

基于 MDOS 的重庆市 2015 年气象观测数据质量分析^①

李奇临¹, 魏麟骁², 吴 翀³, 鲁小平¹

1. 重庆市气象信息与技术保障中心, 重庆 401147; 2. 重庆市气候中心, 重庆 401147;

3. 中国气象科学研究院 灾害天气国家重点实验室, 北京 100081

摘要: 实时与历史结合的气象资料业务系统(MDOS)的运行实施, 完善了气象资料质量控制体系, 调整了气象资料业务布局和流程. 本文利用 2015 年重庆市 35 个国家台站、1 914 个区域自动站的气象常规观测小时数据, 按台站划分、观测要素分类, 以可用率、可疑率、错误率、缺测率等为指标, 基于 MDOS 对数据进行评估与统计分析. 结果表明, 重庆市 2015 年国家站自动观测资料的可用率平均可达 99.56%, 平均缺测率为 0.44%, 平均可疑率、平均错误率均在 0.01% 以下; 区域站自动观测资料可用率平均为 67.82%, 平均缺测率为 32.12%, 平均错误率为 0.06%, 平均可疑率为 0.01% 以下. 据此, 本研究总结了影响气象要素可用率的主要因素.

关键词: 数据质量; 可用率; MDOS

中图分类号: P415.1⁺2

文献标志码: A

文章编号: 1000-5471(2018)11-0135-06

随着气象事业的快速发展, 气象观测数据质量控制的研究取得了重大进展, 在实时数据质量控制方面, 也有较大突破. 窦以文等^[1]针对 2008 年奥运会气象服务需求, 研发了自动气象站观测资料数据质量控制系. 任芝花等^[2]针对全国区域站和国家站实时小时降水数据, 研制了较为全面且有效的质量控制方案. 宋先居等^[3]研发了自动气象站地面气象观测数据实时传输及质控系统. 韩海涛等^[4]研究发现实时数据质量控制结合自动控制 and 人工经验的判断则更为有效. 李志鹏等^[5]开发了“省、市、县三级自动站疑似数据快速处理反馈业务系统”, 促进了观测业务工作的开展. 史静等^[6]研究了相对湿度与风向风速变化间的关联规则.

综上, 科研人员通过多年的努力, 提出了实时资料自动和人工质控相结合的方案, 并研究了某一气象要素的质控效果, 开发了具有一定可靠性和实用性的系统. 然而, 仍缺少一个全国统一的系统来进行全面的质量控制. 有的系统在数据缺测、及时修正方面存在不足, 有的系统质控方案不够全面. 2013 年, 湖北省在以历史地面气象资料为质量控制对象的基础上, 开发了“实时与历史结合的气象资料业务系统”(MDOS), 该系统结合了历史和实时资料, 涵盖了较为全面的质控方案, 现已投入全国气象观测业务中. 郑盐源^[7]对 MDOS 的使用总结了一部分经验. 陈丹凤等^[8]进一步探究了影响地面气象观测数据文件质量的主要因素, 为改进数据质量控制提供了参考依据.

目前 MDOS 功能不断完善, 流程不断调整, 已是现有较全面的数据质控系统, 能为气象预报、气象服务业务以及气象研究提供经质控后的数据. 因此本文基于 MDOS, 对重庆市 2015 年 35 个国家台站、1914

① 收稿日期: 2018-01-19

基金项目: 重庆市气象局青年基金项目(QNJJ-201504).

作者简介: 李奇临(1987-), 女, 硕士, 工程师, 主要从事常规气象观测数据质量控制研究.

个区域自动站的小时气象地面观测数据,按台站划分、观测要素分类,以可用率、可疑率、错误率、缺测率等为评估指标,进行数据质量的统计及分析,评估重庆市气象观测数据质量,并提出提高气象观测质量的建议.

1 MDOS 概况与研究数据

1.1 MDOS 概况

MDOS 将历史资料和实时数据相结合,开展了质量控制、疑误数据快速处理、台站元数据管理和产品实时滚动生成等业务,实现了“台站级—省级—国家级”三级质控的数据传输、处理、查询和反馈流程.如图 1 所示,在业务运行中,MDOS 处理流程包括数据入库、质量控制、统计加工、数据出库,而各流程之间的接洽都需要人工参与,人机交互是数据质控的重要保障.

