

四川省地面气象自动站观测数据质量分析与评估

吕纯月^{1,2}, 吴薇^{1,2}, 黄小梅^{2,3,4*}

(1. 四川省气象探测数据中心, 成都 610072; 2. 高原与盆地暴雨旱涝灾害四川省重点实验室, 成都 610072;
3. 中国气象局成都高原气象研究所, 成都 610072; 4. 四川省气候中心, 成都 610072)

摘要: 利用经过质量控制的国家和区域站逐小时气象观测资料, 针对四川省资料一体化业务运行近 5 年的地面自动站数据质量进行分析与评估。结果表明: 国家站数据质量整体呈优良水平, 随着 2020 年 4 月四川省自动化观测的转变, 可用率存在由高向低的显著突变, 随后逐步上升。区域站数据质量低于国家站, 但总体呈显著上升趋势。从季节变化看, 导致可用率降低的主要原因为数据缺测, 国家站可用率在汛期偏低而非汛期偏高, 区域站则与之相反。从日变化看, 区域站数据可用率在白天大于夜间, 于 17 时达到峰值后直至凌晨均呈下降趋势。从空间分布看, 四川盆地数据可用率最高, 盆地西部与高原过渡区域数据可用率大多偏低。从区域站各要素质量看, 风向风速可用率最高, 降水可用率最低且各月波动较大, 各要素缺测率相关性较好, 主要是通讯故障或软件故障造成的整站资料缺测。总之, 地面自动站质量存在显著的季节波动、日变化和空间分布特征, 但整体上表现较为稳定且呈一定的上升趋势。

关键词: 地面自动站; 气象资料业务系统; 质量控制; 质量评估

中图分类号: P413

文献标识码: A

doi: 10.3969/j.issn.1674-2184.2023.03.016

引言

常规地面气象观测数据是在天气预报和气候预测中应用的重要基础性资料^[1-2]。其中, 国家级气象观测站(简称“国家站”)气象资料因建站早、记录相对完整准确被广泛应用于天气气候变化研究, 而区域自动气象观测站(简称“区域站”)观测数据作为国家站观测资料的有效补充, 大大提高了观测密度, 满足了区域性灾害天气预警、精细化预报和服务的需求^[3-8]。地面气象观测数据的正确与否直接影响着天气预报和气候预测的质量, 因此, 气象观测资料的质量控制与评估显得尤为重要^[9-10]。

目前, 气象资料业务系统(MDOS)实现了地面气象观测数据实时质量控制, 大大提升了地面资料的应用质量, 已投入全国气象观测业务中, 四川省也于 2015 年 5 月开始了基于 MDOS 的资料一体化业务。已有多项研究^[11-15]对不同区域地面气象资料质量进行了分析与评估。其中, 吴薇等^[16]针对 2018 年四川省区域自动气象站资料进行了质量评估, 指出该年区域站资料完整性较好, 可用性、可靠性较高, 能更精细地反映区域气候特征, 应用于天气预报、气候预测、

防灾减灾和决策服务中。但随着观测自动化的推进, 观测数据的质量问题与质量管理呈现出一些新的特点和需求^[17-18]。

目前, 针对近几年四川省国家站和区域站观测数据质量情况, 尤其是全面自动化观测以后的情况, 尚缺少较为系统全面的分析与评估。本文以近 5 a 四川复杂地形下地面气象自动站观测数据为研究对象, 利用经过 MDOS 质量控制的国家和区域站逐小时气象要素观测资料, 从不同时间尺度、不同区域、不同观测要素等多个维度进行质量分析与评估, 以期科学合理地应用四川省地面观测资料提供客观有效的科技支撑。

1 资料与方法

1.1 资料

研究选取经过 MDOS 质量控制后的 156 个国家气象观测站和 4930 个区域自动站的逐小时气象要素观测资料, 国家站数据质量统计要素包括降水、气温、气压、相对湿度、风向风速、地温、能见度、蒸发等 47 个要素, 区域站数据质量统计要素包括降水、气温、气压、相对湿度、风向风速等 17 个要素, 统计时段为

