

# 地面实时气象数据质量控制方法研究进展

韩海涛,李仲龙

(甘肃省气象信息与技术装备保障中心,甘肃 兰州 730020)

**摘要:**实时气象数据的质量控制是确保高质量的气象预报与气候预测必不可少的环节,本文对国内外气象数据质量控制方法进行总结分析,发现传统的质量控制方法仍然是实时气象数据质量控制的主要方法,实时质量控制在自动控制的基础上还需要辅以人工的经验判断才更为有效,还需要加强台站级质量控制、空间质量控制方法、综合判别方法以及质量控制信息的标识问题的研究。

**关键词:**实时数据;质量控制方法;研究进展

**中图分类号:**P413

**文献标识码:**A

## 1 质量控制的必要性

实时气象数据是制作天气预报和气候预测的基础资料,其质量好坏直接影响着天气预报和气候预测的正确与否。我国自1990年代末开始试验在部分台站布设地面自动气象观测系统,随着软硬件技术的成熟,自动站建设速度明显加快,至2011年,全国已经安装了2500多个地面自动气象站。台站气象观测手段的自动化和数据传输速度持续的提高,观测频率的提高,海量数据被采集和传输、保存。因此,快速和高效的证实和标注错误的或者可疑的观测记录的质量控制方法是很必要的,这样才能快速的为用户提供尽可能可靠的观测信息,才能使预报员做出尽可能准确的业务决定。

有关实时气象数据质量控制方法的研究,国内外已经开展了一些工作<sup>[1-8]</sup>,比如北欧国家的4级质量控制流程、美国的3级监控与质量控制、美国国家环境预报中心(NCEP)的质量控制方法研究、我国的3级质量控制方案等。在质量控制工作的研发过程中,有的考虑实时气象数据,有的既考虑实时气象数据也考虑了历史气象数据。目前自动站观测的数据质量还存在不少问题。其原因之一是我国现有的地面气象观测资料的质量控制方法(QC方法)和业务规程主要是针对人工观测站且历史数据而制定的,缺乏从台站、省到国家完整的实时质量控制系

统。许多工作者<sup>[9-15]</sup>已开始对实时数据质量控制方法和技术进行研究和开发。鉴于地面自动站的数据的复杂性、实时性等,目前还没有一个规范的业务化的实时数据质量控制系统,因此需要对实时数据质量控制方法的研究现状、进展及需要开展的工作进行分析。

## 2 国内外研究现状

质量控制是确保高质量的气象观测数据必不可少的重要环节。气象数据实时质量控制对航空活动和防灾减灾的气象保障意义重大。因此,各国都在积极开展质量控制技术的研发工作。

### 2.1 传统方法的应用

传统的QC方法基于资料所反映的大气变量的物理、气候特征,是资料必须遵循的基本规律。因此各国的质量控制流程中占据主要的地位。北欧5个国家<sup>[9]</sup>对实时和非实时气象资料进行质量控制,并对数据质量情况进行了标识。从观测台站到数据中心经历了QC0、QC1、QC2和HQC4个级别的质量控制流程,QC0是台站质量控制,包括极值检查、时变检查、缺测检查、格式检查以及一致性检查;QC1是对实时数据在入库前进行的质量控制,主要是对报文代码进行内部一致性检查,极值检查和完整性检查,方法基本类似于QC0,不同的是QC1利用更多更好的统计值,检查要素也比QC0多;QC2是对库

