

任芝花, 余予, 邹凤玲, 等. 部分地面要素历史基础气象资料质量检测. 应用气象学报, 2012, 23(6): 739-747.

部分地面要素历史基础气象资料质量检测

任芝花* 余予 邹凤玲 许艳

(国家气象信息中心, 北京 100081)

摘 要

为深入了解地面基础气象资料中存在的问题, 进一步提高资料质量, 综合利用国家级和省级气象资料部门存储的 1951—2009 年 2474 个国家级地面气象站观测的气温、气压、水汽压、相对湿度、风向、风速、降水量 7 种要素信息化基础数据, 检测并分析了数据中存在的问题。结果显示: 国家级和省级气象部门存储的资料中均存在大量与实际观测数据不符的信息化问题, 包括资料的替代问题、要素数据类似缺测问题以及数据录入错误等; 还存在国家级和省级气象部门保存的基准基本站资料不一致现象, 包括资料序列长短不同、对外服务时提供自动还是人工观测数据不一致、更正不同步造成的数据不同等。该文针对上述资料问题给出了详细的检测方法及检测结果。为了确保数据的正确性, 有必要在此次数据质量检测经验的基础上, 对所有历史月报数据文件中的所有要素观测值进行彻底检测与更正。

关键词: 地面; 基础气象资料; 数据质量; 质量检测

引 言

地面基础气象资料是指地面气象台站直接观测的各要素值, 包括各要素定时值、日极值或日累计值。中国地面历史基础气象资料包括 2400 多个由基准基本气象站(以下简称基准基本站)和一般气象站(以下简称一般站)组成的国家级地面气象站(以下简称国家站)建站以来的气温、气压、湿度、风、降水、蒸发、日照、地温、天气现象、云、能见度等共 20 类观测项目数据^[1]。台站直接观测的气温、气压、湿度、风、降水等要素定时值、日极值或日累计值等地面基础气象资料, 是研究区域乃至全球气候变化与预测、天气动力分析、数值天气预报模式研究、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、农业决策的重要依据。若使长期保存在纸质报表上的资料发挥作用, 首先要实现资料的信息化。1979 年开始, 各省气象局按照全国地面气象资料信息化模式的有关规定^[2], 将地面历史气

象观测记录月报表中的 19 类观测项目(不含雪温)观测数据以纸带为数据存储载体进行了信息化。20 世纪 80 年代后期以来的人工观测数据陆续改为软盘和磁带作为数据存储载体, 同时也将以前归档纸带上的信息转储到磁带上。一份比较好的信息化历史基础气象资料首先应与纸质报表一致。但是在对信息化后的资料处理、应用过程中发现, 除了含有纸质资料中本身存在的观测和抄录错误数据外, 还存在信息化过程中(包括转储过程中)引起的资料错误和缺失(无数据)现象。只有清楚并解决信息化历史基础资料中所存在的质量问题, 才能保证业务应用、研究分析以及资料统计与加工的准确性。

质量控制一直是资料工作的重点, 国内外针对地面基础气象资料质量控制技术的研究与应用^[3-25]涉及到的质量控制技术主要包括气候学界限值检查、区域界限值检查、要素间内部一致性检查、时间一致性检查和空间一致性检查等。上述质量控制技术可用于观测台站、数据传输部门进行实时质量控制, 也可用于数据处理部门进行非实时质量控制。

2012-04-23 收到, 2012-08-29 收到再改稿。

资助项目: 国家重点基础研究发展计划(2010CB951600), 公益性行业(气象)科研专项(GYHY201106038), 中国气象局基础气象资料业务能力建设专项“中国气候资料整编”, 国家气象信息中心青年基金项目“中国地面历史气温序列插补”

* E-mail: rzh@cma.gov.cn

中国地面历史基础气象资料在日常业务中经历了台站、省级和国家级气象资料部门三级质量控制^[5], 历次气候资料整编过程中曾进行过多次质量检查, 但上述质量控制技术只可解决部分数据问题, 有的数据问题仍得不到检测与解决。例如, 任芝花等^[6-7]进行 1971—2000 年 30 年气候资料整编时, 通过对月统计资料的质量控制, 发现有部分明显的信息化数

据录入错误或数据替代现象(见表 1), 但是非明显的的数据替代现象, 如某站当年 1 月的观测数据用前 1 年 1 月数据替代、某站观测数据用同期邻近站数据替代、信息化时数据遗漏或因录入不符合规定而造成数据类似缺测现象等, 很难通过常规的质量控制技术检测出。

