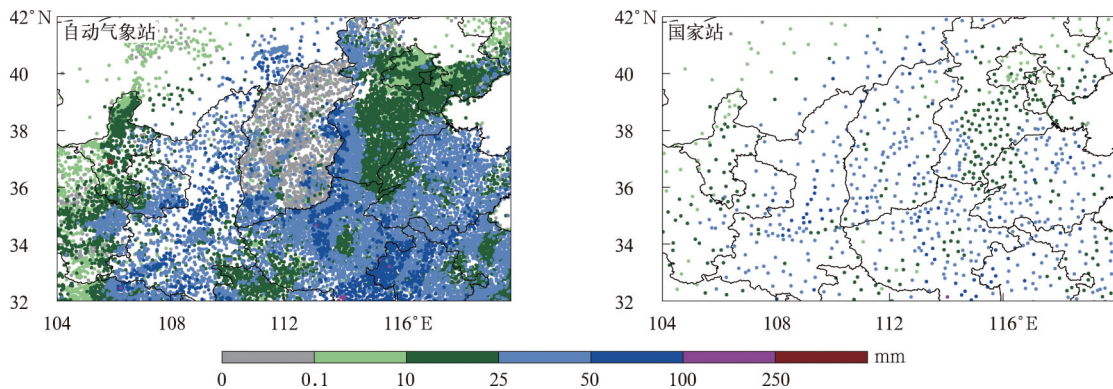


图 11 2023 年 4 月 2—4 日自动气象站和国家站 72 h 累积降水量空间分布
 Fig. 11 Spatial distribution of 72 h accumulated precipitation at automatic weather stations and national stations from 2 Apr to 4 Apr in 2023

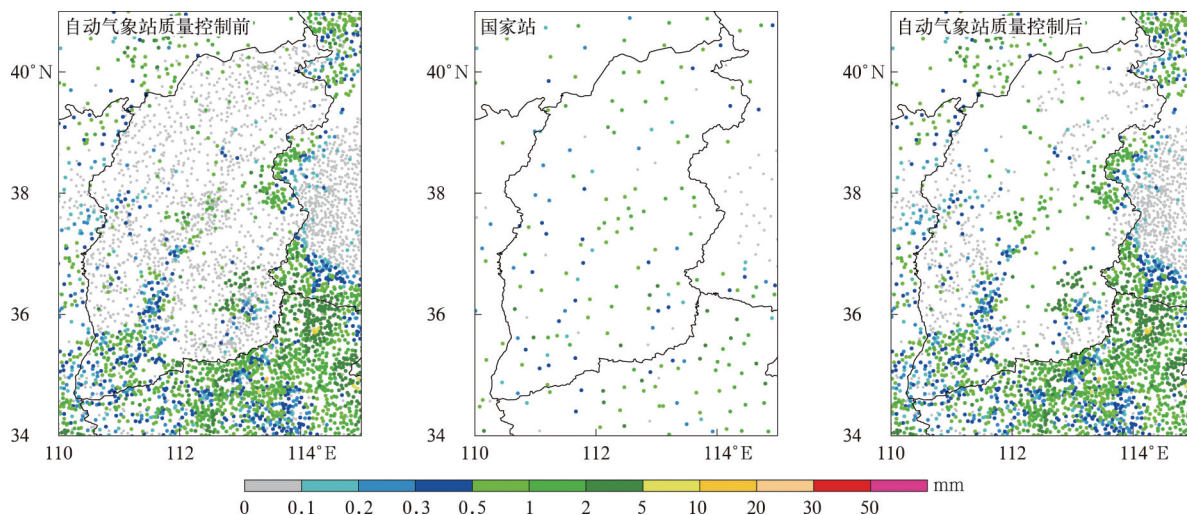


图 12 2023 年 4 月 3 日 05:00 自动气象站质量控制前、国家站和自动气象站质量控制后 1 h 降水量空间分布

Fig. 12 Spatial distribution of hourly precipitation data at automatic weather stations before quality control, national stations and automatic weather stations after quality control at 0500 UTC 3 Apr 2023

