

全国高空资料质量控制和建库方法的研究*

周尚河

(国家气象中心, 北京 100081)

提 要

通过一个实例, 阐述我国高空资料质量控制的主要方法及建立高空气候资料数据集的工作流程、主要技术问题的处理和系统设计思想, 对于我国高空气候资料现代化业务建设, 具有重要的参考价值, 最后讨论了高空气候资料数据集的技术特征。

关键词: 质量控制与数据集 高空资料 气候

引 言

高空气象历史资料是我国主要的气候资料类型之一, 近 50 年来, 已被广泛应用在我国气候研究和国民经济建设中, 并发挥了重要的作用。为了满足四化建设和科技兴国的需要, 建立以气候研究为目的我国高空气候历史资料库, 把资料的利用从繁琐、枯燥的资料预处理中解脱出来, 已显得十分迫切。

全国高空气象资料自 1959 年起每 10 年整编一次。最近 30 年(1961~1990)的整编统计结果, 在印刷出版后已保存在磁带上, 但尚未形成真正意义上的可以共享的气候资料库。

作者利用我国 30 个站高空历史资料, 与英国哈得莱(Hadley)中心 D. P. N. Cullum 先生合作, 建立起来的高空气候历史资料数据集, 通过该中心近几年在季风和气候研究中的应用, 业已证明, 不仅具有较强的实用性和可靠性, 其数据集结构、设计思想和对一些技术问题的处理, 在高空资料管理、应用和现代化业务建设中也具有重要参考价值。

1 原始信息化资料的组成

选用的我国 30 个站高空原始信息化资料由 5 部分构成: ①全国高空压、温、湿卡片信息化资料(1950~1979); ②增补的全国部分台站高空压、温、湿软盘信息化资料(1950~1979); ③全国高空测风卡片信息化资料(1950~1979); ④全国高空压、温、湿、风天气报告磁带资料(1980~1987); ⑤全国高空压、温、湿、风天气报告磁带资料(1988~1990)。

2 质量控制

原始资料的质量直接影响到各项气候统计成果的可靠性和正确性, 因而质量控制成

* 1999-01-21 收到, 1999-04-14 收到修改稿。

为高空资料数据集建立过程中一个十分重要的工作环节. 根据上述资料的质量状况, 质量控制采取以下步骤: 增补资料因完成信息化后未作检查, 确定为统计前质量控制的重点, 边检查边纠错; 所有的资料在统计前的质量检查中, 凡错误或可疑的数据一律不参加气候统计.

质量检查的内容和方法^[1]主要有以下几个方面:

信息化格式检查 高空信息化资料是有格式资料, 通过对资料中各种控制符号的识别, 重点检查资料的数据结构和信息化规定, 以确保数据格式正确.

句合计值检查 以要素为单位, 重新计算各句合计值, 通过与资料中原有的各句合计值的比较, 初步检查资料中有无明显的错误.

气候极值检查 各标准等压面的高度、温度的气候极值范围见表 1, 露点温度的气候极值范围为 $-140^{\circ}\text{C} \sim +35^{\circ}\text{C}$, 风向的范围为 $0^{\circ} \sim 360^{\circ}$, 风速的气候极值范围为 $0 \sim 150 \text{ m/s}$, 凡超出气候极值范围的要素值, 都认为是错误记录.

表 1 各等压面高度、温度气候极值范围表

规定层 (hPa)	位势高度(m)		温度($^{\circ}\text{C}$)		气压(hPa)	
	下限	上限	下限	上限	下限	上限
地面			- 70	+ 45	600	1060
850	900	1700	- 40	+ 35		
700	2300	3300	- 40	+ 20		
500	4600	6000	- 50	+ 10		
300	7800	10000	- 70	- 20		
200	10000	13000	- 80	- 35		
150	12000	15000	- 90	- 35		
100	14000	18000	- 90	- 35		
50	18000	22000	- 95	- 30		
30	21000	25000	- 95	- 25		
20	23000	28000	- 95	- 20		
10	28000	33000	- 95	- 20		

上、下层等压面高度逻辑关系检查 当上下两个相邻等压面的高度不满足关系式 $H_{p_1} < H_{p_2}$ 时, 则认为上、下等压面的高度中至少有一个错误, 其中 H_{p_2} 、 H_{p_1} 分别为上、下层等压面的实际高度(下同).

