

熊安元 等 编著



气象出版社
China Meteorological Press

中国地面和高空气候变化 数据产品研发技术

熊安元 等 编著



 气象出版社
China Meteorological Press

内容简介

本书归纳总结了近几年来作者在研发气候变化数据产品时涉及的数据分析处理技术,包括我国地面和探空资料的质量控制技术,序列均一化技术和多源降水融合技术。在分析国内外相关领域研究进展的基础上,较详细地介绍了如何应用这些数据处理技术获得相关数据产品,评估了数据产品的基本质量和存在的问题,并与国外相关产品进行了比较分析。

读者可通过阅读本书学习气象资料质量控制、气候序列均一化和多源降水融合等技术及其研究进展,系统了解国家气象信息中心在该领域的研究现状以及相关的数据产品。

图书在版编目(CIP)数据

中国地面和高空气候变化数据产品研发技术/熊安元等编著. —北京:气象出版社, 2015. 4

ISBN 978-7-5029-6124-4

I. ①中… II. ①熊… III. ①地面-气候变化-数据处理-技术开发-中国
②高空-气候变化-数据处理-技术开发-中国 IV. ①P467

中国版本图书馆 CIP 数据核字(2015)第 073570 号

出版发行:气象出版社

地 址:北京市海淀区中关村南大街 46 号

总 编 室:010-68407112

网 址: <http://www.qxcbs.com>

责任编辑:白凌燕 陈 红

封面设计:易普锐创意

印 刷:中国电影出版社印刷厂

开 本:787 mm×1092 mm 1/16

字 数:262 千字

版 次:2015 年 6 月第 1 版

定 价:55.00 元

邮政编码:100081

发 行 部:010-68409198

E-mail: qxcbs@cma.gov.cn

终 审:黄润恒

责任技编:吴庭芳

印 张:10.25

印 次:2015 年 6 月第 1 次印刷

本书如存在文字不清、漏印以及缺页、倒页、脱页等,请与本社发行部联系调换。

中国地面和高空气候变化数据产品研发技术

主 编:熊安元

副 主 编:沈 艳 李庆祥 任芝花 陈 哲 阮 新

编著人员:唐国利 廖 捷 曹丽娟 杨 溯 朱亚妮

宇婧婧 徐文慧 余 予 远 芳 郭艳君

前言

IPCC 第五次评估报告指出,观测资料的不确定性是近百年全球气候变化不确定性的主要原因之一,观测资料的不确定性主要表现为观测资料中的误差和数据产品研制技术的不确定性。全球气候变化观测事实的分析主要基于全球气候观测资料及其数据分析产品。例如,IPCC 第四次和第五次评估报告对全球近百年气温变化的评估结果均是基于 GHCN、CRUTEM、Berkeley 和 GISS 四套全球百年地面气温数据产品得到的,且四套数据产品获得的气温变化趋势各有差异。显然,数据产品的质量会造成全球温度变化分析结果的不确定性,这些数据产品的质量依赖于观测资料的误差、观测资料的处理方法和分析技术。

观测资料分析和模式模拟结果均证实了近百年全球变暖的事实,但变暖的幅度和变暖的区域差异仍然存在较大的不确定性。与全球变暖相一致,我国近百年年平均温度在增加,现有的多数研究结果表明我国地面气温近百年增加幅度超过全球平均。我国对高空气候变化的研究结果不多,有限的一些研究结果之间也存在较大的差异,主要原因是高空观测资料序列在时间上的非均一性问题对研究结果的影响。我国是全球气象观测网络最密集的区域之一,对气候变化关键变量的实地观测有上百年的历史,为我国气候变化研究提供了很好的基础。我国现有观测资料的主要问题是资料质量问题,表现为观测资料的不完整以及资料中的系统误差和粗大误差。例如,由于台站迁移、仪器变更、数据处理方法变化等原因导致的不同时间段资料间的系统差异(即非均一性)。通过对我国观测资料的科学加工处理,获得可直接用于我国区域气候变化研究的基础数据产品一直是我国从事地球科学和全球气候变化研究人员的迫切希望。为此,国家重点基础研究发展计划(973 计划)项目设立了“全球变化数据的评估、同化、融合与应用”(2010CB951600)项目,旨在获得全球变化分析研究所用的数据产品。本书是该项目第二课题“构建地面和高空常规观测资料气候变化数据库”(2010CB951602)的主要研究成果。

本书共分 5 章,第 1 章描述中国近 60 年地面气温和降水观测资料台站观测数据的处理和质量控制以及数据产品,由任芝花主笔;第 2 章是关于中国近 60 年高空探测资料的加工处理和质量控制技术以及数据产品的研究成果,由阮新和廖捷主笔;第 3 章是关于对中国近 60 年地面气温和降水观测资料的均一化检测订正技术和数据产品,由李庆祥和唐国利主笔;第 4 章是关于对中国 1979 年以来高空