收稿日期:2011-11-02;改回日期:2012-03-19

基金项目:甘肃省气象局科研项目“基于数据挖掘的气象数据质量控制系统(2008-01)”资助

作者简介:韩海涛(1979-),女,汉,山西晋城人,工程师,理学硕士,主要从事气候变化和气候资料分析应用研究。E-mail:hht19981@163.com

中的非实时数据进行的质量控制,方法更多;QC0、QC1、QC2 均是由计算机程序自动实现的,HQC 是在QC0、QC1、QC2 处理后对资料进行的人工质量控制。北欧各个国家根据用户对象的不同制定了不同的质量控制标识方案,尽可能的把每种方法的质量控制码都显示出来。我国国家气象信息中心设计了台站、省、国家级质量控制方案,国家级质量控制方案以气候界限值检查、极值检查、要素内部一致性检查、时间一致性检查为主,在此基础上对有疑误的数据进行空间一致性检查;对由业务软件确认可疑的数据,通过人机交互修改,进行人工审核。质量控制的结果不是改变观测数据,而是通过质量控制,分别给出每个数据确定的质量控制码。近几年在中国气象局预报与网络司的组织下,由国家气象信息中心联合广西、湖北、江西等省级单位共同建设完成了“省级自动站实时数据质量控制系统”,该系统目前正在各省级单位试运行。

北京市气象局<sup>[10]</sup>设计的自动气象站观测数据质量控制的方案主要是物理极值检查、历史极值检查、内部一致性检查、时间一致性检查、空间一致性检查等方法,并应用于北京地区2007年奥运气象服务演练,取得良好的效果,结果也显示了该系统能实时反映北京市气象局自动站观测数据具有很高的可靠性,而数据缺测和及时性是主要的问题。数据的质量信息标注参考 Shafer<sup>[11]</sup>的标注方法,建立了自己的质量控制标志,具体为:0 要素数据没有审核,1 要素数据可靠,2 要素数据怀疑,3 要素数据不可用,4 要素数据缺测,5 没有此观测要素,6 人工修改要素数据,7~9 保留使用。采用了综合决策算法综合分析各质量控制方法的结果。

王海军等<sup>[12]</sup>采用气候极值范围检查、内部一致性检查、时间一致性检查、空间一致性检查,对湖北省近80个自动气象站15个月的实时气象数据进行了质量控制,省级实时资料质量控制分2个阶段,第一阶段(QC1)为每小时一次,不需人工干预,计算机自动完成,另一阶段(QC2)就是在QC1基础上,质量控制人员利用计算机质量控制后生成的疑误信息进行人工判断,最后决定数据质量。对于数据的质量信息的标注,参考 Shafer<sup>[11]</sup>的标注方法分为10种,对应10类质量控制码(0~9)。考虑到实时资料实际情况,QC1只设置4个质量控制码(0~3),0表示正确,1表示可疑,2表示警告,3表示数据错误,数据质量随着控制码数值的增加而降低。质量

控制码与检查方法有关,即对同一数据,每种方法都有相应控制码,最后综合各种方法,确定每个数据最后的质量控制码。对于单项检验为可疑的数据,如何进一步判别他的错误与正确,采用了综合检查的方法,综合检查是对单项检查结果进行综合判断,确定每个数据最终的质量控制码。如果某要素内部一致性、时间一致性和空间一致性检查标识码之和 $\geq 4$ ,且空间一致性控制码 $\geq 1$ ,则认为该数据错误。

孙娟等<sup>[13]</sup>对上海地区自动站采用单站控制和空间控制方法进行了实时质量控制,单站控制方法包括极值控制、时间连续性和内部一致性;空间控制仅对空间相关程度高的温度和湿度要素进行了质量控制,通过质量控制使得自动站实时观测数据的可靠性有了保障,还为自动站建设、维护等提供经验,并认为合理有效的选取阈值也是质量控制有效性的关键所在,恰当的选取阈值才能准确的反映数据的正确与否,阈值还要动态更新。