异;经多源数据协同检测,这些零值降水被甄别剔除,降水范围与国家站一致。根据《地面气象自动观测规范》^[37],在 11 月—次年 4 月,山西暂停区域站降水观测任务。但在本次降水过程中,降水观测被记录为零值并进入应用,导致降水范围出现严重偏差。

4 结论与讨论

本研究针对超出界限值的异常大降水、虚假降水和虚假零值降水等多种错误数据类型,研制了适用于地面自动气象站小时降水数据的质量控制方法(MDC)。MDC 包含自检、多源数据协同检测和动态黑名单 3 个模块。应用 MDC 对 2021—2023 年中国自动气象站小时降水数据开展质量控制与评估,得到以下主要结论:

1) 以中国地面小时降水数据集(V3.0)2021—2023 年数据为样本,评估质量控制方法对正确数据的判识能力。结果表明:MDC 判识正确率达到 99.92%,误剔率为 0.08%,其中大部分误剔数据为 [0.1 mm, 1 mm) 的弱降水,占比为 60.72%。不低于 10 mm 小时降水量的误剔数据集中在西南地区的高山站。

2) 以 2022—2023 年 1.97×10^5 条更正前数据为错误数据样本,评估质量控制方法错误数据甄别能力。MDC 的平均错误数据命中率为 39.8%,较现行 MDOS 实时质量控制提高了 39.3%。

3) 不同类型错误数据质量控制效果个例分析结果表明:MDC 显著提升了晴空虚假降水、融雪性虚假降水和虚假零值降水等错误数据的甄别能力,有效弥补了传统方法的不足。

需要指出的是,MDC 在实际应用中仍面临一定挑战。首先,该方法的有效性受限于协同数据的空间分辨率,因此在观测稀疏区域,其适用性和效果显著降低。其次,雷达和天气现象观测数据的质量问题可能干扰质量控制效果。此外,弱降水数据和高山台站降水大值仍存在误剔现象。下一步将致力于引入卫星降水产品等空间全覆盖的数据作为参照,结合人工智能技术,以提升观测稀疏区、复杂地形以及雨雪冰冻等极端天气过程数据质量控制能力。

参考文献

- [1] 林佳璐,李英,柳龙生. 风暴-低涡影响下青藏高原一次强降水过程. 应用气象学报, 2023, 34(2): 166-178.
Lin J L, Li Y, Liu L S. A heavy precipitation process over the Tibetan Plateau under the joint effects of a tropical cyclone and vortex. *J Appl Meteor Sci*, 2023, 34(2): 166-178.
- [2] 王国荣,卞素芬,王令,等. 用地面加密自动观测资料对北京地区一次飑线过程的分析. 气象, 2010, 36(6): 59-65.
Wang G R, Bian S F, Wang L, et al. Analysis on a typical squall line case with surface automatic weather observations. *Meteor Mon*, 2010, 36(6): 59-65.
- [3] 邢楠,仲跻芹,雷蕾,等. 基于 CMA-BJ 的北京地区短时强降水预报试验. 应用气象学报, 2023, 34(6): 641-654.
Xing N, Zhong J Q, Lei L, et al. A probabilistic forecast experiment of short-duration heavy rainfall in Beijing based on