温度和湿度观测资料的均一化检测订正技术和数据产品,由陈哲主笔;第5章是关于中国降水分析产品和多源数据融合产品,由熊安元和沈艳主笔。全书由熊安元统稿和技术审核。项目第二课题“构建地面和高空常规观测资料气候变化数据库”(2010CB951602)的全体研究参与人员均对本书的编著做出了贡献。

本书编著过程中,得到了编著者所在单位国家气象信息中心的大力支持。本书的编著和出版主要由国家重点基础研究发展计划(973计划)项目“全球变化数据的评估、同化、融合与应用”(2010CB951600)资助。

熊安元

2014年12月于北京

目 录

前言

第 1 章 中国地面气温和降水观测资料的质量控制	(1)
1.1 地面基础资料重建	(1)
1.2 地面月资料的质量控制	(10)
第 2 章 中国无线电探空资料的质量控制	(17)
2.1 中国高空历史观测数据的重建	(18)
2.2 中国探空基础数据集质量控制方法	(20)
2.3 新版中国探空基础数据集数据完整性和质量状况	(26)
2.4 小结	(33)
第 3 章 中国地面气温和降水气候序列的均一化	(35)
3.1 地面气候资料均一性研究技术进展	(35)
3.2 中国近 60 年地面气温序列的均一性检验订正	(45)
3.3 中国近 60 年地面降水序列的均一性检验订正	(51)
第 4 章 中国探空温度和湿度气候序列的均一化	(61)
4.1 高空气候序列均一性检验技术研究进展	(61)
4.2 中国近 30 年高空温度序列的均一性检验订正	(63)
4.3 中国近 30 年高空湿度序列的均一性检验订正	(82)
第 5 章 中国降水分析产品研发	(92)
5.1 全球降水分析产品的研究进展	(92)
5.2 中国区域基于地面雨量器观测的降水格点分析	(101)
5.3 中国区域基于地面和卫星观测的融合降水产品	(122)
参考文献	(147)

第 1 章 中国地面气温和降水观测资料的质量控制

1.1 地面基础资料重建

气象台站观测资料是研究区域和全球气候变化与预测、天气动力分析、数值天气预报、资料同化与再分析、公共气象服务的基础,是雷达与卫星定标、水文设计、农业决策的基础资料,是资料加工与分析的根本。

从 1979 年开始,通过采用人工录入的手段,全国大规模地开展地面、高空、辐射等气象观测月报表以及台站元数据的数字化工作,建立了我国基础气象资料数字化资料集。

但是由于早期资料数字化工作中标准规范和管理工作不健全,加上数据文件因存储介质的变化(如卡片、纸带、软盘、磁带、光盘,以及数据库系统)而频繁转储,实时资料与历史资料不衔接等诸多原因,使得地面气象资料的质量问题陆续显现,甚至导致一些明显的错误数据直接出现在气象预报预测、公共气象服务的业务平台上。2011 年,我们开始对基础气象数据集进行全面的、系统的质量检测、控制和更正,旨在通过系统的质量梳理,获得一套高质量的可用于气候变化研究的地面基础资料数据集。

基础资料重建工作将进一步加强基础资料业务的标准化和规范化建设,通过国家级与 31 个省级业务技术人员联合攻关,系统解决 2400 余个国家级地面站建站以来的基础气象资料质量问题,建立上述台站完整的、高质量的基础数据文件和台站元数据信息。通过重建,确保国家级和省级归档保存的基础气象资料在质量和完整性等方面的一致性,解决不同气象业务部门间目前存在的资料服务、资料应用的不一致问题,完善各级资料审核、修改和反馈业务机制,确保业务服务数据库和基本业务服务平台中的数据达到同步修正。

1.1.1 重建前基础资料状况

(1) 资料状况

2000 年以前,我国气象台站观测的地面历史基础气象资料记录在纸质报表上,通过邮寄的方式上交归档。2000 年之后,尤其自动站业务运行以来,地面观测数据由台站测报软件自动形成电子化文件上传归档。

按照过去的任务分工,国家级归档保存了地面基准、基本站的观测月报表,省级保存一般站的月报表。由于站网调整、台站观测任务的变更,国家级保存的基准、基本站历史资料并不完整。

(2) 资料问题

一份比较好的数字化观测资料首先应与纸质报表一致。但在对数字化后的资料处理、应用过程中,发现它除了含有纸质资料中本身存在的观测错误和抄录错误数据外,较为严重的是

还存在历史上数字化过程(包括转储过程)中引起的资料错误和缺失(无数据)现象。虽然在历次资料整编过程中,发现和纠正了数据文件中大量错误数据,但是一直没有对地面、高空、辐射历史基础资料进行全面、系统的质量梳理,一些错误数据尚没被发现和纠正。

在我国各类气象观测资料中,地面气象观测资料是记录年代最长、最重要的组成部分,而基于《地面气象观测月报表》数字化而形成的“地面月报数据文件”(简称 A 文件)(中国气象局预测减灾司,2001;中国气象局,2005)又是最基本、最完整的地面观测数据组织单元,其质量问题在几种基础资料中也最多。限于篇幅,本章对于重建前基础资料质量问题、资料质量控制技术及其结果分析主要针对地面月报数据文件。