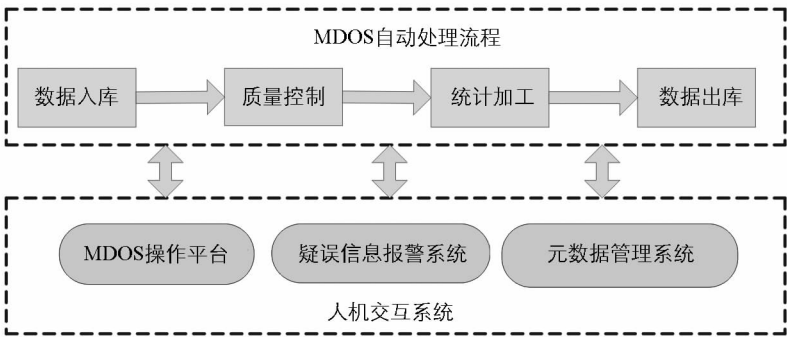


图 1 MDOS 自动处理流程与人机交互系统关系图

MDOS 的质控内容包括以下 9 个方面:格式检查、缺测检查、界限值检查、主要变化范围检查、内部一致性检查、时间一致性检查、空间时间一致性检查、质量控制综合分析和质量控制标识.经 MDOS 质控后,数据处理员需对显性错误数据、可疑数据和缺测数据进行人工判断及处理.

1.2 研究数据

本文使用的是重庆市 2014 年 12 月 31 日 21 时至 2015 年 12 月 31 日 20 时 35 个国家台站、1 914 个区域自动站的小时气象地面观测数据,包括:降水、风、气温、气压、相对湿度、地温、蒸发量以及能见度 8 项观测要素.图 2a、图 2b 分别为重庆市国家自动观测站点和区域自动观测站点的空间分布图,国家站的空间分布要比区域站均匀,区域站则分布在人口相对密集的地区.这是由于国家站的布局除了考虑气象服务外,还考虑了天气气候的代表性和空间分布上的广泛性,而区域站建立主要是为地方预警服务.

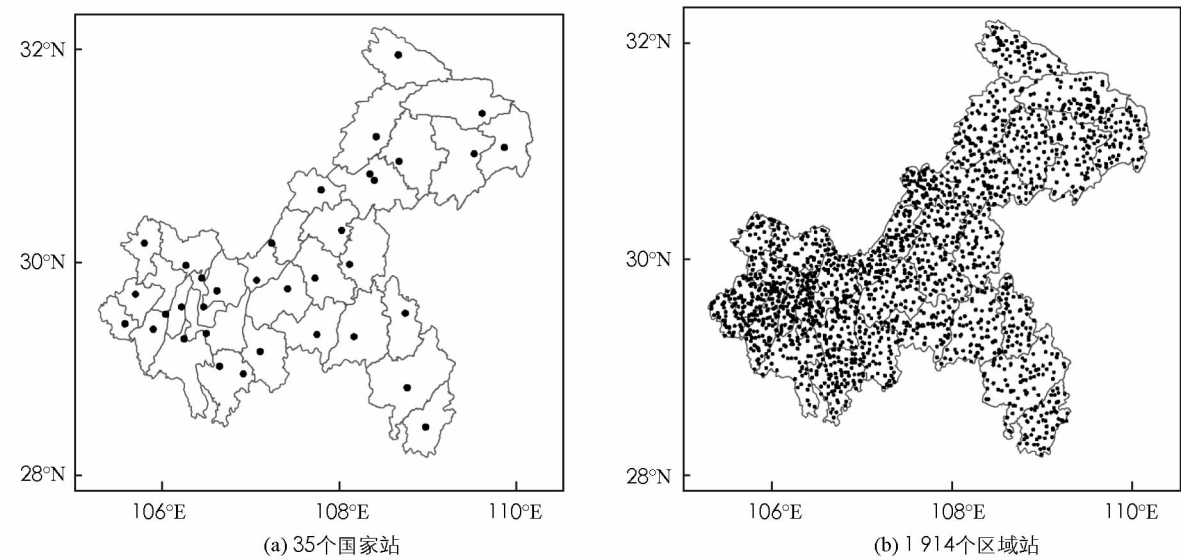


图 2 重庆地区自动站分布图

2 数据质量控制的结果及分析

2.1 国家自动站常规气象观测资料质量检验结果及原因分析

根据 2015 年重庆市 35 个国家气象站自动观测数据, 计算得出平均可用率为 99.56%, 平均缺测率为 0.44%, 平均可疑率、平均错误率均在 0.01% 以下, 详细统计结果如表 1 所示.