收稿日期: 2022-11-15

资助项目: 高原与盆地暴雨旱涝灾害四川省重点实验室科技发展基金项目(SCQXKJYJXMS202221); 四川省气象探测数据中心川气探数课题(2022-9); 高原与盆地暴雨旱涝灾害四川省重点实验室科技发展基金项目(SCQXKJQN202221)

作者简介: 吕纯月, 助理工程师, 主要从事气象资料质量控制与应用研究。E-mail: 645327001@qq.com

通讯作者: 黄小梅, 副研究员, 主要从事气候与气候变化研究。E-mail: hxmlovely@163.com

2017年1月—2021年12月。所用到的地形数据来源于90 m数字高程模型(Model Digital Elevation, DEM)数据,利用ArcMap软件,提取各站点坡度、坡向等地形因子数据。

1.2 方法

利用上述资料,基于质量控制码(表1),对四川省近5 a小时地面观测数据按台站、观测要素分类,以数据可用率、缺测率、错误率、可疑率等为评估指标,进行资料可用性的统计分析,研究国家站和区域站数据质量整体趋势变化、季节波动、日变化以及空间分布特征。可用率为质控码为0、3、4的要素数据个数之和占参与统计的总要素数据个数的比例。可疑率、错误率分别为质控码为1、2的要素数据个数占参与统计的总要素数据个数的比例。缺测率为两部分之和:一是质控码为8的要素数据个数占参与统计的总要素数据个数的比例,二是整个时段缺测所导致的数据缺测量占参与统计的总要素数据个数的比例。

表1 质量控制码及其含义

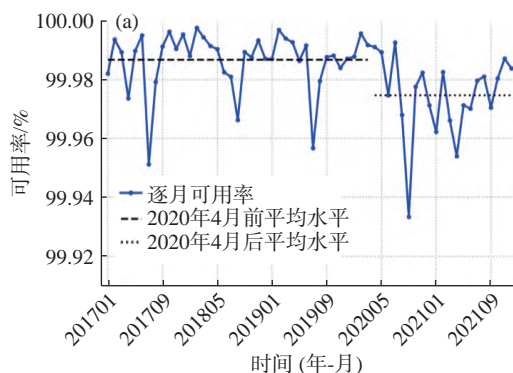
质控码	含义	质控码	含义
0	数据正确	5	预留
1	数据可疑	6	预留
2	数据错误	7	无观测任务
3	数据为订正值	8	缺测
4	数据为修改值	9	数据未作质量控制

依托MDOS系统,文中所有国家站、区域站资料经过了气候学界限值、区域气候极值、内部一致性、时间一致性、空间一致性等质量控制或检查。本文还采用了Mann-Kendall趋势检验(简称M-K检验)^[19]、线性倾向估计^[20]、相关分析及显著性检验等统计方法,分析地面站数据质量变化趋势与特征。

2 国家站数据质量评估分析

2.1 总体情况

分析近5 a四川省国家站数据可用率曲线(图1a)



可知,国家站数据可用率较高,且整体较为稳定,平均值为99.98%。其中,数据可用率高于平均值的月份总数达42个,占比70%;可用率低于平均值的月份总数为18个,占比30%,说明少数月份可用率异常低值(如2017年7月、2019年7月、2020年9月、2021年4月可用率均异常偏低,均小于99.96%)对整体平均值的影响较大,拉低了可用率平均值水平。

2020年4月,四川省地面气象观测自动化改革开始正式业务化运行。为探究该观测方式的转变对数据可用率是否存在影响,对近5 a国家站逐月数据可用率序列进行M-K检验(图1b)。结合图1a分析可知,数据可用率在2020年4月存在由高向低的显著突变,2020年4月后可用率进入相对偏低的阶段,可用率平均值相较于2020年4月前偏低0.02%,随后可用率呈逐步上升趋势。自动化观测可以对气象数据进行自动收集、自动处理、自动存储、自动传输,大大减少了人工投入,提高了气象观测的实时性和准确性。然而,自动化观测业务容易在不同因素的影响下出现各种故障问题,比如电源、通信、仪器故障,若维护不及时,或无法及时发现异常情况并对其进行处理,则会导致地面气象观测数据质量显著下降,因此需要采取针对性措施进行管理和控制。