(3) 地面资料问题

2011 年,国家气象信息中心组织对 1951 年以来地面数字化文件中常用气象要素(气温、气压、湿度、风和降水)逐日数据进行了质量梳理,发现了如下资料问题(任芝花,余予,邹凤玲等,2012):

①不同数字化文件中,存在资料替代现象,具体情况见 1.1.2 节。

②国家级与省级保存的地面基准、基本站资料存在不一致现象,具体情况见 1.1.2 节。

③因数字化错误或遗漏而导致部分要素整月无数据现象。

④地面数字化文件中的 19 个要素中,应用频率相对较低的云、能见度、天气现象、地温、日照、蒸发、冻土、积雪、电线积冰等 10 多个要素数据,存在问题可能比气温、气压、湿度、风和降水等要素要严重些。

1.1.2 地面基础历史观测资料的质量检查和分析

1.1.2.1 质量检测过程和方法

本次研究主要检测信息化过程中引起的国家级 2400 多个台站温、压、湿、风、降水等要素历史基础气象资料质量问题、检测国家级与省级存档而对外服务资料的不一致问题。

我们知道,国家级和省级保存的基准、基本站信息化历史基础资料均经过业务质量审核,并且在日常的数据处理与应用过程中,一旦发现数据问题均会对其进行维护修改。但是历史基础数据从未进行过系统的质量检测,因此数据问题修改总是不彻底,而且国家级和省级数据问题修改不同步,必然导致国家级和省级一方数据正确、另一方数据错误的现象存在。在质量检测前,我们依据以前质量控制工作中发现的问题数据的种类以及国家级和省级保存的基准、基本站历史基础资料的对比结果,梳理出基准、基本站温、压、湿、风、降水历史基础气象资料质量问题的种类,并分别给出其检测方法如下:

(1)不同信息化月报文件中,存在数据替代现象

这种现象表现为用其他站资料或本站其他月资料替代本站本月观测资料。

检测方法:利用 2400 多站 1951—2009 年温、压、湿、风、降水基础数据集,通过软件输出各要素本站本月本日各定时值、日极值与本站其他日或其他站日相应值完全相同的记录,通过归纳后,人工核查报表,并进一步核查是否整月所有观测要素数据被替代。

(2)要素资料中,部分要素整月无数据现象

这种现象表现为某要素实际上可能有观测数据,但是由于信息化错误或遗漏,造成整月无数据现象。在数据服务时,表现为类似缺失现象。

检测方法:当有信息化月报数据文件时(如国家级保存了基准基本站数据文件),若文件中某要素表现为整月无数据现象时,则软件输出该信息,人工对照报表进行核实。当无信息化数据文件时,则利用 2400 多站 1951—2009 年温、压、湿、风、降水基础数据集,假设温、压、湿、风、降水观测项目应同时存在的情况下,当某个月台站有气温记录,筛查压、湿、风或降水要素却整月无记录现象,另外当某个月台站有其他要素记录,筛查气温要素却整月无记录现象,软件结合两者统计筛查,最后人工对照报表进行核实。

(3)人工测风资料信息化错误现象

这种现象表现为风速数据未按规定扩大 10 倍录入,风向实际上为 8 方位观测,却按 16 方位方式录入。

检测方法:由于多数情况下人工定时观测风速的分辨率为 1 m/s,而信息化文件要求风速扩大 10 倍录入,因此风速数据个位一般为 0,通过软件输出个位数非 0 的风速数据,核查报表后,确定数据正确与否;另外,人工测风向普遍为 16 方位,但也曾出现过风向只为 8 方位观测的情况,因此对于 8 方位观测却用 16 方位表达的数据信息需要更正。通过软件输出整月风向用 16 方位方式表达,却未出现 8 方位之外的风向信息,人工核查后确定是否正确。

(4)省级与国家级保存的相应站资料不一致现象

这种现象表现为国家级和省级保存的资料中,同要素同观测时次的数据却不同,缘自日常数据维护不同或数据来源不同。

检测方法:利用 2007 年省级上报的 1951—2006 年 731 个基准、基本站数据集及中国气象局气象档案馆保存的相应站信息化文件,通过软件输出 1951—2006 年省级和中国气象局气象档案馆保存的基准、基本站 7 要素资料中不一致的数据,进行人工查阅报表来进一步核对。

1.1.2.2 重建前地面基础资料质量问题

在气象科学数据共享项目的支持下,2008 年 6 月国家气象信息中心初步建成“1951—2007 年中国地面基本气象要素数据集(1.0 版)”,但在试用过程中发现存在要素项不全、资料缺失、个别站连续几年降水为 0 等问题。2010 年 9—12 月我们对上述数据集以及国家级保存的 756 个基准、基本站温、压、湿、风和降水观测数据进行了质量检测,发现如下数字化资料问题(任芝花等,2012):

(1)资料替代问题

对 1951—2009 年 2400 多站信息化资料的质量检测中,发现 837 对月报数据文件中整月所有要素数据完全相同,另外发现 58 对月报文件中部分要素整月数据相同、或连续某几天各要素或部分要素数据相同现象(表 1.1)。

