

## 海流(连续站)资料中过失误差的判别

王 骥

(国家海洋信息中心, 天津 300171)

过失误差的判别方法有多种, 作者提出, 按照样本个数的多少, 应采用不同的判别准则, 同时给出了相应的判别计算方法。

过失误差是指在观测中或在以后的资料整理中由于人为因素或其它原因(如仪器故障或恶劣的环境条件)造成的明显错误。

过失误差的判别方法很多。方国洪(1981)提出的一个准则如下: 假定海流北、东分量的余差<sup>1)</sup>服从正态分布且相互独立, 它们的均值为零, 方差为 $\sigma^2$ 。并规定全部( $n$ 个)观测值的余差的方差均不大于一个正数。

$$M = \mu^2 \sigma^2 \quad (1)$$

的概率为0.90。则当一个观测值的余差满足:

$$\Delta w^2 = \Delta u^2 + \Delta v^2 > M \quad (2)$$

时, 就认为观测值中存在过失误差。式中 $\Delta u$ 和 $\Delta v$ 分别是海流北、东分量的余差。根据定义,  $\mu$ 值应当由下式确定:

$$\left[ \frac{2}{\sqrt{2\pi}} \int_0^{\mu} e^{-t^2/2} dt \right]^n = 0.90 \quad (3)$$

由于实际上我们只能计算出样本方差, 所以式(1)中的总体方差 $\sigma^2$ 应当用样本方差( $S^2$ )代替, 即:

$$S^2 = \frac{1}{n - \lambda - 1} \sum_{i=1}^n (\Delta u_i^2 + \Delta v_i^2) \quad (4)$$

其中:

$$\lambda = 2m \quad (5)$$

这里 $m$ 是计算潮流时所取分潮的个数。式(3)中的 $\mu$ 值可由下式近似计算:

$$\mu^2 = 2.56 + 1.738 \ln(n) + 0.0096 \ln^2(n) \quad (6)$$

除上述方国洪提出的准则外, 在判别小样本海流(连续站)资料中是否存在过失误差时, 最好选用格拉布斯(Grubbs)准则(简称G准则)或t分布准则。利用G准则或t分布准则判别海流观测值中过失误差的具体步骤如下。

### 1. 资料预处理

设一次连续观测中, 于时刻 $t_1, t_2, \dots, t_n$ 测得 $n$ 个流向和流速:

收稿日期: 1990年1月10日。

1) 余差定义为在实测海流中消去潮流和平均余流之后的剩余部分。

$$\left. \begin{array}{l} \theta_1, \theta_2, \dots, \theta_n \\ u_1, u_2, \dots, u_n \end{array} \right\} \quad (7)$$

把它们分为北、东分量:

$$\left. \begin{array}{l} u_i = w_i \cos \theta_i \\ v_i = w_i \sin \theta_i \\ (i = 1, 2, \dots, n) \end{array} \right\} \quad (8)$$

## 2. 对北、东分量分别作调和分析

下面只以北分量为例给出计算步骤(东分量的分析过程与此完全相同)。令:

$$u_i = R_0 + \sum_{j=1}^3 R_j \cos(\omega_j t_i - \eta_j) \quad (9)$$

式中  $t_i$  ( $i = 1, 2, \dots, n$ ) 从观测开始日零时算起。若观测延续到下一天, 则次日的时刻应加上24。一般地说, 若观测开始日期为  $d_0$ , 以后的日期依次为  $d_1, d_2, \dots$ , 则:

$$t_i = t_1 + (d_i - d_0) \times 24 \quad (10)$$

其中  $d_k$  为第  $k$  次观测的所在日, ( $k = 0, 1, \dots$ )。式(9)中  $w_i$  取如下数值:

$$\left. \begin{array}{l} w_1 = 0.252\,934\,024 \text{ (rad/平太阳时)} \\ w_2 = 2w_1 \\ w_3 = 4w_1 \end{array} \right\} \quad (11)$$

