

地面气象资料质量控制方法研究概述

刘小宁 任芝花

(国家气象局气象信息中心,北京 100081)

摘要 为了促进我国地面气象资料质量控制方法的研究,以几个有代表性的数据集,如 GHCN、GDCN、ISH 为例,概述了国外地面气象资料质量控制技术的进展及新动态。研究表明:传统的质量控制方法仍是主要工具。在计算机技术迅速发展的时代,可以有更好的条件,使用自动控制和人机交互、气候背景资料和统计检验相结合的技术,并适当应用空间检验方法,来设计我国自动站地面气象资料质量控制业务流程。

关键词 地面资料 质量控制 控制技术

引言

地面气象观测记录是人类有器测以来最长的气象资料。如地面气温是目前积累的最好的仪器观测的气候资料之一,也是气候状况的关键参数^[1]。随着气候变化研究的深入,科学家更加迫切需要高质量、长序列的资料。然而,资料质量受到观测仪器、观测技术、测站位置、观测时间、观测方法的影响,尤其是长期观测形成的地面气象资料,各种非气候因素造成的影响更大,使资料质量大打折扣^[2]。

气象观测资料质量控制的重要性已经为所有使用气象资料的科学家所公认。地面气象观测记录必须具有代表性、准确性、比较性^[3]。

各种观测值在统计意义上都可以看为是随机变量。观测值采集时通常会对真值产生误差。根据误差的性质和产生的原因,气象观测资料可能存在 3 类性质完全不同的误差:随机误差、系统性误差、过失误差(偶然误差)^[4]。

质量控制的目的是确保提供应用的资料符合各种要求(包括不确定性、分辨率、连续性、均一性、代表性、时限、格式等)。好的资料无需特别出色,重要的是其质量应该是已知的和可证实的。

“全球气候观测系统资料与信息计划”特别提出,为 GCOS (Global Climate Observing System)

所建立的数据库系统中的数据“必须经过严格的质量控制”,使“数据质量符合标准,以保证使用此类数据的研究人员能够接受”^[5]。

因此,随着气候研究的深入,各国都进一步加强对气象资料的质量控制研究。近年来资料质量控制技术有了新的进展。我国在这方面的研究还有待于加强。本文试图以几个有代表性的数据集为例,概述近年国外在地面气象资料质量控制中的进展及最新动态,并对我国该技术的发展进行探讨。

1 国外地面气象资料质量控制技术的进展

传统的质量控制(QC)主要根据气象学、天气学、气候学原理,以气象要素的时间、空间变化规律和各要素间相互联系的规律为线索,分析气象资料是否合理。其方法包括:范围检查、极值检查、内部一致性检查、空间一致性检查、气象学公式检查、统计学检查、均一性检查。这些方法被普遍应用到地面气象资料的质量控制中。

近年来国外地面气象资料质量控制技术有了很大发展,如美国对全球月、日、小时地面资料均进行了较完善的质量控制,制作了有代表性的数据集。本文以几个有代表性的数据集为例,概述近年来 QC 技术的进展。

科技部基础性工作专项基金项目“2002DEA30041”、“气象资料共享系统建设”资助

作者简介:刘小宁,女,1947年生,正研级高工,从事气象资料处理及分析,Email:liuxn@cma.gov.cn

收稿日期:2004年3月1日;定稿日期:2004年5月23日

1.1 全球历史气候网资料

美国国家气候资料中心(NCDC)所制作的全球历史气候网 GHCN (Global Historical Climatology Network) 是很有影响的台站每月观测的数据集。目前版本 2 的气温数据集包括 7300 个站的月平均气温和 5100 个站的月平均最高/最低气温,来自于 30 多个源数据集。GHCN (版本 2)改进的不仅仅是数据量的增加,还使用了特殊均一性检测订正技术 (Peterson 和 Easterling, 1994; Easterling 和 Peterson, 1995; Easterling 等, 1996) 和一套新的 QC 方法^[6,7]。

QC 过程包括对源数据集的检查、台站时间序列检查、单个数据点的检查 3 个阶段。

在第 1 阶段中对源数据的检查包括明显的数据问题(如文件被切断、格式化错误、读不出来的记录等)。

在第 2 阶段,台站时间序列检查时,检测平均值变化使用了累积总和检测法(CUSUM, van Dobben de Bruyn, 1968),通过找出时间序列平均数的变化来检查。方差变化检查使用了尺度累积量(SCUSUM),该量是 CUSUM 的一个变量,用来查找台站时间序列的方差变化。

在第 3 阶段单个数据点的检查中,辨别时间奇异值使用了双权重标准差用作计算方差,这种方法既能很好地抵抗奇异值,又能保持整体的功效(Lanzante, 1996)。

