高空电报资料压缩库系统

申学勤　谢成开

（重庆市气象局，630039）

提　要

历史天气图高空逐日资料量大、层多，存取极不方便。该文介绍一种高效压缩的高空资料库系统，将高空资料和相应的存取软件结合在一起，形成一个微机高空资料处理系统，降低了80%以上存储资料所需的介质空间，且方便了资料的查阅、提取、添加、插补等。

关键词： 高空资料　压缩　编码　软件

引　言

高空历史电报资料在天气预报以及气象科研中的地位是不言而喻的。随着资料年限的不断增加，资料的存储已成为省以下各级气象部门的困难之一。例如：重庆市有1980—1992年亚欧地区850hPa以上各层逐日资料，总量为350M字节左右，并以每年约30M字节左右的速度增加。目前存储高空资料主要用磁带和磁盘，磁带要配备专门的磁带机，大型磁带还需要在小型以上计算机上才能进行存取，软磁盘存储容量小。读取磁带和软磁盘上的资料速度慢，且使用也不方便。我们开发的高空电报资料压缩库系统，全部资料存放在微机硬盘上，经压缩后，每年亚欧地区各层逐日资料仅占5M字节左右的硬盘空间，且资料存取快速、方便。

1　系统结构及功能模块

系统包括资料添加、查阅、提取、插补、图显5个功能模块及一个主控模块。各功能模块相互独立工作，互不影响；主控模块控制各功能模块的运行与退出，提供用户界面。系统结构如图1所示。

1.1　资料查阅模块

该模块完成各层资料的查阅，且可通过菜单按钮选择，将资料输出到打印机。屏幕查阅时，站号及各要素值用不同颜色显示，使查阅者一目了然，每显示一屏自动暂停，便于用户抄录和详细对照。

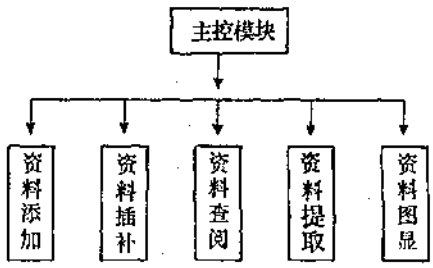

图1　系统结构图

1.2　资料添加模块

该模块含有要素提取、压缩编码、库添加3个步骤。要素提取负责报文检错，并把要素值从报文中分离出来；压缩编码是把要素值按设计的压缩编码方式重新编码，形成资料的压缩码，并对缺报站编成特定的缺报码，为以后资料插补留好存储位置；库添加是以二进制文件方式，将资料的压缩码追加到文件尾部。

1.3　资料插补模块

该模块是为保证资料的完整性而设计的。当资料库需补充资料时，可以通过该模块将所缺资料补入库中。

1.4 资料提取模块

该模块根据常规用户要求，设有整年资料提取和月份资料提取两个功能。资料提取灵活、方便，系统设有一站号文件，用户将所需资料的站号集输入该文件中，便可提取到这些站号的资料。提取到的资料，按输入的站号顺序存放在一指定文件中，并标注有资料的日期。整年提取是指提取历年的逐日资料，用户只需在菜单按钮上选择起始年份和后续年数；月份提取是指用户提取历年指定月份的逐日资料，只需在菜单按钮上选择输入月份，起始年份，后续年数。

1.5 图显模块

在查阅资料时，如需站经、纬位置，采用该模块。它在一张标有经纬度的屏幕图形上显示有站号，并对普通站、省级站、区域气象中心以不同颜色的圆圈表示。查阅资料时，在输入日期、层次、要素信息后，将鼠标指向指定站号的站圈内，执行操作，便可在信息窗内看到该站资料。