由式(9), 对北分量可得如下方程:

$$\sum_{i=0}^6 a_{ij} x_j = f_i, \quad (i = 0, 1, \dots, 6) \quad (12)$$

式中:  $a_{00} = \pi$  (13)

$$a_{0j} = \begin{cases} \sum_{k=1}^5 \cos \omega_k t_k, & (j = 1, 3, 5) \\ \sum_{k=1}^5 \sin \omega_k t_k, & (j = 2, 4, 6) \end{cases} \quad (14)$$

$$a_{ij} = \begin{cases} \sum_{k=1}^5 \cos \omega_k t_k \cos \omega_j t_k, & (i, j = 1, 3, 5, j \geq i) \\ \sum_{k=1}^5 \cos \omega_k t_k \sin \omega_j t_k, & (i = 1, 3, 5, j = 2, 4, 6, j > i) \\ \sum_{k=1}^5 \sin \omega_k t_k \sin \omega_j t_k, & (i, j = 2, 4, 6, j \geq i) \\ \sum_{k=1}^5 \sin \omega_k t_k \cos \omega_j t_k, & (i = 2, 4, 6, j = 1, 3, 5, j > i) \end{cases} \quad (15)$$

$$a_{ij} = a_{ji}, \quad (i = 1, 2, \dots, 6, j = 0, 1, 2, \dots, i-1) \quad (16)$$

$$f_0 = \sum_{i=1}^n u_i \quad (17)$$

$$f_i = \begin{cases} \sum_{k=1}^5 u_k \cos \omega_k t_k, & (i = 1, 3, 5) \\ \sum_{k=1}^5 u_k \sin \omega_k t_k, & (i = 2, 4, 6) \end{cases} \quad (18)$$

方程(12)可用主元消去法求解。由于东分量方程的系数矩阵和北分量完全相同,所以在用主元消去法求解之前应保存 $a_{1j}$ 。这样在分析东分量时就可只计算右端项(17)和(18)。解出 $x_j$  ( $j=0, 1, \dots, 6$ )之后,由下式计算振幅和初位相:

$$R_0 = x_0 \quad (19)$$

$$\left. \begin{aligned} R_k &= \left[ x_{2k-1}^2 + x_{2k}^2 \right]^{\frac{1}{2}} \\ \eta_k &= \begin{cases} \sin^{-1}(x_{2k}/R_k), & (x_{2k-1} \geq 0) \\ \pi - \sin^{-1}(x_{2k}/R_k), & (x_{2k-1} < 0) \end{cases} \end{aligned} \right\} \quad (20) \\ (k=1, 2, 3)$$

### 3. 计算余差及其方差

经调和分析后算出 $R_0$ ,  $R_k$ 和 $\eta_k$  ( $k=1, 2, 3$ )后,用下式计算余差:

$$\Delta u_i = u_i - \left[ R_0 + \sum_{k=1}^3 R_k \cos(\omega_k t_i - \eta_k) \right] \quad (21) \\ (i=1, 2, \dots, n)$$

同时类似地算出东分量的余差 $\Delta v_i$ ,余差的样本方差由式(4)计算。

### 4. 判别 $u_i$ 和 $v_i$ 中是否存在过失误差

(1) 若 $n > 25$ ,建议采用方国洪提出的准则。计算时,可先把 $\Delta w^2$ 〔见式(2)〕按从大到小的顺序排列为:

$$\Delta w_{(1)}^2, \Delta w_{(2)}^2, \dots, \Delta w_{(n)}^2 \quad (22)$$

然后对 $\Delta w_{(1)}^2$ 作判别,〔见式(1)、(2)、(4)和(6)〕。若 $\Delta w_{(1)}^2$ 满足条件(2),则剔除 $\Delta w_{(1)}^2$ ,并对剩下的 $(n-1)$ 个 $\Delta w$ 值重新计算 $S^2$ 〔注意:式(4)中的 $n$ 应当改 $(n-1)$ 〕,然后对 $\Delta w_{(2)}^2$ 作判别。这一过程直到 $\Delta w_{(n)}^2$ 不满足条件(2)为止。此时记下以下信息:①含有过失误差的测量值的个数:

$$C_g = K - 1 \quad (23)$$

②含有过失误差的观测值所对应的时间:

$$t_{(1)}, t_{(2)}, \dots, t_{(C_g)} \quad (24)$$

当 $n > 25$ 时,也可试用肖维勒准则(Chauvenet)。这一准则的临界系数 $\mu$ 由下式近似计算:

$$\mu^2 = 0.45 + 1.187 \ln(n) + 0.1642 \ln^2(n) - 0.01956 \ln^3(n) + 0.000899 \ln^4(n) \quad (25)$$

其余步骤和前述相同。肖维勒准则规定 $n$ 次观测中出现过失误差的概率小于 $\frac{1}{2n}$ 。因而由式

(25)计算的 $\mu^2$ 值要比式(6)的结果小。也就是说,它的允许误差界限要比方国洪提出的准则小。

(2) 若 $25 \geq n \geq 20$ ,建议采用格拉布斯准则。G准则的临界系数 $\mu$ 与 $n$ 和显著性水平有关。若规定显著性水平为0.01,则 $\mu$ 值可用下式近似计算:

$$\mu^2 = -0.33 - 0.793 \ln(n) + 2.7395 \ln^2(n) - 0.65729 \ln^3(n) + 0.051088 \ln^4(n) \quad (26)$$

其余步骤和前述相同。

(3) 若  $n < 20$ , 建议采用  $t$  分布准则。这一准则与前述不同之处是, 在判别  $\Delta w^2_{(1)}$  时, 式 (4) 中就不能包含  $\Delta w^2_{(1)}$ , 同时应当用  $(n-1)$  代替  $n$ 。判别  $\Delta w^2_{(2)}$ ,  $\Delta w^2_{(3)}$ , ..., 也是这样。除此之外, 临界系数  $\mu$  与自由度和显著性水平有关, 它的近似值可用下式计算 (对应显著性水平为 0.01) :

$$\mu = \exp\{7.8403 - 8.49928 \ln(\nu) + 4.198469 \ln^2(\nu) - 0.9498411 \ln^3(\nu) + 0.08142562 \ln^4(\nu)\} \quad (27)$$

式中:  $\nu = n - \lambda - k - 1$  (28)

$\lambda$  由式 (5) 定义;  $k$  是  $\Delta w^2_{(k)}$  的序号。例如, 判别  $w^2_{(1)}$  时,  $k = 1$ ; 判别  $w^2_{(2)}$  时,  $k = 2$ ; 等等。

#### 5. 观测期间较短时, 调和项的调整

若  $T = t_n - t_1 < 24\text{h}$  时, 应考虑式 (9)、(12) 和 (21) 中的调和项选取问题。一般可采用下列原则:

- (1) 若  $18 \leq T < 24$ , 仍取  $\omega_1$ ,  $\omega_2$  和  $\omega_3$ ;
- (2) 若  $8 \leq T < 18$ , 只取  $\omega_2$  和  $\omega_3$ ;
- (3) 若  $T < 8$ , 则只取  $\omega_3$  一项。

致谢: 此项工作得到马继瑞同志的大力协助, 作者谨致衷心的感谢。

#### 参 考 文 献

- [1] 方国洪, 1981, 潮汐分析和预报的准调和分潮方法, 海洋科学集刊, 18.
- [2] 罗南星, 1984, 测量误差及数据处理, 长度计算测试丛书, 第二分册, 计量出版社.

## EXAMINATION OF ACCIDENTAL ERROR IN QUALITY CONTROL OF CURRENT DATA FROM SERIAL OBSERVATIONS

Wang Ji

(National Marine Data and Information Center, SOA, Tianjin)

#### ABSTRACT

There are several methodes of examining accidental error in quality control of current data of serial observations. It is suggested in this paper that different criteria be adopted according to the sample number. A set of corresponding calculations are advanced.