

谐波分析中波参数的误差分析

黄忠恕

(长江流域规划办公室)

提 要

谐波分析中由于采用离散数据的有限差分代替对函数的连续积分以及实际离散数据取样方式和数据性质的不同等原因,谐波参数或多或少存在一定的计算误差和拟合误差。本文着重讨论了离散化效应对截断频率谱参数的影响以及气象中各种常用离散数据的谐波参数误差情况和校正误差的方法。

一、引 言

谐波分析是富里叶谱系中最基础而又广泛应用的一种谱方法,它源于函数的富里叶级数展开。由于计算过程中以离散数据的有限差分代替对函数的连续积分以及实际离散数据取样方式和数据性质的不同等原因,致使谐波参数或多或少存在一定的计算误差和拟合误差。本文重点讨论了离散化效应对截断频率谱分量的影响以及气象中各种常用数据的谐波参数误差情况和校正误差的方法。由于计算条件(采用不同的计算方法和计算工具)造成的误差,本文不作讨论。

二、误差因子分析

函数式的富里叶级数展开是采用连续积分算式,即欧拉-富里叶公式来确定各展开项系数的,因此无计算误差问题。在展开区间,无穷级数的收敛性服从狄里希利(Dirichlet)充分条件;展开为有限项级数时,拟合误差有最小二乘方性质。

谐波分析被“展开”的是一批离散数据,一般称为时间序列,以 f_j ($j=1, 2, \dots, N$)表示。有谱展式和谐波参数算式如下:

$$f_j = \frac{a_0}{2} + \sum_{k=1}^m \left[a_k \cos \frac{2\pi k}{N} (j-1) + b_k \sin \frac{2\pi k}{N} (j-1) \right] + \varepsilon_j \quad (1)$$

$$\left. \begin{aligned} a_k &= \frac{2}{N} \sum_{j=1}^N f_j \cos \frac{2\pi k}{N} (j-1) \\ b_k &= \frac{2}{N} \sum_{j=1}^N f_j \sin \frac{2\pi k}{N} (j-1) \end{aligned} \right\} \quad (2)$$

$$\left. \begin{aligned} c_k &= (a_k^2 + b_k^2)^{\frac{1}{2}} \\ \varphi_k &= \arctg \frac{a_k}{b_k} \end{aligned} \right\} \quad (3)$$

式中 a_k 、 b_k 为富氏系数, c_k 为波幅, φ_k 为初相, k 为波数, ε_j 为误差项。

离散(数字)化产生的谱失真表现在两方面: 一是高频混淆现象, 二是谐波参数的计算误差。前者已有详细分析^[1]。根据有限差分代替连续积分的精度分析可知, 计算误差的大小与差分格距有关。对于一确定的差分格距, 不同的谐波分量其计算误差也不同。具体来说, 误差大小取决于谐波分量数据样本的多少, 即谐波的周期数 T_k ($T_k = N/k$)。 T_k 愈大, 计算误差愈小, 因而高频分量的误差总比低频分量的要大些。

离散数据的采样方式, 气象中一般有瞬时值(格点值)、平均值(时空格距间的均值)和累积值(如累积距平)几种形式(见图1)。但是, 任何谐波过程总是时间(或空间)变化上的连续周期过程, 不同的采样方式必将对谐波参数带来系统性的影响。我们关心的是这些影响到底有多大, 能否消除这些影响从而得到比较真实可靠的谱估计?

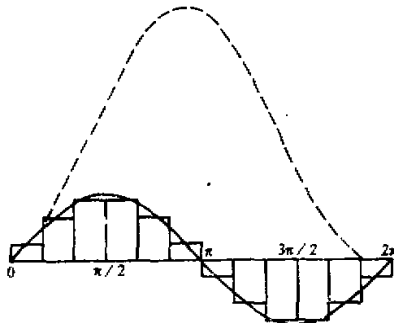


图1 谐波过程的几种离散数据

实线: 瞬时值, 直方: 平均值, 虚线: 累积值

从富氏级数展开定理可知, (2)式是在波数 k 取正整数条件下产生的, 谱频 $k\omega$ 是基频 ω ($\omega = 2\pi/N$) 的整数倍, 在数学中称这样的谱分量为正交分量。但是, 实际离散数据中可能存在的周期分量多数是非整数波分量, 应用正交分解公式(1)和(2)展开非整数波分量, 也会产生计算误差和拟合误差, 造成谱图中的“虚假成份”。本文通过对非整数波分量的个例分析, 试验了一种消除其波参数误差的分析方法。

