

金莉莉,何清,李振杰,等. 2016. 乌鲁木齐市气象塔梯度观测资料质量控制与处理. 气象, 42(6):732-742.

# 乌鲁木齐市气象塔梯度观测资料质量控制与处理<sup>\*</sup>

金莉莉<sup>1,3</sup> 何 清<sup>1,3</sup> 李振杰<sup>4</sup> 缪启龙<sup>2</sup>

买买提艾力·买买提依明<sup>1,3</sup> 赵逸舟<sup>5</sup>

1 中国气象局乌鲁木齐沙漠气象研究所, 乌鲁木齐 830002

2 南京信息工程大学应用气象学院, 南京 210044

3 新疆塔克拉玛干沙漠大气环境观测试验站, 塔中 841000

4 云南省临沧市气象局, 临沧 677099

5 新疆乌鲁木齐市气象局, 乌鲁木齐 830002

**提 要:** 本文介绍了乌鲁木齐市建成区内 5 座 100 m 气象塔地理位置, 根据逻辑极值检查、僵值检查、时间一致性检查、空间一致性检查和人工干预检查等组成的一套针对气象塔观测资料的质量控制方法和四点滑动平均插值、要素垂直分布拟合和线性回归方法组成的数据插值方法, 对乌鲁木齐市从南至北 5 座 100 m 气象塔 2012 年 4 月 1 日至 2014 年 4 月 30 日的资料进行质量控制和数据插值。结果表明: 该套方案能很好地找出缺测、错误和可疑的数据, 结合人工干预, 使得检验结果更为可靠。5 座气象塔资料的质量是比较好的, 正常数据占总数据的 97.17%, 其中红光山数据质量最好(正常数据占总数据的 99.01%)。非正常数据只是少数情况(占总数据的 2.83%), 包括缺测和错误数据, 其中缺测数据占非正常数据的 6.23%(出现在米东和燕南立交), 错误数据占非正常数据的 93.77%。虚假数据占错误数据的 89.99%(大部分为风速、风向), 僵值数据占错误数据的 5.47%(大部分为气温和湿度), 超出逻辑极值数据占错误数据的 0.41%(只出现在水塔山气温), 不符合一致性数据占错误数据的 4.13%(主要为湿度, 主要在燕南立交)。利用四点中央插值法、每座气象塔要素垂直分布拟合和不同气象塔之间线性回归方法插值气温和相对湿度的效果较好。

**关键词:** 乌鲁木齐市, 气象塔, 质量控制, 插值, 观测资料

**中图分类号:** P413

**文献标志码:** A

**doi:** 10.7519/j.issn.1000-0526.2016.06.009

## Quality Control Method and Treatment for Urumqi Meteorology Tower Gradient Observation Data

JIN Lili<sup>1,3</sup> HE Qing<sup>1,3</sup> LI Zhenjie<sup>4</sup> MIAO Qilong<sup>2</sup> ALI Mamtimin<sup>1,3</sup> ZHAO Yizhou<sup>5</sup>

1 Institute of Desert Meteorology, CMA, Urumqi 830002

2 College of Applied Meteorology, Nanjing University of Information Science and Technology, Nanjing 210044

3 Taklimakan Desert Atmosphere and Environment Observing and Experimental Station of Xinjiang, Tazhong 841000

4 Lincang Meteorological Office of Yunnan, Lincang 677099

5 Urumqi Meteorological Bureau of Xinjiang, Urumqi 830002

**Abstract:** This paper introduced geographical position of 5 100 m high meteorological towers in Urumqi. A series of quality control methods were advanced to deal with observation data, such as logical extreme check, static check, time consistency check and similarity check. Besides, a series of data interpolation methods were advanced to interpolation controlled data, such as four point moving average interpolation,

\* 中国沙漠气象科学研究基金(Sqj2015009)、中央级公益性科研院所基本科研业务费专项资金(IDM201505)、公益性行业(气象)科研专项(GYHY(QX)201506001-14)共同资助

2015 年 8 月 17 日收稿; 2016 年 2 月 21 日收修定稿

第一作者: 金莉莉, 主要从事大气边界层和陆面过程模拟研究. Email: jinlili1984@126.com

通信作者: 何清, 主要从事沙漠气象研究. Email: qinghe@idm.cn

elements of vertical distribution fitting and linear regression. Especially in the treatment process, using manual intervention can make the test results more reliable. The quality control method was applied to data from the 5100 m high meteorology towers from 1 April 2012 to 30 April 2014 in Urumqi. The results showed that the method can differentiate the missing, error and dubious. The data of 5 towers are reliable with normal data accounting for 97.17% of the total data, which is the best quality of Hongguangshan tower (normal data accounting for 99.01% of the total data). Abnormal data only account for a small proportion (accounting for 2.83% of the total data), including missing data and error data. Missing data accounts for 6.23% of the abnormal data, which mainly occurred in Midong Tower and Yannanlijiao; error data accounts for 93.77% of the abnormal data. Among error data, false data accounts for 89.99% of the error data (mainly wind speed and wind direction); static data accounts for 5.47% of the error data (mainly air temperature and relative humidity); extreme error data accounts for 0.41% of the error data (mainly air temperature in Shuitashan Tower); consistency error data accounts for 4.13% of error data (mainly relative humidity in Yannanlijiao Tower). The effect of temperature and relative humidity interpolation are better by using the four points moving average interpolation method, elements of vertical distribution fitting method and linear regression method between different meteorological towers.