表 1.1 2400 站 1951—2009 年间地面月报文件相同情况(站月)

类别	总数	迁站	替代	全月观测值相同
整月相同	837	703	116	18
部分要素	58	0	57	1

上述相同现象缘自一个文件中的数据是用其他文件中的数据替代的结果。被替代文件的时空分布见图 1.1。从图中可见替代文件主要属于我国东部,20 世纪 60 年代替代现象较多。

837 个月报数据文件替代现象通过征询省级资料部门后发现,703 个月报文件替代现象是因为迁站或台站任务变更造成的,替代的目的是为了资料序列的连续应用和完整性。但作为档案资料这种替代是不恰当的,实际上只需记录迁站、台站任务变更等元数据信息即可,在后期的数据处理与服务时可考虑通过插补技术保证资料序列的连续性和完整性。另外 134 个月信息化数据文件被错误地用其他站或本站其他月资料替代,其中一种替代是本站资料缺测而人为地用邻近站资料替代,另一种是明显的信息化错误。58 对月报文件中部分要素整月数据、或连续某几天各要素或部分要素数据替代现象,也是由于资料缺测用邻近站数据替代或明显的信息化错误造成的。

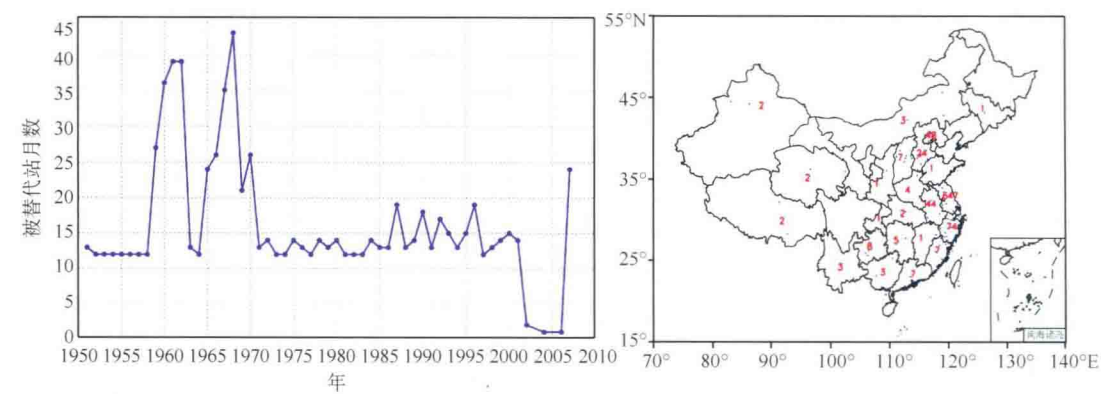


图 1.1 1951—2009 年被替代的信息化数据文件时空分布
(右图中圆点表示台站位置,数字表示所属省份的站月数)

(2)各要素整月无数据现象

这项检测综合利用了国家级和省级制作的 1951—2009 年温、压、湿、风、降水基础资料数据集、国家级归档保存的基准、基本站数字化月报数据文件和基准、基本站 1954—2000 年月报表。检测结果见表 1.2。从表中可见,国家级保存的基准、基本站月报数据文件中各要素共检测到 12859 个站月无数据现象,根据国家级存档的月报表情况,实际仅对其中 6279 个站月核对了纸质报表,共核查到 1749 个站月数据应补录或更正。

表 1.2 国家级存储的 1951—2009 年 756 个基准、基本站资料各要素整月无数据情况(站月)

要素	整月无数据现象发生的时间	整月无数据量	已核查报表量	无数据原因及相应站月数		
				数字化遗漏或错误	有观测数据,按规定可不数字化	无观测任务或缺测
气温	1951—1965 年	20	20	8	0	12
气压	1951—1984 年	7525	3613	22	1276	2315
水汽压	1951—1969 年	465	160	142	0	18
相对湿度	1951—1969 年	314	154	139	0	15
风向风速	1951—1997 年	3062	881	6	859	16
降水	1951—2006 年	1473	1451	1433	0	18
合计		12859	6279	1750	2135	2394

从表 1.2 及图 1.2 可见,国家级存储基准、基本站信息化文件中,降水信息化问题比较明显,共 1433 个站月数据需要补录或更正。其中除了 5 个站月有明显的降水现象而未信息化外,剩下的 1428 个站月整月降水量实际均为 0 或微量(小于 0.05 mm),而信息化月报数据文件中均作了缺测处理。可能省级在日常数据维护中,对信息化数据文件做了更正。

