

吉林省区域自动气象站雨量数据的质量分析与评估

于翠红

(吉林省人工影响天气办公室, 长春 130062)

摘要: 区域自动气象站雨量资料对于气象预警、决策服务、效果检验都能提供有效的数据支撑, 建立科学有效的雨量资料质量评估与控制方法, 对于提高雨量数据的应用性至关重要。基于吉林省区域自动气象站5年雨量观测历史数据, 提出了适用于本省的雨量数据质量控制方法, 并利用此方法对雨量数据进行质量分析与评估, 最终得到错误及异常数据。根据错误及异常数据占比以及对站点占比情况, 可初步判断样本数据的数据质量相对较高, 可用性较强。

关键词: 区域自动气象站, 雨量, 质量分析

DOI: 10.3969/j.issn.2095-1973.2019.06.016

Quality Analysis and Assessment of Rainfall Data at Regional Automatic Weather Stations in Jilin Province

Yu Cuihong

(Jilin Weather Modification Office, Changchun 130062)

Abstract: The rainfall data at regional automatic weather stations can provide effective data support for meteorological warning, decision-making services and effect inspection. Establishing scientific and effective methods for quality assessment and control of rainfall data is essential for improving the applicability of rainfall data. Based on the historical data of 5-year rainfall observations at the regional automatic weather stations in Jilin Province, this paper presents a method for quality control of rainfall data to be suitable for the province, then uses this method to analyze and evaluate the rainfall data, and finally reveal error and abnormal data. According to the proportion of errors and abnormal data and the proportion of corresponding sites, it can be initially judged that the data quality is relatively high and the usability is good.

Keywords: regional automatic weather stations, rainfall, quality analysis

0 引言

区域自动气象站是气象防灾减灾预测预警系统的重要组成部分, 具有速度快、观测密度大、自动化程度高等特点。吉林省2006年开始使用区域自动气象站, 截至2014年吉林省在用的区域自动气象站数量已达1100多个, 其中大部分站点为无人值守的区域站。区域站的观测资料已实现了实时数据上传, 这些观测数据包括雨量、气温、湿度、风速、气压等数据, 它们对气象决策服务、气象预警、气象预报、效果检验等都提供了重要的数据支撑作用^[1]。众多站点的观测数据不可能完全无误, 往往存在观测数据明显异常的情况, 造成数据异常的原因是多方面的, 建立科学有效的资料质量评估与控制方法, 能在数据源头严把数据质量关, 进而提高数据的可用性, 为气象业务提供有效的支撑。

目前, 国内外学者对地面气象站雨量数据的质

量评估方法做过一些研究, 如: 任芝花等^[2]研究了自动站小时降水资料的质量控制方法, 莫遵元^[3]通过极值、空间连续性、时间连续性检查对区域气象站雨量数据异常进行分析处理, Sciuto等^[4]利用神经网络算法对日降水数据进行质量控制, 针对小时观测资料尤其是小时雨量观测资料质量评估方法的研究是比较有限的, 应用到小时雨量资料的质量评估方法主要是气候学界限值检查法, 诸如台站或区域极值检查、要素间内部一致性检查等应用于地面气象资料质量评估的方法基本尚未开展。

本文研究的开展主要基于区域自动气象站对气象决策服务、气象灾害预警等的支撑作用, 着眼于吉林省雨量数据的质量分析及评估, 提出适用于本省的雨量数据质量控制方法, 并在质量控制方法中将区域特点以及设备系统特点融入进来, 进而提升评估方案的全面性及可用性。