**Key words:** Urumqi, meteorology tower, quality control, interpolation, observation data

## 引言

第二次世界大战后,世界人口经历了前所未有的发展,同时城市也经历了急剧的扩展,城市面临着巨大的资源与环境的压力,出现了诸多城市环境问题,成为全球环境问题的重要组成部分(Tyler et al, 2007; Peter et al, 2007)。近年来,随着城市气象发展进展表明,城市大气条件和过程的观测是提高城市天气和气候认识的基础。城市气象观测已经有近200年的历史(Grimmond, 2006)。国际和国内开展了许多大型的城市气象综合观测试验和计划,这些观测试验和计划主要集中在美国、欧洲、中国和日本。

城市低层大气结构是大气边界层的重要组成部分,充分表现出下垫面动力和热力作用的直接影响结果以及中小尺度天气系统的部分特征,它们反过来又作用于大气污染、人类生活和生产活动的许多方面(解以扬等, 2003)。研究低层大气结构,可以满足城市气象预报业务和污染治理的需要。乌鲁木齐市建成区内从南至北建立了5座100 m气象塔,充分利用这5座气象塔全天候的连续观测资料,有助于深入细致地了解乌鲁木齐市低层大气结构变化规律和精细的三维流场特征。

然而,资料质量可靠才能保证结论的正确性,因此需要对铁塔气象要素资料进行质量评价和质量控制。气象资料经历了仪器观测、记录、传输,甚至编

码、解码等重要环节,各个环节都可能引起数据错误,而气象塔出现错误资料的几率远远高于常规地面观测资料(王超等, 2010)。质量控制主要是对原始资料修正或剔除异常值(WMO, 2004)。气象资料质量控制包括:地面常规观测资料(周笑天等, 2012; 王轶等, 2013; 徐安伦等, 2013; 王海军等, 2014; 李岩松等, 2014; 刘莹等, 2015; 任芝花等, 2015; 杨云等, 2015; 李明华等, 2007)、高空常规观测资料(李平等, 2013)、雷达观测资料(肖艳姣等, 2012; 朱立娟, 2015; 董德保等, 2015; 赵瑞金等, 2015; 张旭斌等, 2015)和卫星观测资料(薛湛彬等, 2013)等。而边界层资料质量控制工作是十分复杂的,由于整个边界层物理学的复杂性使得资料质量控制难以选择一个精确的判断标准。因此,本文结合了地面观测资料常用的质量控制方法,提出在乌鲁木齐市建成区内设立的5座100 m气象塔资料进行质量控制分析研究,建立了一套适合该区域较为完整的资料质量控制系统。

## 1 观测地点和观测仪器

### 1.1 观测地点

观测地点位于乌鲁木齐市  $42^{\circ}45'32'' \sim 44^{\circ}08'00''N$ 、 $86^{\circ}37'33'' \sim 88^{\circ}58'24''E$ , 海拔 580 ~ 920 m, 中温度大陆性干旱气候。乌鲁木齐市三面

环山,市区呈开口朝北的“喇叭口”地形(王春华等, 2010)(图 1)。

本研究所用资料的站点为乌鲁木齐市 5 座 100 m 气象铁塔,其分布位置如图 2 所示,具体信息见表 1。5 座 100 m 气象铁塔分别位于乌鲁木齐市南郊(燕南立交)、北郊(米东)、近北郊(红光山)、城区(鲤鱼山和水塔山)。

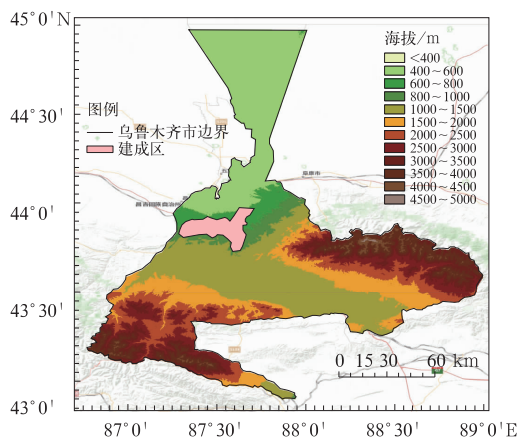


图 1 乌鲁木齐市地理位置  
(粉色区域为乌鲁木齐市建成区)

Fig. 1 The location of Urumqi and the urban area of Urumqi (pink area)

## 1.2 观测仪器

5 座 100 m 气象塔各气象传感器在安装架设前均进行了标定,塔上的仪器采取分层设置,仪器间隔分别为 10、13、17、22、28、36、46、60、77、100 m,共有 10 层。具体观测设备见表 2。

表 2 观测设备主要参数一览表

Table 2 Observational sensors and equipment

观测要素	仪器型号	测量范围	测量精度	厂家
风速	ZQZ-TF	0~75 m·s <sup>-1</sup>	±(0.3+0.02 v) m·s <sup>-1</sup>	江苏省无线电科学研究所有限公司
风向	ZQZ-TF	0~360°	±5°	江苏省无线电科学研究所有限公司
温度	HMP45D	-40~+60℃	±0.1℃	VAISALA
湿度	HMP45D	0.8~100%	±1%	VAISALA
气压	PTB220 型	500~1100 hPa	0.1 hPa	VAISALA
数据采集系统	ZQZ-CII-SE			江苏省无线电科学研究所有限公司

## 2 铁塔常规气象要素的数据质量控制

气象铁塔资料质量控制问题对实现边界层实测资料的质量控制具有相当的代表性和实用价值。在原始资料中,由于仪器的系统误差,仪器发生故障以及传输和记录过程中的其他原因,会出现一些虚假

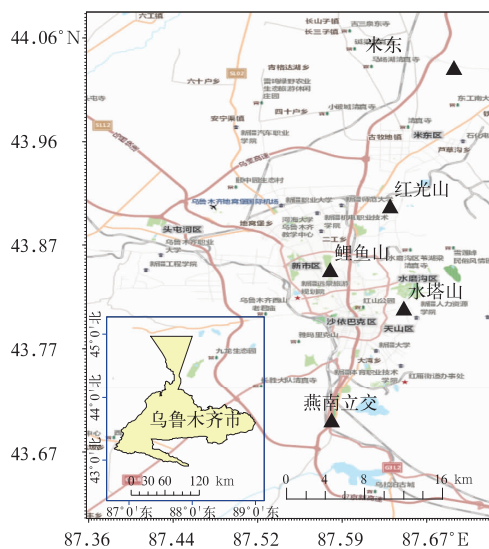


图 2 乌鲁木齐市 5 座 100 m 铁塔地理位置

Fig. 2 The location of 5 100 m height towers sites in Urumqi

表 1 乌鲁木齐市 5 座 100 m 气象铁塔信息

Table 1 The information of 5 100 m height towers

站点位置	铁塔高度 /m	经度/°E	纬度/°N	海拔高度 /m
燕南立交	100	87°34'47"	43°42'19"	1009
水塔山	100	87°38'50"	43°49'06"	890
鲤鱼山	100	87°34'45"	43°50'41"	821
红光山	100	87°38'05"	43°54'14"	717
米东	100	87°41'37"	44°01'58"	568