表 1 1971—2000 年中国基准基本站地面信息化资料错情^[6]

Table 1 Information about surface meteorological data in error of base stations in China from 1971 to 2000^[6]

要素	错误资料站数及月数	资料年份	错误原因
地面所有要素	4 站共 4 个月全月地面所有要素资料均错	1981, 1984, 1992	用其他站或其他月资料替代了本站本月的资料
	1 站 1 个月从云量开始所有要素资料均错	1994	用其他站或其他月资料替代了本站本月的资料
风速	10 站共 85 个月全月风速资料错	1980—1994	资料均缩小 10 倍录入
小型蒸发	1 站 1 个月全月蒸发资料错	1990	资料均扩大 10 倍录入
地温	3 站 10 个月 0 cm 地温全月错	1989, 1994	已取消地温观测, 但全月该资料却错用 0 表示
	3 站 14 个月极端最高地面温度全月错	1989, 1994	已取消地温观测, 但全月该资料却错用 0 表示
	1 站 1 个月 5~15 cm 地温全月错	1984	地温超刻度, 将符号位穿错
	1 站 2 个月 20 cm 地温全月错	1994	地温表坏, 资料应缺测处理, 而信息化资料却全为 0
	1 站 8 个月 40 cm 地温比气候平均值偏高 7.5℃	1971	地温表瓷板断裂, 资料仅供参考
日照时数	1 站 1 月 1 d 资料错	1976	录入错误
日照百分率	1 站 1 月 27 d 资料错	1993	录入错误
极端最高气温	1 站 1 月 1 d 资料错	1995	资料缩小 10 倍录入
相对湿度	1 站 1 个月全月相对湿度方式位错	1995	录入错误
	1 站 1 个月 11 d 相对湿度错	1992	录入错误
最大冻土深度	3 站 3 个月共 15 d 资料错	1980, 1987, 1992	录入错或缩小 10 倍录入
最大积雪深度	2 站 2 个月共 2 d 资料错	1977, 1985	资料扩大 10 倍录入

2010 年 9—12 月作者首先尝试对建站至 2009 年全国 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与质量分析, 以探索数字化过程引起的地面历史基础气象资料质量问题以及国家级与省级气象部门对外服务资料的不一致问题及其解决措施, 为基础气象资料建设工作^①奠定了技术基础。

1 资料

本研究所用的资料包括各要素每日定时值、日极值或日累计值。数据分别来源于国家级气象资料部门(以下简称国家级)和省级气象资料部门(以下简称省级)制作的数据集。

①国家级制作的 756 个基准基本站 1951—2006 年温、压、湿、风、降水基础资料数据集和 2474 个站 2007—2009 年温、压、湿、风、降水基础资料数据集, 来源于国家级存档维护的国家站历史信息化月报数据文件。该信息化文件由省级上报而来, 由

国家级负责进行日常质量维护。

②省级制作上报的 2474 个国家站 1951—2006 年温、压、湿、风、降水基础资料数据集, 来源于省级存档维护的基准基本站和一般站历史信息化月报数据文件。该信息化文件由省级信息化或台站上报而来, 由省级负责进行日常质量维护。

另外, 2474 个国家站中, 国家级只存档了 1954—2000 年大部分基准基本站的纸质月报表和所有站年报表。基准基本站其他年份月报表以及所有一般站纸质月报表均存放在省级。因此, 本研究仅利用国家级存档的基准基本站纸质月报表、部分站年报表, 核对了软件检测出的基准基本站基础数据质量问题。部分问题数据由省级人员对照月报数据文件和月报表进行了核实。

2 数据质量问题的整理及其检测方法

本次研究主要检测信息化过程中引起的基准基本站温、压、湿、风、降水历史基础气象资料质量问

①气预函[19 号]. 关于开展基础气象资料发展与改革专项工作的通知. 2011.

