

随机读取标准数据研究

王丽萍

(北京气象中心)

提要

本文简述了标准数据的存贮格式，主要研究了怎样读取标准数据更加快速、灵活、方便，且节省空间。提出了计算随机访问起始记录号的两个公式和利用 WRITE 语句组装灵活多变的数据文件名等方法，并给出了一个访问程序的框图。该访问程序已应用于“地面气象年鉴整编程序”中，取得了较好的效果。

一、引言

建国以来，我国积累了大量的气象观测数据，这些数据不仅应用于中、长期天气预报，也应用于我国工业、农业、国防建设和科学研究等各个部门。充分发挥它们的经济效益和社会效益，是我们气候资料工作者的重大责任。随着计算机技术的发展，我国地面气象数据自动化加工处理的工作已基本纳入轨道。近几年来，已逐步形成一套较完整的业务软件系统，从而对数据的收集、整理、检查、标准处理、加工统计到表格输出，完全由计算机来完成。本文所介绍的随机读取标准数据，是在利用微机对地面气象资料进行自动加工、整编及对外服务中的一个关键性技术问题。

二、标准数据的存贮格式

标准数据是以无格式方式存贮的随机文件，其文件名为 WIIiiimm.Xnn。其中 W 为保留字，IIii 为区站号，mm 为标准数据的开始年份，nn 为标准数据的总年数，X 为气象要素符号。该文件以二进制数的形式存放在载体上，每个数据占两个字节，每个文件为单站多（单）年月单要素文件。一个月的气象资料的集合称为一个记录。这里的“记录”是指逻辑记录，它是不依赖于实际

环境的项的集合。在地面气象资料中，每个要素的逻辑记录长度是不一样的。如气温要素的记录长度为 484 个字节，日照要素的记录长度为 88 个字节。在 CCS-400 微机上配有的 FORTRAN-80 语言中规定一个记录的最大长度为 128 字节，这里的“记录”指的是物理记录。为了避免混淆，我们把这种记录称为“页”。这样，气温要素只需要 4 页来存放（128 的整倍数，最后一页多余部分为空白），日照要素只需 1 页来存放。尽管 IBM-PC 机上无此限制，但是，为了标准数据存贮方式的一致性，我们把 IBM-PC 机上经过标准处理后的数据也以 128 字节为一页来进行存放，这样，两种不同机型上的标准数据就达到了完全兼容。

每个要素的数据个数、字节数和页数之间的关系如下：

$$\text{数据个数} = \text{首部域} + \text{数据域}$$

$$\text{字节数} = \text{数据个数} \times 2$$

$$\text{页数} = (\text{字节数} + 128 \div 2) \div 128$$

(小数部分舍掉)

其中首部域是固定的，为 12 个数据，包括区站号、经度、纬度、年份等项目。

只知道以上信息还不够，还需知道每个记录中数据的具体存放格式，以便有目的的调用某要素的定时值、自记值、平均值或极值等数据。各要素的详细存放格式可参见国

家气象局颁发的“地面气象资料标准处理程序说明”，这里仅以温度为例说明（见附表）：

附表 气温 T(230)

日期	定时段(186)						日平均段 (31)
	02时	08时	14时	20时	最高	最低	
1	1	2	3	4	5	6	187
2	7	8	9	10	11	12	188
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	181	182	183	184	185	186	217

月 值 段

月平均	02时	08时	14时	20时	日平均	机动
序号	218	219	220	221	222	229
月极值	最高	日期1	日期2	最低	日期1	日期2
序号	223	224	225	226	227	228

从附表我们看到，温度要素的数据域共有230个数据，由上式可得到：

数据个数 = 242

字节数 = 484

页数 = 4

三、标准数据的随机读取

标准数据是用无格式方式存贮的，必须用无格式方式读取。由于无格式记录是值的序列，没有系统转换和解释，因此它的存取速度快，适用于我们这种需要大量保存中间结果的数据处理。目前，我们已有在CCS-400和IBM-PC两种微型机上实现的标准处理程序。由于这两种机器的类型不同，配备的高级语言的文本也不尽相同，故读取数据的程序有所差别，本文以IBM-PC机FORTRAN-77语言为例，来研究标准数据的随机读取问题。

