

确定 Fourier 展开式最佳截断项数的探讨*

修艾军 廖洞贤

(国家气象局国家气象中心)

提 要

本文用理论估计及数值实验两种方法研究了有限调和和分析中考滤资料误差情况下如何确定 Fourier 展开式最佳截断项数的问题。理论估计表明最佳截断项数与物理量本身的变化及资料的误差有关。数值实验的结果证实了这一点。

一、引 言

如今 Fourier 级数在气象上得到广泛的应用。然而在实际中能得到的是函数在有限个点上的值,因此只能对这些离散值做有限调和和分析。由于计算条件的限制以及出于物理上的考虑,使得人们往往不是取最大可能的项数进行计算而是截取展开式的前 n 项作近似。加之所得到的资料中不可避免地包含有误差,因而使得我们在实际中所能得到的 Fourier 展开式的值中包含有各种误差(见图 1),由此引出以下几个问题:

1. 对于不包括误差的真实场,用带有误差的资料作谱展开,其截断项数 n 为最佳的客观标准是什么?

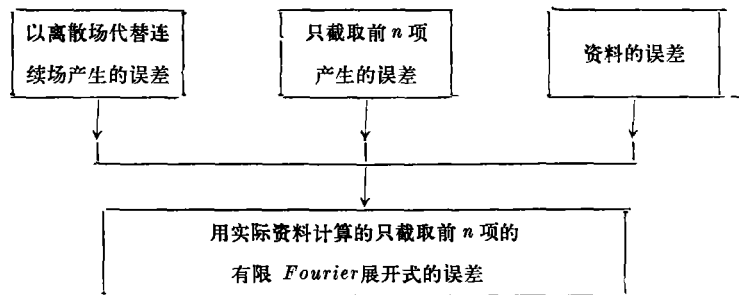


图 1 误差的来源

2. n 是不是越大越好? 怎样才能在保证精度的前提下尽可能地减少工作量?
3. 资料误差在展开式中起什么作用? 它们是怎样影响最佳截断项数的?

对第二个问题, 华罗庚、王元^[1]提出并不是截取的项数越多越好,但他们没有对此作定量分析。马淑芬^[2]曾利用有次序的取冒险函数极小的方法解决上述三个问题,得到了许多有意义的结果,但文章中确定最佳截断项数的方法不够直观,从中无法得知资料误差所起的作用。

今天谱方法(其中必须进行有限调和和分析)的广泛应用使得上述问题越来越显示出它

本文于1988年1月25日收到。

们的重要性。Simmons^[3]曾用 ECMWF 的谱模式做过实验,从其文中图 12 可以看到有时候截取项数多的效果反而比截取项数少的要差。Manabe 等人^[4]用 GFDL 的谱模式研究了谱截断对气候模拟的影响。虽然总的说来分辨率越高模拟效果越好,但也有许多例外,如对海平面气压而言,截断项数少可使副高的位置与实际更接近,使副极地低压带得以保留。本文就是试图从最简单的一维、二维 Fourier 展开式入手解决上述三个问题。

二、理论估计式

1. 一维的情形

考虑一维物理量 $F(x)$, 由观测得到它在 $N (= 2n' + 1)$ 个点上的值为 $f_{ob}(x_k)$, ($k = 0, 1, \dots, N-1$), $x_k = \frac{2\pi k}{N}$, 而在这些点上该物理量的真值为 F_k , 并且 $F_k = f_{ob}(x_k) + E_k$, 其中 E_k 为误差。

设实函数 $F(x)$ 的周期为 2π , 满足狄里希利条件, 则它可以展成 Fourier 级数:

$$F(x) = A_0 + \sum_{m=1}^{\infty} (A_m \cos mx + B_m \sin mx) \quad (1)$$

其中

$$\begin{cases} A_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(x) dx \\ A_m = \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \cos mx dx \\ B_m = \frac{1}{\pi} \int_{-\pi}^{\pi} F(x) \sin mx dx \end{cases} \quad (2)$$