的数据,这些数据一旦被诊断出来,就必须进行修正或者剔除,否则在下面的理论分析中出现很多错误的结果,影响结果的可靠性(陈明等,1993)。但对数据做到恰如其分的订正是很不容易的,为了最大限度地保护原始资料,在处理数据的时候只对明显的错误的数据进行剔除和订正处理。

边界层常规气象要素数据的质量控制方法与地

面、高空常规观测数据质量控制方法有相似之处,都要考虑极值检查、内部一致性检查、时间一致性检查和空间一致性检查等方法(王新华等,2006;何志军等,2010;李志鹏等,2012;张乐坚等,2016),但并不能将地面、高空常规观测数据质量控制方法完全用于边界层常规气象要素数据的质量控制。其中,边界层数据极值检查方法与地面、高空观测数据的极值检查方法一样,都要检查某气象要素值是否超过该站该要素历史上出现过的最大值和最小值。对于地面观测数据,需要进行内部一致性检查,如总云量必须大于等于低云量等。同地面观测资料的时间一致性检查一样,边界层常规气象要素也要考虑气象要素在一定时间内的变化是否具有特定规律。地面观测资料的空间一致性检查是将某一站点的观测资料与其周围邻近站点的资料进行比较分析,从而判别该要素是否正常(王新华等,2006);边界层常规气象要素数据空间一致性检查的方法与高空常规观测资料的相似(李平等,2013),但不能完全参照。

边界层常规气象要素数据的质量控制可以参考地面、高空常规观测数据质量控制的方法,但由于时间分辨率和空间分辨率的不同,因而需要根据资料本身的特点来选择判别数据是否正常的阈值。

本文数据质量控制和插值流程如图 3 所示,该流程主要针对 5 座气象铁塔不同高度和不同气象要素的质量控制和插值。

### 2.1 质量控制方法

本文所用的观测资料时间分辨率均为 10 min,在资料质量控制的时候,遵循资料的完整性、原始性和一致性原则,尽量对缺测数据做插补,使修正数据最少,使廓线尽可能地光滑。同时参考彭珍(2005)和王超等(2010)资料质量控制方法。其质量控制包括以下几个方面:

(1) 逻辑极值检查。检查数据采集时间的合法性;进行极值诊断。极值是根据观测仪器的范围和乌鲁木齐市历年气象资料统计出的各个气象要素的最大值和最小值(本文极值诊断标准参考乌鲁木齐市气象站和米东气象站 1981—2010 年地面气候要素平均与极值)。如果当观测值超过极值,认定为错误值。具体判断指标见表 3。

此外,由于缺测、数据乱码或故障造成数据错误而被用 0 和 99999 取代,因此,如果所测的风速、风向、温度和湿度数据都为 0 和 99999,那么该层数据作

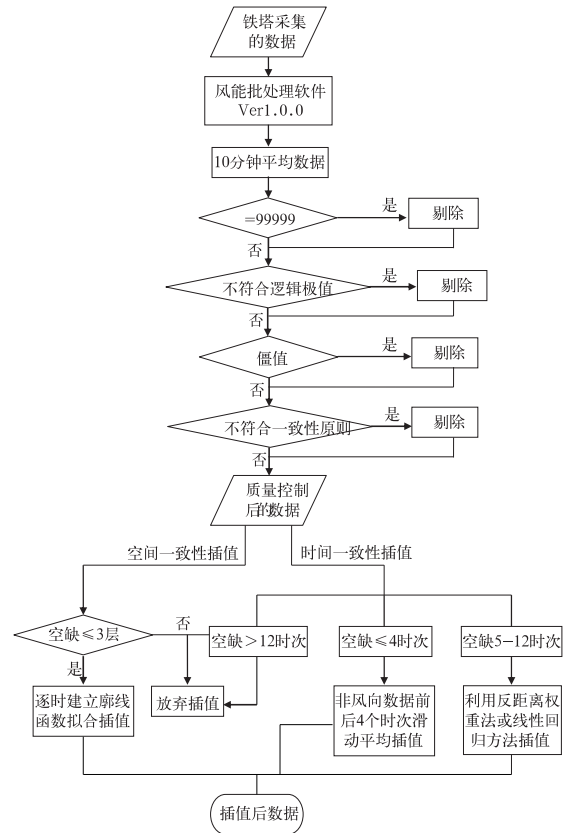


图 3 铁塔资料质量控制和插值方法流程

Fig. 3 The process of quality control and interpolation method

表 3 逻辑极值检查指标

Table 3 The threshold of logical extreme value check

观测要素	范围
风速	0~40 m·s <sup>-1</sup>
风向	0~360°
温度	春季-25~40℃、夏季 0~50℃、 秋季-30~40℃、冬季-35~20℃
湿度	0~100%
气压	600~1000 hPa

为虚假数据予以剔除。

(2) 僵值检查。僵值检查也称最小变率检查。僵值是由于观测仪器发生故障以及传输和记录过程中的其他原因,使观测值长时间不变,引起记录不真实。考虑到观测数据频率比较快,我们规定,将连续出现某个观测值的数目超过 10 个的数据段定为僵值。如果找出连续相同数据(连续个数>10,其数值假设为 N),及连续相同数据之后第一个与 N 之差大于特定值 D 的数据(M),则 M 之前连续相同数据则认为是僵值,将其剔除,否则予以保留(这是为了最大限度保留原始数据)。D 值取值为仪器测量精

度的 1~2 倍(风向  $\Delta \text{deg} > 5$ , 风速:  $\Delta U > 0.3$ , 温度:  $\Delta T > 0.1$ , 湿度:  $\Delta f > 1$ , 气压:  $\Delta P > 0.1$ )。

(3) 时间一致性检查。根据要素在某一时段内可能变化范围判断该要素值的质量, 风向不具有连续性, 故不做此检查。判断依据: 将某一要素  $X_i$  与其前一时刻和后一时刻观测值分别作差的绝对值的积的大小  $<$  规定阈值  $D$ , 即  $|X_i - X_{i+1}| |X_i - X_{i-1}| < D$  记为 0(正常), 若该值  $> D$  则记为 1(怀疑), 若该值  $> 2D$  则记为 2(错误), 当前后有缺测时记为 8(无法判断)。本文时间一致性  $D$  值的取值为: 风速:  $U = 45$ , 温度:  $T = 25$ , 湿度:  $f = 100$ , 气压:  $P = 10$ 。