2.2 季节波动与空间分布特征

从近5 a平均各月数据可用率变化情况(图2)来看,各月数据可用率均在99.97%以上,但呈现出明显的季节波动特征。数据可用率在汛期(6—9月)相对偏低,非汛期相对偏高。数据缺测率与可用率呈反相变化,汛期偏高而非汛期偏低,是导致数据可用率降低的主要原因。数据可疑率在3、4月相对偏高,数据错误率趋近于0。

分析国家站数据可用率空间分布(图3)可知:四川盆地可用率最高,接近100%;位于四川盆地西南边缘的峨眉山站,为单站可用率最低(99.85%),该站由于地形原因,易受雷暴天气影响,常因雷击导致仪器

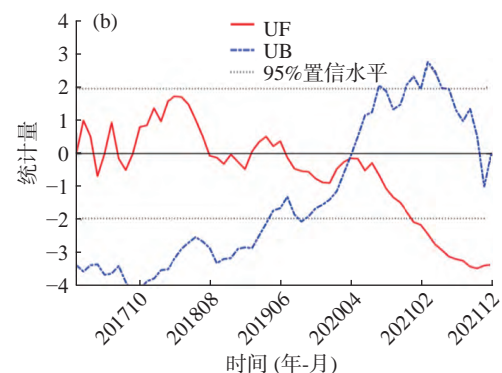


图1 2017年1月—2021年12月四川省国家站数据可用率(a)及M-K检验曲线(b)

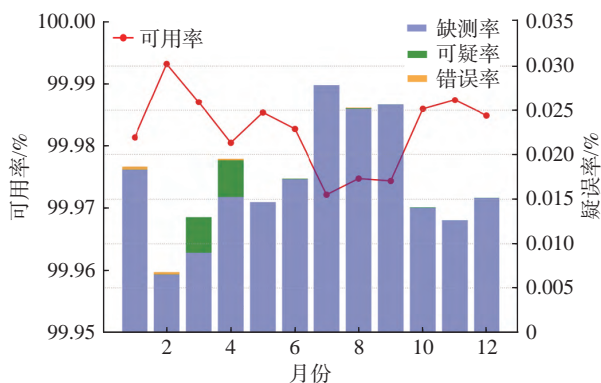


图2 四川省国家站数据可用率、缺测率、可疑率及错误率逐月变化

损坏,数据缺测率较高,严重影响了数据平均可用率;阿坝州、甘孜州、凉山州地区(简称“三州”)数据可用率相对偏低,且空间差异较大,这是由于三州地区地势起伏大,地形环境、气象条件复杂,对气象观测仪器的适用性、稳定性、可靠性都有很高的要求,加之高原台站条件相对艰苦,设备通讯、电力保障相对不足,某些恶劣情况下(如遇雷暴、降雪等天气)台站设备维护较为困难所致。

3 区域站数据质量评估分析

3.1 总体情况

分析近 5 a 四川省区域站数据可用率曲线(图 4a)可知,区域站数据可用率相比国家站有所偏低,数值介于 90.06%~98.75%,平均值为 96.14%,各月份数据可用率波动较大。如图 4b 所示,各月份可用率波动较大,但整体呈显著上升趋势,且通过了 95% 水平的置信度检验,但 UF 与 UB 曲线交点位于置信水平之外,故上升不具有突变性。

3.2 季节波动、日变化与空间分布特征

3.2.1 季节波动

从近 5 a 平均各月数据可用率变化情况(图 5)来看,除 1 月外,其余各月数据可用率均在 95% 以上。数据可用率在汛期(6—9 月)整体较高,平均值为 96.73%,在冬季(12 月—次年 2 月)相对偏低,平均值