利用有限个离散值 F_k ($k = 0, 1, \dots, N-1, N = 2n' + 1$) 可以近似地把 $F(x)$ 展成有限 Fourier 级数:

$$\bar{F}(x) = a_0 + \sum_{m=1}^{n'} (a_m \cos mx + b_m \sin mx) \quad (3)$$

其中由最小二乘法原理, (3) 式中的系数应取成

$$\begin{cases} a_0 = \frac{1}{N} \sum_{k=0}^{N-1} F_k \\ a_m = \frac{2}{N} \sum_{k=0}^{N-1} F_k \cos \frac{2\pi km}{N} \\ b_m = \frac{2}{N} \sum_{k=0}^{N-1} F_k \sin \frac{2\pi km}{N} \end{cases} \quad (4)$$

由上式可以看出, $b_{n'}$ 总等于零, 故只须算 $n' + 1$ 个 a_m (包括 a_0), $n' - 1$ 个 b_m 。如只截取 (3) 式的前 n 项, 则

$$\bar{F}_n(x) = a_0 + \sum_{m=1}^n (a_m \cos mx + b_m \sin mx) \quad (5)$$

而系数仍由 (4) 式确定, 只有这样才能保证使 $[F_k - \bar{F}_n(x_k)]^2$ 达最小^[1]。

由于在实际工作中只能得到 $f_{ob}(x_k)$, 所以利用 $f_{ob}(x_k)$ 得到带有误差的 $F(x)$ 的 Fourier 展开式

$$\tilde{F}(x) = \tilde{a}_0 + \sum_{m=1}^{n'} [\tilde{a}_m \cos mx + \tilde{b}_m \sin mx] \quad (6)$$

由 $f_{ob}(x_k) = F_k - E_k$ 又有

$$\tilde{F}(x) = a_0 - e_0 + \sum_{m=1}^{n'} [(a_m - e_{1m}) \cos mx + (b_m - e_{2m}) \sin mx] = \bar{F}(x) - \bar{E}(x) \quad (7)$$

其中
$$\bar{E}(x) = e_0 + \sum_{m=1}^{n'} (e_{1m} \cos mx + e_{2m} \sin mx) \quad (8)$$

$$\begin{cases} e_0 = \frac{1}{N} \sum_{k=0}^{N-1} E_k \\ e_{1m} = \frac{2}{N} \sum_{k=0}^{N-1} E_k \cos \frac{2\pi km}{N} \\ e_{2m} = \frac{2}{N} \sum_{k=0}^{N-1} E_k \sin \frac{2\pi km}{N} \end{cases} \quad (9)$$

如只截取(6)式的前 n 项, 则

$$\tilde{F}_n(x) = \tilde{a}_0 + \sum_{m=1}^n [\tilde{a}_m \cos mx + \tilde{b}_m \sin mx] = \bar{F}_n(x) - \bar{E}_n(x) \quad (10)$$

同样(10)式中的系数仍由(4)及(9)式确定。

设 $\sigma_1 = \sum_{i=0}^{N-1} [F_i - \tilde{F}_n(x_i)]^2$ F_i 为真值。 σ_1 表用实际资料计算的只截取前 n 项的有限 Fourier 展开式在所有格点上误差平方和。如果没有资料误差并且展开式取满 n' 项, 那么 σ_1 为零。因此 σ_1 与 n 及资料误差有关。

由(10)式可以推出

$$\sigma_1 = \sum_{i=0}^{N-1} \{ [F_i - \bar{F}_n(x_i)]^2 + 2[F_i - \bar{F}_n(x_i)] \cdot \bar{E}_n(x_i) + [\bar{E}_n(x_i)]^2 \} \quad (11)$$

