

海洋气候资料数据集及其质量控制

马有哲 许松 李淑春 李晶

(国家气象中心, 北京 100081)

提 要

详细介绍了全球海洋气候资料数据集的处理过程及其初步的质量控制方法。

关键词: 海洋气候 数据集 质量控制

引 言

占地球面积 71% 的海洋与人类的生活息息相关, 丰富的生物和矿产, 用不尽的潮汐能源, 激励着人类开发和研究海洋, 在全球气候变化中, 海洋起着关键性的作用, 它不仅是地球的重要储热库而且是全球热量传递和再分配的基本渠道, 人类为了分析和预测地球的气候变化, 必须认识海洋并且揭示海洋与陆地、大气之间的关系, 在诸多的研究领域中海洋气象资料发挥了巨大的作用。

早在 60 年代, 美国成立了 6 人小组, 历时 5 年, 耗资 700 万美元, 对海洋资料进行了系统整理, 然后再由 NCDC 等 4 个单位联合起来, 将上亿份报告整理成 COADS, 并建立了海洋资料加工处理系统。与之相比, 我国的海洋资料加工处理水平存在着明显的差距。为了提高我国气候资料处理及应用服务的总体水平, ‘八五’ 期间, 中国气象局投资建成了“气象资料处理和应用服务系统”, 该系统收集了 1854~1995 年的全球海洋资料, 并且进行了统一的处理和初步的质量控制, 建立了全球海洋气候资料数据集, 该数据集对于全国乃至全球性的气候分析和预测, 无疑具有显著的社会和经济效益。

1 数据集的建立(见附图)

1.1 资料的收集

全球海洋资料的特点是: 数据量大, 格式多样, 内容繁杂, 目前现存的全球海洋资料可分为两类: ① 从美国 NCDC 引进的 1854~1980 年的全球海洋资料。② 从我国 GTS 系统收集的 1980~1995 年的全球海洋资料。