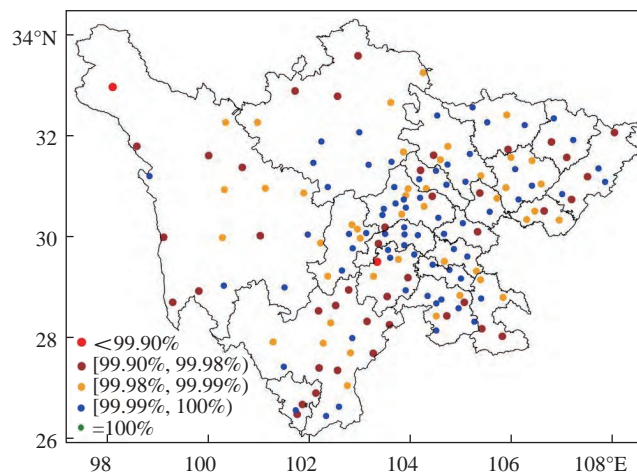


图3 国家站数据可用率空间分布

为 95.26%。这与国家站的统计结果相反,主要是由于汛期各市州县对区域站保障力度增强,缺测率有所下降,而冬季部分台站为避免固态降水、融雪滞后降水数据影响预报服务,对雨量传感器进行加盖处理,影响了整体降水数据的可用率。数据缺测率与可用率呈反相变化,仍是导致数据可用率降低的主要原因。各月均存在一定量的数据错误,平均每月的数据错误率为 0.06%,数据可疑率趋近于 0,横向对比缺测率(3.8%),两者对可用率的影响较小,可忽略不计。

3.2.2 日变化特征

利用区域站逐小时气象观测资料,统计一天中各时次的数据可用率,分析四川省区域站数据可用率和缺测率的日变化特征。如图 6 所示,四川省区域站平均数据可用率在白天(08—20 时)高于夜间(20 时—次日 08 时),在 03 时、15 时存在 2 个显著的低值,分别为 95.64%、95.94%,在 17 时达到可用率最高值,为 96.36%。可用率在 07—14 时呈上升趋势,17 时达到峰值后直至凌晨均呈下降趋势。数据缺测率与可用率呈反相变化特征,两者相关性较好。

