

# 我国近百年温度降水资料现状及数据集的研制

王丽萍

郑瑞清

(国家气象中心,北京 100081) (内蒙古气象探测资料中心)

## 提 要

主要介绍了我国 1949 年以前温度、降水资料的情况和“中国近百年温度、降水数据集”的制作情况,如数据集的处理原则、特殊数据的处理、存储格式等。本数据集包含全国 725 个气象台站(包括香港站)的月平均温度、月平均最高温度、月平均最低温度和月降水量资料。最早有记录的台站为北京,开始时间为 1841 年 1 月。最长连续有记录的台站为上海龙华,为 1873 年 1 月~2001 年 12 月。数据集截止时间为 2001 年 12 月。

关键词: 百年 温度 降水 数据集

## 引 言

全球气候变化一直是当前的热门研究课题。随着经济的发展,洪涝、干旱、热浪、冷害的影响也越来越大。在全球气候变化研究和减灾、防灾的工作中,长序列气候资料是不可或缺和重要的。以往大部分采用的只是 1949 年以来观测的数据,本数据集首次将解放前和解放后的数据同时处理,形成百年温度、降水数据集。该数据集在气候预测、气候变化研究、气候服务等工作中具有实际应用价值。

### 1 1949 年以前地面观测站点的情况

我国 1949 年以前的温度、降水资料比较复杂,这是因为气象台站建设情况的复杂性所造成的。1949 年以前气象台站遍布于全国,除西藏、新疆、海南站点较少外,其他各省均有不同数目的站点;其观测资料的长度也随其建站时间的不同而不同<sup>[1]</sup>;由于战乱不断,导致观测资料时有缺测,同一站点观测的完整资料很少。这些情况在台站信息文件里(表略)有全面的反映,在此不再赘叙,仅简要列出我国解放前已有的各种形式的地面气象台站如下:帝国主义列强在中国建立的气象台站;民国政府创办的气象台站;民办气象台

站;解放区创建的气象台站。

众多的气象台站导致出现了站名相同而观测点不同(即经纬度不同)的情况,众多的观测时段导致出现了重复的时段和部分重复的时段(见表1)。

表1 北京站和天津站的降水资料

区站号	纬度	经度	海拔高度	年代
				1841~1855,1860~1861, 1869~1884,1886,1889~1900, 1903,1905~1908,1910~1911, 1914~1939
54511	3954	11628	375	
54511	3954	11628	520	1914~1921
54511	4000	11622	845	1932~1936
54511	3956	11620	513	1940~1950
54511	3956	11628	500	1940~1944
54527	3908	11711	190	1891~1931
54527	3910	11710	40	1906~1936,1946~1949
54527	3909	11711	150	1919~1920
54527	3908	11711	160	1921~1950
54527	3906	11712	99999	1920~1949
54527	3910	11710	99999	1931~1934
54527	3908	11700	80	1918~1939
54527	3855	11726	99999	1931~1934

其中:99999为缺测。

## 2 对1949年前温度、降水资料进行订正

我国在80年代曾经组织了人力对1949年以前的温度、降水日值资料进行了信息化和处理,最后形成了803个温度月值文件和1195个降水月值文件。但由于这段时间资料的复杂性,因此仍存在着一些问题。本人在制作数据集的过程中,对资料进行了检验和处理,并对出现的问题进行了订正,以保证资料的正确性和完整性,为数据集的制作奠定了良好的基础。

## 3 数据集的制作

### 3.1 处理原则

①选取1949年以后仍有资料的台站放在主序列,1949年以后没有资料的台站放在参考序列;②如果资料有多段,且年份不重叠,则各段资料均放在主序列中,年代从前到后排列;③如果资料有多段,且年份有重叠,则取年代最长的资料放在主序列,其余的放

在参考序列,主序列年代不完整时,用参考序列中的补齐;④如果资料有多段,且年份有重叠,年代也相同,则取经度、纬度最接近现在台站经纬度的资料放在主序列,其余的放在参考序列。

### 3.2 特殊数据的处理

在1949年以前的温度、降水信息化资料中,数据采用字符方式存储,其中含有空格、?、\、+等符号,该符号均取自原始报表;另外在处理后的信息化资料中,也含有特殊字符。