## 2.2 新方法的研究

随着气象观测手段的自动化,气象数据的采集、传输等都是自动实现,数据的时间密度大,数据量也相应变大,传统的质量控制主要针对人工观测数据进行的人工质量控制,已经不能满足业务的需求,需要研究新的质量控制方法实现实时气象数据的质量控制。有的研究者<sup>[3]</sup>针对地面观测中不同的错误源(观测和编码错误、人为错误、计算机过程的错误)来研发实时质量控制程序进行时间和空间检查,并且作了订正,而订正数据的标准主要还是靠业务经验。李学坤<sup>[14]</sup>介绍了2种海洋水文气象实时数据的质量控制方法:极值控制法和一致性检验控制法,极值控制方法主要是用“莱因达”准则计算极值的方法,一致性检验控制主要是通过要素之间的关系的方法,同时也指出了一致性检验控制的自动化实现有一定的困难。李良富<sup>[15]</sup>等介绍一个基于黑板模型的地面气象观测数据实时质量控制系统,该系统应用领域本体分析技术,集成多方面专家的经验知识,确定气象观测数据质量控制的知识点,通过异构知识体之间的协作,实时推断气象观测数据的正确性,实现对地面气象观测数据的实时自动错误检测、警示并给出更改建议,并以地面航空气象观测数据实时质量控制为例,提出了用“错误召回率”和“警示正确率”指标对质量控制进行评测的方法。考虑到微气象误差,王海军等<sup>[12]</sup>研制了单站大幅降温事件检测方法,减少了看似“异常”,实际为正确

数据的误检率。美国国家环境预报中心(NCEP)<sup>[3]</sup>发展了基于网络的实时数据监控系统,这个过程分为3个阶段,第一阶段主要集中在数据的数量上,确保每个台站的数据在指定时间到达,并且要业务化,这一阶段在2005年已经实现;第二阶段主要集中在数据的质量上,目前还在研制中;第三阶段主要集中在卫星辐射数据。在气象数据管理、质量控制等都比较成熟先进的北欧国家已经研究了空间质量控制技术,并应用于其质量控制流程中,主要方法有:Madsen-Allerup方法、DECWIM方法、数值预报模式插值方法、Kriging统计插值方法、MESAN方法。刘小宁等<sup>[16]</sup>介绍了一种国外常用的空间回归检验方法,并对其在我国的自动站气象资料质量检验中的适用性进行检验,结果发现该方法能有效地检验出可疑数据,适用于单一要素的检验;对降水、风速等空间变化较大的要素也有较好的检验效果;该方法也与地理环境、台站密度、周边台站分布等有关。针对常用的空间一致性检验方法都是假设被检站与邻近站都处于同一天气系统中,对实时资料分析易造成错误的判断,何志军<sup>[17]</sup>等采用4方位水平空间一致性综合分析法将被检测站的资料与其周围4个不同方位上的邻近站资料分别进行分析,然后再进行综合分析判断,可以避免错误的判断。

### 3 国内外目前采用和可以采用的质量控制方法

目前国内外采用的气象数据质量控制方法可以分成3类。

(1)在自动化装备中与仪器有关的错误的自诊断技术,主要在台站完成。

(2)对单个要素或多个要素的实时质量控制、非实时质量控制。质量控制方法主要有格式检查、缺测检查、极值检查、时间一致性检查、内部一致性检查、空间一致性检查、综合判断检查等。极值检查包括气候界限值检查和台站极值检查,有的合并二者检查,极值检查关键是合理选择极值上下界值,目前采用的方法是从台站历史资料中挑出各月最大值和最小值,再加减 $n$ 倍标准差作为极值的上下界限值, $n$ 值根据地区不同采用不同的值,该方法还需要动态更新台站历史极值表。时间一致性检查是指与要素时间变化规律性是否相符的检查,有5 min时变检查、1 h时变检查、3 h时变检查等方法。内部一致性检查是气象要素之间是否符合一定的规律,