具体来看, 有 33 个站的可用率达到 99% 以上, 其中潼南站和武隆站的可用率相对偏低. 通过分析发现, 潼南站 2015 年 7 月迁站后, 存在台站海拔升高、观测仪器变动的情况, 从而导致该站部分气压数据缺测; 武隆站在 2015 年 7 月能见度由人工转换自动观测业务切换过程中, 出现了数据缺测. 可见, 台站变迁、观测仪器的变动以及观测方式的变化都是造成数据可用率偏低的因素.

表 1 2015 年重庆市国家自动站观测资料质量可用率统计表 %

台站	可用率	缺测率	台站	可用率	缺测率	台站	可用率	缺测率
江津	99.99	0.01	黔江	99.9	0.1	石柱	99.21	0.79
丰都	99.98	0.02	酉阳	99.9	0.1	万盛	99.16	0.83
大足	99.97	0.03	云阳	99.86	0.14	渝北	99.14	0.86
荣昌	99.97	0.03	奉节	99.84	0.16	北碚	99.13	0.87
涪陵	99.97	0.03	合川	99.8	0.2	彭水	99.1	0.9
綦江	99.97	0.03	长寿	99.75	0.25	城口	99.09	0.9
南川	99.96	0.04	沙坪坝	99.7	0.3	璧山	99.08	0.92
永川	99.95	0.05	秀山	99.48	0.52	开县	99.06	0.94
万州	99.94	0.06	巴南	99.46	0.54	巫溪	99.02	0.98
梁平	99.93	0.07	天城	99.42	0.58	武隆	98.64	1.36
忠县	99.93	0.07	巫山	99.37	0.62	潼南	98.62	1.38
铜梁	99.93	0.07	垫江	99.36	0.64	合计	99.56	0.44

可用率与缺测率具有较好的对应关系, 即数据缺测率高, 则可用率低, 反之亦然. 图 3a 和图 3b 分别为 2015 年重庆市国家自动站观测资料可用率和缺测率的空间分布图, 图中圆形面积大小与该要素所占比例呈正比. 图 3a 显示出大部分区域可用率较高, 尤其是江津、丰都站. 图 3b 能较明显反应出潼南和武隆的缺测率相对其他台站较高. 整体来看, 我市 35 个国家站的气象数据质量可靠性较高.