辨别时间奇异值使用 3 个极限值。使用空间检验来验证奇异值,空间检验有很多不同的方法。Eischeid 等(1995),用了 6 种不同的方法来预测或估计某个站的值。这些方法是:正常比率法;简单反距离加权法;最佳内插法;使用最小绝对偏差标准的多元回归法;单个最佳估计算子;前 5 种方法的综合。根据分析,空间检验中邻近站与被检验站的数据要逐个检查,因此通常要借助于人工进一步判断。

1.2 全球日气候资料数据集

美国 NCDC 制作的全球日气候资料数据集 GDCN (Global Daily Climatology Network V1.0) 是最近开发出的数据集,其包括最高/最低气温、24 h 降水量,年代为 1840~2001 年^[8]。

对最高/最低气温资料进行了极端值检查、内部一致性检查。该数据集的检查也使用了双权重标准

差的统计检验方法,并使用了月平均值与格点资料比较,进行了空间检验。

对于日降水资料进行了极端值检查。所用的极端值为目前观测到的全球曾出现过的日极端记录:极端最高气温:57.8;极端最低气温:-89.4;最大 24 h 降水量:1828.8 mm。

1.3 全球地面小时数据集

美国 NCDC 所制作的全球地面小时数据集 ISH (Intergated Surface Hourly Database),共有全球 20000 个台站,早从 1900 年开始,台站类型有自动站(ASOS)、自动天气站(AWOS)、天气站等^[9]。QC 分为两个步骤,第 1 步进行了各种源数据集的集成、整理、转换、统一格式、统一时间(格林威治时间)。第 2 步去除随机及系统误差。通过一系列 QC 的算法,进行自动和人工的质量控制。共使用 54 种 QC 方法,包括:变量检查、极值检查、内部一致性检查。

但目前该数据集检查中未使用空间检验方法,NCDC 计划今后进一步发展。

2 国外地面气象资料质量控制方法发展新动态

2.1 自动观测资料质量控制对传统方法的挑战

随着观测自动化技术的发展,产生了大量的自动观测资料。自动化资料与传统资料的 QC 方法主要差别在于:

传统的 QC 主要面向人工观测的数据,数据的时间跨度大,方法侧重于对单个数据的检查。电子传感器及自动传输的发展,对传统 QC 方法是一种挑战。这是因为电子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,对资料的连续性检查比单个数据的检查更重要。

电子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的自动 QC 方法。

2.2 新质量控制技术的使用

(1) 气候资料质量控制的空间概率方法 (Probabilistic-spatial approach to quality control of climate observations)^[10]。

该方法是美国俄勒冈州立大学空间气候分析所在 1990 年建立的降水图基础上,新发展的气候图的一种新“空间质量控制”工具。新的 QC 系统针对自动观测的变化及需要而产生,被称为 PRISM (Pa-

parameter-elevation Regressions on Independent Slopes Model, 基于独立倾斜模式的参数-高度回归)。它使用了气候图和气候统计技术对每个观测值提供连续、定量的概率值。其 QC 过程由两个嵌套的循环组成。在“日循环”中, PRISM 对每个站逐日资料进行计算, 首先将被检查站去掉, 使用所有邻近站观测资料, 利用 PRISM 回归函数, 反复计算, 得出被检查站的“预报值”、偏差、回归标准差。

再计算每日的“长期平均”气候值, 计算每年逐日的平均值、标准差。

将每日观测值、“预报值”、长期平均的气候统计值用 T 检验法比较, 计算其概率值。

因此, 该方法以气候资料为背景、以统计检验为依据、加空间检验, 是一种综合检验方法。

(2) 自动控制技术与人机交互可视化技术的结合。

该技术在加拿大太阳辐射资料质量自动控制系统中应用^[11]: 该系统对 RCS(气候参考站)的每小时资料和 CORE(有人观测的大气成分和辐射研究站)的每周资料进行质量控制, 系统由界面层、核心层、存储层 3 个部分组成。

界面层由绘图用户界面组成, 用户使用人工 QC 图形交互或 Web 浏览; 核心层由运算法则、Java 代码、分析设计模式组成; 存储层由 Oracle 数据库构成, 包括台站和仪器详细信息、QC 标识、错误信息等。

在自动控制中的运算法则使用了传统 QC 方法。

由于自动控制的运算法则不能替代有经验的技术人员的判断, 对一些数据是正确还是错误, 在界面层通过 Gui 和 Web 平台, 可进行一周资料和一日的编辑浏览, 用交互处理和图表进一步人工判断。

总之, 该系统融合了自动控制技术和交互式应用技术, 提高了自动进程及近实时的能力, 也允许在必要时对特殊资料进行详细的人工判断。

2.3 美国地面自动观测系统

美国地面自动观测系统(ASOS)^[12]已经有超过 1100 个观测站点, 能实现数据的自动采集、质量控制、运行监控和数据传输和调用。在 ASOS 的运行中, 对资料质量进行了 3 级监视和控制。