本数据集中,将字符型数据均处理为数值型数据,采用整型方式存储。由于年降水量中有大于39000的值,超过了2个字节整型数所能存储的最大限度,故采用4个字节整型数存储。对其中的特殊符号和特殊数据做如下处理:

温度中:‘ ’——32766;‘±xxx?’——±xxx+31000;‘±xxx\’——±xxx+32000

降水中:‘99999’——32766;‘00000’——32700;‘00000?’——42700(32700+10000);‘xxxxx?’——xxxxx+40000;‘xxxxx+’——xxxxx+负号

例如:温度中字符型数据为‘-118?’则数值型数据为30882;温度中字符型数据为‘-12\’则数值型数据为31988。

### 3.3 数据集存储格式

温度和降水各为一个数据集。每个数据集中区站号、年、月均按升序排列。温度要素的记录长度为36个字符,包括区站号、年份、月份、平均温度、平均最高温度、平均最低温度。降水要素的记录长度为24个字符,包括区站号、年份、月份、降水量。数据类型为整型。

### 3.4 台站信息存储格式

本数据集有2个台站信息文件,一个是温度要素台站信息文件,一个是降水要素台

站信息文件。台站信息文件分别包含温度、降水全部台站的区站号、中文站名、省名、纬度、经度、拔海高度、记录年代和观测场拔海高度特征码。其中1949年以前的经度、纬度、拔海高度取自与其数据观测时段相对应的文件,1949年以后的经度、纬度、拔海高度取自于1998年。年代按实有年代记录。

台站信息按区站号大小升序排列,由于1949年以前资料来自不同观测站点,故将其按年代顺序排列,以供使用人员参考。

#### 4 数据集的构成

本数据集处理的1949年以前的资料共由1998个文件组成,其中温度803个文件,降水1195个文件;1949年以后的资料共由4个文件组成。该数据集对2002个文件进行加工、处理后形成6个文件。它们是:BNTT.DAT(长年代温度数据集文件)、BNRR.DAT(长年代降水数据集文件)、TZXXQHT.DOC(温度要素台站信息文件)、TZXXQHR.DOC(降水要素台站信息文件)、CKXLT.DAT(温度要素参考序列文件)、CKXLR.DAT(降水要素参考序列文件)。

#### 5 主要技术难点

①本数据集所处理的1949年以前的资料文件数多、数据量大,资料来源不同,站段

重复出现,文件资料段序号多达15种(如降水资料中的序号分别有01、02、03、04、05、06、07、08、20、21、22、23、24、25、26),北京站(54511)有5个站段,天津站(54527)有8个站段,各个站段又有多个不连续的时间段,因此处理起来比较复杂。

②1949年以后的原始资料数据量太大,约40多兆,分成几个文件后仍有十几兆之多,给程序的调试带来困难,资料的查找和合并等工作既花费时间又非常繁琐。

③资料格式说明和数据错误种类太多,因此在程序调试中不断出现各种情况,对此需要进行分析、判断、查找并处理,既增加了程序调试的难度,对错误数据的查找和修改也增加了时间消耗。

#### 6 结语

本数据集首次将我国解放前和解放后的温度、降水资料进行统一处理,站点之多、年代之长、资料之完整,前所未有的。该数据集可在气候预测和气候分析研究等领域中发挥重要作用。

#### 参考文献

- 1 吴增祥. 中国近代地面气象台站史略. 科技年报. 国家气象中心, 1996: 303~311.

## Data Set of Temperature and Precipitation in Last 100 Years in China

Wang Liping

(National Meteorological Center, Beijing 100081)

#### Abstract

The data set of the temperature and precipitation last 100 years in China are introduced, for example, the handling principle of the data set and special data, stock form and so on. This data set contains mean monthly temperature, mean monthly maximum temperature, mean monthly minimum temperature and monthly precipitation at 725 meteorological stations. It is Beijing station that has the most early records, which start from January 1841. It is Shanghai Longhua which has the longest and continuous record, the record are from January 1873 to December 2001. This data set's end time is December 2001.

**Key Words:** temperature precipitation data set