(4) 空间垂直一致性检查。当相邻两观测高度间要素值之差  $>$  给定的阈值, 则认为数据可疑, 除非在特殊天气条件下(如锋面过境), 气象要素不可能在有限的时空范围内发生极剧烈的变化。在此只做温度、湿度和风速空间一致性检查。判断依据: 将 10 层要素分为 9 组, 即 10、13 m 为第一组, 13、17 m 为第二组, 17、22 m 为第三组, 22、28 m 为第四组, 28、36 m 为第五组, 36、46 m 为第六组, 46、60 m 为第七组, 60、77 m 为第八组, 77、100 m 为第九组, 分别求其差的绝对值  $A$ , 然后规定阈值  $D$ ,  $A < D$  标记为 0(可信),  $D < A < 2D$  标记为 1(怀疑),  $A > 2D$  标记为 2(错误), 再利用与其相邻的一组或多组判断出错误数据归属的高度层, 从而知道具体的错误数据分布; 如果一组中有一个数据没有通过前面的检查, 则标记为 8(无法判断)。本文空间一致性  $D$  值的取值为(见表 4)。由于水塔山 1~5 层资料与 6~10 层资料不在同一年份, 因此做空间一致性检查时, 我们分为八组, 即 10、13 m 为第一组, 13、17 m

为第二组, 17、22 m 为第三组, 22、28 m 为第四组, 36、46 m 为第五组, 46、60 m 为第六组, 60、77 m 为第七组, 77、100 m 为第八组, 方法同上, 其中第一组至第四组阈值分别采取表 4 中第一组至第四组的阈值, 第五组至第八组阈值分别对应采取表 4 中第六组至第九组阈值。

表 4 中阈值的选取, 主要是根据近地层气象要素特征、仪器的采样精度来确定的。参照陈明等(1993)对北京 325 m 铁塔气象资料的控制方法, 熊安元(2003)的北欧气象观测资料控制方法, 王超等(2010)对敦煌戈壁多层气象资料的控制方法, 用最小二乘法结合各层权重等方法, 来确定每一组气象要素的阈值。

(5) 人工干预检查。质量控制不可能解决所有的数据质量问题, 对于一些诸如因中小尺度现象或局地气候使得某一观测要素产生较大变化及似是而非的资料, 还需要进行人工检查。

综上所述, 表 5 给出各要素采用的质量控制方法。

表 4 空间一致性检查的阈值

Table 4 The threshold of similarity check

	风速	温度	湿度
第一组	0.5	0.3	1.5
第二组	0.7	0.5	2
第三组	0.8	0.7	3
第四组	0.8	0.7	3
第五组	0.8	0.7	3.5
第六组	1	0.8	4
第七组	1	1	5
第八组	1	1.2	6
第九组	1.5	1.5	7

表 5 各观测值采用的质量控制方法

Table 5 Quality control method of all observation data

要素	逻辑极值检查	僵值检查	时间一致性检查	空间一致性检查	人工干预检查
风向	✓	✓			✓
风速	✓	✓	✓	✓	✓
温度	✓	✓	✓	✓	✓
湿度	✓	✓	✓	✓	✓
气压	✓	✓	✓		✓

## 2.2 数据的检查与结果分析

利用上面所述的数据质量控制方法, 对本文中所使用的乌鲁木齐市 5 座 100 m 气象铁塔 10 min 平均数据进行了全面的质量检查, 结果见表 6。

乌鲁木齐市气象铁塔的观测质量还是比较好的, 非正常数据出现较少。5 座铁塔正常数据占总数据的 97.17%, 非正常数据占总数的 2.83%。气压均为正常。非正常数据中, 缺测数据只占很小部分(6.23%)。由于仪器或软件问题, 会出现整天没

有数据的情况,主要是米东和燕南立交,其余 3 座铁塔无缺测数据。错值数据中(图 5c),大部分是因为出现了虚假数据和僵值数据而导致的错值。风向、风速出现虚假值的次数最多,其次为气温和湿度;而气温和湿度出现僵值的次数最多,其次为风向。僵值数据以鲤鱼山的最多,该错误的大部分出现在 60 和 10 m 风向。超出逻辑极值的数据占错误数据的

0.41%,该错误数据只出现在水塔山 100 m 夏季气温(1062 个)和秋季气温(5 个)。一致性错误数据主要是湿度,且主要出现在燕南立交。对于经过上述所有质量控制步骤进行检查后,我们进行人工判断,经过人工干预检查出的错误数据主要是温度和湿度,且所占比例非常小。

表 6 各铁塔各气象要素质量控制结果(单位:%)

Table 6 Quality control results of various meteorological elements of each meteorological tower(unit:%)

铁塔项目		燕南立交				水塔山				鲤鱼山				红光山				米东			
正常(占总数 97.17)		96.96				97.64				95.09				99.01				96.04			
		风速	风向	气温	湿度	风速	风向	气温	湿度	风速	风向	气温	湿度	风速	风向	气温	湿度	风速	风向	气温	湿度
非正常 (占总数 2.83)	虚假(占错误 89.99)	4.93	4.49	3.61	3.61	0	6.7	6.7	0	21.71	9.7	1.65	1.68	3.6	2.54	0.01	0	8.58	7.81	4.48	4.47
	极值(占错误 0.41)	0	0	0	0	0	0	0.41	0	0	0	0	0	0	0	0	0	0	0	0	0
	僵值(占错误 5.47)	0	0.17	0.13	0.39	0	0.12	0.44	0.56	0	0.79	0.45	0.22	0	0.09	0.58	0.14	0	0.01	0.39	0.2
	不一致(占 错误 4.13)	0.3	0	0.12	1.40	0.03	0	0.34	0.39	0.01	0	0.05	0.51	0.02	0	0.06	0.38	0.03	0	0.16	0.3
	人工干预	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0
缺测(占非正常 6.23)		14.44				0				0				0				11.07			