3.2.3 空间分布特征

统计区域站近 5 a 平均的数据可用率情况,分析

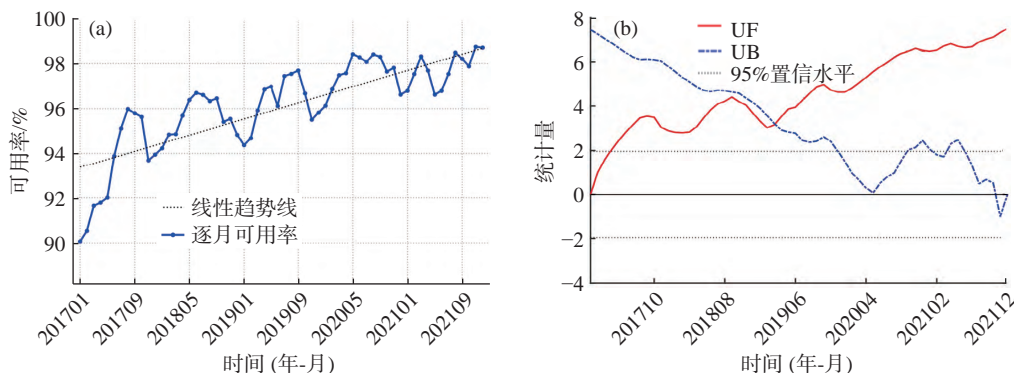


图4 同图1,但为区域站

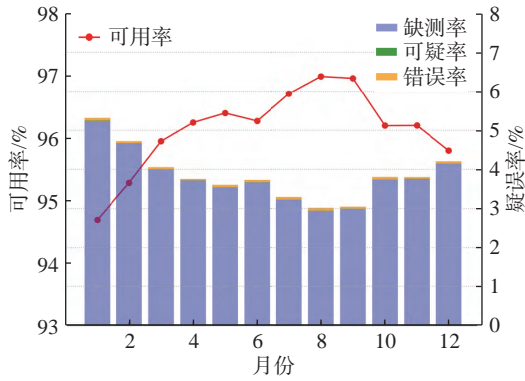


图 5 同图 2, 但为区域站

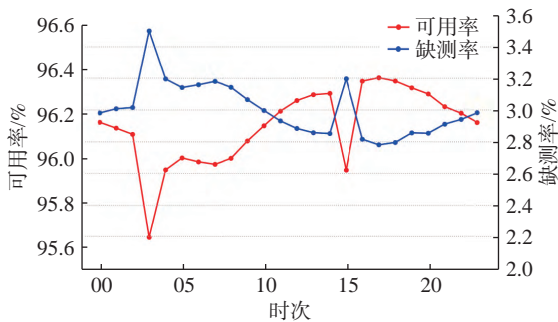


图 6 四川省区域站数据可用率、缺测率日变化

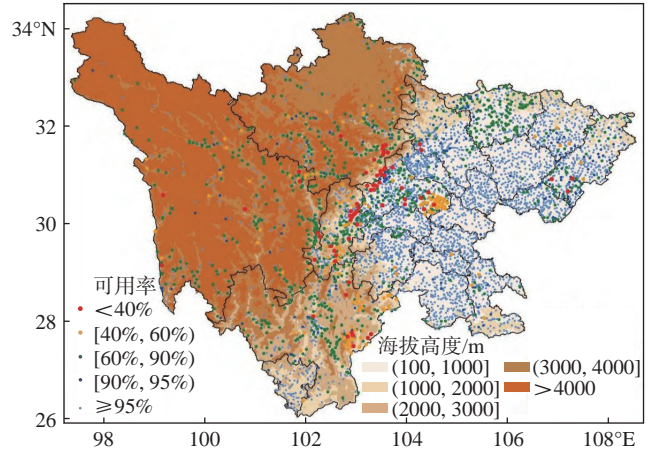


图 7 区域站海拔高度和近 5 a 平均数据可用率空间分布

四川省区域站可用率空间分布及地形特征。如图 7 所示, 数据可用率整体呈西部低、东部高的分布特征,

盆地大部及攀枝花地区数据可用率高于 95%。可用率低于 90% 的站点主要分布在三州以及广元市、简阳市等地。可用率低于 40% 的站点主要分布于四川盆地西部与川西、攀西高原交界的过渡区以及高原地区地势低洼带。表 2 给出了部分可用率低于 40% 的区域站地形因子(海拔高度、坡向、坡度)及地理位置信息。分析可知: 这些低可用率站点大多坡度较大, 说明地势起伏显著; 三州地区的低可用率站点大多位于沟谷处, 坡向不固定; 绵阳、成都、雅安等地的低可用率站点大多位于高原与盆地的地势过渡区, 坡向大

表 2 部分可用率低于 40% 的区域站地形及位置信息

市(州)	区站号	县(区)	海拔高度/m	坡度/°	坡向/°	特殊地理位置	可用率/%
阿坝	S8488	汶川	2008	27.86	92.60	沟谷	39.73
	S8501	汶川	1860	18.47	106.97	沟谷	23.08
	S8148	白玉	3001	7.69	53.13	沟谷	28.05
甘孜	S8221	巴塘	3408	6.55	17.49	沟谷	22.22
	S8280	丹巴	2379	18.11	0.00	山腰	20.43
	S8348	泸定	1799	33.49	77.16	沟谷	28.60
凉山	S9190	布拖	2540	21.81	220.44	山腰	20.83
	S9197	布拖	2835	2.62	66.80	沟谷	27.19
	S9350	金阳	1642	14.17	256.95	山腰	26.31
绵阳	S2424	安县	984	25.91	96.03	地势过渡区	23.46
	S2431	安县	643	0.73	135.00	地势过渡区	24.83
	S1648	大邑	1114	35.71	170.75	地势过渡区	10.03
	S1649	大邑	1265	4.90	216.47	地势过渡区	10.28
成都	S1656	崇州	1683	18.30	96.51	地势过渡区	9.90
	S1699	崇州	1003	13.45	342.47	沟谷	10.27
	S1700	崇州	843	14.35	146.96	地势过渡区	10.05
雅安	S7274	汉源	1530	24.51	285.45	沟谷	29.94
	S7280	汉源	1239	23.03	160.41	地势过渡区	22.91
	S7285	汉源	1092	15.19	106.03	地势过渡区	24.64