当等压面的风向风速中有一个缺测, 或者等压面 850 hPa 及其以下各层出现风向为 0 风速不为 0, 或者风速为 0 风向不为 0 情况时, 则认为该风向风速中有错误.

露点温度与温度相互关系检查 当同一等压面的露点温度(T_d)和温度(T)之间不满足关系式 $T_d \leq T$ 时, 则认为温度、露点温度中有错误.

静力学检查 等压面厚度计算公式^[2]为:

$$H_{p_1}^{p_2} = 18422.7 \times \frac{\bar{T}}{273} \times \lg \frac{P_1}{P_2} + 18422.7 \times \frac{\bar{T}}{273} \times 0.378 \times \frac{\bar{E}_T}{P} \times \lg \frac{P_1}{P_2}$$

$$\Delta H = (H_{p_2} - H_{p_1}) - H_{p_1}^{p_2}$$

式中 P_2, P_1 分别为上、下两个相邻等压面的标准气压值(hPa), $H_{P_1}^{P_2}$ 为上下层等压面厚度计算值(m), \bar{T} 为气层平均温度(K), \bar{P} 为气层平均气压(hPa), $E_{\bar{T}}$ 为温度在 \bar{T} 时的饱和水汽压(hPa), ΔH 为上下层等压面实际厚度与计算厚度之差(m).

用上面给出的公式计算出上下层等压面厚度 $H_{P_1}^{P_2}$ 及厚度差 ΔH , 当 ΔH 绝对值大于 30 m 时, 则认为上下层等压面的高度、温度和露点温度中至少有一个错误.

风切变检查 风切变计算公式^[1]为:

$$S_d = \sqrt{(V_{EB} - V_{EA})^2 + (V_{NB} - V_{NA})^2}$$

式中 $V_{EA}, V_{NA}, V_{EB}, V_{NB}$ 分别为上、下两个相邻等压面的东风分量和北风分量, 当计算出的风切变绝对值大于 20 m/s 时, 则认为上、下层的风向风速中至少有一个错误.

3 气候统计

气候统计的主要任务是根据气候研究的需要, 通过对原始资料的加工处理, 向高空气候资料数据集提供建库用的统计成果资料. 从参加气候统计的等压面中选出 12 层, 即地面、850、700、500、300、200、150、100、50、30、20、10 hPa 等.

(1) 统计项目

气候统计分为两大部分, 即历年值统计和累年值统计.

历年值统计: 历年值统计项目共有 6 项, 即等压面的月平均高度、温度、露点温度、稳定性和平面合成风向、风速. 累年值统计: 参加累年值统计的要素共有 5 个, 即等压面的月平均高度、温度、露点温度、东风分量、北风分量; 累年值统计项目共有 6 项, 即累年月平均、累年月最高、累年月最低、累年月中位值、累年月标准差和该月参加累年统计的年数.

(2) 关于 1980 年以前等压面风计算问题

在我国 1980 年以前的高空历史资料中, 高空压、温、湿和高空风自有史以来一直分开存放, 其主要原因是它们不是在同一类型观测高度上的资料, 前者为等压面上资料, 观测高度是可变的, 后者观测高度是固定的. 为了从规定高度上的风向风速中计算出其间标准等压面的风向风速, 计算的方法一般有两种, 即内插法和向量合成法, 本文采用向量合成算法^[3], 计算出的等压面上的东风分量、北风分量和风速, 经检查误差较小.

4 数据集的逻辑结构和物理结构

中国高空气候资料数据集是由两个基本数据集, 即历年值基本数据集和累年值基本数据集组成. 与以往其他气候资料库不同, 它是一个成果气候资料库, 它的逻辑结构和物理结构设计是否科学、合理将直接关系到我国高空成果资料的管理和使用.

(1) 数据集的逻辑结构

数据集的逻辑结构是指可以用图形来表示数据之间的逻辑关系, 它独立于数据的存储介质, 是数据存取、处理和调用的基础.