2 资料组织方式的选择

资料的组织方式影响到资料的安全性、软件的复杂程度、对微机系统的性能要求，以及使用上的方便与否等诸方面的问题。我们选择了分层次组织的方式，这种组织方式有如下特点：①处理方便，由于目前各台站对每日收到的报文都进行了分层分检，各层次的资料形成了不同的文件，这对分层资料组织提供了方便；②程序设计简单有效，每一层次的资料在同一个文件中要素齐全，时间上连续完整，在编写资料查阅、提取等软件时，只需打开一个文件，计算一次文件指针即可；③资料的安全性较好，当因磁盘故障或人为操作对某层次资料文件损坏时，只需对该层资料进行再处理，不影响其它层次资料的正常使用；④对微机硬件设备要求低，每层资料经压缩后，10 年也只占用 5M 字节左右的硬盘空间，这样，在只能配置小容量硬盘的微机上也能使用长年限的资料。相反，如将所有层次资料组织在一起，则要求硬盘容量大，微机档次也相应提高；⑤使用方便、灵活，由于分层组织方式不要求资料在层次上的完整性，各站可根据自己的需求，装配自己的资料库，不影响软件系统的使用。

3 压缩技术

在设计压缩方法时，一方面要考虑压掉尽可能多的冗余信息，另一方面要考虑能正确、简单地恢复。在高空电报资料中，主要的冗余信息有层次信息、站号、要素识别码、码组间隔以及要素值编码方式引起的多余字节数。层次信息已通过分层资料组织方式压缩掉了，其余的冗余信息我们通过记录方式选择和要素值压缩编码来解决。

3.1 记录方式选择

为了说明数据记录方式的作用，我们将数据库的结构模型示于图 2。

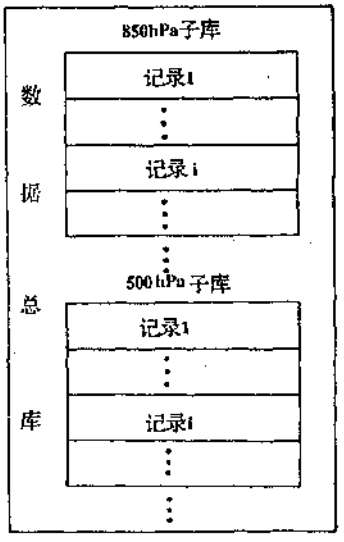

图 2 总数据库结构模型

总数据库由多个相对独立的子库组成，每个子库由许多记录组成，每一个子库中的一条记录，记载有相应层次一天所选各站的要素资料，记录号代表该记录距资料起始日期的天数。为了压缩掉站号、要素识别码及码组间隔，我们选用固定站集的定长记录方式，设 f_{ij} 代表 i 站 j 要素的资料值，则记录结构如表 1 所示。该记录结构要求入库资料的站集固定，且站序的排列也保持不变，要素顺序一致，这需要设置一张站号检索表和要素名检索表。从记录结构表中不难看出：这种记录方式可以省掉站号信息、要素识别码和码组间隔。

表 1　记录结构

f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	…	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	…

上述记录方式的不足之处是，只能将指定站集的资料入库，因此，在选择站集时应选择那些来报率较高的站，以提高资料库的有效率。在记录要素值时，为每个要素都分配了存储位置，而不管该要素是否来报。这牺牲了一些压缩效率，但为资料的插补提供了可能。

3.2　要素值的压缩编码

在高空电报资料中，所有字符采用的是 ASCII 编码，每个要素值占 3 个字节。对风速值，它的取值 255m·s^{-1}以下，方便地以二进制方式编码为一个字节；温度值进行规定的五舍六入成为整形数，因温度值在－100—＋100℃之间，对温度作加 100 处理，这样，取值范围也在一个字节的二进制表示范围内，可编码为一个字节；露点温度可将原报中的数值，直接以二进制编码为一个字节；风向值的范围是在 0—360 之间，但由于观测规范规定，它总是 5 的整数倍，可通过一个简单的线性转换，将取值范围映射到一个字的二进制码表示范围内；各层的高度值差异很大，直接采用二进制的编码方式将占用两个字节长度。经分析方知，各层的高度值变化幅度范围在一个字节的表示范围内，只要选择好一个适当的常数，在原高度值上减掉该常数后，再以二进制方式编码，可将高度值压缩为一个字节。对不同的资料层次，选择的常数不同。