多为 $90^{\circ}\sim 180^{\circ}$, 符合地势由西向东递减的特征。上述地区地势起伏大且地形地貌复杂, 观测站易受周边环境影响, 设备通讯、电力保障条件不稳定, 维护成本、难度较高, 进而导致观测站数据可用率较低。从数据可用率季节分布(图略)来看, 夏季数据可用率最高, 春、秋季次之, 冬季最低, 盆地西部与高原交界地带在不同季节均是主要的可用率偏低区域。

3.3 分要素评估

为进一步分析各观测要素的数据质量情况, 对四川省区域站按观测要素分类, 分别针对气温、气压、相对湿度、风向风速、降水这 5 个基本气象观测要素进行质量评估。

从四川省区域站各气象观测要素总体质量情况(图 8)来看, 风向风速数据的可用率最高(96.71%), 其次分别为湿度(95.09%)、气温(94.46%)、气压(91.1%)、降水(90.7%)。对各要素数据可用率影响较大的主要是缺测率, 风向风速的缺测率最小, 仅为 3.08%, 气温、降水的缺测率较大, 分别为 4.94%、4.65%。气压的错误率、可疑率相比其他要素均明显偏大, 分别为 2.70%、0.44%, 共同导致气压数据的可用率偏低。

图 9 给出了四川省区域站各气象观测要素质量逐月变化。如图 9a 所示, 各要素可用率均为汛期高于非汛期, 缺测率汛期低于非汛期。其中, 风向风速、湿度、气温的各月可用率较为稳定, 均在 92.5% 以上, 气压可用率除 1 月相对偏低(87.7%)外, 其余月份均在 90% 以上。降水的各月可用率波动最大, 汛期(6—9 月)平均值为 95.98%, 而冬季(12 月—次年 2 月)平均值只有 83.43%, 这主要是冬季部分台站对雨量传感器进行加盖所致。如图 9b 所示, 各要素缺测率随月份的变化走势基本一致, 各要素缺测率之间的相关性较好, 均能通过 95% 水平的显著性检验。上述特征说明台站数据缺测主要是通讯故障或数据采集软

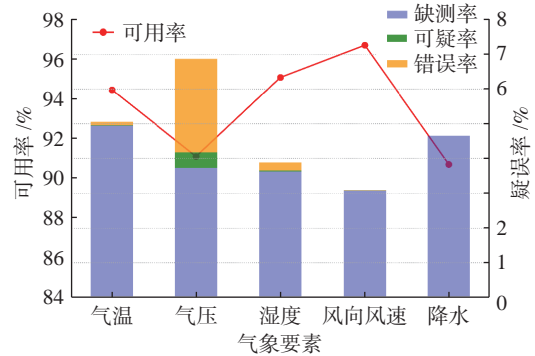


图 8 四川省区域站各气象观测要素总体质量情况

件故障等造成的整站资料缺测, 要素仪器设备故障导致的缺测率占比较小。

4 结论

本文对四川省 2017—2021 年地面自动站观测数据按台站和观测要素分类, 以可用率、缺测率、错误率、可疑率等为评估指标, 从多个时空维度对数据质量进行评估分析, 得到以下主要结论:

(1) 四川省国家站数据质量整体呈优良水平。随着 2020 年 4 月四川省自动化观测的转变, 可用率存在由高向低的显著突变, 随后可用率呈逐步上升趋势。自动化观测因其特殊性需采取针对性措施进行管理和控制。

(2) 四川省国家站数据可用率季节波动明显, 汛期低于非汛期, 数据缺测率与之呈反相变化, 是导致可用率降低的主要原因。四川盆地可用率最高, 阿坝州、甘孜州、凉山州地区数据可用率相对偏低, 且空间差异较大。