题、检测国家级与省级存档而对外服务资料的不一致问题。假设提前知道基础资料中数据问题的种类及其表现形式,那么有针对性地检测与解决数据中的质量问题效果会更好。

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,因此数据问题修改得不彻底,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、一方数据错误的现象存在。在质量检测前,依据以前质量控制工作中发现的问题数据的种类(如表1)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,整理出基准基本站温、压、湿、风、降水历史基础气象资料质量问题的种类,并分别给出如下检测方法。

①不同信息化月报文件中,存在数据替代现象。该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2474个国家站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。该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,表现为类似缺测现象。

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。当检测时无信息化数据文件,则利用2474个国家站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。该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 $1\text{ m}\cdot\text{s}^{-1}$,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。该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国家级保存的相应站信息化文件,通过软件输出1951—2006年省级和国家级保存的基准基本站7种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。

3 检测结果

3.1 资料替代问题

对1951—2009年2474个站信息化资料的质量检测发现,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。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。由图1可见,替代现象主要出现在我国东部地区,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

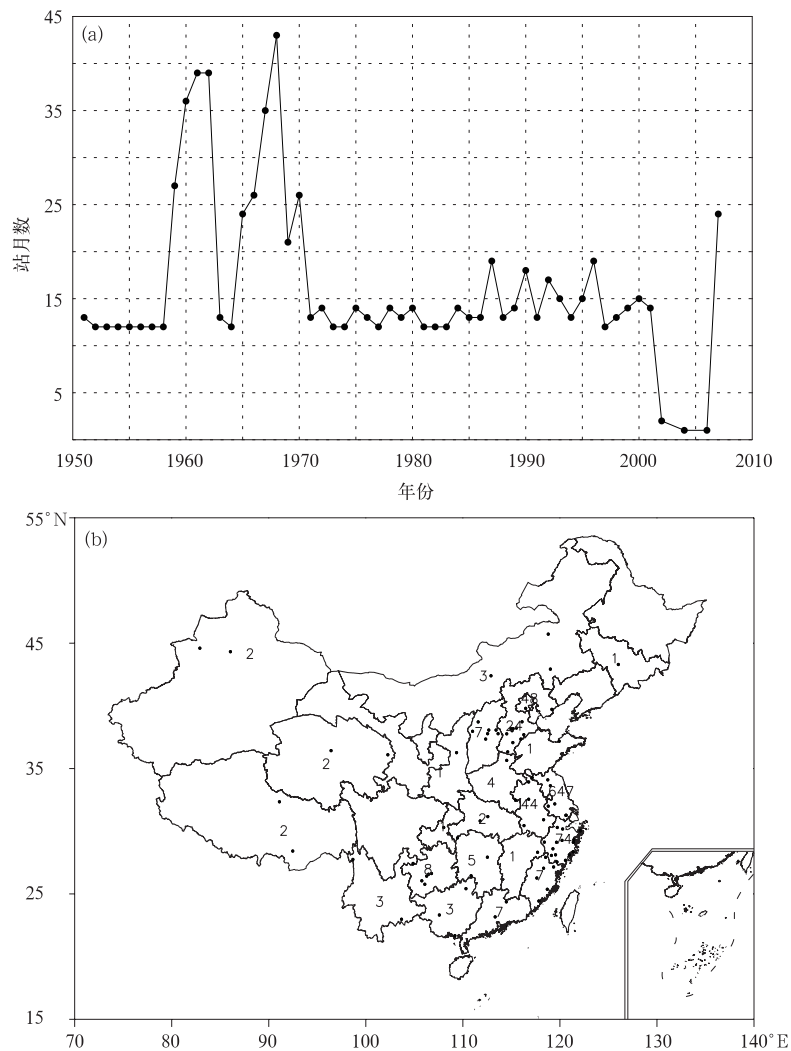


图 1 1951—2009 年被替代的信息化数据文件时间(a)和空间(b)分布
(图 b 中圆点表示台站位置,数字表示所属省份的站月数)

Fig. 1 Temporal distribution(a) and spatial distribution(b) of the number of replaced electronic monthly data files from 1951 to 2009
(dots denote station location and figures denote the total number of replaced electronic monthly data files in a province in Fig. 1b)

信息化错误造成的。

3.2 各要素整月无数据现象

该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,利用了国家级归档保存的基准基本站信息化月报数据文件和基准基本站 1954—2000 年月报表。

国家级和省级保存的基准基本站、一般站各要素整月无数据情况如图 2 所示。由图 2 可见,各要素整月无数据现象普遍集中在 20 世纪 70 年代之前,但 20 世纪 90 年代一般站湿度资料(包括水汽压和相对湿度)整月无数据现象比较高,由于国家级无