1 数据和方法

1.1 数据

本研究利用的数据为吉林省区域自动气象站的

收稿日期: 2018年12月25日; 修回日期: 2019年6月24日
第一作者: 于翠红(1985—), Email: 12379549@qq.com

2007—2011年每年4—9月的雨量数据。这些数据来自吉林省区域自动气象站雨量数据库。所利用的数据格

式详见图1。

首先，需要对数据的可用性做出初步筛查，筛

StationID	观测时间	状态	插入时间	分钟雨量	气温	电池电压
1	2007-10-06 18:11:00	0	2007-10-06 18:15:00	.1	14.2	12.8
1	2007-10-26 16:00:00	0	2007-10-26 16:04:00	.2	15.4	13.1
1	2007-10-26 16:02:00	0	2007-10-26 16:05:00	.4	15.4	13.1
1	2007-10-27 8:17:00	0	2007-10-27 8:21:00	.2	4	12.7
2	2007-10-06 14:29:00	0	2007-10-06 14:33:00	.1	17.4	12.9
2	2007-10-06 14:37:00	0	2007-10-06 14:41:00	.1	17.3	12.9
2	2007-10-06 14:44:00	0	2007-10-06 14:48:00	.1	17.2	12.8
2	2007-10-06 14:49:00	0	2007-10-06 14:52:00	.1	17.1	12.8
2	2007-10-06 14:54:00	0	2007-10-06 14:57:00	.1	17.1	12.8
2	2007-10-06 15:02:00	0	2007-10-06 15:06:00	.1	17.1	12.9
2	2007-10-06 15:25:00	0	2007-10-06 15:29:00	.1	17.2	12.8

图1 样本数据格式

Fig. 1 Sample data format

查结果为：本次数据抽取范围内所研究的区域自动气象站站点数量为797，样本数据总数据条数为1233500条，其中观测时间为空的数据为0条，重复插入的数据为0条，分钟雨量为空的数据也已剔除，数据初步可用。

在正式开始雨量数据的评估流程之前，需要将分钟样本数据转换为小时样本数据，这样才能满足评估方案的要求。转换原则：（1）同一站点同一自然小时内的分钟雨量数据相加和得到小时雨量数据；

（2）同一站点同一自然小时内的分钟温度数据求平均值，得到小时温度数据^[5]。

1.2 方法

1.2.1 界限值检查

1) 气候学界限值检查

气候学小时降水数据的界限值为：0~150 mm/h，超越该限值的数据记为错误数据。

2) 区域界限值检查

根据吉林省的经纬度和降水量空间分布情况，参考全国区域划分及国家气象信息中心发布的区域小时降水界限值分布，吉林省区域的小时降水界限值为0~100 mm/h，超越界限值的数据记为错误数据^[6]。

1.2.2 数据连续无变化检查

雨量计漏斗部分堵塞，承水器收集的降水以匀速渗漏的方式进入翻斗计量；测量仪器故障等情况均会造成数据连续无变化，这是一种数据异常情况。但是同时细雨绵绵似的降水也会造成降水数据连续无变化的情况，所以参考相关研究论著^[6]，提出以下检查方法：

连续6个及以上小时雨量数据相等，且满足以下3类情况之一（其中， $N1 < N2 < N3$ ），视为数据

异常。

1) 连续 $N1$ 个及以上小时雨量（简称 R ，下同）数据相等且 $R \geq 1.0$ mm。

2) 连续 $N2$ 个及以上小时雨量数据相等且 $R \in [0.5, 1.0)$ mm。

3) 连续 $N3$ 个及以上小时雨量数据相等且 $R \in (0, 0.5)$ mm。

1.2.3 针对电压和数据采集系统的检查

区域自动气象站的雨量计和温度计使用同一电源和数据采集系统，温度计在其测量过程中受外界因素的干扰较少，因此可以利用对异常温度数据的检查来判断自动气象站的电源和数据采集系统的工作情况^[7]。结合吉林省的温度变化特点，如果吉林省在4—9月份出现 -5 °C以下低温和 40 °C以上高温时视为异常值；如果相邻两个小时的温度值差值 > 8 °C，则视为异常数据。