(3) 四川省区域站数据质量整体低于国家站, 平均值为 96.14%, 但总体呈显著上升趋势。导致区域站数据可用率降低的主要原因仍是数据缺测, 错误率和可疑率对可用率的影响较小。从季节变化来看, 可用

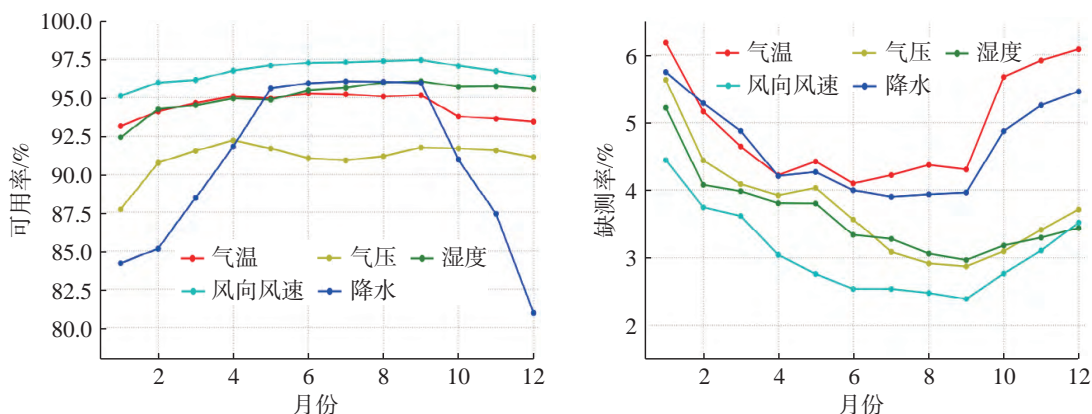


图 9 四川省区域站各气象观测要素质量逐月变化 (a. 可用率, b. 缺测率)

率在汛期高于非汛期,主要是由于汛期各州市县对区域站保障力度增强,而冬季部分台站对雨量传感器进行加盖处理,影响了整体降水数据的可用率。

(4)四川省区域站数据可用率在白天大于夜间,07—14时呈上升趋势,在17时达到峰值后直至凌晨均呈下降趋势。盆地西部与高原交界的过渡区以及高原地带地势低洼地区可用率较低。

(5)四川省区域站风向风速的数据可用率最高,降水可用率最低。风向风速、湿度、气温的各月可用率较为稳定,而降水的各月可用率波动较大。各要素缺测率随月份的变化走势基本一致且相关性较好,说明台站数据缺测主要为通讯故障或软件故障等造成的整站资料缺测。