图 4 给出了不符合逻辑极值的数据,这是水塔山 2012 年 6 月 2 日 00—23 时 100 m 气温情况,气温记录在 01、03、05 时、08—12 时毫无根据地为 -81~-29℃,这在 6 月是不可能出现的。我们根据季节的极值判定就能轻易地辨别出来。

图 5 是鲤鱼山 2013 年 6 月 4 日—5 日 60 m 风向僵值数据,这几天 60 m 的风向记录一直保持为同一个值(248°),如此长时间的观测值不变,可认定为僵值。

图 6 是南郊 2014 年 1 月 9—11 日湿度不符合一致性的数据,36~100 m 湿度在 9 日 09:40—10:10,10 日 05:00—10:20 猛降到 0%,而同时段的

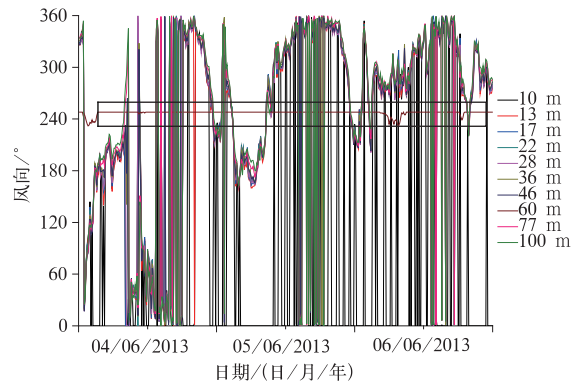


图 5 典型的僵值数据

Fig. 5 Typical static error data

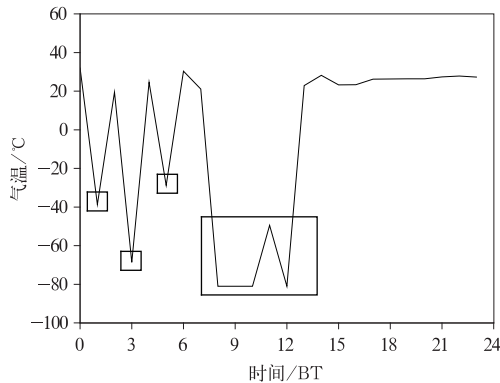


图 4 典型的超过逻辑极值数据

Fig. 4 Typical extreme error data

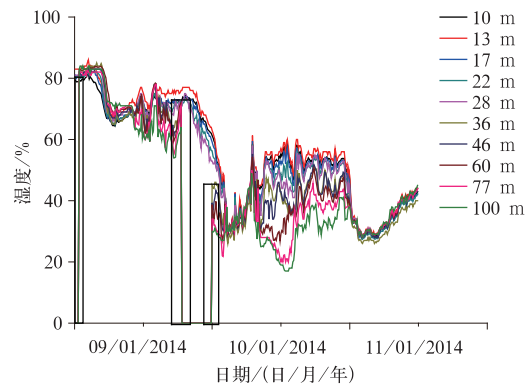


图 6 典型的不符合一致性数据

Fig. 6 Typical consistency error data

其他高度的湿度均在 50% 以上, 所以根据一致性原则认定为错值。

### 3 数据的插值

为了保证数据的完整性, 需要对缺测的数据插值, 具体解决方式如下: 首先采用四点滑动平均插值法。当前后两个数据存在缺测或被剔除值而无法采用该插值方法, 且其他高度层数据完好时, 根据空间连续性原则, 采用垂直分布拟合方法插值, 对于仍不能满足每一座塔各自插值的, 则采用反距离权重法插值, 或者利用其他塔同步完整数据的线性回归方法插值出本塔数据。具体介绍如下:

#### 3.1 要素垂直分布拟合方法

每一座塔各自采用逐时拟合特征廓线函数插值法把其他高度的缺测数据补上, 所用函数形式为  $Y$

$= a \log X + bX + c$ , 其中  $X$  为缺少的高度,  $a$ 、 $b$  为其他层次拟合出来的函数的系数,  $c$  为对应常数。

该方法主要是依据气象要素在垂直方向上的分布规律(一般为对数分布规律), 插值步骤为: (1) 判定 10 层数据中缺少的层次少于或等于 3 层; (2) 逐时次拟合出垂直方向对应的对数函数廓线形式; (3) 将对对应缺测的高度值代入拟合出的函数; (4) 插值出缺测高度的气象要素数据; (5) 根据前述第 2 节的数据质量控制方法, 再次进行质量控制, 剔除本方法拟合出的不符合条件的数据。

根据上述方法, 进行垂直方向插值的数据量所占比例较小, 如表 7 所示。

#### 3.2 反距离权重方法分析

各铁塔距待估铁塔的距离不同(表 8), 其气象要素值对待估铁塔的影响程度也不同。

本文研究中选用幂次为 2 来进行编程插值。

表 7 气象塔空间垂直方向插值数量及比例统计

Table 7 The number and proportion of interpolation spatial vertical direction of meteorological towers

站点	空缺层数	风速		气温		相对湿度	
		时次	比例/%	时次	比例/%	时次	比例/%
红光山	1	1807	3.8	454	0.9	267	0.5
	2	910	1.9	179	0.4	6	0
	3	357	0.7	95	0.2	30	0.1
鲤鱼山	1	—	—	495	1.0	495	1.0
	2	1192	2.5	194	0.4	36	0.1
	3	213	0.4	64	0.1	28	0.1
米东	1	2213	4.6	380	0.8	377	0.8
	2	935	1.9	158	0.3	76	0.2
	3	375	0.8	66	0.1	0	0
燕南立交	1	774	1.6	147	0.3	835	1.7
	2	200	0.4	62	0.1	74	0.2
	3	64	0.1	19	0	10	0

表 8 铁塔之间的距离

Table 8 Distance between the 5 towers

铁塔距离组合	距离/m
红光山 鲤鱼山	7965.9
红光山 米东	15066.5
红光山 燕南立交	22519.6
红光山 水塔山	9573.3
鲤鱼山 米东	22829.2
鲤鱼山 燕南立交	15488.5
鲤鱼山 水塔山	6206.4
米东 燕南立交	37522.5
米东 水塔山	24122.8
燕南立交 水塔山	13681.3