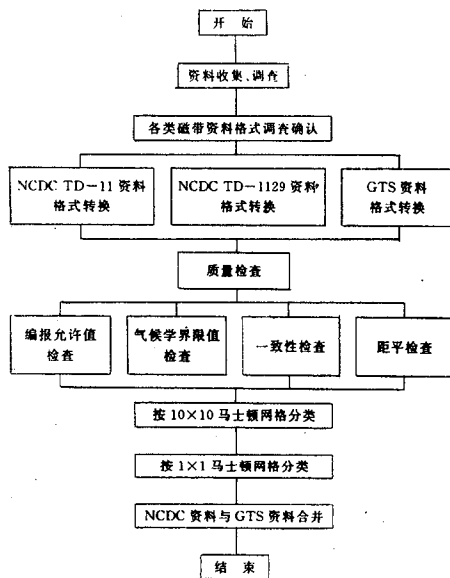


图 1 海洋气候资料数据集系统流程图

从上述的资料来源和年代可以看出, 数据集覆盖了 1854~1995 年的时空范围, 其连续性和完整性相对较好。

1.2 确定标准数据集格式

为了能全面体现海洋资料的特点,使数据格式既有充实的内容又有灵活性。为此,选择“七五”攻关课题所研制的《全球海洋气象

资料格式》作为基础,并对格式进行了校正和补充,正式确定为全球海洋气候资料的标准格式(详见表 1)。

表 1 全球海洋资料标准格式

序列	要素	字符数	序列	要素	字符数
01	报告标识	XXX		总云量 N	X
02	报告长度	XXX		低云量 Nh	X
03	资料来源	XXX	36	云 低云状 Cl	X
04	马士顿 10'×10'方格号	XXX		低云高 h	X
05	马士顿 1'×1'方格号	XX		中云状 CM	X
06	象限	X		高云状 CH	X
07	纬度	XXX	37	云质量控制	X
08	经度	XXXX	38	风浪周期	XX
09	年	XXXX	39	风浪高	XX
10	月	XX	40	风浪质量控制	X
11	日	XX	41	涌向	XX
12	时	XX	42	涌浪周期	XX
13	风向	iXX	43	涌浪高	XX
14	风速	iXXX	44	涌浪质量控制	X
15	风质量控制	X	45	第二涌向	XX
16	能见度	iXX	46	第二涌浪周期	XX
17	V 质量控制	X	47	第二涌浪高	XX
18	现在天气现象 W	XX	48	第二涌浪质量控制	X
19	W 质量控制	X	49	航向	X
20	过去天气现象 W1W2	XX	50	航速	X
21	W1W2 质量控制	X	51	船舶积冰	X
22	海平面气压 P	XXXXX	52	船舶积冰厚度	XX
23	P 质量控制	X	53	积冰速率	X
24	气压倾向 a	X	54	国家和地区代号	XX
25	a 质量控制	X	55	船舶呼号	XXXXXXXXX
26	气压变量 P	XXX	56	冰情	XXXXX
27	P 质量控制	X	57	6 小时降水量 R ₆	XXX
28	温度单位	X	58	R ₆ 质量控制	X
29	气温 T	SXXX	59	24 小时降水量 R ₂₄	XXX
30	T 质量控制	X	60	R ₂₄ 质量控制	X
31	露点 Td	SXXX	61	最高气温 T×	SXXX
32	Td 质量控制	X	62	T× 质量控制	X
33	表层海温 Tw	SXXX	63	最低气温 Tm	SXXX
34	Tw 质量控制	X	64	Tm 质量控制	X
35	海温观测法	X	65	实测云组	8XXXX
			66	特殊天气组	9XXXX

① 数据类型: EBCDIC 码, 即扩充二~十进制交换码。

② 磁带记录型式: 可变长块记录 (VB), 块长 (BL) 为 32000 字节, 记录长 (RL) 为 160 字节。

③ 记录排列方式: 月资料按马士顿方格号、日期、时次的升序排列。

1.3 重报的处理

在通讯传输过程中, 海洋天气报告重复出现的现象比较严重, 约占全部报量的 20%。如何确定重报的客观标准呢? 船舶呼

号是一个重要的依据, 如果呼号与时间相同则可判为重报; 没有呼号时确定重报要困难一些, 但是一般说来, 如果地理位置、时间和报文的内容完全相同时则可判为重报, 进行

剔除。

1.4 温度处理中的难点

通常情况下,温度以4位表示(包括符号位),但是在1854~1969年的美国资料中,温度以3位表示,如14K、01J等等,K、J表示什么呢?使用手册中没有这方面的说明,这给资料处理带来了一定的困难。后来发现,每个字母有两种含义:一是表示温度的+、-号,二是表示温度的末位数(如表2)。

表2 1854~1969年美国温度资料字母含义

	0	1	2	3	4	5	6	7	8	9
+	{	A	B	C	D	E	F	G	H	I
-	}	J	K	L	M	N	O	P	Q	R

这些字符的含义得到了确认,温度复原的问题才得以解决。

如14K = -14.2℃, 01J = -01.1℃等等。

2 初步的质量控制

由于观测、编码、发报和传输中的诸多因

素,造成海洋资料存在各种各样的错情,这些错情直接影响资料的使用价值,所以在资料处理过程中必须对有关的要素进行质量控制。

2.1 要素编报允许值和气候学界限值的检查

海洋天气报告有其自己的编报规定,要素值超过其编报允许范围则为错,如温度符号的编报,只能是0或1,如果出现其它的符号则为错,考虑到该项检查是以电码的规定为依据,所以发现的错情多数舍去,少数以不明处理。

气候学界限是指要素气候学变化的可能极端限度,例如:水温在-2℃~40℃之间变化,超过这个范围则为错。气候学界限的检查是比较粗略的,发现可疑的记录可由人工分析检查,也可由计算机进行自动控制,但不宜太复杂,能保证一个可以接受的数据质量水平就行了。