利用离散情况下的三角函数正交关系式有:

$$\begin{aligned} \sum_{i=0}^{N-1} [F_i - \bar{F}_n(x_i)] \cdot \bar{E}_n(x_i) &= 0 \\ \sum_{i=0}^{N-1} [\bar{E}_n(x_i)]^2 &= N e_0^2 + \frac{N}{2} \sum_{m=1}^n (e_{1m}^2 + e_{2m}^2) \end{aligned} \quad (12)$$

由 Cauchy 不等式 $(\sum_{k=1}^n a_k b_k)^2 \leq \sum_{k=1}^n a_k^2 \cdot \sum_{k=1}^n b_k^2$ 便有

$$\begin{aligned} \sum_{i=0}^{N-1} [\bar{E}_n(x_i)]^2 &= N e_0^2 + \frac{N}{2} \sum_{m=1}^n \left[\left(\frac{2}{N} \sum_{k=0}^{N-1} E_k \cos \frac{2\pi k}{N} \right)^2 + \left(\frac{2}{N} \sum_{k=0}^{N-1} E_k \sin \frac{2\pi k}{N} \right)^2 \right] \\ &\leq N e_0^2 + \frac{N}{2} \cdot \frac{4}{N^2} \sum_{m=1}^n \left\{ \sum_{k=0}^{N-1} E_k^2 \cdot \left[\sum_{k=0}^{N-1} \left(\cos^2 m \frac{2\pi k}{N} + \sin^2 m \frac{2\pi k}{N} \right) \right] \right\} \\ &= N e_0^2 + 2n N \sigma_E^2 \end{aligned}$$

其中 $\sigma_E^2 = \frac{1}{N} \sum_{k=0}^{N-1} E_k^2$ 为资料误差之方差。

如资料误差是随机的则 $e_0 = 0$, 此时

$$\sum_{i=0}^{N-1} [\bar{E}_n(x_i)]^2 \leq 2nN\sigma_E^2 \quad (13)$$

根据文献[1]§ 8 中的定理, 设 $F(x)$ 在 $[-\pi, \pi]$ 中有二阶连续微商, $F''(x)$ 亦有周期 2π , 且 $|F''(x)| \leq M$, 则

$$|F(x) - \bar{F}_n(x)| \leq 2M \left(\frac{1}{n'} + \frac{1}{n} \right) \quad (14)$$

因此从(11)式有

$$\sigma_1 \leq N \left\{ 4M^2 \left(\frac{1}{n'} + \frac{1}{n} \right)^2 + 2n\sigma_E^2 \right\} \quad (15)$$

上式即为 σ_1 的定性估计式。其中不等号右边第一项表示只截取前 n 项所造成的误差平方和, 第二项表示资料误差展开式截取前 n 项后的平方和。第一项与截断项数 n 成反比, 而第二项与 n 成正比, 因而必存在一个最佳截断项数 n_{op} , 使 σ_1 达到最小。从(15)式可看出 n_{op} 与表示函数本身变化的量 M 和总的点数或分辨率成正比, 与资料误差的方差成反比。物理量值的波动起伏变化越大, 分辨率越高, 越要截取更多的波逼近它。同时, 当资料的精确度较低时截取过多的项数只会把更多的资料误差当作合理的成分考虑进去, 使展开式的精度下降。所以不能以为靠截取更多项数的办法可以提高展开式的精度, 这样做只会适得其反。这与文献[2]利用统计方法得出的结论是一致的。

如果资料误差包括系统误差和随机误差, 即 $e_0 \neq 0$, 这时

$$\sigma_1 \leq N \left\{ 4M^2 \left(\frac{1}{n'} + \frac{1}{n} \right)^2 + e_0^2 + 2n\sigma_E^2 \right\} \quad (16)$$