例, 验证 5 座铁塔间的反距离权重法插值效果。比如插值鲤鱼山 10 m 气温, 将运用红光山、米东、燕南立交和水塔山 10 m 高度同步气温数据。利用其他 4 座铁塔反距离权重法插值成的 5 座铁塔 10 m 高度气温与实测数据的对比如表 9 所示。

由表 8 和表 9 对比可知, 利用反距离权重法进行拟合, 造成的误差较大, 因此, 本文的插值方法不选用此方法。

#### 3.3 线性回归方法分析

由于红光山铁塔数据完整, 因此建立其他铁塔和红光山铁塔气象要素相关函数。建立的 10 m 气

限于篇幅, 本文仅以 5 座塔 10 m 气温插值为

温全部现有数据一元线性函数及误差参数如表 10 所示。红光山铁塔和鲤鱼山、米东、燕南立交、水塔山铁塔 10 m 气温都具有很好的一元线性相关。 $R^2$  能达到 0.98 以上,且各项误差系数较小,符合插值要求。其中,离红光山最远的燕南立交,比其他 3 座塔的相关性略小。

考虑到气温的年内变化,采用逐月数据进行一元线性相关性分析,并对拟合出来的值和实测值进行比较,方程和各参数如表 11 所示(逐月建立的一元线性函数不再列出)。由表 10 与表 11 可知,采用逐月回归分析后,建立的一元线性函数,其相关性改善很小, $R^2$  仅增加 0.0002。 $Bias$ (平均偏差)提高 2

个数量级左右, $SEE$ (标准差)增加 0.033, $NSEE$ (相对标准差)减少约 0.002。此外,我们也进行了多元一次函数的相关性分析(表略),相关性的改进更小。因此,对气温采用全序列一元线性函数进行插值,即满足相关性要求。

相对湿度也具有明显的日变化,但不如气温变化明显。对红光山和其他 4 座塔之间 10 m 相对湿度进行一元线性相关分析,各项参数如表 12 所示。红光山与鲤鱼山、米东、水塔山的相对湿度相关性很好,但燕南立交略差,这可能与燕南立交与红光山距离较远,且地形地貌差别最大有关。 $R^2$  均能达到 0.93 以上,具有很好的相关性,各项误差量值均较小。

表 9 5 座塔 10 m 高度气温反距离权重插值结果和实测值气温相关性分析

Table 9 Correlation analysis between the results of the 10 m height temperature by inverse distance weighted interpolation results and the measured value of temperature in the 5 towers

气温	函数	$R^2$	$Bias$	$SEE$	$NSEE$
红光山	$y=0.961x-0.0626$	0.9930	0.3628	1.3739	0.0858
鲤鱼山	$y=1.015x-0.6492$	0.9975	0.5307	0.9110	0.0565
米东	$y=0.9235x+0.6075$	0.9895	-0.0356	1.8884	0.1165
燕南立交	$y=1.024x+0.7501$	0.9692	-0.9055	2.6823	0.1670
水塔山	$y=0.996x+0.4697$	0.991	-0.4412	1.4194	0.0885

表 10 铁塔 10 m 气温一元线性相关性分析(全序列)

Table 10 Linear correlation analysis of the 10 m height air temperature (all data)

10 m 气温	函数	$R^2$	$Bias$	$SEE$	$NSEE$
鲤鱼山 & 红光山	$y=0.9961x+0.0297$	0.9980	-4.74E-16	0.9043	0.0557
米东 & 红光山	$y=0.9932x+0.0485$	0.9966	-2.41E-14	1.2975	0.0750
燕南立交 & 红光山	$y=0.9640x+0.2257$	0.9819	-5.06E-15	2.6507	0.1756
水塔山 & 红光山	$y=0.9930x+0.0475$	0.9965	-2.75E-14	1.2070	0.0757

表 11 铁塔 10 m 气温一元线性相关性分析(逐月)

Table 11 Linear correlation analysis of the 10 m height temperature (monthly)

10 m 气温	函数	$R^2$	$Bias$	$SEE$	$NSEE$
鲤鱼山 & 拟合值	$y=0.9964x+0.0272$	0.9982	-3.71E-14	0.8668	0.0534
米东 & 拟合值	$y=0.9935x+0.0465$	0.9968	-4.19E-14	1.2699	0.0734
燕南立交 & 拟合值	$y=0.9684x+0.1982$	0.9841	-3.59E-14	2.4837	0.1643
水塔山 & 拟合值	$y=0.9938x+0.0420$	0.9969	-3.89E-14	1.1360	0.0712

表 12 铁塔 10 m 相对湿度一元线性相关分析(全序列)

Table 12 Linear correlation analysis of the 10 m height relative humidity (all data)

10 m 相对湿度	函数	$R^2$	$Bias$	$SEE$	$NSEE$
鲤鱼山 & 拟合值	$y=1.0577x-0.8474$	0.98	-4.8869	7.4860	0.1127
米东 & 拟合值	$y=0.9942x+1.1122$	0.97	-1.5512	5.9346	0.0927
燕南立交 & 拟合值	$y=0.9612x+1.938$	0.93	0.2620	4.1569	0.0680
水塔山 & 拟合值	$y=1.0256x-0.8293$	0.97	-1.3078	6.1523	0.0956

同样,可以采用逐月数据建立一元线性函数进行相关性分析,各项参数如表 13 所示。鲤鱼山、米东、水塔山、燕南立交相对湿度插值效果较好,经过

逐月建立函数,计算得到的拟合值和实测值之间,相关性参数有所提高,但提高幅度不大。 $R^2$  仅提高 0.01~0.02, $Bias$ 、 $SEE$ 、 $NSEE$ 与全序列的相差不



表 13 铁塔 10 m 相对湿度一元线性相关性分析(逐月)

Table 13 Linear correlation analysis of the 10 m height relative humidity (monthly)

10 m 相对湿度	函数	相关性 $R^2$	Bias	SEE	NSEE
鲤鱼山 & 拟合值	$y=0.9691x+1.7252$	0.98	-1.1365	4.1567	0.0687
米东 & 拟合值	$y=0.9559x+2.5228$	0.98	-6.3525	5.2226	0.0839
燕南立交 & 拟合值	$y=0.9589x+2.337$	0.96	0.0029	4.9071	0.0796
水塔山 & 拟合值	$y=0.9517x+2.7763$	0.98	-1.0776	5.3080	0.0854