我们已经知道标准数据的存贮特点是①随机存贮；②无格式；③所有记录具有相同长度；④每个记录前有一记录号。对于这

些，在编程时必须在OPEN语句和READ语句中给予指明。

目前，在我国气象部门已投入运行的一些访问程序中，存在着以下几点弊端：

1. 由OPEN语句打开的数据文件，其文件名限制了数据的站点、年代的长度和固定的要素。如：

OPEN(7,FILE = 'W5451181·T02')

换言之，含有该语句的程序只能读北京站1981—1982年的温度数据。

2. 对地面资料所包括的气压、气温、降水、风等19个要素各建立了一个访问程序，使得整个源程序既重复又冗长。

3. 对需跨年的温度、降水、地温、天气现象四个要素没有上跨功能或上跨功能不灵活。

4. 读语句中记录号的初始值限定太死。

如：READ(7, REC=1, END=55, ERR=50) IB

由此，限定了所使用的数据文件的起始年代和月份，使用很不方便。

如何解决这些弊端呢？笔者对此进行了研究，采用了下面的几种手段：

1. 利用系统提供的人机对话功能对区站号、数据的起始年代、总年数、要素项等值进行随机输入，然后用WRITE语句组装文件名（见图1）。从而在字符变量A中就形成了一个灵活多变的数据文件名。

2. 要素名既可以通过上述的随机方式获得，也可以通过DATA语句得到，如图2所示。

这样，在一个程序里，用不同的ELE_i (_i=1, 2, 3, ..., 19)去组成不同要素的数据文件名，我们就可以达到使用同一个访问程序，调用多至19个要素的标准数据的目的，避免了程序的冗余，节省了内存空间。

3. 利用人机会话功能选择所访问数据是否要跨年，并根据气候特点选择上跨的月

```

CHARACTER A*12, B, DI, ELE, II*5, NS*2
DATA B/'W'/, DI/'.'/
WRITE (*, 10)
10 FORMAT(3X, 'STATION, FILE-YEAR, ELEMENT, SUM-OF-YEAR=')
READ (*, 20) II, NW, ELE, NS
20 FORMAT (A5, 1X, I2, 1X, A1, 1X, A2)
WRITE (A, '(A1, A5, I2, A1, A1, A2)') B, II, NW, DI, ELE, NS

```

图 1

份（见图 3）。

4. 随机文件既可以顺序读也可以随机读，当我们想要访问其中某些记录而不是全部记录时。采用随机读方式则比顺序读要快得多。因此在我们的业务应用中，不必为做长年代整编建立一个数据文件，也不必为做年鉴整编而另外再建立一个文件。为此，通过总结规律，笔者建立了计算随机访问起始记录号的两个公式，有了这个公式，无论数据的年代多长，都可以根据需要从任意年代、任意月份调用，不必从头进行扫描，提高了读取速度。

公式如下：

$$N = (NUB - NW) \times (NY \times 12) + 1 \quad (1)$$

$$N = (NUB - NW - 1) \times (NY \times 12) + [NY \times (MYF - 1)] + 1 \quad (2)$$

其中 N：随机访问的起始记录号；

```

CHARACTER ELE1, ELE2, ELE3, ---
DATA ELE1/'P'/, ELE2/'T'/, ELE3/'I'/, ---

```

图 2

NW：标准数据文件的起始年份；

NUB：所调用的标准数据的年份（不含上跨的年份）；

NY：某要素一个月记录所占的页数；

MYF：上跨的月份。

公式（1）用来计算不跨年时随机访问的起始记录号；公式（2）用来计算跨年时随机访问的起始记录号，通过人机对话来获得NW、NUB和MYF的随机值。

例如，我们有文件 W5451151 · T37，

```

      WRITE (*, 30)
30 FORMAT(3X, 'WHETHER WANT ACROSS YEAR?(Y-1/N-0)')
      READ (*, 40) KN
40 FORMAT (I2)
      IF (KN, EQ, 1) THEN
          WRITE (*, 50)
50 FORMAT(3X, 'START MONTH=')
      READ (*, 40) MYF
      ELSE
          :
      ENDIF