此外, 数据中的随机误差(观测误差、噪音振动等)也能产生谱估计误差。这类误差问题比较复杂, 本文不作讨论。

三、瞬时值数据谐波参数误差分析

以单一谐波过程的瞬时值序列

$$f_j = c \sin \left[\frac{2\pi k}{N} (j-1) + \varphi_k \right] \quad (j=1, 2, \dots, N) \quad (4)$$

为例, 应用公式(1)一(3)展开 f_j 。 (4)式中谐波参数 c , φ_k 和 N 分别取定值, 分析其计算值 $\hat{c}$, $\hat{\varphi}_k$ 与给定值 c , φ_k 的误差以及谐波展开的拟合方差变化情况。结果表明, 除截断频率 ($N/k=2$) 外, 波幅和初相的计算误差一般都十分微小, 波幅误差为 $c \times 10^{-9}$, 初相误差为 10^{-5} 弧度, 它们随 T_k 的变化在一个较大范围内不超出一个量级。表1是拟合误差的方差随 T_k 和 φ_k 的变化, 与波参数的计算误差一样, 截断频率的拟合误差也最大, 在 $\varphi_k = \pi/2$ 处达到极大值。截断频率分量拟合误差的方差量级为 10^0 , 其余各频率则为 10^{-9} 。

四、截断频率分量的波幅、初相误差

截断频率亦称奈奎斯特(Nyquist)频率、折叠频率或香农(Shannon)极限, 是谱分析

的极限频率, 其频率为 $1/2 \Delta t$, 波数 $k = [N/2]$ ¹⁾, 周期为 $2\Delta t$ (Δt 为数据取样的时

表 1 谐波拟合误差的方差变化

方差 T_k	$\frac{\pi}{18}$	$\frac{\pi}{6}$	$\frac{5\pi}{18}$	$\frac{7\pi}{18}$	$\frac{\pi}{2}$	量级
2	0.030	0.250	0.587	0.883	1.000	$\times 10^0$
3	7.494	7.295	7.307	7.122	7.236	$\times 10^{-9}$
4	3.404	3.339	3.431	3.173	3.207	$\times 10^{-9}$
5	2.489	2.396	2.341	2.314	2.340	$\times 10^{-9}$
6	1.873	2.056	1.833	1.912	2.023	$\times 10^{-9}$
7	1.698	1.704	1.713	1.637	1.599	$\times 10^{-9}$
8	1.526	1.689	1.679	1.626	1.643	$\times 10^{-9}$
9	1.526	1.611	1.457	1.554	1.512	$\times 10^{-9}$
10	1.585	1.531	1.464	1.488	1.456	$\times 10^{-9}$
11	1.417	1.462	1.395	1.491	1.429	$\times 10^{-9}$

间或空间之间隔, 一般取作 1).

1. 截断频率分量波参数的多值性

可以证明, (1) 式中当展开的最大波数 $m = [N/2]$ 时谱展式收敛于序列 $f_j (j = 1, 2, \dots, N)$, 即 $\varepsilon_j \equiv 0$. 但实际计算表明, 这时对序列的拟合仍是近似的, 误差项 ε_j 为一个截断频率振动.

对 (1) 式取截断频率分量可得:

$$f_j = \frac{a_0}{2} + c_{N/2} \sin [\pi(j-1) + \varphi_{N/2}] + \varepsilon_j \quad (5)$$

$c_{N/2}$ 和 $\varphi_{N/2}$ 分别为截断频率的波幅和初相, 以 $f'_j = f_j - \frac{a_0}{2}$ (距平序列) 代入 (5) 式并化简得:

$$f'_j = (-1)^{j-1} c_{N/2} \sin \varphi_{N/2} + \varepsilon_j \quad (6)$$

即

$$(-1)^{1-j} f'_j = c_{N/2} \sin \varphi_{N/2} + (-1)^{1-j} \varepsilon_j \quad (7)$$