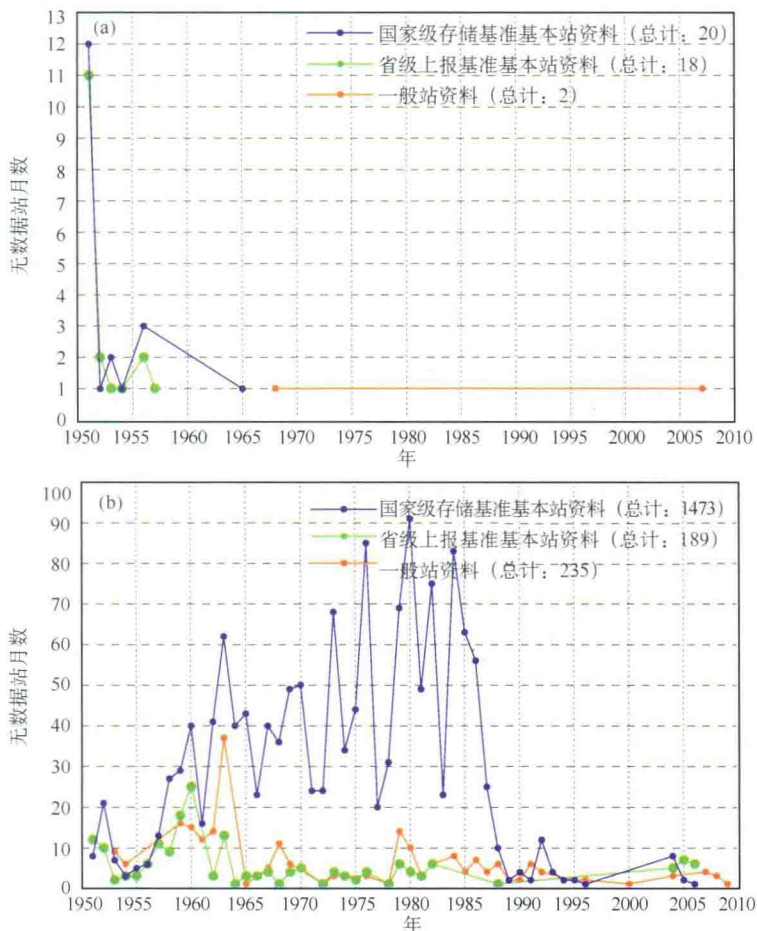


图 1.2 1951—2009 年气温、降水整月无数据站月数历年变化情况

(a)气温;(b)降水

(3)人工测风问题

- ①国家级和省级保存的基准、基本站数据检测中,10 个站月整月风速资料录入错误;
- ②1979个站月整月风向记录属于规定的 8 方位却用 16 方位方式表达。

(4)省级与国家级保存的同一站点资料不一致现象

从省级存储的基准、基本站资料与国家级存储的相应站资料对比结果来看,除存在上述一方整月无数据,而另一方数据存在外,还有以下 3 类不一致现象:

①国家级存储基准、基本站资料与省级上报资料在时间跨度上的差异

总体来讲,国家级存储的基准、基本站资料在时间跨度上短于省级存储的数据。将省级上报的 731 站 1951—2006 年的数据与国家级存储的同站数据进行对比,结果显示国家级比省级缺少 28298 个站月。共有 133 站国家级存储的资料时间长度短于省级资料,其全国分布如图

1.3 所示,主要集中在华中、华南地区。

造成国家级保存的基准、基本站资料少的原因主要是由于台站任务变更造成的。由于2007年之前,按规定国家级资料部门只收集归档基准、基本站观测资料,当某个站由过去的基准、基本站变为一般站时,资料不再上传,而当某站由一般站升为基准、基本站时,才开始上传升级后的观测资料。

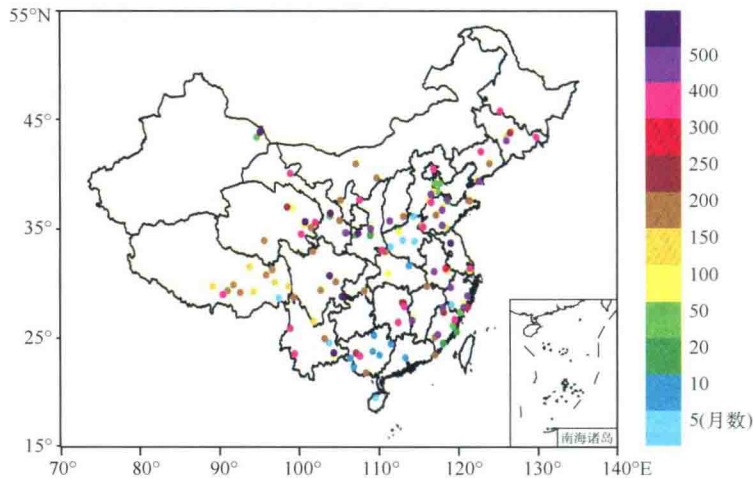


图 1.3 国家级与省级存储基准基本站资料时间跨度差异分布

②国家级资料与省级上报资料在观测数据源上的差异

自动站业务试运行或运行以来,国家级和省级对外服务存在用自动观测资料还是人工观测资料上的差异。从国家级和省级分别制作的2001—2006年731个基准、基本站基础数据集中可见,有173个站因为人工和自动观测数据来源上的差异,使得国家级存储的业务数据和省级数据出现不同。其中,9个站差异小于6个月,136站差异为12个月,1站差异为24个月。2001—2006年间,只有2002年未出现观测数据源的差异,出现差异站月数较多的年份为2004年和2005年。