第 1 级 QC 在观测站点内进行, 在实时资料发送前完成。在一次观测完成后, ASOS 根据 QC 算

法自动对传感器探测的原始数据进行质量控制, 防止错误数据编入报文。如果发现问题, 则将有问题的数据从报文中去掉, 在报文中加入相应的标志。

第 2 级 QC 是在州级, 对通过通信网收集的自动站资料进行常规监测。当发现问题时, 通知 A-SOS 维护人员进行检查处理, 并对资料进行更正或停止使用有问题的资料。

第 3 级 QC 在国家级完成。将带标志的资料、缺发的报文进行检查和处理。

通过这样 3 级质量控制, 能及时发现观测数据的质量问题, 保证实时资料的及时准确, 并监测自动站运行。

2.4 强调质量评估专家的作用

俄克拉何马州气候调查(Oklahoma Climatological Survey)是美国的一个州级气象机构, 特别强调质量评估专家的作用。软件对由自动观测产生的每天超过 1 百万数据量的观测资料进行质量控制, 但是也有一些特殊现象被判断错误, 因此使用了人工技术补充自动 QC 的不足。

对于降水资料, 评估专家通过和雷达资料的对比进行判断。同时, 对月资料用被检查站与周围邻近站比较, 使用了双质量分析方法(double mass analysis), 以此判断其变化及原因(是否有仪器的漂移)。每月分析月统计资料来判断是否有错误资料。计算各要素的平均值和距平, 通过图来分析, 并与相邻站比较。找出仪器有问题的站, 并人工重做出标识^[13]。

3 关于我国地面气象资料质量控制方法发展的探讨

3.1 地面自动站资料的质量控制方法

2003 年后我国近 700 个基本(准)站全部安装了自动气象站。大部分站在进行人工和自动站双轨运行。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 QC 方法和业务规程主要是针对人工观测站而制定的。缺乏从台站、省到国家级完整的质量控制系统, 没有针对自动站新资料的 QC 方法和流程。

建立流程的关键是制定 3 级质量控制方案: 自动站实时采集数据的质量控制方法、省级准实时资料的质量控制方法、国家级质量控制方法。下面是自动站资料质量控制方案流程设计时要注意的几个问题。

(1) 充分认识自动观测与人工观测资料在质量控制中的区别。

自动气象站在观测手段、资料密度、数据格式、信息量上,都与人工观测站有很大的差别。在设计质量控制流程中要充分考虑其差别。我国地面自动站资料具有以下特点:

复杂性 地面自动站是一个实时采集数据的自动观测系统,具有实时采集数据、自动形成观测报告、自动生成和传输数据文件的功能。在其形成的文件中也有部分人工观测数据。同时,除地面资料外,有些站还有辐射观测,要素项目多。由于自动站观测产生错误的来源有传感器、采集器、传输、人工操作等多种途径,不易发现,因此,对台站观测资料的质量控制是比较复杂的。

实时性 自动站数据是每分、每小时、每24小时等不同时间尺度的数据。在每小时除有自动观测外,有些时次还有人工采集的数据,要不间断地对资料质量进行监控,实时性很强。

难于更改性 由于是自动采集数据,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,在省或国家级质量控制中发现时,该资料很难再更改。

(2) 强化台站级质量控制。

由于自动站资料的特点,观测站级的质量控制就显得更加重要。应当充分发挥自动采集数据、计算机自动检查资料的自动化优势,使观测站传输出的资料绝大部分得到很严格的控制。

台站的实时质量控制是自动站资料质量控制系统的核心。因此,要在质量控制方案的设计中将所有能在台站进行的质量控制全部做好。对于采集的每一个实时数据,都要及时进行检查,有可疑数据,显示提示信息,以人工判断。在台站上报的数据文件中要有质量控制的标识。

(3) 要有不同层次的质量控制。

观测站、省、国家级对资料质量控制应该有不同层次。在观测站强调对逐时、逐日资料,使用传统的QC方法;在省及国家级对准实时的月、日资料更多地使用统计判断方法。

(4) 注意自动站仪器的校准漂移等特殊问题。

根据国外经验,自动站仪器的校准漂移会使资料产生问题,在质量控制流程设计中要予以考虑。仪器使用一段时间后要对其资料进行均一性检查,以

判断观测仪器漂移对资料的可能影响。

(5) 重视质量评估专家的作用。

国外成功的质量控制方法表明,完全自动化的质量控制不可能解决所有的数据质量问题,对于一些特殊情况或问题的判断还要辅以人工检查。必须采取计算机自动控制和人工相结合的方式。在各级,尤其是省、国家级要重视质量评估专家的作用。

通过分析国外地面气象资料质量控制方法发展动态,结合我国自动站资料的特点,在使用QC方法上,我们可以明确以下几点:

(1) 传统的QC方法仍是主要的工具。国外经验表明,即使针对自动观测资料,其质量控制仍大量使用了传统的QC方法。因为,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^[14]。

(2) 自动控制和人机交互技术的结合。以自动控制为主,结合数据库和人机交互技术形成方便调用的数据及良好用户界面,辅以人工判别,是非常重要的。

(3) 气候背景资料和统计检验相结合。计算技术及计算机能力的提高,使统计检验有了实现的可能。因此,使用大量的气候背景资料,用统计检验的方法,对于判别资料的奇异点和仪器漂移,是一种好的方法^[15]。

(4) 空间检验方法的适当应用。若有条件,可应用空间检验方法。但由于资料计算量大,使用该方法难度较大。不同时间尺度的气候变量的空间变化特征不同,对小时资料,其空间变化太大。美国只在月、日资料(GHCN、GDCN)的检验中使用,在小时资料(ISH)中的质量控制还未使用该方法。

(5) 各种检验方法应用后的综合判别需要进一步研究。

3.2 月资料的质量控制方法

我国在业务中还未系统地对月资料进行质量控制。因此,在设计月资料的质量控制时,可借鉴国外经验,如GHCN资料的质量控制。除进行资料的范围、气候极端值检查外,可更多地使用统计方法,如双权重方差判别方法。过去由于计算条件的限制,我国还未使用空间检验方法。在计算条件极大改善的今天,可进一步利用数据库技术挑选邻近站,使用空间检验的方法对每个站进行统计判别,以检查每个站月资料。

4 小结

近年来国外在地面气象资料质量控制技术方面有了显著进展,但是在质量控制方法的运用时,传统的QC方法仍是主要工具。在计算机技术迅速发展的时代,我们可以有更好的条件,使用自动控制和人机交互、气候背景资料和统计检验相结合的技术,并适当应用空间检验方法,来设计我国地面自动站资料质量控制业务流程。

参考文献

- 叶笃正,曾庆存,郭裕福(主编). 当代气候研究. 北京:气象出版社,1991. 211
- 李庆祥,刘小宁,张洪政,等. 定点观测气候序列均一性研究. 气象科技,2003,31(1):2-11
- 中国气象局. 地面气象观测规范. 北京:气象出版社,2003
- 么枕生,丁裕国(编著). 气候统计. 北京:气象出版社,1990. 776
- 全球气候观测系统中国委员会办公室. 全球气候观测系统中国委员会成立大会暨委员会专家组第一次全体会议文集. 北京:气象出版社,1997. 67
- Peterson T C, Vose R, Schmoyer R, et al. Quality control of monthly temperature data: The GHCN experience. Asheville: National Climatic Data Center/NOAA, 1994
- National Climatic Data Center. Global Historical Climatology Network (GHCN) Quality Control of Monthly Temperature Data. <http://www.ncdc.noaa.gov>
- Geason B E. For Data Set 9101 Global Daily Climatology Network V1.0. Asheville: National Climatic Data Center Data Documentation, 2002
- Lott N, Baldwin R, Jones P. The FCC integrated surface hourly database: a new resource of global climate data. National Climatic Data Center Technical Report No. 2001(01). Asheville: National Climatic Data Center, 2001
- Daly C, Gibson W, Doggett M, et al. A probabilistic spatial approach to the quality control of climate observations. AMS Annual Meeting, Seattle, 2004
- Sopoco H T, Boudreau L D, Armstrong M. A new automated and interactive system for solar radiation data processing and quality control. AMS Annual Meeting, Seattle, 2004
- 周林,李湘. 关于美国自动地面观测系统(ASOS)的考察报告. 气象科技合作动态, 2002, (4): 18
- Martinez J E, Fiebrich C A, Shafer M A. The value of a quality assurance meteorologist. AMS Annual Meeting, Seattle, 2004
- Abbott P F. Guidelines on the Quality Control of Surface Climatological Data. WCP-85, 1986
- Lanzante J R. Resistant, robust and nonparametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. Int. J. Climatol., 1996, (16): 1197-1226

Progress in Quality Control of Surface Meteorological Data

Liu Xiaoning Ren Zhihua

(National Meteorological Information Centre, Beijing 100081)

Abstract: For facilitating the researches in quality control (QC) of surface meteorological data in China, the QC methods of several representative data sets (e. g. GHCN, GDCN and ISH) are introduced. The latest international development and trends of the correlative technologies are expounded. It is indicated that the conventional QC methods are still the basic tools. Along with the development of computer technology, several new technologies, such as auto-control, HCI (Human Computer Interaction), the spatial test method and the technology of combining climate data and statistical test, etc., will be applied to the QC of surface meteorological data to help design the operation flow of quality control of surface meteorological data from autonomous weather stations.

Key words: surface meteorological data, quality control, QC technology