主要有同一时刻不同要素之间的一致性检查和同一时刻相同要素不同项目之间的一致性检查。空间一致性检查是气象要素在空间上的相关性的检查,主要方法有空间插值法、空间回归检验、Madsen-Allerup方法、气候统计比较法等。综合判断检查主要是对单项检验的结果进行综合的检查判断,确定数据的最终质量情况,目前使用的方法主要有综合判别法、综合决策算法、Decision-Making Aglorithm (DMA)等方法。实时质量控制主要是对传统的质量控制方法中的格式检查、缺测检查、极值检查、时间一致性检查、内部一致性检查、空间一致性检查以及综合判断检查等实现了计算机自动完成质量控制,实时质量控制主要在台站和省级数据中心完成。非实时质量控制对非实时资料进行的更为精细化的质量控制,不仅包括实时质量控制的各种方法,而且还采取更多的空间质量控制方法、统计分析和内插方法。非实时质量控制针对的资料也是多时间尺度的,主要在省级和国家级数据中心完成。

(3)用在长时间序列数据方面的均一性检验方法,应用比较广泛的有:标准正态均一化检验<sup>[18-19]</sup>;二相回归方法、序列均一性的多元分析、多元线性回归和历史元数据检查<sup>[20]</sup>,如:气象站的检查报告。这些方法主要在省级和国家级数据中心完成。

### 4 我国地面实时质量控制存在的问题及解决方法

#### 4.1 实时数据质量控制需加强台站级质量控制

由实时数据的定义可知,从观测站自动采集数据并传输到资料收集中心的资料称为实时资料。而且这些数据还分为分钟数据、小时数据和日数据等不同时间尺度的数据,而且小时数据不仅有自动观测的数据而且还有人工采集的数据,还有报文数据。这些不同时间尺度的数据,不同类型的数据都应该进行实时自动质量控制,才能发挥实时数据的效益。WMO<sup>[21]</sup>认为,自动站资料实时质量控制应在2个层面进行,一是在观测台站进行基本质量控制;二是在数据处理中心进行,数据处理中心除采用综合时间一致性、内部一致性检查外,还应对仪器传感器失效、长期漂移进行评估。目前大部分质量控制研究和系统都是在省级或国家级的基础上进行的,台站级的实时质量控制研究工作还很薄弱,因为实时资料是自动采集的数据,如果在台站观测中数据发生错误或有疑义,在台站没有及时发现,在省级或国家

级质量控制中发现时,该资料很难再更改。因此,实时气象数据质量控制主要还是在台站和省级完成,尤其是在台站。

#### 4.2 重视传统质量控制方法的同时不断研究开发新的质量控制方法

自动化观测资料与传统资料的 QC 方法的主要差别在于:传统的 QC 方法主要是面对人工观测的数据,数据的时间跨度大,方法侧重于单个数据的检查。而自动化观测系统更多的误差是连续性的漂移,而不是孤立的误差,因此,对资料的稳定和连续性检查比单个数据的检查更重要。自动化观测仪器产生了更多的资料量,具有高时间分辨率,使得自动观测资料比人工观测资料有更多自动的 QC 方法,响应速度要快,质量控制过程尽可能全自动完成。基于实时数据的实时性很强,需要开发新的自动质量控制算法。国内外研究现状表明传统的质量控制方法仍是进行实时气象数据质量控制的主要工具,目前实时气象数据质量控制主要还是应用传统的质量控制方法,实现了自动化的质量控制,基本的质量控制流程如图 1。实时资料的质量控制是以自动控制为主,但是自动控制不可能解决所有的数据的质量问题,对一些特殊的情况,还是需要技术人员的经验进行判断,因此实时资料的质量控制采取自动控制为主人工控制为辅的人机交互模式更为有效。

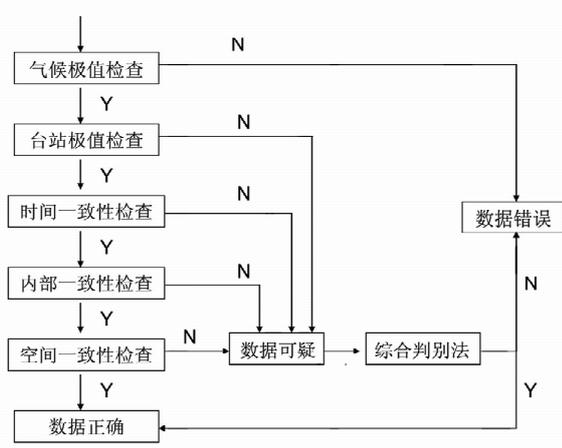


图 1 地面实时气象数据质量控制流程简图

Fig. 1 The flow diagram of quality control of real time surface meteorological data