③其他不一致

除了发现①、②整月数据不一致现象外,还发现部分零散的数据不一致现象,主要原因是因为原数据录入时发生错误,但省级或国家级仅一方做了修正。对比结果如表1.3所示。731个基准、基本站资料中有292263个(组)数据不同,通过报表核查了其中约52%的数据。从核查结果看,省级、国家级正确的比例基本各半。

表 1.3 731 个基准、基本站 1951—2006 年不一致数据核查结果

要素	不一致总数(个)	核查数(个)	省级正确(%)	国家级正确(%)	均不正确(%)
气温	15672	12633	37.6	62.3	0.1
气压	22053	16387	44.9	55.0	0.1
水汽压	109963	14082	36.1	63.6	0.3
相对湿度	34838	16835	43.6	56.2	0.2
风向	53762	47869	64.8	35.1	0.1
风速	35809	26894	21.3	78.1	0.6
降水量	20166	18299	27.3	72.5	0.2
合计	292263	152999	43.33	56.45	0.22

(5) 基准站月报文件 A1 与 A0 中相应数据不一致现象

由于历史原因,基准站常规要素每日有 24 次观测数据(保存在 A1 月报文件中),在由 A1 生成 A0 文件时,存在少许数据不一致现象;平时数据维护时,侧重于 A0 文件,造成基准站 A1 与 A0 文件相应数据不一致现象。

(6) 台站参数错误现象

月报数据文件中的区站号、观测时间、台站经纬度、观测场海拔高度存在数字化错误现象,尤其是台站经纬度、观测场海拔高度和气压感应器海拔高度信息在一段时期内出现频繁交替或变更现象,这与迁站造成的变更是不一样的,需要纠正。

(7) 数据格式及要素方式位表达错误现象

地面历史月报数据文件中数据格式非常复杂,文件中要素方式位的准确表达非常重要,它关系到数据的录入、以后的数据读取是否正确,如降水方式位中,‘R=’表示整月降水缺测或无观测任务,而‘R0=’表示整月有观测,但每日降水量为 0 mm。

(8) 大、小型蒸发数据录入顺序颠倒现象

在月报文件中,两种蒸发均观测时,应先录入全月小型蒸发量,后录入大型蒸发量,但是文件中存在两种资料录入顺序相反的现象。

(9) 因传感器接反或数据记录错误导致的各层地温数据间异常现象

人工站:地温数字化录入错页、错格。比如,纸质报表中抄录错格,将报表中 0.5 cm 栏手工改写为 5.0 cm 顺序,在数字化录入时仍然按照 0.5 cm 的顺序录入,导致两层地温数据录反。

自动站:地温表套管进水或接触不良引起的记录错误;各层地温传感器间接反。如 53533 站,2005 年 1 月—2010 年 12 月连续 6 年地温传感器接反且延续至 2011 年。

(10) 风向传感器疑似雷击造成的风向异常现象

多数自动站中,由于雷击造成的风向传感器故障时并不是表现为某些风向缺测,而是表现为长时间某些风向缺少。

1.1.2.3 重建后的地面基础资料

下面重点介绍两个重建后的地面基础资料产品:“中国国家级地面气象站基本气象要素定时值数据集(V3.0)”和“中国国家级地面气象站基本气象要素日值数据集(V3.0)”

两个新的数据集都包含了中国基本、基准气象站、一般气象站在内的主要 2474 个站点 1951 年 1 月以来地面基本气象要素观测数据以及相应时期的台站信息(台站经纬度、海拔高度)。国家级地面气象站点数由原 2.0 版本的 2425 个增加至 2474 个。时间序列实现自动更新延长,更新频次不低于 1 次/月。

对于地面基本气象要素每日 4 次定时值数据集:各要素数据实有率均超过 99.4%,正确率接近 100%。

相对于原 2.0 版本的本站气压、气温、相对湿度、风 4 种要素定时值,数据修正量为 7628724 个,数据修正比例在 7.5‰~9.5‰之间。

对于地面基本气象要素逐日值数据集:日最大风、日极大风、大型蒸发量数据实有率在 98.1%以上,其他要素项目数据实有率均超过 99.2%,各要素数据正确率接近 100%。

相对于原 2.0 版本的本站气压、气温、相对湿度、风、降水量、蒸发量、日照 7 种要素 17 类项目,数据修正量为 15500438 个,数据修正比例在 0.5%~4%之间。

1.1.2.3.1 中国 198 站逐日气温和降水资料数据集与 GHCND 资料的对比分析结果

GHCND 是美国国家气候资料研制的全球地面气象台站逐日资料数据集,包含逐日最高、最低气温和逐日降水量。为了评估中国逐日气温资料和日累积降水资料质量控制后的数据集质量,我们将之与 GHCND_V3.0 逐日数据集(ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily)中所包含的中国 198 个国际交换站资料分别从资料完整性、数据一致率及其质量进行了对比分析。