根据所选定的评估方法，制定雨量数据质量评估流程，详见图2。

1.2.4 数据类型标记

1) 不满足气候学小时降水数据的限值的界数据记为错误数据，标记为W1。

2) 不满足区域界限值检查的数据记为错误数据，标记为W2。

3) 除W1和W2之外的存在温度值异常的，将这些数据标记为异常数据S1。

4) 除W1、W2、S1外的存在数据连续无变化的，将这些数据标记为异常数据S2^[8]。

5) 未做标记的可直接视为正常数据。

在进行错误及异常数据标记过程中，可同样将对应的站点也归入同一类型的错误及异常情况中，进行

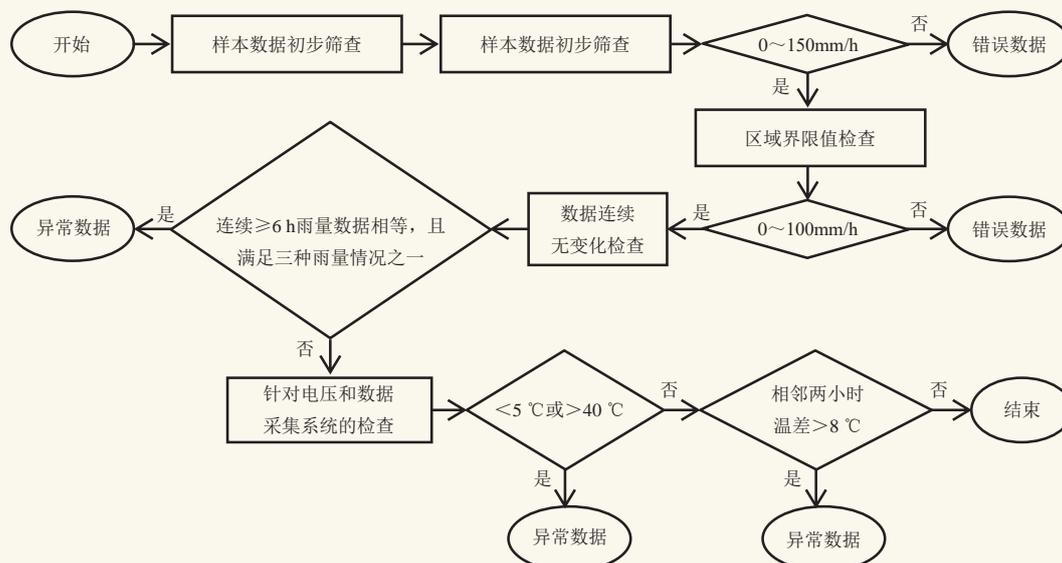


图2 雨量数据质量评估流程图

Fig. 2 Flow chart of the rainfall data quality assessment flow chart

后续分析。

2 结果与分析

根据上述质量评估方法，基于转换后的小时样本数据，编写数据库查询脚本，对各类错误及异常数据进行分析与评估。

依据质量评估方案实施过程的分类情况，样本数据的错误及异常情况已初步得到判定，如表1为站点质量评估结果，表2展示出样本数据质量评估结果，其中正确站点及正确样本数据的占比相对较高，可以初步得出本次雨量资料质量分析与评估过程所抽取的样本数据的可用性还是相对较高的。

表1 站点质量评估结果

Table 1 Results of the site quality assessment

站点总数	正确站点数量	错误站点数量	异常站点数量
797	574	8	215

表2 样本数据质量评估结果

Table 2 Results of the sample data quality assessment

总数据量	正确数据量	错误数据量	异常数据量
1233500	1201932	8	31560

因异常及错误数据条数较多，现以异常温度为例，选取典型的异常或错误数据进行截图展示，图3为吉林省区域自动气象站2006年4月四要素数据中，气温<-5℃或气温>38℃的异常数据；图4为吉林省区域自动气象站2006年4月四要素数据中相邻两小时温度数据差值绝对值>8℃的异常数据。

形成以上错误及异常数据的原因应该是多方面的，建议站点观测人员首先应检查观测设备的运转情

图3 吉林省区域自动气象站四要素 (气温<-5℃或气温>38℃数据)

Fig. 3 The four elements data of Jilin Province automatic weather station (temperature <-5℃ or temperature >38℃)

况，分析设备本身是否存在设备松动、老化、运转不正常等情况，进而导致观测错误及异常；其次，观测人员还应分析设备配套的软件系统，查找是否存在数据获取异常、数据叠加等问题。最后，加强定期巡检及相关理论学习。