月报表,原因有待基础资料建设工作完成后查明。20 世纪 50 年代末至 70 年代中,省级保存的风速资料无数据现象比国家级保存的资料严重,主要因为 9 个台站共 898 个站月的风向、风速均有观测数据,省级可能在由纸带向磁带进行信息转储时,造成风数据丢失而产生类似缺测现象。国家级保存的基准基本站信息化文件中,降水资料整月无数据现象比省级保存的资料明显严重。造成上述各要素整月无数据的原因包括下列 3 种情况:①台站早期无该要素观测任务或观测仪器故障;②观测仪器质量较差,按规定可不信息化;③信息化时遗漏或信息化

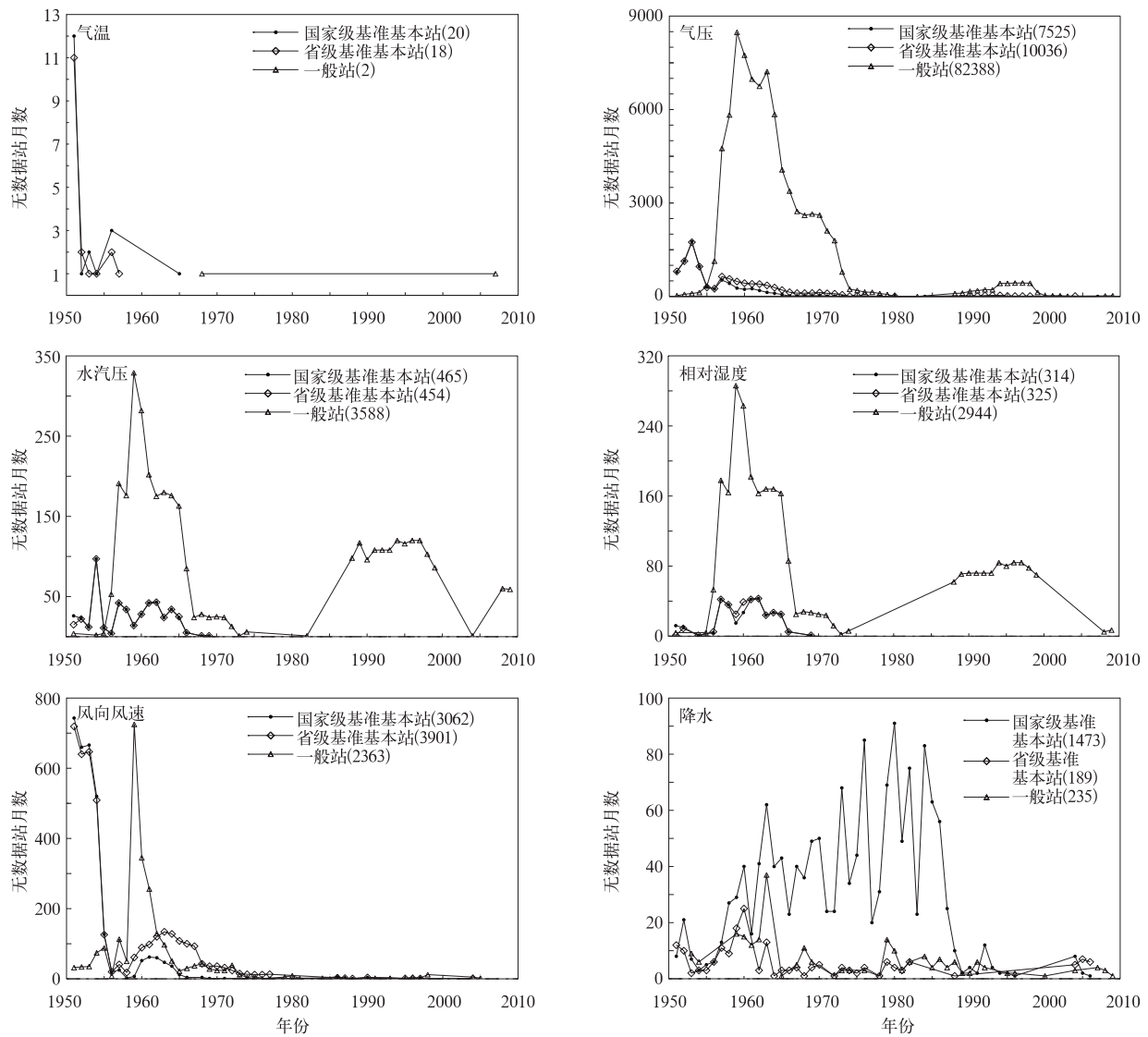


图 2 1951—2009 年各要素整月无数据站月数历年变化情况

Fig. 2 Annual number of whole-month data missing of air temperature, air pressure, vapor pressure, relative humidity, wind and precipitation amount from 1951 to 2009