现将要素编报允许值和气候学界限值列于表3。

表3 要素编报允许值和气候学界限值

要素	编报允许值	气候学界限	要素	编报允许值	气候学界限
LaLaLa	000~999		TnTnTn		-90~60℃
LoLoLoLo	0000~1800		PPPP	百位仅允许0,8,9	880~1080hPa
Qo	1,3,5,7或(0,1,2,3,5,6,7,8)		a	0~8	
YY	1~31		PPP		-40~40hPa
GG	00~23,30~53,60~83		VV	90~99	
Iw	0,1,3,4		TwTwTw		-2~40℃
dd	00~36,51~86,99		HwHw		00~68(0.5m)
ff	00~99	<70m·s ⁻¹	dwdw	00~36,49,99,	
TTT		-90~60℃	hshs	00~50,56~99	
TdTdTd		-90~60℃	Is	1~5	
TxTxTx		-90~60℃	Rs	0~4	

2.2 一致性检查

天气报告中的各要素之间存在着或多或少的联系,利用这种关系而进行的相互配合的检查,就是一致性检查。例如:吹雪往往伴

有强风,降雪往往伴有低温,强风往往伴有大浪等等。一般说来,发现矛盾不是太困难,解决矛盾则是很困难的。例如,当 $N=0$ 而 $N_h=3$ 时,显然不合理,仅仅根据 N 和 N_h 难以

做出处理,如果再检查一下其它的要素也许能得以解决,当 $C_l = C_m = C_h = 0$ 时,便可判定 $N_h = 3$ 是错误的,可修改为 $N_h = 0$; 当 $C_l \neq 0$ 而 $C_m = C_h = 0$ 时便可判定 $N = 0$ 是错误的,可修改为 $N = 3$ 。一致性检查发现的矛盾现象,除非有充分的根据可以做出修改之外,大多数予以保留,根据错误程度给出该要素的质量控制标记(表略)。

2.3 距平检查

首先求出数列中各个变数与算术平均数的差值,然后取这个差值的绝对值,再根据这个绝对值的大小,来检查资料的可信度。

例如:检查某网格点某月的气温序列。

①求气温平均值: $T = \frac{1}{n} \sum_{i=1}^n T_i$

② 求序列中每个气温与平均值的差值的绝对值: $|d_i| = |T_i - \bar{T}|$

③ 最初可以提出这样的指标:

当 $|d_i| < 150$ 时,为允许值

当 $150 \leq |d_i| \leq 250$ 时,为可疑值

当 $|d_i| > 250$ 时,为错误值

根据数据可信度,就可以知道该数据的使用价值。需要强调一点,指标的确认,并非一次成功,可根据经验和原始资料的处理情况,进行修改,最终选择一个合适的指标。这种检查方法,适用于时空变化具有一定连续性的要素,如:气温、露点温度、水温、气压等

等。

3 数据集概况

根据多数用户的需要和先进国家的经验,海洋资料按马士顿方格号(五位数)、年、月、日、时的升序方式进行排序。

数据集覆盖了 1854~1995 年的全球海洋时空范围,共有高密度磁带 90 多盘,总存储量约为 12000MB,是一份相当完整的海洋资料。

4 结语

(1) 海洋气候变化对全球气候变化所产生的影响,越来越受到人们的重视,在气候变化研究中,海洋气候资料数据集具有重要的使用价值。

(2) 对该数据集的质量控制虽然是初步的但是效果显著,资料的质量有了明显提高。

(3) 数据集为建立海洋资料数据库和自动检索服务打下了良好的基础。

参考文献

- 1 中国气象局编译. 国际气象电码. 1970, 6.
- 2 裘国庆, 陶士伟编译. 气象观测资料的质量控制. 1993.
- 3 王伯民, 花灿华著. 气象资料的整理和统计方法. 北京: 气象出版社, 1982, 10.
- 4 MARINE DATA USERS REFERENCE. USA, 1989, 1.

Data Set of the Maritime Climate and Its Quality Control

Ma Youzhe Xu Song Li Shuchun Li Jing

(National Meteorological Center, Beijing 100081)

Abstract

The processing and methods of quality control about the Global Maritime Climate Data Set are introduced.

Key Words: Maritime Climate data set quality control