当要素值缺报或为错报时，我们将它编为特定的无报码，也以一个字节表示，各要素的无报码有所不同。

4　信息恢复

4.1　层次信息的恢复

由于我们是以分层方式组织资料库的，在查阅和提取资料时，只打开相应层次的数据文件，这里已经包括了资料的层次信息。

4.2　日期信息的恢复

在记录方式中已经确定，一日资料为一个记录，因此，资料值所处的记录号便是距起始日期的总天数。设记录序号为 R_i，入库资料站数为 Z，资料值距文件头的长度为 P 字节，则有：

$$R_i = \mathrm{INT}(P/5Z) + 1 \qquad (1)$$

有了 R_i，可方便地根据各月天数，资料起始日期以来的闰年情况计算相应的年、月、日。

4.3　站号信息的恢复

在资料库中，每一个记录均记载相同站集的资料，且排列顺序不变，因此，只要求出资料值在库中距记录头的相对位置，便可查出它的隶属站号。表 2 为站号与记录坐标的对应关系，表中，S_i 为站序号，S_n 为站号，L 为距记录头的相对长度，亦记录坐标，单位为字节，它与 P 的关系为：

$$L = P - \mathrm{INT}(P/5Z)5Z \qquad (2)$$

因已知 5 个要素压缩编码后共占 5 个字节，则有：

$$S_i = \mathrm{INT}(L/5) + 1 \qquad (3)$$

根据 S_i，可从表 2 查出对应的站号。

表 2　记录坐标与站号对应关系

S_i	1	2	……	n	……
S_n	57516	56492	……	S_n	……
L	0-4	5-9	……	5(n−1)～5n−1	……

4.4　要素名信息的恢复

在一个记录中，每站资料占有相同字节数，且要素的排列顺序一致，要素序号与要素名对应关系如表 3。设要素序为 F_i，则有：

$$F_i = \mathrm{MOD}(P,5) + 1 \qquad (4)$$

表 3　要素排列顺序

要素序号	1	2	3	4	5
要素名称	高度	温度	露点	风向	风速

根据 F_i，可由表 3 查出对应的要素名。

4.5　要素值的恢复

要素值的恢复是压缩编码的逆运算，对于以二进制直接编码的要素，直接将所取字节作为一个短整形数处理，采用了转换关系的要素，再按压缩编码时的转换关系作反转换。

5　压缩效率计算

设压缩效率为 R，一个要素压缩前编码为 L 字节，压缩后编码为 K 字节；则：

$$R = (1 - \frac{K}{L}) \times 100\% \qquad (5)$$

在高空电报资料中，每个要素编码为 5 个字节，加上组间间隔，共计 6 个字节。经压缩后，每个要素编码为 1 个字节，由此可计算出 R 为：

$$R = (1 - \frac{1}{6}) \times 100\% = 83\%$$

事实上，在电报资料中还存在有站号、报头、报尾等其它信息，实际压缩效率还约高于上述计算值。

6　结语

高空电报资料压缩库系统使用了高效、简洁的方法，使大量的电报资料存放在一个硬盘上，为编制通用资料处理程序提供了可能，亦为历史资料查询、天气图相似分析等提供了方便。

A Database of Compressed Upper Air Data

Shen Xueqing　Xie Chengkai

(ChongQing City Meteorological Bureau, 630039)

Abstract

The storage and extraction of upper-air data is unconvenient due to its large amount. An efficient database system combined compressed upper-air data and software storing and drawing to form a compression and processing system of upper-air data is introduced. Efficiency of compression to upper-air data is more than 80 percent, the inquiry and extraction of upper-air data is convenient.

Key Word:　upper-air data　compression　code　software