(6) 式两边对 j 求和最后得:

$$\overline{\Delta f} = 2 c_{N/2} \sin \varphi_{N/2} \quad (8)$$

式中 $\overline{\Delta f} = 2/N [(f'_1 + f'_3 + \dots) - (f'_2 + f'_4 + \dots)]$ 为距平序列的奇数序数据与偶数序数据的平均变差. 对于任一已知序列 $f_j (j = 1, 2, \dots, N)$, $\overline{\Delta f}$ 有定值. 因此, (8) 式表明截断频率分量的波幅和初相互为函数关系. 换言之, 任一已知序列都对应着无穷多个截断频率振动, 其波幅和初相彼此不同, 但对序列的拟合效果相同 (见图 2). 也就是说, 截断频率分量不具备其它谱分量所具有的“展开的唯一性”. 这是离散化效应的结果.

[1] $[N/2]$ 为对 $N/2$ 取整.

2. 误差分析

既然离散化效应给截断频率谱参数带来了多值解, 因而存在着如何确定比较合适的谱参数和可能存在的误差问题.

1) 波幅误差

根据三角函数的和差化积公式, 可以得到富里叶系数 a_k 、 b_k 与波参数 C_k 、 φ_k 有以下的基本关系

$$\left. \begin{aligned} a_k &= c_k \sin \varphi_k \\ b_k &= c_k \cos \varphi_k \end{aligned} \right\} \quad (9)$$

(9) 式与 (3) 式等价, 对离散序列展开的一切谐波分量都成立, 其中也包括截断频率. 但是, 差分公式 (2) 则不同, 它不适用于截断频率分量的计算, 因此会产生明显的计算误差. 证明如下:

以 $k = N/2$ 代入 (2) 式得截断频率的富氏系数

$$\begin{aligned} a_{N/2} &= \frac{2}{N} \sum_{j=1}^N f_j \cos \pi(j-1) = \frac{2}{N} \sum_{j=1}^N f_j (-1)^{j-1} \\ &= \frac{2}{N} [(f_1 + f_3 + \dots) - (f_2 + f_4 + \dots)] = \overline{\Delta f} \end{aligned}$$

根据 (8) 式得:

$$\left. \begin{aligned} a_{N/2} &= 2 c_{N/2} \sin \varphi_{N/2} \\ b_{N/2} &= 0 \end{aligned} \right\} \quad (10)$$

不难看出, 由差分公式推得的截断频率系数关系式 (10) 不符合基本公式 (3) 和 (9), 其误差是显而易见的.

2) 初相误差

谐波分量的初始位相取决于时间序列第一个观测数据与该分量变化的相对位置. 因此, 截断频率分量同其他任何分量一样, 初相值获得 $0 - 2\pi$ 之间的任意值都是可能的. 但是以差分算式 (2) 和基本关系式 (3) 得到的截断频率分量的初相仅有两种可能取值: $\pi/2$ 或 $-\pi/2$ (正弦波) 和 0 或 π (余弦波). 以正弦波为例, 当实际初相在 $0 - \pi$ 之间时, 有 $a_{N/2} > 0$ 和 $b_{N/2} = 0$, 得计算初相 $\varphi_{N/2} = \pi/2$; 当实际初相在 $\pi - 2\pi$ 之间时, 有 $a_{N/2} < 0$ 和 $b_{N/2} = 0$, 得计算初相 $\varphi_{N/2} = 3\pi/2$ (或 $-\pi/2$). 因此, 计算得到的初相是实际初相可能取值的中位数, 初相误差在 $-\pi/2 - \pi/2$ 之间.

3. 误差修正

根据以上分析, 截断频率波参数的误差修正有两重意义: 一是修正由波参数误差造成的谱展式与原序列的拟合误差, 二是在无拟合误差情况下, 求得具有最小计算误差的波参数.