错误。本次检测主要检查信息化遗漏或信息化错误造成的整月无数据现象。

国家级保存的基准基本站月报数据文件中,各要素整月无数据现象统计与核查结果见表 2。由表 2 可见,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。其中 1276 个站月气压由空盒气压表观测,859 个站月风速采用风力等级或轻便风速表观测,按规定均可不信息化。

由图 2 及表 2 可见,国家级保存基准基本站信

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,剩余 1428 个站月整月日降水量实际均为 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,因此降水信息化问题较轻。

3.3 人工测风问题

3.3.1 风速检测结果

经核查报表,国家级和省级保存的基准基本站数据中,共发现 10 个站月整月风速资料录入错误。

表 2 国家级存储的 1951—2009 年 756 个基准基本站资料各要素
整月无数据核查结果(单位:站月)

Table 2 Verification results of whole month data missing of 6 meteorological elements observed
by 756 national surface base stations from NMIC during 1951—2009(unit: month)

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	原因及相应站月数		
				信息化遗 漏或错误	有观测数据,按 规定可不信息化	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	1749	2135	2395

其中,新疆 5 个站月、西藏和海南各 1 个站月整月风速比观测值缩小了 10 倍。而福建出现 3 个站月整月风速大于 $10.0 \text{ m} \cdot \text{s}^{-1}$ 时,应在末位补 0,错将 0 补在高位,如应为 110,错为 011。

3.3.2 风向 8 方位检测结果

国家级保存的基准基本站风向资料中,整月风向记录只属于规定的 8 方位却用 16 方位方式表达的有 1979 个站月,主要集中在 1970 年之前。对照报表只核查了其中的 1511 个站月,核查结果表明均属于 8 方位观测,因此应更正信息化文件中的数据记录方式。1511 个站月风向 8 方位观测资料中,有 300 个站月省级资料记录正确。

3.4 省级与国家级保存的相应站资料不一致现象

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 个站 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,其分布如图 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。

②国家级资料与省级上报资料在观测数据源上的差异。

自动气象站业务运行或试运行以来,国家级和省级对外服务存在提供自动观测资料还是人工观测资料的差异。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 个站差异为 12 个月,1 个站差异为 24 个月。2001—2006 年,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 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、国家级正确比例基本各半。

上述地面历史基础气象资料质量问题检测结果分析表明,利用常规的质量控制方法很难检查出资料替代、要素整月无数据、观测数据源不同造成的国家级与省级资料不一致等问题。例如,某站 2002 年 1 月的气温数据用该站 2001 年 1 月的观测数据替代,或用同期邻近站观测数据替代现象,气候极值检