空间质量控制相对于单站质量控制来说,具有更重要的作用,也具有更多的数学技巧和难度。我国各级气象部门也在积极研制和开发实时气象数据质量控制方法及系统,但相比而言在空间质量控制

方法方面应用很少,还有待于进一步的研究和应用。空间插值方法不仅用来弥补气象站点不足,通过空间插值得到需要的值,而且作为空间质量控制的新的研究方向。该方法的主要原理就是通过比较观测值与期望值进行空间质量控制的一种方法,期望值是利用临近台站资料经统计插值计算得到的,当站点要素的空间插值与观测值相比超过一定的阈值时,就认为该观测值有疑问。空间插值方法也比较多<sup>[8,22-23]</sup>,还需要研究选择合适的空间插值方法。

#### 4.3 各种检验方法应用后的综合判别需要进一步研究

从图 1 也可以看出,综合判别法决定数据的最终质量水平,因此对于单项检验为可疑的数据,如何进一步判别它的错误与正确,是气象资料质量控制中一个重大的技术问题,而且也关系到质量控制码的标识问题。王伯明<sup>[24]</sup>参考国外研究的思路,研制了一种综合判别法,是以一个观测时次数据或者一日数据为单位,将进行各种单项质量检验的结果(质量控制数据)建立一个质量控制数据表,对表数据采用高可疑数据逐步确认淘汰再分析的方法,完成对各种单项质量检验的综合判别。同时,也指出了该方法的缺点是不能对正确的极端值进行判断。目前国内大部分实时质量控制采用该综合判别法,也有采用综合决策算法的,国外文献中综合判别法有 the decision-making algorithm(DMA)方法。因此可以研究比较,多种决策算法在综合判别中的应用,确定一种决策算法,弥补综合判别法的缺点。

#### 4.4 质量控制码的标识

质量控制码的标识主要是为资料的使用者提供有关资料的质量信息,同时在对数据进行质量控制过程中也应该详细标识数据的质量情况,并最终给用户以简单明了的质量标识。一般而言,质量控制码应该包括反映数据的质量水平、质量控制方法、可疑或者错误的类型和是否订正的信息。如何把数据的这些质量控制信息进行标识,如何在数据文件中附加质量控制码而又能让数据使用者一目了然,仍是气象资料质量控制中亟待解决的一个重要问题。

目前各国对质量控制码的标识也各不相同,有的很复杂有的只是进行简单的标识。国内研究者对数据的质量信息标注已开始重视,有的研究者参考了国外研究者的质量标注码,建立了自己的质量标注码,但是还没有统一质量标注码。我国在质量控制码的标识问题上,需要进一步的搜集和参考先进

的标识方法来规范质量控制,发展一套便于理解的标注系统,能显示数据的质量水平;使得用户很容易确定可疑或者错误的的数据,突出正确的数据,最终能形成一个标准的质量控制码。