```

图 3

即北京站从1951—1987年的温度要素的标准数据，假如我们要做1981年的年鉴，如不跨年，用公式（1），可以求得随机访问的起始记录号为：

$$N = (81 - 51) \times (4 \times 12) + 1 = 1441$$

如上跨到1980年的9月份，利用公式（2），可以求得随机访问的起始记录号为：

$$N = (81 - 51 - 1) \times (4 \times 12) + [4 \times (9 - 1)] + 1 = 1425$$

解决了以上几点弊端，下面以调用要素T、W、F为例，给出一个完整的随机读取标准数据的程序框图(见图4)。

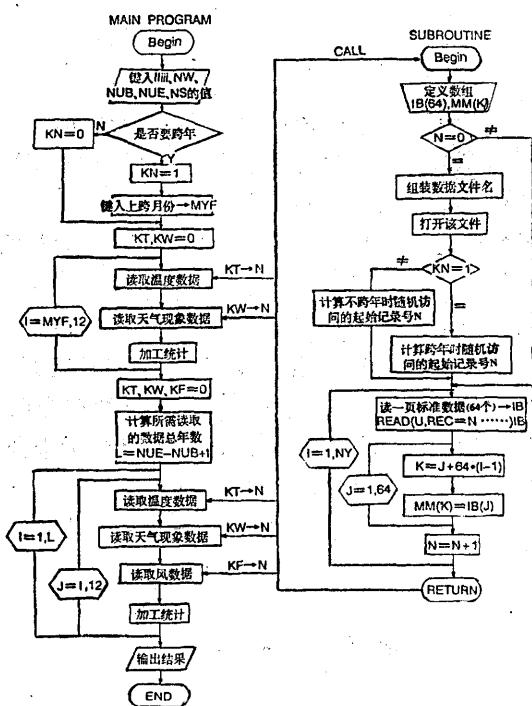


图 4

其中，NUE：所需调用的标准数据的结束年份；

MM (K)：存放标准数据的可调数组。下标 K 为哑元，其值随各要素的逻辑记录大小而定；

$KT = \begin{cases} 0: & \text{打开标准数据文件, 计算随机访问的起始记录号。} \\ KW = N: & \text{标准数据文件已打开, 直接调用记录N。} \end{cases}$

NS：标准数据的总年数。

其它变量说明同前。

如果要调用其它要素，只需在此基础上做些增改即可。

该程序可以调用标准数据文件中任意站点、从任意年代、任意月份开始的任意要素的标准数据。调出的数据可以跨年，也可以不跨年，既可以做年鉴整编，也可以做长年代资料统计及其它方面的应用，非常灵活。

在CCS-400微机上用FORTRAN-80语言访问标准数据时，程序与此大致相同，只是个别语句有所差别，现将不同之处简单说明如下：

(1) FORTRAN-77中字符型变量用语句CHARACTER说明，而FORTRAN-80中没有字符型变量，只能用字节说明语句BYTE来说明。

(2) FORTRAN-77与FORTRAN-80中OPEN语句的形式不一样，所带的变元也不相同，如物理设备说明、文件名说明、存取方式说明等。

(3) FORTRAN-80中，用ENCODE语句把变量表列中各项其内存中的数据按FORMAT语句中的规定转换成ASCII代码格式，而FORTRAN-77中，用WRITE语句即可实现。

(4) FORTRAN-80语言的读写语句中，终端设备号为“1”而FORTRAN-77中，终端设备号为“*”。

以上仅简单地列出两种型号的微机上FORTRAN语言的部分不同之处，读者可据此将IBM-PC机上的访问程序修改成CCS-400机上的访问程序。