此时前述定性结果不变。

再利用正交关系式, (11)式右端第一项可以写成

$$\begin{aligned} & \sum_{i=0}^{N-1} [F_i - \bar{F}_n(x_i)]^2 \\ &= \sum_{i=0}^{N-1} F_i^2 - 2N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^n (a_m^2 + b_m^2) \right] + N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^n (a_m^2 + b_m^2) \right] \\ &= \sum_{i=0}^{N-1} F_i^2 - N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^n (a_m^2 + b_m^2) \right] \end{aligned}$$

由离散形式的巴什瓦定理(附录1)有:

$$\sum_{i=0}^{N-1} F_i^2 = N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^{n'} (a_m^2 + b_m^2) \right]$$

则

$$\sum_{i=0}^{N-1} [F_i - \bar{F}_n(x_i)]^2 = \frac{N}{2} \sum_{m=n+1}^{n'} (a_m^2 + b_m^2) \quad (17)$$

假设资料误差是随机的, 由(12)式有

$$\sum_{i=0}^{N-1} [\bar{E}_n(x_i)^2] = \frac{N}{2} \sum_{m=1}^n (e_{1m}^2 + e_{2m}^2)$$

$$\therefore \sigma_1 = \frac{N}{2} \sum_{m=n+1}^{n'} (a_m^2 + b_m^2) + \frac{N}{2} \sum_{m=1}^n (e_{1m}^2 + e_{2m}^2) \quad (18)$$

上式等号右边第一项随 n 的增大而减小, 第二项随 n 的增大而增大, 这与前述定性结果一致。由(18)式就可以对给定的函数值及误差值定量地计算出不同 n ($n \leq n'$) 时的 σ_1 值, 并根据其最小值确定出最佳截断项数。

如采用复数形式的 Fourier 展开式亦有类似的结果(证明见附录 2):

$$\sigma_1 = N \sum_{m=n}^{N-1} |\bar{c}_m|^2 + N \sum_{m=0}^{N-1} |\bar{c}_{Em}|^2 \quad (19)$$

其中 \bar{c}_m 和 \bar{c}_{Em} 分别为物理量和误差的展开系数。

2. 二维的情形

考虑二维实函数 $F(x, y)$, 它可以展成二维 Fourier 级数

$$F(x, y) = \sum_{i=0}^{\infty} \sum_{m=0}^{\infty} c_{im} e^{i(i x + m y)} \quad (20)$$

$$c_{im} = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} F(x, y) e^{-i(i x + m y)} dx dy \quad (21)$$

但在实际中只能得到函数在有限个点上的离散值。设在 x, y 方向上各有 L, M 个格点, $F(x, y)$ 在这些点上的值为 f_{jk} ($j = 0, 1, \dots, L-1$; $k = 0, 1, \dots, M-1$), 利用它们可以近似地把 $F(x, y)$ 展成有限 Fourier 级数

$$f(x, y) = \sum_{i=0}^{L-1} \sum_{m=0}^{M-1} \bar{c}_{im} e^{i(i x + m y)} \quad (22)$$

$$\bar{c}_{im} = \frac{1}{LM} \sum_{j=0}^{L-1} \sum_{k=0}^{M-1} f_{jk} e^{-i 2\pi (i j / L + m k / M)} \quad (23)$$

如在 x, y 方向分别截取 \bar{L}, \bar{M} 项, $\bar{L} \leq L, \bar{M} \leq M$, 则

$$\bar{f}(x, y) = \sum_{i=0}^{\bar{L}-1} \sum_{m=0}^{\bar{M}-1} \bar{c}_{im} e^{i(i x + m y)} \quad (24)$$

\bar{c}_{im} 仍由(23)式决定。

然而实际得到的值只是 f_{0b} , $f = f_{0b} + E$, E 为误差。若在展开式中 x, y 方向分别截取 \bar{L}, \bar{M} 项则在实际资料进行的 Fourier 展开式为

$$\tilde{f}(x, y) = \sum_{i=0}^{\bar{L}-1} \sum_{m=0}^{\bar{M}-1} \bar{c}_{im} e^{i(i x + m y)} = \bar{f}(x, y) - \bar{E}(x, y) \quad (25)$$

$$\bar{E}(x, y) = \sum_{i=0}^{\bar{L}-1} \sum_{m=0}^{\bar{M}-1} \bar{c}_{Em} e^{i(i x + m y)} \quad (26)$$