查、空间一致性检查等常规质量控制技术一般很难检测发现。此外,资料工作是一个不断深入的过程,

只有切实了解基础资料中可能存在的问题,才能有针对性地制定有效的检测方法。

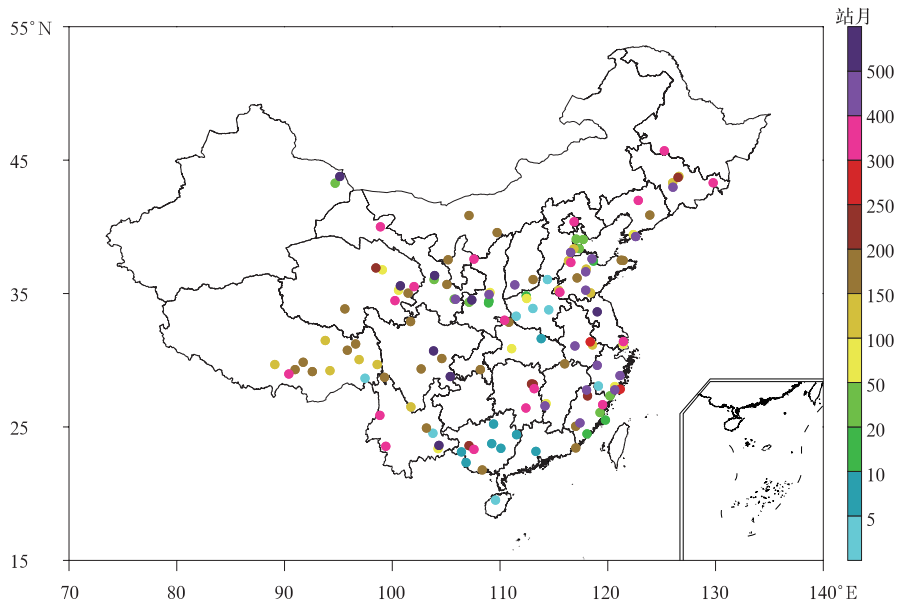


图3 国家级与省级存储基准基本站资料时间跨度差异分布

Fig. 3 Distribution of the time span difference between NMIC and provincial data observed by national base stations

表3 1951—2006年731个基准基本站不一致数据核查结果

Table 3 Verification results of the difference between NMIC and provincial data observed by 731 national base stations from 1951 to 2006

要素	不一致总数	核查数	省级正确率/%	国家级正确率/%	均不正确率/%
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

4 结论与展望

通过温、压、湿、风、降水基础资料数据质量检测,发现国家级和省级存储的资料中均存在大量与实际观测数据不符的信息化问题,包括:①1951—2009年2474个国家站中存在837个月数据替代现象;②国家级存储的1951—2009年756个基准基本站资料中,1749个站月要素资料信息化遗漏、10个站月风速资料错误、1979个站月风向表达方式不准确;③国家级和省级存储的731个基准基本站资料,有133站国家级资料的时间长度短于省级、173站

存在人工和自动观测数据来源上的差异、因数据维护不同步造成国家级和省级存储的数据中约29万不同。上述资料问题通过常规的质量控制技术很难发现,只有针对特殊问题采用特殊的检测手段方可达到目的。

检测出的问题数据量比较大,但是占整个资料量的比例并不高。如按照每个站平均观测时间50年计算,2474个站中,约0.6‰的数据存在整月资料被替代现象;国家级存储的756个基准基本站降水资料中,约3.4‰的整月实有观测数据在信息化文件中表现为类似缺测现象。

本次检测只是针对月报信息化文件20类观测

项目中的气温、气压、水汽压、相对湿度、风向、风速、降水量数据开展的工作,仅通过基准基本站月报表进行了部分数据问题核实。有必要在本次数据质量分析与检测经验的基础上,对所有历史月报数据文件中的所有要素观测值进行彻底检测与更正。广泛发动省级资料主管部门的力量,利用其存档月报数据文件和纸质报表,对发现的所有问题数据进行核查与更正。在核查与更正工作完成后,国家级和省级统一保存一套完整、一致的历史基础数据,同时制定有效的数据更正机制,确保今后国家级和省级在数据存储与对外服务中资料的一致性。

致谢:由于国家级无一般站月报表,因此在检测出的一般站数据替代问题时,通过电询省级资料审核人员方得到核实。在此,特别感谢各省级审核人员的支持!