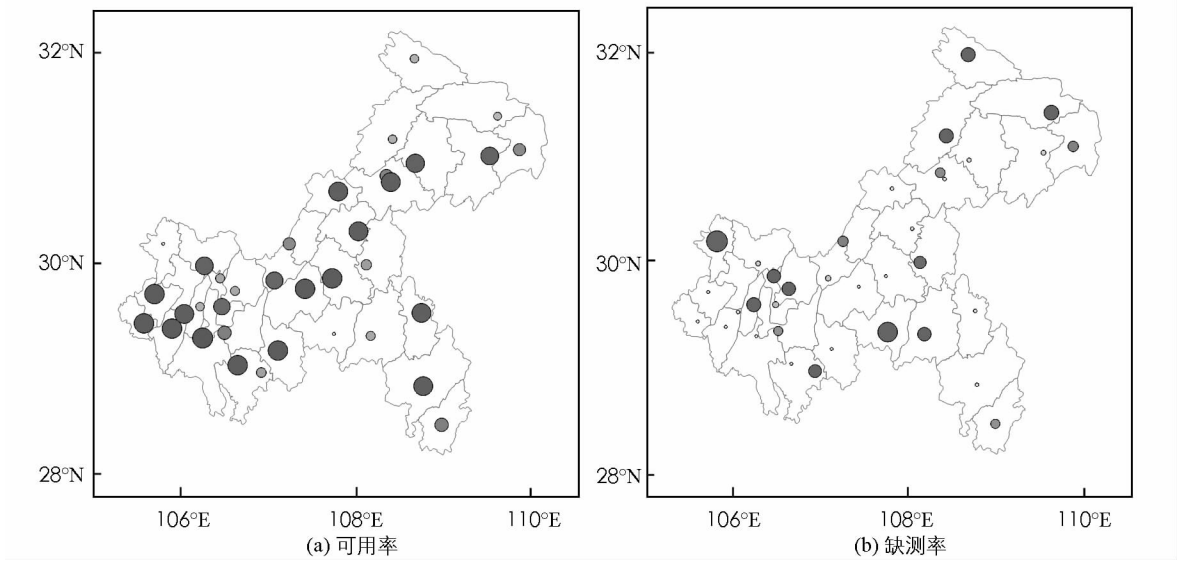


图 3 2015 年重庆市国家自动站观测资料质量统计分布图

2.2 区域自动站常规气象观测资料质量检验结果及原因分析

通过对 2015 年重庆市 1 914 个区域自动站常规气象观测数据分析可见, 数据可用率平均为 67.82%,

平均缺测率为 32.12%，平均错误率为 0.06%，平均可疑率在 0.01% 以下。图 4 为区域自动站观测资料质量统计分布图，由图 4a 所示，万盛、渝北所在区域内的自动站观测数据可用率相对较高，分别为 79.29%，75.59%，石柱区域的观测数据可用率较低，为 59.07%。由图 4b 所示，万盛区错误率为 0.46%，相对其余台站较高。

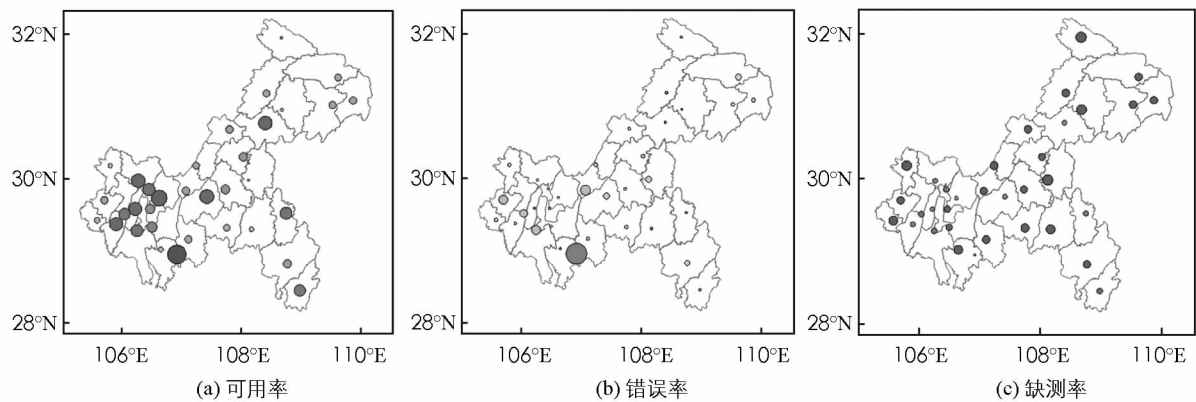


图 4 2015 年重庆市区域自动站观测资料质量统计分布图

分析发现，2015 年龙鳞石海北站的温度仪器长期存在故障，且无备用数据，因而导致气温数据错误，但万盛区域站的可用率仍然较高、缺测率较低，由此可见，评估数据的准确度不仅要看可用率，还要看错误率。部分区域自动站因地处偏僻，一旦仪器故障，维护所需时间较长。因此，加强区域自动站的数据质量，应安装稳定性较高的仪器设备，以提高数据准确度。

2.3 气象要素的质量统计及分析

按观测要素分类，分别对降水、风、气温、气压、相对湿度、地温、蒸发量以及能见度等 8 种观测要素的小时数据的可用率和缺测率进行统计，结果如图 5 所示，降水、风、气温、气压和相对湿度的可用率均达到 99% 以上，降水量的可用率最高，为 99.96%，地温、蒸发和能见度要素的可用率相对偏低，分别为 98.12%，98.09% 和 96.84%。这可能与我国自动气象观测要素仪器的研发周期有关系，应对仪器进行更新换代，使不同气象要素观测的稳定性达到一致的水平。

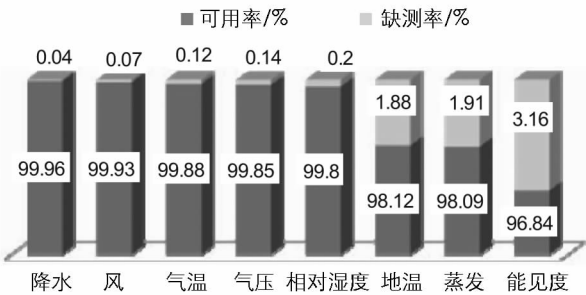


图 5 2015 年重庆市国家自动站和区域自动站的观测要素质量统计图

另外，对于某个时段可用率突然下降的情况，如巴南花溪城市区域站，其地温有较多疑误数据，提示为“未通过空间一致性的检验”，数据处理员按缺测处理，由于缺测数据较多，可用率降低。通过调研发现，该站地温仪器所在地不是纯土壤，情况特殊，地温偏高属正常，可以按照正确数据处理。又如，2015 年 7 月有 6 个站点可用率急剧下降，通过调研发现，这几个台站的蒸发、能见度观测均由人工转为自动，自动观测前期没调试好，因此造成转换起始期蒸发、能见度等数据部分缺测。可见，蒸发和能见度等观测要素的可用率大都受仪器故障、操作失误等影响。