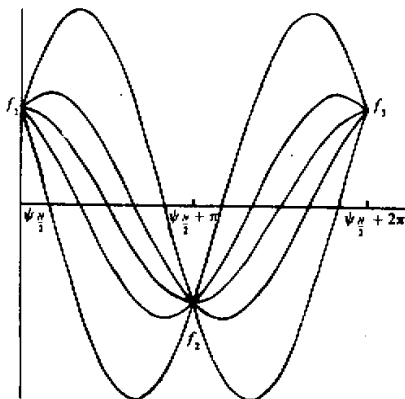


图 2 截断频率谱参数的多值性

1) 修正拟合误差

以系数 $a_{N/2}$ 为例, 下面我们证明截断频率的无拟合误差算式.

将谱展式(1)中波数 k 代之以 n , 在等式两边乘以 $\cos(2\pi k/N)(j-1)$ 后对 j 求和, 应用三角函数的正交性化简可得:

$$\sum_{j=1}^N f_j \cos \frac{2\pi k}{N}(j-1) = a_k \sum_{j=1}^N \cos^2 \frac{2\pi k}{N}(j-1) \quad (11)$$

当 $k \neq N/2$ 时, 有

$$\sum_{j=1}^N \cos^2 \frac{2\pi k}{N}(j-1) = \frac{N}{2}$$

代入(11)式可得到(2)式, 当 $k = N/2$ 时, 有

$$\sum_{j=1}^N \cos^2 \frac{2\pi k}{N}(j-1) = \sum_{j=1}^N \cos^2 \pi(j-1) = N$$

代入(11)式得:

$$a_{N/2} = \frac{1}{N} \sum_{j=1}^N f_j \cos \pi(j-1) = \frac{1}{2} \Delta f$$

同样可得:

$$b_{N/2} = \frac{1}{N} \sum_{j=1}^N f_j \sin \pi(j-1) = 0$$

最后根据(3)式得出无拟合误差的截断频率富氏系数和波参数算式如下:

$$\left. \begin{aligned} a_{N/2} &= \frac{1}{N} \sum_{j=1}^N f_j \cos \pi(j-1) \\ b_{N/2} &= 0 \end{aligned} \right\} \quad (12)$$

$$\left. \begin{aligned} c_{N/2} &= |a_{N/2}| \\ \varphi_{N/2} &= \begin{cases} \pi/2 & a_{N/2} > 0 \\ -\pi/2 & a_{N/2} < 0 \end{cases} \end{aligned} \right\} \quad (13)$$

2) 修正波参数误差

根据上述截断频率多值性的分析可知, 无拟合误差的波参数算式(12)和(13)是多值解公式(8)中的一组特解, 这组特解的初相是无穷多组解初相的中位数. 由于特解的 $|\sin \varphi_{N/2}|$ 有极大值, 因而波幅 $c_{N/2}$ 的特解值是无穷多组解中的极小值. 因此, 应用这组特解作为截断频率的波参数是不妥的, 因它与实际波参数的平均误差最大. 我们认为, 在无法确定真实波参数的情况下, 以平均误差最小为条件选定波参数的特解是可取的. 为此, 作者建议用 $\sin \varphi$ 在 $0 - \pi$ (或 $\pi - 2\pi$) 之间的平均值所对应的位相为截断频率的初相.

因为

$$\frac{1}{\pi} \int_0^{\pi} \sin \varphi d\varphi = \frac{2}{\pi} \quad \frac{1}{\pi} \int_{\pi}^{2\pi} \sin \varphi d\varphi = -\frac{2}{\pi}$$

所以

$$\overline{\varphi} = \arcsin \left(\pm \frac{2}{\pi} \right)$$

代入(8)式,最后得到既无拟合误差且波参数的平均误差又极小的截断频率谱估计值为

$$\left. \begin{aligned} \bar{\varphi}_{N/2} &= \arcsin \left(\pm \frac{2}{\pi} \right) \\ \bar{c}_{N/2} &= \frac{1}{4} \pi \Delta \bar{f} \end{aligned} \right\} \quad (14)$$

五、平均值数据波参数误差分析

天气气候分析中常用时间(或空间)平均值数据或滑动平均值数据来达到消除噪音和高频振动、突出低频分量的目的。但是,这样做的结果对低频分量的强度也有一定的削弱作用。Malin 和 Chapman^[2]曾研究了大量平均值数据的谐波展开,得出实际波幅 c 和平均值数据的计算波幅 c_r 之间有以下关系:

$$c = c_r \frac{\pi / T_k}{\sin (\pi / T_k)} \quad (15)$$

上式右边的分式称为校正因子,起着修正误差的作用。从(15)式还可知, $T_k = 2$ 是谐波分析的最小周期,此时的校正因子有最大值为 1.571。当 $T_k \rightarrow \infty$ 时,校正因子的极限值为 1,表示波幅不受平均的影响。实际上,当 $T_k > 13$ 时,修正值不足 1%,校正因子的作用已经不显著了。