历年值基本数据集的逻辑结构是根据数据项之间的关连和累年值统计的需要, 设计

为树形层次结构(见图 1), 从根部往下分为 6 层, 依次是台站、月、时次、年、等压面、要素历年值统计项, 其中“年”放在第 4 层, 与正常次序有所不同, 因为累年值统计是在历年值基础上完成的, 这样设计使后面将要进行的累年统计容易一些。

累年值基本数据集的逻辑结构是根据数据项之间的关连, 设计为树形层次结构(见图 2), 从根部往下分为 6 级, 依次是台站、月份、时次、等压面、要素项、要素累年值统计项。

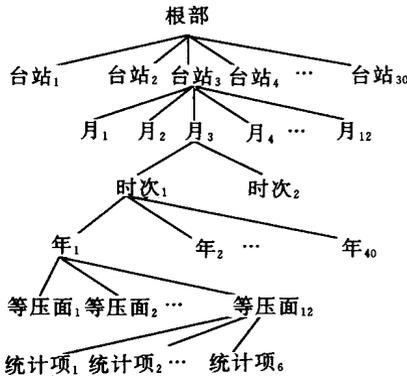


图 1 历年值基本数据集的逻辑结构

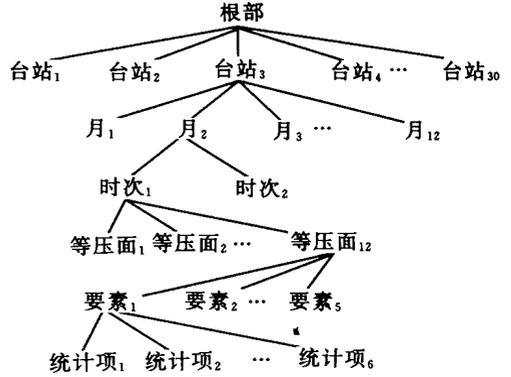


图 2 累年值基本数据集的逻辑结构

(2) 数据集的物理结构

数据集的物理结构, 即数据的存储结构, 用来描述数据在物理存储介质上的组织方式。

历年值基本数据集的物理结构是历年值基本数据集物理记录的概念, 把 1 个站 1 个月 1 个时次所有的要素历年统计值定义为 1 个历年值记录, 其结构如下:

```

m month year block station
surf time 0 T1 Td1 M12 S1 d1 f1 C11 C12 C13 C14 C15
850 h2 M21 T2 Td2 M22 S2 d2 f2 C21 C22 C23 C24 C25
700 h3 M31 T3 Td3 M32 S3 d3 f3 C31 C32 C33 C34 C35
:
10 h12 M121 T12 Td12 M122 S12 d12 f12 C121 C122 C123 C124 C125

```

这是一个准二维表, 横向为各个要素历年值统计项和质检码, 纵项为标准等压面。表中 m month、year、block、station 分别表示资料的月、年、台站的区号和站号, surf、850... 10 分别为地面实际气压值和各等压面标准气压值, time 表示资料的观测时间, 表中 0 无意义, 相当于空格, h_i、T_i、T_d_i、S_i、d_i、f_i 分别表示等压面的高度、温度、露点温度、稳定性、合成风向和风速, i 从 1 到 12 分别对应 12 个等压面(下同), M_{1i}、M_{2i} 分别表示等压面温度和风每月缺测的天数, C_{1i}、C_{2i}、C_{3i}、C_{4i}、C_{5i} 分别表示等压面的高度、温度、露点温度、风向、风速的质检码。

累年值基本数据集的物理结构即累年值基本数据集物理记录的概念, 是把 1 个站 1 个月 1 个时次所有的要素累年统计值定义为 1 个累年值记录, 其结构如下:

```

station month time
surf HH mean highest lowest median deviation years
      TT
      TD
      VE
      VN
850  HH
      TT
      TD
      VE
      VN
      :
10   HH
      TT
      TD
      VE
      VN

```

这是一个准二维表, 横向为累年值统计项, 纵项为各等压面参加累年值统计的要素项. 表中 station、month、time 分别表示资料的台站号、月份和观测时间, mean、highest、lowest、median、deviation 分别表示各等压面各要素累年月平均值、累年月最高值、累年月最低值、累年月中位值和累年月标准差, years 表示该月参加累年统计的年数, HH、TT、TD、VE、VN 分别表示各等压面的高度、温度、露点温度、东风分量和北风分量等.

5 数据集建立的工作流程

高空资料数据集建立的工作流程如图 3 所示, 它由历年值统计、累年值统计和数据检索等 3 个基本部分组成.