- CMA-BJ. *J Appl Meteor Sci*, 2023, 34(6): 641-654.
- [4] 东高红,刘黎平,雷达与雨量计联合估测降水的相关性分析. *应用气象学报*, 2012, 23(1): 30-39.
Dong G H, Liu L P. Correlation analysis on estimating rainfall using radar-rain gauge calibration. *J Appl Meteor Sci*, 2012, 23(1): 30-39.
- [5] Zhang J, Howard K, Langston C, et al. National mosaic and multi-sensor QPE(NMQ) system: Description, results and future plans. *Bull Amer Meteor Soc*, 2011, 92: 1321-1338.
- [6] 刘非凡,郑永光,罗琪,等.京津冀及周边一般性降水与短时强降雨特征对比. *应用气象学报*, 2023, 34(5): 619-629.
Liu F F, Zheng Y G, Luo Q, et al. Comparison of characteristics of light precipitation and short-time heavy precipitation over Beijing, Tianjin, Hebei and neighbouring areas. *J Appl Meteor Sci*, 2023, 34(5): 619-629.
- [7] 任芝花,冯明农,张洪政,等.自动与人工观测降雨量的差异及相关性. *应用气象学报*, 2007, 18(3): 358-364.
Ren Z H, Feng M N, Zhang H Z, et al. The difference and relativity between rainfall by automatic recording and manual observation. *J Appl Meteor Sci*, 2007, 18(3): 358-364.
- [8] 宝兴华,夏茹娣,罗亚丽,等.“21·7”河南特大暴雨气象和天文雨量观测对比. *应用气象学报*, 2022, 33(6): 668-681.
Bao X H, Xia R D, Luo Y L, et al. Comparative analysis on meteorological and hydrological rain gauge observations of the extreme heavy rainfall event in Henan Province during July 2021. *J Appl Meteor Sci*, 2022, 33(6): 668-681.
- [9] 齐道日娜,何立富,王秀明,等.“7·20”河南极端暴雨精细观测及热动力成因. *应用气象学报*, 2022, 33(1): 1-15.
Chyi D, He L F, Wang X M, et al. Fine observation characteristics and thermodynamic mechanisms of extreme heavy rainfall in Henan on 20 July 2021. *J Appl Meteor Sci*, 2022, 33(1): 1-15.
- [10] 常国旭,董秀辉,刘红艳,等.加密自动气象站实时短信报警查询系统. *安徽农业科学*, 2009, 37(1): 427-428.
Chang G X, Dong X H, Liu H Y, et al. Real-time short message alarming and information query system for intensive automatic weather station network. *J Anhui Agric Sci*, 2009, 37(1): 427-428.
- [11] Martinaitis S M. Effects of Multi-sensor Radar and Rain Gauge Data on Hydrologic Modeling in Relatively Flat Terrain. Florida: Florida State University, 2008.
- [12] Sevruk B. Rainfall Measurement: Gauges. *Encyclopedia of Hydrological Sciences*. Anderson M G, Ed. 2005. DOI:10.1002/047084-8944.hsa038.
- [13] Habib E, Krajewski W F, Kruger A. Sampling errors of tipping-bucket rain gauge measurements. *J Hydrol Eng*, 2001, 6(2): 159-166.
- [14] Kondragunta C R, Shrestha K. Automated Real-time Operational Rain Gauge Quality-control Tools in NWS Hydrologic operations. 20th Conf on Hydrology, Boston, MA, Amer Meteor Soc, 2006: P2. 4.
- [15] Sieck L C, Burges S J, Steiner M. Challenges in obtaining reliable measurements of point rainfall. *Water Resour Res*, 2007, 43(1). DOI:10.1029/2005WR004519.
- [16] 叶柏生,杨大庆,丁永建,等.中国降水观测误差分析及其修正. *地理学报*, 2007, 62(1): 3-13.
Ye B S, Yang D Q, Ding Y J, et al. A bias-corrected precipitation climatology for China. *Acta Geographica Sinica*, 2007, 62(1): 3-13.
- [17] Goodison B E, Louie P Y T, Yang D. WMO Solid Precipitation Measurement Intercomparison. WMO Instruments and Observing Methods Rep No 67, 1998, WMO/TD-No. 872.
- [18] Rasmussen R, Baker B, Kochendorfer J, et al. How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull Amer Meteor Soc*, 2012, 93(6): 811-829.
- [19] Nitu R, and Coauthors. WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012-2015). IOM Rep 131, 2008.
- [20] 陶士伟,徐枝芳.加密自动站资料质量保障体系分析. *气象*, 2007, 33(2): 34-41.
Tao S W, Xu Z F. Analysis of the quality assurance procedures in intensified automatic surface weather observation system. *Meteor Mon*, 2007, 33(2): 34-41.
- [21] 杨萍,刘伟东,仲跻芹,等.北京地区自动气象站气温观测资料的质量评估. *应用气象学报*, 2011, 22(6): 706-715.
Yang P, Liu W D, Zhong J Q, et al. Evaluating the quality of temperature measured at automatic weather stations in Beijing. *J Appl Meteor Sci*, 2011, 22(6): 706-715.
- [22] Fiebrich C A, Crawford K C. The impact of unique meteorological phenomena detected by the Oklahoma mesonet and ARS microneon on automated quality control. *Bull Amer Meteor Soc*, 2001, 82(10): 2173-2187.
- [23] 任芝花,赵平,张强,等.适用于全国自动站小时降水资料的质量控制方法. *气象*, 2010, 36(7): 123-132.
Ren Z H, Zhao P, Zhang Q, et al. Quality control procedures for hourly precipitation data from automatic weather stations in China. *Meteor Mon*, 2010, 36(7): 123-132.
- [24] 任芝花,张志富,孙超,等.全国自动气象站实时观测资料三级质量控制系统研制. *气象*, 2015, 41(10): 1268-1277.
Ren Z H, Zhang Z F, Sun C, et al. Development of three-step quality control system of real-time observation data from AWS in China. *Meteor Mon*, 2015, 41(10): 1268-1277.
- [25] Kim D, Nelson B, Seo D J. Characteristics of reprocessed hydrometeorological automated data system (HADS) hourly precipitation data. *Wea Forecasting*, 2009, 24(5): 1287-1296.
- [26] 陶士伟,仲跻芹,徐枝芳,等.地面自动站资料质量控制方案及应用. *高原气象*, 2009, 28(5): 1202-1209.
Tao S W, Zhong J Q, Xu Z F, et al. Quality control schemes and its application to automatic surface weather observation system. *Plateau Meteor*, 2009, 28(5): 1202-1209.
- [27] 张乐坚,俞小鼎,李峰,等.地面降水的多源数据辅助质量控制方法. *气象*, 2016, 42(3): 363-371.
Zhang L J, Yu X D, Li F, et al. Quality control method for