本文通过各种滑动平均序列的谐波展开,并根据文献[3]的证明得到实际波幅 c 与滑动平均序的计算波幅 C_T 之间满足以下关系

$$C = C_T \frac{\pi L / T}{\sin (\pi L / T)} \quad (16)$$

式中 L 为平滑步长, T 为具有平滑波幅 C_T 的谐波周期。比较(15)和(16)式,可知平均值数据与滑动平均数据有着极其相似的校正因子,滑动平均数据中 T/L 起着谐波周期 T_k 的作用。与平滑步长相等的周期分量经过平滑后完全消失,校正因子值趋于 ∞ 。大于平滑步长的周期分量,在一个相当长的波段内也会受到明显削弱,例如当 $T/L = 3.0$ 时,平滑的校正因子值为 1.209,平滑后的波幅仅为原波幅的 82.7%。应用(16)式可以对任意滑动平均值序列谐波分量的波幅进行修正。

六、累积值数据谐波参数误差分析

与平均值数据不同,累积值数据可以使低频分量得以放大而高频分量则受到抑制,因此,在中长期水文气象预报中经常应用累积距平序列。图 3 是波幅 $c_k = 1.0$, 周期 $T_k = 12$, 初相 $\varphi_k = 0, \pi/2, \pi$ 和 $3\pi/2$ 的单一正弦波过程及其累积值过程。可以证明,单一谐波过程的累积值过程仍为单一谐波过程,其周期不受影响,而波幅、初相和平衡位置都有明显变化。

设谐波过程为:

$$f(t) = c_k \sin \left(\frac{2\pi}{T_k} t + \varphi_k \right) \quad (17)$$

c_k 、 T_k 和 φ_k 分别为波幅、周期和初相,其平衡位置(即零波)为 0。则它的累积过程

为:

$$f(t_n) = \sum_{t=t_0}^{t_n} C_k \sin \left(\frac{2\pi}{T_k} t + \varphi_k \right) \quad (t_n = t_0, t_1, t_2, \dots) \quad (18)$$

当 $\Delta t = t_n - t_{n-1} \rightarrow 0$ 时, 累积过程可写为:

$$\begin{aligned} f(t_n) &= \int_{t_0}^{t_n} C_k \sin \left(\frac{2\pi}{T_k} t + \varphi_k \right) dt \\ &= -\frac{C_k T_k}{2\pi} \cos \left(\frac{2\pi}{T_k} t + \varphi_k \right) \Big|_{t_0}^{t_n} \\ &= \frac{C_k T_k}{2\pi} \left[\cos \varphi_k - \cos \left(\frac{2\pi}{T_k} t_n + \varphi_k \right) \right] \\ &= \frac{C_k T_k}{2\pi} \cos \varphi_k - \frac{C_k T_k}{2\pi} \cos \left(\frac{2\pi}{T_k} t_n + \varphi_k \right) \\ &= \frac{C_k T_k}{2\pi} \cos \varphi_k + \frac{C_k T_k}{2\pi} \sin \left(\frac{2\pi}{T_k} t_n + \varphi_k - \frac{\pi}{2} \right) \\ &= f_\theta + C_\theta \sin \left(\frac{2\pi}{T_k} t_n + \varphi_\theta \right) \end{aligned} \quad (19)$$

累积过程(19)式仍是一正弦过程, 但平衡位置 $f_\theta = (C_k T_k / 2\pi) \cos \varphi_k$, 波幅 $C_\theta = (C_k T_k / 2\pi)$, 初相 $\varphi_\theta = \varphi_k - \pi/2$ 都与原过程不同, 仅周期 T_k 保持不变. 从图3 和(19)式可以