参考文献

- [1] 中国气象局. 地面气象观测规范. 北京:气象出版社,2003.
- [2] 中国气象局. 常规气象资料信息化模式文本汇编. 北京:气象出版社,2001.
- [3] 刘小宁,任芝花. 地面气象资料质量控制方法研究概述. 气象科技,2005,33(3):199-203.
- [4] 刘小宁,鞠晓慧,范邵华. 空间回归检验方法在气象资料质量检验中的应用. 应用气象学报,2006,17(1):37-43.
- [5] 任芝花,熊安元. 地面自动站观测资料三级质量控制业务系统的研制. 气象,2007,33(1):19-24.
- [6] 任芝花,刘小宁,杨文霞. 极端异常气象资料的综合性质量控制与分析. 气象学报,2005,63(4):526-533.
- [7] 任芝花,熊安元,邹风玲. 中国地面月气候资料质量控制方法的研究. 应用气象学报,2007,18(4):516-523.
- [8] 任芝花,许松,孙化南,等. 全球地面天气报历史资料质量检查与分析. 应用气象学报,2006,17(4):412-420.
- [9] 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):123-132.
- [10] 陶士伟,张跃堂,陈卫红,等. 全球观测资料质量监视评估. 气象,2006,32(6):53-58.
- [11] 陶士伟,仲跻芹,徐枝芳,等. 地面自动站资料质量控制方案及应用. 高原气象,2009,28(5):1202-1209.
- [12] 王海军,杨志彪,杨代才,等. 自动气象站实时资料自动质量控制方法及其应用. 气象,2007,33(10):102-106.
- [13] 王海军,刘莹. 综合一致性质量控制方法及其在气温中的应用. 应用气象学报,2012,23(1):69-76.
- [14] 杨萍,刘伟东,仲跻芹,等. 北京地区自动气象站气温观测资料的质量评估. 应用气象学报,2011,22(6):706-715.
- [15] 中国气象局. 中华人民共和国气象行业标准:地面气象观测资料质量控制. 2010.
- [16] Eischeid J K, Baker C B, Karl T R. The quality control of long-term climatological data using objective data analysis. *J Appl Meteor*, 1995,34:2787-2795.
- [17] Fillipov V V. Quality Control of Meteorological Data. World Weather Watch Planning Report, WMO-No. 26, 1968.
- [18] Igor Zahumensk. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Expert Team on Requirements for Data from Automatic Weather Stations, Third Session, WMO, 2004.
- [19] Peterson T C, Vose R S, Schmoyer R, et al. Global historical climatology network (GHCN) quality control of monthly temperature data. *Int J Climatol*, 1998,18:1169-1179.
- [20] Sciuto G, Bonaccorso B, Cancelliere A, et al. Quality control of daily rainfall data with neural networks. *J Hydro*, 2009, 364:13-22.
- [21] Song Feng, Qi Hu, Qian Weihong. Quality control of daily meteorological data in China, 1951—2000: A new dataset. *Int J Climatol*, 2004,24:853-870.
- [22] Vejen F, Jacobsson C, Fredriksson U, et al. Quality Control of Meteorological Observations Automatic Methods Used in the Nordic Countries. Climate Report, No. 8, 2002.
- [23] WMO. Guide on the GOS. WMO-No. 488, PART VI, 1989.
- [24] WMO. Manual on GDPS. WMO-No. 485, Volume 1: II, 1-4, 1992.
- [25] WMO. Guide on the GDPS. WMO-No. 305, Chapter 6, 1993.

Quality Detection of Surface Historical Basic Meteorological Data

Ren Zhihua Yu Yu Zou Fengling Xu Yan

(National Meteorological Information Center, Beijing 100081)

Abstract

Surface basic meteorological data are observational values from surface stations, including hourly values, daily extreme and cumulative values. Chinese surface historical basic meteorological data consists of observations from more than 2400 national stations since as early as 1951, including 20 kinds of elements, such as air temperature, air pressure, humidity, wind, precipitation, evaporation and so on. These data are the basis of regional and global climate change researches and predictions, synoptic dynamic analysis and public meteorological services. The digitization of these data is started by China Meteorological Administration in 1979. High quality digital data should be faithfulness to the paper reports. But incorrect data and data missing problems caused by digitization and restoring have been found, besides those caused by error observations. Resolving these problems will contribute to the improvement of operational application accuracy, scientific researches and data processing.

In order to identify quality problems of Chinese surface basic meteorological data and improve the data quality, several methods are applied to detect the data quality of air temperature, pressure, humidity, wind and precipitation observed by 2474 national surface stations in China from 1951 to 2009. The above data are collected based on National Meteorological Information Center (NMIC) and provincial archived electronic data files. Results show that a large number of error data different from actual measurements are stored in both NMIC and provincial archived data files, which are caused by incorrect digitization. For example, some electronic data files are replaced by other station measurements, some meteorological elements aren't observed for a period, and some data are miss-typed. Some flaws between NMIC and provincial archived data files also shows, including the data file differences in time span and various data source (automatic or manual observation), and data difference induced by asynchronous correction of wrong data. It is impossible to detect all the above data problems using conventional data quality control methods, so some special methods are proposed for each kind of problem. The proportion of incorrect measurements of the whole data is low, though its number is huge. For example, for more than 2400 national stations, only 0.06 percent of data are wrong copied, and for more than 700 baseline stations, only 0.34 percent of precipitation data are missing. The data quality detecting method mainly focuses on temperature, pressure, humidity, wind and precipitation data from about 700 baseline stations, and data problems are verified by checking the monthly paper reports but not corrected. With this work experience, thorough data problem detections and corrections on the whole elements from historical electronic data files are carried out by China Meteorological Administration. A review of this program and the results will be published soon.

Key words: surface; basic meteorological data; data quality; quality detection