大。故采用全序列建立函数进行插值即可。

由于燕南立交与其他铁塔相距较远,且周围地形差异较大。经上述一元线性函数拟合,效果在 5 座塔中较差。因此,需要尝试其他方法对相对湿度进行插值。在此选用多元线性回归模型进行相关性分析。利用函数  $y=a_0+a_1x_1+bx_2+cx_3+dx_4$  建立燕南立交铁塔与其他 4 座铁塔气象要素值的多元一次函数关系,进行分析,并验证效果。其中  $a_0$  到  $a_5$  为多项式系数,  $x_1$  到  $x_4$  为自变量,代表燕南立交外的其他 4 座铁塔的气象要素值,  $y$  为因变量,即燕南立交的气象要素值。

经过拟合,得到相关函数为  $y=9.7045-0.1990x_1-0.1026x_2+0.0993x_3+1.0160x_4$ ,  $R^2=0.8810$ ,  $Bias=-1.87E-13$ ,  $SEE=10.5179$ ,  $NSEE=0.1760$ ,对比表 8 可知,利用多元一次函数,相关性各项指标均有所提到,但提高幅度不大。如逐月构建多元一次线性函数,最终则得到燕南立交 10 m 相对湿度与拟合值的相关函数为  $y=0.8x+11.2876$ ,  $R^2=0.8944$ ,  $Bias=1.99E-14$ ,  $SEE=9.9405$ ,  $NSEE=0.1661$ ,相关性各项参数有所改进,但效果不很明显。

由于受地形影响,风向和风速具有很大的瞬时性和局地性。建立每座铁塔之间 13 m 风向、风速的一元线性相关函数,每座铁塔之间的风向和风速数据相关性较差。同样,我们利用邻近地面观测站 10 m 风速和各个塔 10 m 风速,得到的风向、风速数据效果都较差。这可能与风向、风速受乌鲁木齐复杂地形的影响有关。因此,我们只能对连续缺测达到 12 个时次(2 h, 10 min 为一个时次)以上的风向和风速数据进行剔除。

#### 4 应用效果分析

乌鲁木齐市 5 座铁塔的气象资料经过上述数据资料质量控制,包括逻辑极值检查、僵值检查、时间一致性检查、垂直空间一致性检查及人工干预检查

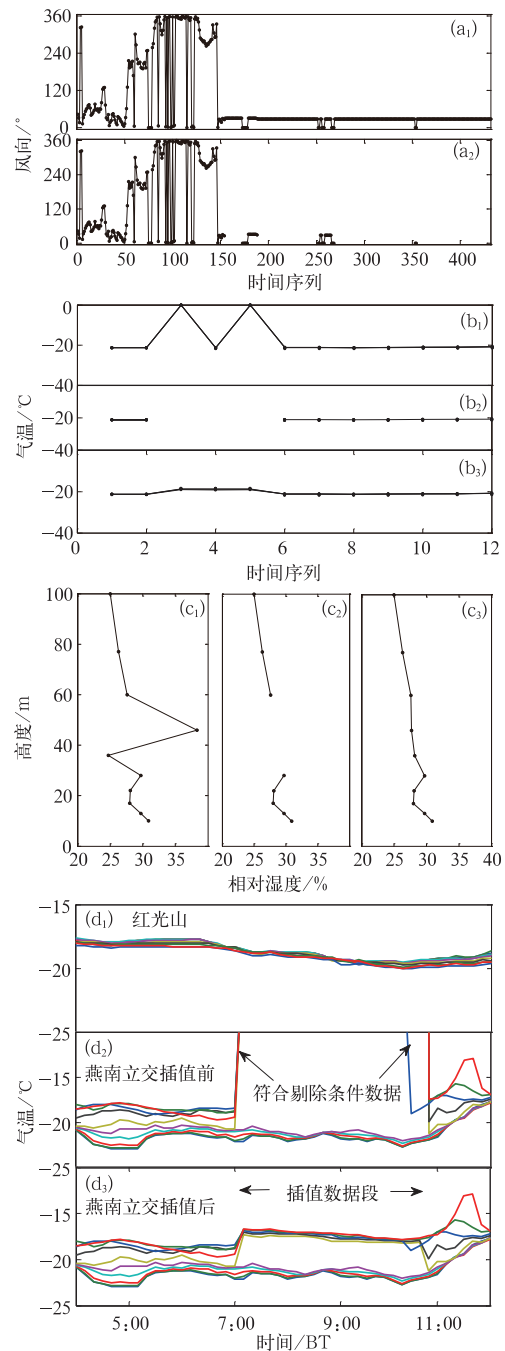


图 7 数据僵值处理、时间一致性处理、空间一致性处理前后效果对比(以燕南立交为例)

Fig. 7 Contrast of static check, time consistency check and similarity check before and after data treatment

之后,将错误数据进行剔除。剔除部分对应时次数据,将在后续的工作中进行插值。质量控制之后的效果,(1) 保证数据都控制在对应的要素极值范围内(100%);(2) 不再有僵值数据;(3) 数据在时间序列上保证是连续的(除缺测时次和剔除时次外);(4) 在垂直空间分布上,也保证了100 m内各层的垂直方向一致性。由于气象要素多样、观测层次多、数据时次多、数据量大,在此仅选取部分数据作为代表,进行分析。

图7为经过上述一系列数据处理过程,得到的燕南立交铁塔气象要素部分结果。图7中 $a_1$ 、 $a_2$ 分别为2013年12月3—4日13 m高度风向数据僵值处理前和处理后;图7中 $b_1$ 、 $b_2$ 、 $b_3$ 分别为2014年2月7日10—11时28~100 m高度的气温数据在时间一致性处理前、删除后和插值后;图7中 $c_1$ 、 $c_2$ 分别为2014年4月24日12:20垂直高度的相对湿度数据在空间一致性处理前、删除后和插值后。数据经过上述处理,将不符合质量控制标准的时次进行清理,在接下来的工作中,将对空缺位置的数据进行插值。