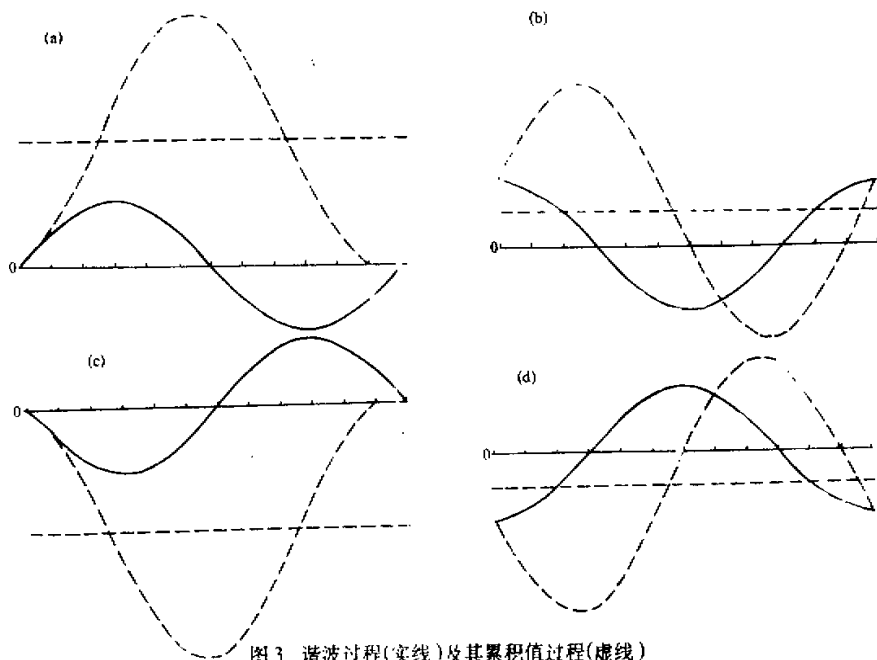


图3 谐波过程(实线)及其累积值过程(虚线)
a: $\varphi_k = 0$; b: $\varphi_k = \frac{\pi}{2}$; c: $\varphi_k = \pi$; d: $\varphi_k = \frac{3}{2}\pi$

清楚看出, 累积值数据谐波参数的变化与实际谐波过程的波参数紧密相关. 例如图 4 是图 3a 的谐波过程当 $T_k = 2-12$ 时累积值过程的波幅 C_θ 和初相 φ_θ 的演变曲线, 当 C_k 不变时, 累积过程的波幅 C_θ 与周期 T_k 呈线性关系. 当 $T_k = 2\pi \approx 6$ 时, 累积过程的波幅 C_θ 与实际波幅 C_k 相等; 当 $T_k < 6$ 时发生抑制而 $T_k > 6$ 时产生放大.

此外, 累积过程的零波(平衡位置)也与 C_k 、 T_k 和 φ_k 有关.

七、非整数波分量波参数的误差分析

根据谐波展开式(1)和(2)得到的谐波分量具有相互正交的性质, 因为谱展式是在波

数 k 取正整数的条件下产生的. 正交分量的

展开顺序是任意的, 分量间不产生相互影响.

但是, 整数波的假定不符合实际观测

数据所反映的自然周期振动. 实际上, 在展

开区间存在正交分量是不多的. 而大量的

为非整数波分量. 应用一些改进的谐波分析

方法得到非整数波已被广泛采用^[4], 但由于

非整数波之间的相互影响, 容易导致波参

数的误差和谐失真现象. 本文通过对非整数

波分量的计算表明, 分量间相互影响所产生

的计算误差和拟合误差非常显著, 波幅的加

强或减弱, 初相的超前或落后, 取决于非整

数波分量之间周期长度和初相的相互配置,

变化十分复杂. 修正其误差比以上几种情况

要困难得多.

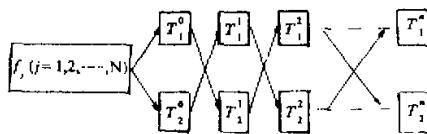
通过以上的各种误差分析可以看到, 波参数中比较稳定的是周期, 最不稳定的是波幅和初相. 为此, 作者提出一种保持非整数波的周期不变, 反复消除波幅或初相误差的逐步近似算法, 可以达到修正非整数波参数误差的目的. 下面以两个非整数波分量的叠加序列为例, 说明这种方法的计算步骤及其修正情况.