(1)资料完整性对比分析

分别对中国和 GHCND 数据集中中国 198 个站自建站日至 2010 年 12 月 31 日的实有资料长度(表 1.4)和资料完整性的分布情况进行统计,获取了两套资料的完整性的详细情况。中国日最高气温 198 站实有数据总量为 4193167 个,序列完整(缺测率为 0)的台站数为 111 个,占总台站数的 56.5%;序列完整性较好(缺测率低于 0.5%)的台站有 70 个;序列完整性一般(缺测率 0.5%~3%)的台站有 14 个;还有 3 个台站序列完整性较差(缺测率大于 3%)。中国日最低气温 198 站实有数据总量为 4193152 个,其中 97 个台站序列完整,84 个台站序列完整性较好、12 个台站序列完整性一般,5 个台站序列完整性较差。中国 20—20 时日累积降水资料实有数据总量为 4196144,198 站中有 146 个台站序列完整,37 个台站序列完整性较好,12 个台站序列完整性一般,3 个台站序列完整性较差。

GHCND 日最高气温 198 站实有数据量为 4067567 个,序列完整的台站为 0 个,24 个台站序列完整性较好,136 个台站序列完整性一般,38 个台站序列完整性较差。GHCND 日最低气温 198 站实有数据量为 3812583 个,序列完整序列完整性较好的台站均为 0 个,1 个台站序列完整性一般,197 个台站序列完整性较差。20—20 时日累积降水资料实有数据总量为 3421853,198 个台站缺测率均大于 3%,序列完整性较差。

表 1.4 中国和 GHCND 数据集实有数据量个数(单位:个)

数据项	日最高气温	日最低气温	20—20 时降水
中国数据集	4193167	4193152	4196144
GHCND 数据集	4067567	3812583	3421853
GHCND/中国	97.0%	90.9%	81.5%

注:GHCND 资料开始时间,以中国为准统计。

通过以上分析,表明中国数据集的实有数据长度和台站序列完整性都明显优于 GHCND 数据集。

(2)资料一致性分析

分别计算二者均不缺测时日最高气温、日最低气温和日累积降水的差值。通过统计差值为 0 的资料长度与实有资料长度的比值,得到两套资料的一致率。

由图 1.4 可见,日最高气温一致性较好,有 195 个站的一致率达到 90%以上,其中云南腾冲站与重庆沙坪坝站一致率为 100%,仅有 3 个站一致率低于 90%,上海徐家汇站一致率最低为 65.80%;日最低气温一致性较日最高气温差,一致率在 90%以上的站为 164 个,另外 34 站的一致率均低于 50%,其中徐家汇站的一致率仅为 3.63%;日累积降水一致率比最高和最低气温都差,在 90%以上的仅有 109 个站,50%以下的有 16 个站。与气温不同,日累积降水还有一种不一致性存在,即中国数据集数据为缺测时,GHCND 数据集为 0,这样的记录一共达

到 4763 条,涉及 36 个台站。通过进一步核查报表,我们发现这 4763 条记录实际均为缺测。

对日最低气温一致率低于 50% 的 34 个站进行核查发现,1961 年 1 月 1 日—1990 年 12 月 31 日,GHCND 数据集中的日最低气温资料相对于中国数据集的资料均发生错位,即 GHCND 数据超前中国数据一日。核查报表之后,确认中国数据集记录正确,而 GHCND 数据集超前一日。