#### 参考文献:

- [1] Vejen F, Jacobsson C, Fredriksson U, et al. Quality Control of Meteorological Observations, Automatic Methods Used in the Nordic Countries R[R]. ISSN 0805-9918, KLIMA Report, 2002. 8
- [2] Krishna Kumar V, Bradley Ballish, Jeffrey Stoult. Real time data monitoring at NCEP[EB/OL]. [http://ams.confex.com/ams/Annual2006/techprogram/paper\\_103227.htm](http://ams.confex.com/ams/Annual2006/techprogram/paper_103227.htm)
- [3] Wolfson N, Erze J, Alpers Z. Automatic Real-time Quality Control of Surface Synoptic Observations[J]. Journal of applied meteorology, 1978(17):449-457.
- [4] 刘小宁,任芝花.地面气象资料质量控制方法研究概述[J]. 2005,33(3):199-203.
- [5] Evans D, Conrad C L, Paul F M. Handbook of Automated Data Quality Control Checks and Procedures of the National Data Buoy Center R[R]. NDBC Technical Document 03-02, 2003.
- [6] Danrel Y, et al. New Techniques in Quality Assurance of Hourly Meteorological Data: Resolving Multiple Flags Through a Decision Tree[A]. Northeast Regional Climate Center, Cornell University, 2004 84th AMS Annual Meeting.
- [7] David B Gihousen. Improved real-time quality control of NDBC measurements.
- [8] Eischeid J K, Baker C B, Karl T R. The quality control of long term climatological data using objectived data analysis[J]. Journal of applied meteorology, 1995,34:2787-2795.
- [9] 熊安元.北欧气象观测资料的质量控制[J]. 气象科技,2003,31(5):214-320.
- [10] 窦以文,屈玉贵,陶士伟,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.
- [11] Mark A Shafer, Christopher A Fiebrich, Derek S Arent, et al. Quality Assurance Procedures in the Oklahoma Mseonetwork[J]. J Atmo Oceanic Techenol, 2000:474-494.
- [12] 王海军,杨志彪,杨代才,等.自动站实时资料自动质量控制方法及其应用[J]. 气象,2007,33(10):102-109.
- [13] 孙娟,胡平.上海自动气象站实时质量控制系统的研究[J]. 气象水文海洋仪器,2009(4):38-41.
- [14] 李学坤,李凤金.海洋水文气象实时数据质量控制[J]. 海洋预报.1997,14(3):71-79.
- [15] 李良富,王汉杰,刘金玉,等.基于黑板模型的地面气象数据质量控制[J]. 气象科技,2006,34(2):199-204.
- [16] 刘小宁,鞠晓慧,范绍华.空间回归检验方法在气象资料质量检验中的应用[J]. 应用气象学报,2006,17(1):37-43.
- [17] 何志军,封秀燕,何利德,等.气象观测资料的四方位空间一致性检验[J]. 气象,2010,36(5):118-122.
- [18] 张志富.国外气候资料整编方法介绍[J]. 干旱气象,2009,27(4):395-401.
- [19] 韩海涛,胡文超,刘积林,等.兰州站气候资料序列均一性的初步分析[J]. 干旱气象,2008,26(4):34-38.
- [20] 胡文超,李晓苹,韩海涛.气象元数据应用研究进展[J]. 干旱气象,2008,26(4):12-15.
- [21] Igor Zahumensky. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations[A]. Expert team on surface technology and measurement techniques first Session Geneva, Switzerland, 13-16 October 2004.
- [22] 徐建华.现代地理学中的数学方法[M].北京:高等教育出版社,2002.
- [23] 岳文泽,徐建华,徐丽华.基于地统计方法的气候要素空间插值研究[J]. 高原气象,2005,24(6):974-980.
- [24] 王伯民.基本气象资料质量控制综合判别法的研究[J]. 应用气象学报,2004,12(15):50-59.

## Review of Quality Control Methods of Surface Real Time Meteorological Data

HAN Haitao, LI Zhonglong

(Meteorological Information and Technical Equip Safeguard Center of Gansu Province, Lanzhou 730020, China)

**Abstract:** Quality control of real time meteorological data is an indispensable link to ensure high quality of weather forecast and climate prediction. This paper introduced the domestic and foreign meteorological data quality control methods, and found that the traditional method is still main important tool of real time meteorological data quality control. Real time quality control system can be more effective to combine automatic control with experience judgment. In the future, it still need to strengthen research of station-level quality control, statistical test methods and space quality control methods, as well as comprehensive criterion and information identification of quality control.

**Key words:** real time data; quality control methods; research situation