给定两个正弦波分量如下: 第一个正弦波分量周期 $T_1 = 30$, 波幅 $C_1 = 1.0$, 初相 $\varphi_1 = 0$; 第二个正弦波分量 $T_2 = 40$, $C_2 = 1.0$, $\varphi_2 = 0$. 序列样本 $N = 46$, 因而它们对于序列都是非整数波分量, 即:

$$f_j = \sin \frac{2\pi}{30} (j-1) + \sin \frac{2\pi}{40} (j-1)$$

图5的计算步骤如下:

(1) 首先, 应用非整数波计算方法直接从原序列中分别计算出两个非整数波分量 T_1 和 T_2 的波参数, 在图 5 中分别标以 T_1^0 和 T_2^0 , 上角标 0 表示它们相互间的影响在



这一步未作修正。

(2) 第一次修正。从原序列中分别扣除另一个影响分量的计算值 T_1^0 和 T_2^0 后再作“非整”分析, 分别得到第一次修正值 T_2^1 和 T_1^1 。上角标 1 表示第一次修正。可知 T_1^1 、 T_2^1 与 T_1^0 、 T_2^0 不同。

(3) 第二次修正。作法同第一次修正, 但从原序列中分别扣除的影响分量为 T_1^1 和 T_2^1 。它们是上一步的计算结果, 得到的修正值记为 T_2^2 和 T_1^2 。同样, T_1^2 、 T_2^2 也不同于 T_1^1 、 T_2^1 。如此重复下去, 得到第 n 次修正值 T_1^n 和 T_2^n 。

用这种方法对上述有两个非整数波分量的序列波幅修正结果见图 6, 横坐标为修正次数, 纵坐标为波幅。第 0 次修正得到 T_{30}^0 的波幅为 1.60, 初相为 $-28^\circ 44'$, T_{40}^0 的波幅为 1.57, 初相为 $31^\circ 46'$, 可见它们与原分量的偏差是显著的。经过消除相互影响的计算后, 第一次修正得 T_{30}^1 的波幅为 1.32, T_{40}^1 的波幅为 0.40。经过多次修正, 两个非整数波的波幅各自趋近于一个极限值: T_{30} 的为 0.75, 尚有 25% 的误差; T_{40} 的为 1.00, 与原值完全吻合。实际应用中, 上述修正计算不可能也不必要太多, 需要考虑修正到哪一步比较合适。从上述实例分析, 误差的修正量都是最初几步最显著, 随着修正次数增加, 修正量愈来愈小。以波幅叠加值曲线(见图 6 中 $C_{30} + C_{40}$ 曲线)看, 其极大值点对应于第 4 次修正, 以后的修正意义不大, 到此可以终止计算。

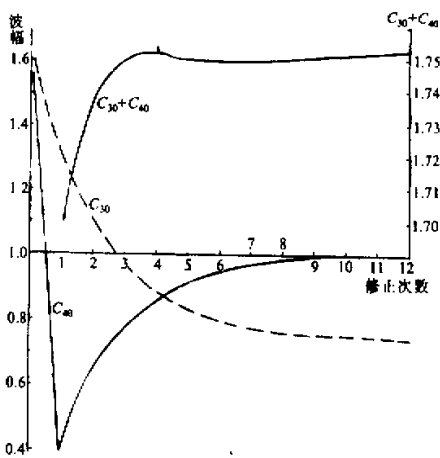


图 6 非整数波参数逐步近似计算结果

以上消除非整数波分量波参数误差的逐步近似算法原则上也适合于两个以上的非整数波分量, 但计算量将随分量的增加而增大。

参 考 文 献

- [1] 巴特, 1978, 地球物理学中的谱分析, 地震出版社。
- [2] Malin, S. R. C. and Chapman, S., 1970, The determination of lunar daily geophysical variations by the Chapman-Miller method, *Geophys. J. R. Astron. Soc.*, vol. 19, p. 15—35.
- [3] 黄嘉佑, 李黄, 1984, 气象中的谱分析, 气象出版社。
- [4] 黄忠恕, 1983, 波谱分析方法及其在水文气象学中的应用, 气象出版社。