参考文献

- [1] 李尚锦,周学云,高文良,等. 1991—2020年雅安地区降水特征研究[J]. 高原山地气象研究, 2022, 42(S2): 8—13
- [2] 刘冬梅,黄少平,王钰. 江西地面气象资料质量评估[J]. 气象水文海洋仪器, 2009, 26(3): 94—97
- [3] 边茜. MCC引发的暴雨天气过程垂直螺旋度分析[J]. 高原山地气象研究, 2022, 42(S2): 19—24, 35
- [4] 郭旭,龙柯吉,范江琳,等. 四种降水融合产品在四川持续性强降水过程中的对比评估[J]. 高原山地气象研究, 2021, 41(2): 42—52
- [5] 周懿,郭云云,余芳. 四川盆地一次西风槽和台风共同作用暴雨过程分析[J]. 高原山地气象研究, 2020, 40(2): 11—17
- [6] 王佳津,王彬雁,曹萍萍. 基于西南区域模式的四川强降水开始时间订正试验[J]. 高原山地气象研究, 2022, 42(2): 63—68
- [7] 何光碧,屠妮妮,高笃明,等. 西南区域自动站资料对区域模式预报影响的同化试验分析[J]. 高原山地气象研究, 2017, 37(3): 1—9
- [8] 夏菲,夏利. 2021年达州“8.8”区域性特大暴雨成因分析[J]. 高原山地气象研究, 2022, 42(S2): 14—18
- [9] 廖捷,周自江. 全球常规气象观测资料质量控制研究进展与展望[J]. 气象科技进展, 2018, 8(1): 56—63
- [10] 叶小岭,周建华,熊雄. 一种基于GEP的地面气温观测资料的质量控制方法[J]. 热带气象学报, 2014, 30(6): 1196—1200
- [11] 李奇临,魏麟骁,吴翀,等. 基于MDOS的重庆市2015年气象观测数据质量分析[J]. 西南师范大学学报(自然科学版), 2018, 43(11): 135—140
- [12] 侯灵,杨玉红,陈晓庆. 广东省区域自动站数据的质量分析与评估[J]. 广东气象, 2019, 41(5): 51—53
- [13] 杨萍,刘伟东,仲路芹,等. 北京地区自动气象站气温观测资料的质量评估[J]. 应用气象学报, 2011, 22(6): 706—715
- [14] 杨萍,刘伟东. 北京地区加密自动气象站数据的质量分析[J]. 气象科技进展, 2013, 3(6): 27—34
- [15] 岳艳霞,智利辉,王磊,等. 石家庄市区域自动气象站气温数据适用性分析[J]. 气象与环境学报, 2017, 33(2): 53—62
- [16] 吴薇,郭旭,尹娟,等. 四川省区域自动气象站资料的质量评估[J]. 气象与环境科学, 2021, 44(1): 104—110
- [17] 张晓岚. 当前地面气象观测自动化存在的问题及发展建议[J]. 北京农业, 2015, (12): 199—200
- [18] 封秀燕,何志军,王荷平,等. 自动气象站实时资料质量控制开放式平台设计[J]. 应用气象学报, 2010, 21(4): 506—512
- [19] Mann H B. Nonparametric tests against trend[J]. Econometrica, 1945, 13(3): 245—259
- [20] 魏凤英. 现代气候统计诊断与预测技术[M]. 北京: 气象出版社, 2007: 37—41

Quality Analysis and Assessment of Observation Data of Automatic Ground Meteorological Stations in Sichuan Province

LYU Chunyue^{1,2}, WU Wei^{1,2}, HUANG Xiaomei^{2,3,4*}

(1. Sichuan Provincial Meteorological Observation and Data Centre, Chengdu 610072, China; 2. Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, Chengdu 610072, China; 3. Institute of Plateau Meteorology, CMA, Chengdu 610072, China; 4. Sichuan Provincial Climate Centre, Chengdu 610072, China)

Abstract: Using the hourly meteorological observation data of national stations and regional stations, the data quality of ground automatic stations of Sichuan data integration business operation for nearly five years is analyzed and evaluated. The results show that the overall data quality of Sichuan national stations is at an excellent level. With the transformation of automatic observation in Sichuan in April 2020, there is a significant mutation in data availability rate from high to low, and then gradually increases. The data quality of regional stations is lower than that of national stations, but there is a significant upward trend overall. From the perspective of seasonal fluctuations, the main reason for the decrease of availability rate is data missing, the availability rate of national stations is low in the flood season and high in the non-flood season, while that of the regional station is the opposite. From the perspective of daily variation, the data availability rate of regional stations is greater during the day than that at night, and after reaching a peak at 17 o'clock, it shows a downward trend until the early morning. From the perspective of spatial distribution, the data availability rate of Sichuan Basin is the highest, and the data availability rate of the western Basin and the plateau transition area is mostly low. From the perspective of the quality of each

element of the regional stations, the availability rate of wind direction and wind speed is the highest, the availability rate of precipitation is the lowest and the fluctuation of each month is large. The data missing rate of each element is well correlated, mainly due to the missing of the whole station caused by communication failure or software failure. In general, the data quality of ground automatic stations has obvious seasonal fluctuations, daily changes and spatial distribution characteristics, but the overall performance is relatively stable and shows a certain upward trend.

Key words: Ground automatic station, Meteorological data operational system, Quality control, Quality assessment