综上分析，从国家站与区域站的数据可用率对比看出，数据的准确度差别较大。国家级自动站仪器由中国气象局统一规划、安装，而各省区域自动站仪器选型、安装则各省自定，存在一定的地域差异^[2,9]。整体来看，影响气象要素可用率的因素主要有：台站仪器故障、感应信号遭受干扰、数据表达格式及传输中的不规范以及人工操作误判等。因此，在今后的数据质控工作中，应在仪器准确度、有效数据传输率以及人工介入方面重点改进。

2.4 台站疑误数据反馈情况及分析

查询疑误数据是指 MDOS 质控后产生的疑误数据，经省级数据处理员反馈给台站数据处理员而需要

进一步人工核实的数据. 2015 年, 各台站查询疑误数据的空间分布来看(图 6a), 沙坪坝、綦江、梁平站, 重庆中部区域总体较多, 东北部和西部地区较少. 台站的反馈率方面(图 6b), 綦江、荣昌和万盛 3 站反馈率达到 100%, 15 个台站反馈率达到 90% 以上, 梁平大约 50% 的反馈率, 相对较低.

通过分析, 2015 年 7 月 10 日以前, 部分台站的通讯系统、MDOS 运行的硬件配置调试频繁以及台站对于 MDOS 产生的疑误数据在一段时间内未及时反馈, 是造成反馈率较低的主要原因, 另外, 还存在人工介入不及时的情况. 经 MDOS 质控产生的疑误数据, 人为进行内插或缺测处理后, 修改后的正确数据没及时上传, 未能确保 MDOS 上的数据和本地数据一致.

总的来看, MDOS 在数据质量控制方面发挥了较大优势, 大大提高了质控效率, 找出了以往人工不易察觉的设备问题或者数据一致性的问题, 同时人工与计算机质检结合, 完善了观测过程中的漏洞, 提高了数据的可用率, 减少了缺测率, 形成良性循环.

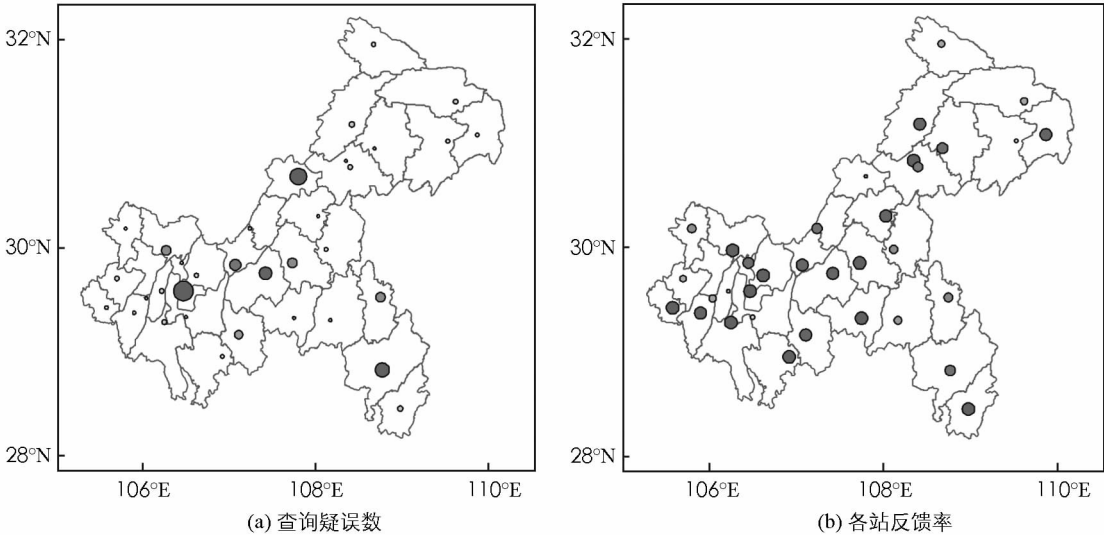


图 6 重庆各台站数据反馈情况的统计空间分布图

3 结论与展望

利用重庆市 2015 年 35 个国家台站、1 914 个区域自动站的小时气象常规观测数据, 按台站划分、观测要素分类, 以可用率、可疑率、错误率、缺测率等为指标, 对数据进行评估与统计分析, 并对 MDOS 按不同类的质控检查进行了研究, 主要得出以下结论:

- 1) 重庆市 2015 年国家站自动观测资料的可用率平均可达 99.56%, 平均缺测率为 0.44%, 平均可疑率、平均错误率均在 0.01% 以下, 江津、丰都站的可用率较高;
- 2) 重庆市区域站自动观测资料可用率平均为 67.82%, 平均缺测率为 32.12%, 平均错误率为 0.06%, 平均可疑率为 0.01% 以下, 万盛、渝北站的可用率较高;
- 3) 按降水、风、气温、气压、相对湿度、地温、蒸发量以及能见度观测要素进行数据分类, 结果表明温、湿、压、降水数据较为准确, 而地温、蒸发、能见度数据的准确度相对偏低;

提高数据质量, 即提高数据可用率. 通过本文分析, 影响气象要素可用率的因素主要有: 台站仪器故障、感应信号遭受干扰、数据表达格式及传输中的不规范和人工介入不及时等. MDOS 在质控检查方面发挥了一定了优势, 数据处理正确性和时效性较好, 充分利用了现有业务系统资源, 实现了台站、省级和国家级的数据传输、处理、查询和反馈流程. MDOS 仍有很多问题需要进一步的实践中去完善, 让观测数据更加准确、及时、完整.

参考文献:

- [1] 窦以文, 屈玉贵, 陶世伟, 等. 北京自动气象站实时数据质量控制应用 [J]. 气象, 2008, 34(8): 77—81.
- [2] 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 [J]. 气象, 2010, 36(7): 123—132.
- [3] 宋先居, 张宗灏, 赵体召, 等. 地面气象观测数据实时传输及质量控制程序实现 [J]. 气象科技, 2011, 39(3): 344—347.
- [4] 韩海涛, 李仲龙. 地面实时气象数据质量控制方法研究进展 [J]. 干旱气象, 2012, 30(2): 261—265.
- [5] 李志鹏, 张 玮, 黄少平, 等. 自动气象站数据实时质量控制业务软件设计与实现 [J]. 气象, 2012, 38(3): 371—376.
- [6] 史 静, 党 岳, 张永欣, 等. 自动站数据质量控制中关联规则挖掘的应用 [J]. 气象科技, 2014, 42(4): 612—616.
- [7] 郑盐源. 地面气象观测软件数据质量控制 [J]. 现代农业科技, 2017, (7): 222—224.
- [8] 陈丹凤, 陈伟昌. 影响地面气象观测数据文件质量的因素分析 [J]. 农业灾害研究, 2017, 7(2): 59—60.
- [9] 周国兵, 向 波, 胡春梅, 等. 基于自动能见度观测的雾和霾天气判别指标研究 [J]. 西南师范大学学报(自然科学版), 2017, 42(10): 78—83.

On Quality Statistics and Analysis of Meteorological Observation Data of Chongqing in 2015 Based on MDOS

LI Qi-lin¹, WEI Lin-xiao², WU Chong³, LU Xiao-ping¹

1. Chongqing Meteorological Information And Technical Support Center, Chongqing 401147, China;

2. Chongqing Climate Center, Chongqing 401147, China;

3. State Key Lab of Severe Weather, Chinese Academy of Meteorological Science, Beijing 100081, China

Abstract: The implementation of the real-time and historical combined meteorological data service system (MDOS) improved the meteorological data quality control system and adjusted the meteorological data business layout and flow. In this paper, we use the hourly observational data of 35 national stations and 1914 regional stations in Chongqing in 2015. According to the station classification and observation factors, we use the availability rate, suspicious rate, error rate, and missing rate as indicators. Based on MDOS data evaluation and statistical analysis. The results show that the availability rate of automatic observation data of national stations in Chongqing in 2015 averages up to 99.56%, the average rate of missed rate is 0.44%, the average suspicious rate and average error rate are all below 0.01%, and the availability of automatic observation data at regional stations is average. For 67.82%, the average missing rate was 32.12%, the average error rate was 0.06%, and the average suspicious rate was less than 0.01%. Based on this, this paper summarized the main factors affecting the meteorological factor availability rate.

Key words: data quality; availability; MDOS

责任编辑 包 颖