(1) 数据集的建立 在进入上面工作流程前, 须根据数据集的逻辑结构、物理结构和信息量的大小, 预先在硬盘上建立两个相互独立、可以直接存取的基本数据集, 其主要作用集中表现在以下两个方面: 一是在历年值和累年值统计过程中, 为各项统计结果自动装入基本数据集提供存储空间; 二是实现气候统计和数据集建立两项工作几乎在同一时刻完成.

(2) 数据检索 在数据集建成后, 通过对用户检索要求分析和技术处理, 实现对历年值和累年值基本数据集中任一数据的检索. 数据检索分为历年值检索和累年值检索.

历年值检索 在一次检索过程中, 可以检索到 1 个站或全部站某年某月某时次所有的要素历年统计值. 检索要求输入的格式为:

YYYY MM T 或者 YYYY MM T IIiii

累年值检索 在一次检索过程中, 可以检索到某站某月某时次某等压面所有的要素

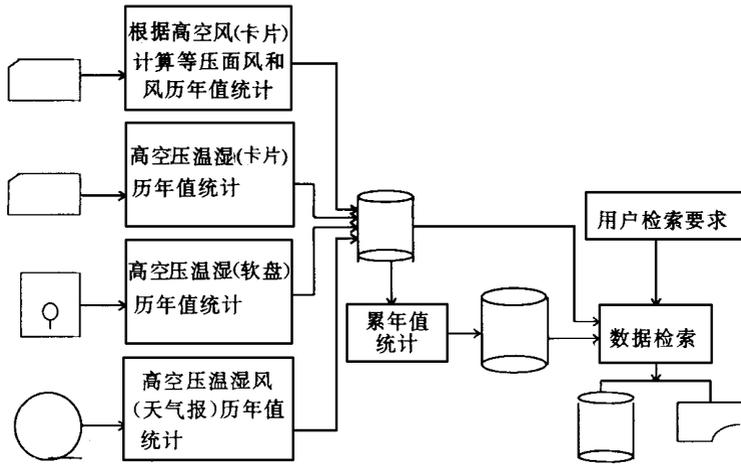


图 3 工作流程逻辑结构图

累年统计值. 检索要求输入的格式为:

IIiii MM T LLLL

在上述格式中, YYYY、MM、T、IIiii 和 LLLL 分别表示资料的年、月、观测时间、台站号和等压面的标准气压.

6 使用与讨论

高空气候资料数据集具有以下主要技术特征:

- (1) 由于在统计前, 已根据气候研究的需要, 对参加统计的等压面和统计项目进行了必要的选择, 因此数据集内各项统计结果针对性比较强, 可以直接使用;
- (2) 为了保证各项统计结果的可靠性和唯一性, 在统计前已用多种检查方法对参加统计的资料进行全面而系统的质量检查, 在历年值统计过程中, 在各类资料统一数据格式的基础上统一了统计程序, 所采取的各项措施基本上达到预期的效果;
- (3) 数据集具有检索功能. 数据集内各项统计资料时间序列长, 地理布局比较合理;
- (4) 建立了一条比较合理的数据集建立的工作流程. 针对不同来源、不同载体的原始信息化资料分散、资料状况差异较大等特点, 在数据集建立过程中, 通过采取分散与集中相结合的方式, 不仅简化了工作过程, 提高了工作效率, 而且大大地降低了工作难度和工作强度, 可操作性强.

参 考 文 献

- 1 Parker D E, Cox D I. Towards a consistent global climatological raw insonde data-base. *Int. J. Climatol.*, 1995, 15: 473~ 496.
- 2 中央气象局. 高空气象观测常用表(内部材料). 1976.
- 3 王树廷, 王伯民, 孙安健, 等. 气象资料的整理和统计方法. 北京: 气象出版社, 1984.

QUALITY CONTROL AND TECHNICAL METHOD FOR PRODUCING DATA SET FOR UPPER-AIR DATA IN CHINA

Zhou Shanghe

(*National Meteorological Center, Beijing 100081*)

Abstract

The quality control of upper air data in China, the working flow and the main technology in processing as well as design ideas for producing the upper air climate data set are described through an example. The experience could be valuable for the operational construction of the upper air climate data in China. In addition, the main features of the upper air climate data set are discussed.

Key words: Quality control and data set Upper air data Climate