- multi-source data of surface rainfall. *Meteor Mon*, 2016, 42(3):363-371.
- [28] 张志强, 仲凌志, 杨和平. 天气雷达在中国自动气象站实时质量控制系统中的应用. *计算机应用*, 2017, 37(增刊 II): 298-300.
Zhang Z Q, Zhong L Z, Yang H P. Application of weather radar in real-time quality control system of hourly gauge precipitation in China. *J Comput Appl*, 2017, 37(Suppl II): 298-300.
- [29] Qi Y C, Martinaitis S, Zhang J, et al. A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS system. *J Hydrometeor*, 2016, 17(6): 1675-1691.
- [30] Yeung H Y, Man C, Seed A, et al. Development of a Localized Radar-rain Gauge Co-Kriging QPE Scheme for Potential Use in Quality Control of Real-time Rainfall Data. The Third WMO International Conference on Quantitative Precipitation Estimation and Quantitative Precipitation Forecasting and Hydrology, Nanjing, China, 2010.
- [31] Yeung H Y, Man C, Chan S T, et al. Application of Radar-rain Gauge Co-Kriging to Improve QPE and Quality Control of Real-time Rainfall Data. Proceedings of the International Symposium on Weather Radar and Hydrology, Exeter, UK, 2011.
- [32] 丛芳, 刘黎平. 新一代天气雷达与地面雨量资料的综合分析. *气象*, 2011, 37(5): 532-539.
Cong F, Liu L P. A comprehensive analysis of data from the CINRAD and the ground rainfall station. *Meteor Mon*, 2011, 37(5): 532-539.
- [33] 王红艳, 王改利, 刘黎平, 等. 利用雷达资料对自动雨量计实时质量控制的方法研究. *大气科学*, 2015, 39(1): 59-67.
Wang H Y, Wang G L, Liu L P, et al. Development of a real-time quality control method for automatic rain gauge data using radar quantitative precipitation estimation. *Chinese J Atmos Sci*, 2015, 39(1): 59-67.
- [34] 吴书成, 魏爽, 吴京生. 雷达估测降水在区域站降水水质控中的应用. *气象科技*, 2015, 43(1): 49-52.
Wu S C, Wei S, Wu J S. Application of radar precipitation estimation to quality control for regional precipitation. *Meteor Sci Technol*, 2015, 43(1): 49-52.
- [35] Seo D J, Breidenbach J P, Johnson E R. Real-time estimation of mean field bias in radar rainfall data. *J Hydrol*, 1999, 223(3/4): 131-147.
- [36] Seo D J, Breidenbach J P. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J Hydrometeor*, 2002, 3(2): 93-111.
- [37] 中国气象局. 地面气象自动观测规范. 北京: 气象出版社, 2020.
China Meteorological Administration. Specification for Automatic Observation of Ground Meteorology. Beijing: China Meteorological Press, 2020.
- [38] 张强, 赵煜飞, 范邵华. 中国国家级气象台站小时降水数据集研制. *暴雨灾害*, 2016, 35(2): 182-186.
Zhang Q, Zhao Y F, Fan S H. Development of hourly precipitation datasets for national meteorological stations in China. *Torrential Rain Disasters*, 2016, 35(2): 182-186.
- [39] 王颖, 刘振. QXT 515—2019 气象要素特征值. 2019.
Wang Y, Liu Z. QXT 515—2019 Meteorological Element Characteristic Values. 2019.
- [40] Yang S, Jones P D, Jiang H, et al. Development of a near-real-time global in situ daily precipitation dataset for 0000—0000 UTC. *Int J Climatol*, 2020, 40(5): 2795-2810.
- [41] Wang R W, Han W, Tian W H, et al. Blacklist design of AM-DAR temperature data and their application in the CMA-GFS. *J Trop Meteor*, 2021, 27(4): 368-377.
- [42] 张博, 张芳华, 李晓兰, 等. “23·7”华北特大暴雨数值预报检验评估. *应用气象学报*, 2024, 35(1): 17-32.
Zhang B, Zhang F H, Li X L, et al. Verification and assessment of “23·7” severe rainstorm numerical prediction in North China. *J Appl Meteor Sci*, 2024, 35(1): 17-32.