3 小结

在对区域自动气象站雨量数据进行质量分析的过程中，也同样发现了异常温度数据，二者是相互检验的过程，相比温度数据有历史同期温度作比较分析，雨量数据的质量分析方法更具有探索性和多元化。目前各省对区域自动气象站的分析研究主要集中在运行维护、故障判断等硬件设备领域，从观测数据角度反

t. Station	t. 观测时间	t. 气温	t1. Station	t1. 观测时间	t1. 气温
2305	2006-04-14 16:15:00		2305	2006-04-14 14:01:00	-20.1
3205	2006-04-26 16:00:00	.9	3205	2006-04-26 14:47:00	15

t. Station	t. 观测时间	t. 气温	t1. Station	t1. 观测时间	t1. 气温
10801	2006-04-29 19:09:00	-6.7	10801	2006-04-29 17:01:00	12.3
10823	2006-04-19 16:36:00	1	10823	2006-04-19 14:34:00	15.9

图4吉林省区域自动气象站四要素200604 (相邻两小时温度数据差值绝对值>8的异常数据)

Fig. 4 The four elements data of Jilin Province automatic weather station 200604 (Abnormal data with an absolute value of the difference between the two consecutive hours of temperature data greater than 8)

推设备观测异常的研究相对较少, 但如广东省、新疆伊犁^[9]、湖南省^[10]等省市也都做了此类研究, 吉林省在区域自动气象站雨量数据质量控制方面的研究相对较少, 希望本文研究内容能对本省区域自动气象站雨量数据质量把控有所帮助。

在充分了解所考察数据所包含的业务内容和数据单位、格式等的前提下, 对可能导致异常数据的原因做了广泛的学习和研究, 并在此基础上建立了有针对性的质量评估方案, 并将此方案进行研究型应用, 主要用于对吉林省区域自动气象站的5年历史数据进行质量评估, 经评估得到该历史数据的正确率、错误率、异常率分别为97%、约为0、3%, 可见该历史数据数据质量相对较高, 对于异常数据及相应站点可进一步分析确定后反馈给有关部门进行跟踪校正。

以上质量评估结果仍存在很多不足之处, 如: 对于异常数据可在后续的研究过程中与雷达数据等进行横向比对, 这样可以增加数据质量评估的可信度; 另外对于历史数据的分析研究, 还不能够实现对实时上传数据的自动化质量控制。

后续将考虑进一步训练质量控制算法, 考虑利用卫星资料或雷达基数据进行定性分析, 从雷达或卫星资料得到可疑数据所属的时间段内区域自动站周围的降雨情况, 从而更深入地对可疑降水量进行分析判断。

参考文献

- [1] 张万东, 龙昭芪, 张凯伦. 浅谈区域自动气象站资料在防灾减灾中的运用. 贵州气象, 2010, 34(5): 42-43.
- [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. 气象, 2010, 36(7): 124-126.
- [3] 莫遵元. 区域气象站雨量数据异常的分析与处理. 农业与技术, 2018, 38(5): 146-148.
- [4] Sciuto G, Bonaccorso B, Cancelliere A, et al. Quality control of daily rainfall data with neural networks. J Hydro, 2009, 364: 14-20.
- [5] 何健, 王潜梅, 钱光明, 等. 广东省区域自动气象站资料的质量控制与评估. 广东气象, 2011, 33(3): 37-39.
- [6] 刘雨佳, 陈洪滨, 金德镇, 等. 加密自动气象站雨量计资料的质量控制及其相关关系的研究. 大气科学, 2014, 38(1): 162-163.
- [7] 冯振家. 区域自动气象站雨量自己部分常见故障分析及其维护. 气象研究与应用, 2010, 31(2): 44-45.
- [8] 王军海, 王忠云, 孙小丽. 区域自动气象站雨量标校应注意的问题. 山东气象, 2009, 29(116): 60-61.
- [9] 邱辉, 詹红霞, 刘振永. 伊犁河谷区域自动气象站资料分析. 宁夏农林科技, 2012, 53(9): 120-125.
- [10] 肖勇, 郭金慧. 区域自动气象站质量控制技术的研究及应用. 气象水文海洋仪器, 2016, 3(1): 98-101.