线性回归插值效果的检验,以燕南立交铁塔10层气温为例,选取2014年1月7日数据进行检验,结果如图7 $d_1$ 、7 $d_2$ 、7 $d_3$ 所示。其中图7 $d_1$ 为红光山10层气温数据,图7 $d_2$ 为燕南立交10层气温原始数据,图7 $d_3$ 为燕南立交10层气温线性回归插值后数据。根据图7 $d_3$ 结果,再次经过前述的数据质量控制标准检验,燕南立交插值后的气温数据质量控制流程标准。

## 5 结 论

(1) 根据逻辑极值检查、僵值检查、时间一致性检查、空间一致性检查和人工干预检查等组成的一套方法讨论了乌鲁木齐市从南至北5座100 m气象塔风、温、湿、压梯度观测资料质量情况,根据实际情况确定了判别指标,对2012年4月1日至2014年4月30日的资料进行了质量控制,剔除错误数据并利用四点滑动平均插值、要素垂直分布拟合和线性回归方法组成的插值方法进行插值处理。

(2) 乌鲁木齐市5座100 m铁塔各层观测资料的质量还是比较好的,正常数据占总数据的97.17%,非正常数据占总数据的2.83%。非正常数据中,缺测数据占非正常数据的6.23%,占总数

据的0.18%,主要是米东和燕南立交出现缺测,其余三个铁塔无缺测数据;错值数据占非正常数据的93.77%,占总数据的2.66%。错值数据中,大部分是因虚假、僵值和不一致而导致的,虚假数据占错值数据的89.99%,僵值数据占错值数据的5.47%,不符合一致性数据占错值数据的4.13%。风向、风速出现虚假值的次数最多,其次为气温和湿度;而气温和湿度出现僵值的次数最多,其次为风向;不符合一致性数据主要是燕南立交湿度。

(3) 用四点滑动平均插值方法无法满足插值所需的,如果缺测层数 $\leq 3$ 层,利用本座塔的气象要素对数拟合结合人工干预方法插值,仍不能满足需求的,利用其他铁塔同步完整数据,进行相关性分析的基础上,回归插值本铁塔数据的方法。

## 参考文献

- 陈明,洪钟祥.1993.大气边界层观测资料的质量控制方案研究及其应用.大气科学,17(1):97-105.
- 董德保,张统明,翁宁泉,等.2015.基于模糊逻辑的风廓线雷达功率谱数据质量控制研究.高原气象,34(2):568-574.
- 何志军,封秀燕,何利德,等.2010.气象观测资料的四方空间一致性检验.气象,36(5):118-122.
- 李中华,范绍佳,王宝民,等.2007.2004年10月珠江口西岸海陆风特征观测研究.中山大学学报(自然科学版),190(2):123-125.
- 李平,徐枝芳,范广洲,等.2013.探空温度资料质量控制技术研究.气象,39(12):1526-1634.
- 李岩松,徐枝芳,范广洲,等.2014.船舶海平面气压观测资料质量控制.应用气象学报,25(2):222-231.
- 李志鹏,张玮,黄少平,等.2012.自动气象站数据实时质量控制业务软件设计与实现.气象,38(3):371-376.
- 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.气象,41(11):1408-1416.
- 彭珍.2005.北京325 m气象塔观测资料的统计分析.中国科学院研究生院大气物理研究所硕士学位论文,7-15.
- 任芝花,张志富,孙超,等.2015.全国自动气象站实时观测资料三级质量控制系统设计.气象,41(10):1268-1277.
- 王超,韦志刚,李振朝.2010.敦煌戈壁气象塔站资料的质量控制.干旱气象,28(2):121-127.
- 王春华,吕爱华,余晓丽,等.2010.乌鲁木齐大气污染现状及影响因素分析.新疆农业大学学报,33(4):349-353.
- 王海军,闫养养,向芬,等.2014.逐时气温质量控制中界限值检查算法的设计.高原气象,33(6):1722-1729.
- 王新华,罗四维,刘小宁,等.2006.国家级地面自动站A文件质量控制方法及软件开发.气象,32(3):107-112.
- 王轶,徐枝芳,范广洲.2013.基于EOF 2 m温度质量控制方法研究.高原气象,32(2):564-574.
- 肖艳皎,王斌,陈晓辉,等.2012.移动X波段双线偏振多普勒天气雷达差分相位数据质量控制.高原气象,31(1):223-230.

- 熊安元. 2003. 北欧气象观测资料的质量控制. 气象科技, 31(5):314-320.
- 解以扬, 刘学军. 2003. 天津气象塔风温梯度观测资料的统计特征. 气象, 29(1):12-16.
- 徐安伦, 李建, 刘辉志, 等. 2013. 大理国家气候观象台地面辐射观测数据的质量控制研究. 高原气象, 32(5):1432-1441.
- 薛湛彬, 龚建东, 何财福, 等. 2013. 静止卫星云导风的质量控制及在同化中的应用. 应用气象学报, 24(3):356-364.
- 杨云, 权继梅, 丁蕾, 等. 2015. 国家级太阳辐射测量标准及其质量控制. 应用气象学报, 26(1):95-102.
- 张乐坚, 俞小鼎, 李峰, 等. 2016. 地面降水的多源数据辅助质量控制方法. 气象, 42(3):363-371.
- 张旭斌, 万齐林, 薛纪善, 等. 2015. 风廓线雷达资料质量控制及其同化应用. 气象学报, 73(1):159-176.
- 赵瑞金, 刘黎平, 张进. 2015. 硬件故障导致雷达回波错误数据质量控制方法. 应用气象学报, 26(5):578-589.
- 周笑天, 褚希, 姚志平. 2012. 一种基于 k-means 聚类的实时气温动态质量控制方法. 气象, 38(10):1295-1300.
- 朱立娟. 2015. 面向资料同化风廓线雷达水平风产品质量初步分析. 气象, 41(12):1494-1502.
- Grimmond C S B. 2006. Progress in measuring and observing the urban atmosphere. Theor Appl Climat, 84:3-22.
- Peter J M, Gordon M. 2007. Scaling urban environmental challenges: from local to global and back. London: Earthscan, 1-17.
- Tyler G M, Scott S. 2007. Environmental science: problems, concepts, and solutions. Belmont: Thomson, 5-6.
- WMO. 2004. Guidelines on quality control procedures for data from automatic weather stations. WMO-No. 305.