Quality Control Method for Land Surface Hourly Precipitation Data in China

Zhu Yani¹⁾ Yang Su¹⁾ Zhang Zhiqiang¹⁾ Qiu Jianhua²⁾

¹⁾ (National Meteorological Information Center, Beijing 100081)

²⁾ (School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044)

Abstract

High spatial-temporal resolution observations of precipitation from automatic weather stations (AWSs) serve as a vital data source, extensively utilized in research activities such as severe weather monitoring, model evaluation, and forecast analysis. Influenced by factors such as observation environments and equipment performance, precipitation observations inevitably contain various forms of random and systematic errors. A quality control method (multi-source data collaborative quality control, MDC) has been established for hourly precipitation data from AWSs in China, based on high spatial-temporal resolution radar data and weather phenomena. The MDC includes three modules: Precipitation self-detection, multi-source data collaborative detection, and dynamic blacklisting. The MDC has been applied to quality control of hourly precipitation data from AWSs from 2021 to 2023. A comprehensive effectiveness assessment of the method has been conducted using a combination of quantitative indicators and case analyses of detection effects on various types of erroneous data. Results indicate that the correct identification rate of the MDC reaches 99.92%, with a false exclusion rate of 0.08%. The majority of falsely excluded data consists of weak precipitation amounts ranging from 0.1–1 mm, accounting for 60.72%. While ensuring a high correct rate, the MDC also demonstrates a high capability in identifying erroneous data. The average error data hit rate of the MDC in China is 39.8%, which represents an improvement of 39.3% over the existing Meteorological Data Operation System (MDOS) real-time quality control system. The ability MDC to identify erroneous data between 0–50 mm is approximately 40%, and this hit rate significantly increases with higher precipitation values. When precipitation amounts exceed 100 mm, the hit rate achieves 100%. MDOS real-time quality control system has an almost zero hit rate for erroneous data with precipitation amounts less than 20 mm but possesses some identification capability for abnormal precipitation of more than 20 mm. The hit rate of the MDC shows significant spatial variation due to the coverage of radar and national station observations. In the eastern region, where observation stations are densely distributed, most stations have an error data hit rate of over 90%. However, in the western and northeastern regions, where observation stations are sparse and do not meet the conditions for multi-source collaborative detection, the hit rate of the MDC decreases significantly, approaching that of the MDOS real-time quality control. Case analyses of the quality control effects on different types of erroneous data reveal that the MDC significantly the identification ability of abnormal data such as clear sky precipitation, snowmelt precipitation, and false zero value precipitation, effectively making up for the deficiencies of traditional methods.

Key words: automatic weather stations; hourly precipitation data; multi-source data collaborative detection; quality control