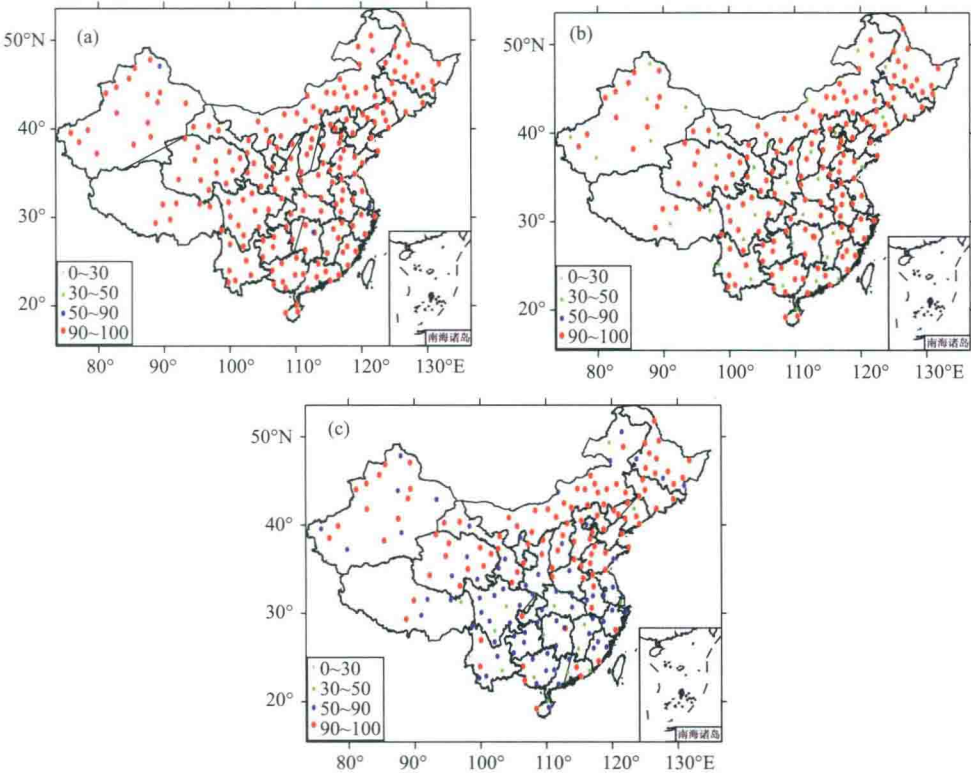


图 1.4 198 个站逐站 GHCND 数据与中国数据的一致率(单位:%)
(a)日最高气温;(b)日最低气温;(c)日累积降水

1.1.2.3.2 质量控制结果分析

中国数据集和 GHCND 数据集由各自的数据制作者进行了质量控制,为了检验中国与 GHCND 数据集中 198 站质量控制效果以及正确率,我们对两套资料中可疑数据和错误数据进行了对比分析,并对错误数据进行分析与核查。两套资料中 198 站的可疑数据率均低于 1%。表 1.5 给出了中国数据集和 GHCND 数据集(二者数据均不缺测时)中 198 站可疑数据的总个数,其中日最高气温中国数据集的错误数为 34 条,而 GHCND(1222 条)约为中国数据集的 40 倍;而就日最低气温而言,中国数据集的总错误记录为 21 条,仅为 GHCND(1253 条)的六分之一;日累积降水中国数据集的可疑数据为 0,而 GHCND 达到了 2457 条。造成这种结果的原因可能是两套数据本身的差异造成的。

表 1.5 中国和 GHCND 中 198 个站总的可疑数据个数

	日最高气温	日最低气温	日累积降水
中国数据集	34	21	0
GHCND 数据集	1222	1253	2457

为了分析造成两套资料质量控制结果差别较大的原因,我们进一步分析了两套资料错误记录的分布特征。结果发现对日最低气温和日累积降水来说,两套数据的质量控制结果的不一致,主要是由于二者数据不一致造成的。当数据存在差异时,GHCND 存在的日最低气温和日累积降水可疑数据分别达到 1081 条和 2292 条。日最高气温则表现了不同的特点,在数值相同的情况下,GHCND 标识的可疑数据高达 1118 个,而数值不同的情况下,其标识的可疑数据仅有 102 个。

通过分析中国数据集和 GHCND 数据集中 198 个国际交换站日最高气温、日最低气温和日累积降水的资料完整性、一致性以及质量控制结果,我们发现中国数据集无论是从完整性还是正确率都远远好于 GHCND 数据集。

1.2 地面月资料的质量控制

1.2.1 气象观测数据的误差分类

任何一种测量所得的观测值通常都存在误差。气象数据相对于其描述要素的真值误差,是决定气候指标能否很好地反映一地气候特征的主要因素(屠其璞等,1984)。存在于气象观测资料中的误差可由许多因子引起。例如,许多气象观测资料是通过遥感或遥测得到的,也就是说,观测设备和观测的对象之间有一定的距离,甚至很大距离。还有一些观测资料是间接观测得到的,如卫星遥感温度探测。有些观测可能因人而异,如能见度的观测。总之,由于观测手段不尽完美,致使观测值与观测对象之间有一定的误差。

所有的气象观测资料的误差可分为三类性质完全不同的误差,即随机误差、系统误差及粗大误差(屠其璞等,1984;翟盘茂,1997;朱江,1995;刘小宁等,2005)。

随机误差是观测资料中普遍存在的固有特性,它是由相互独立的因子引起的,是不相关的。随机误差对于单次测量是无法消除的,其平均值为零。它的分布状态多被假设为正态分布。均方根误差被用作度量随机误差的离散程度。随机误差既然是观测资料中普遍存在的固有特性,因此,在处理单个气象观测资料时,不可能也不必要完全消除它。

与随机误差相反,系统误差的分布相对于零是非对称的,它的平均值显著偏离于零。气象资料的系统误差表现为其气候平均值出现与实际状况明显的偏差,通常与观测仪器、场地变化、观测与资料处理过程中所采用的方法等因素有关。这种误差在观测资料中,大小和符号基本保持不变(屠其璞等,1984)。

粗大误差是没有任何天气学意义的错误资料,通常由观测失误、仪器故障、不正确的编码、抄录和资料传输错误等原因引起。正因为这类资料不代表任何天气意义,因此尽管在整个观测资料中错误资料虽然发生率不高,但也会影响实际数值预报与气候分析结果,其危害性是不可估量的。

1.2.2 质量控制方案

质量控制(QC)的主要目的是检测数据的有效性,并做到对错误资料及时纠正(WMO, 1992)。气候趋势的研究与预测,对气候资料中的错误值和奇异值非常敏感。一直以来,查找资料中的奇异值并对奇异值进行辨别都是质量控制(QC)工作的重点,中国地面历史气象资料