其中

$$\bar{c}_{im} = \bar{c}_{im} - \bar{c}_{Em} \quad (27)$$

$$\bar{c}_{Elm} = \frac{1}{LM} \sum_{j=0}^{L-1} \sum_{m=0}^{M-1} E_{jk} e^{-i2\pi(\frac{Lj}{L} + \frac{mk}{M})} \quad (28)$$

定义

$$\sigma_1 = \sum_{j=0}^{L-1} \sum_{k=0}^{M-1} [f_{jk} - \bar{f}_{jk}]^2 \quad (29)$$

由(25)有

$$\sigma_1 = \sum_{j=0}^{L-1} \sum_{k=0}^{M-1} \{ [f_{jk} - \bar{f}_{jk}]^2 + 2[f_{jk} - \bar{f}_{jk}] \cdot \bar{E}_{jk} + \bar{E}_{jk}^2 \}$$

利用正交关系式有

$$\begin{aligned} & \sum_{j=0}^{L-1} \sum_{k=0}^{M-1} [f_{jk} - \bar{f}_{jk}]^2 \\ &= \sum_{j=0}^{L-1} \sum_{k=0}^{M-1} f_{jk}^2 - LM \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{lm}|^2 \end{aligned}$$

利用巴什瓦定理

$$\sum_{j=0}^{L-1} \sum_{k=0}^{M-1} [f_{jk} - \bar{f}_{jk}]^2 = LM \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{lm}|^2 - LM \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{lm}|^2$$

同时

$$\sum_{j=0}^{L-1} \sum_{k=0}^{M-1} (f_{jk} - \bar{f}_{jk}) \cdot \bar{E}_{jk} = 0$$

且

$$\sum_{j=0}^{L-1} \sum_{k=0}^{M-1} \bar{E}_{jk}^2 = LM \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{Elm}|^2$$

$$\therefore \sigma_1 = LM \left[\underbrace{\sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{lm}|^2}_{A} - \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{lm}|^2 + \underbrace{\sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\bar{c}_{Elm}|^2}_{B} \right] \quad (30)$$

上式 A 项随 \bar{L}, \bar{M} 的增大而减小, B 项随 \bar{L}, \bar{M} 的增大而增大, 因而必存在最佳的 \bar{L}, \bar{M} 使 σ_1 达最小, 这与一维的情况是类似的。由(30)式可对给定的函数值和误差值定量地计算出不同 \bar{L}, \bar{M} ($\bar{L} \leq L, \bar{M} \leq M$) 时的 σ_1 值, 并在 σ_1 达最小的条件下定出最佳截断项数。

三、数值实验结果

1. 用理想函数作的实验

取一形式如下的理想函数

$$f(x) = \begin{cases} 550 + 50 \sin\left(\frac{9}{2}x\right) & 0 \leq x < \frac{4}{9}\pi \\ 550 + 30 \sin\left[9\left(x - \frac{9}{4}\pi\right)\right] & \frac{4}{9}\pi \leq x < \frac{6}{9}\pi \\ 550 + 50 \sin\left[\frac{9}{2}\left(x - \frac{6}{9}\pi\right)\right] & \frac{6}{9}\pi \leq x < \frac{10}{9}\pi \\ 550 + 30 \sin\left[9\left(x - \frac{10}{9}\pi\right)\right] & \frac{10}{9}\pi \leq x < \frac{12}{9}\pi \\ 550 + 50 \sin\left[\frac{9}{2}\left(x - \frac{12}{9}\pi\right)\right] & \frac{12}{9}\pi \leq x < \frac{16}{9}\pi \\ 550 + 30 \sin\left[9\left(x - \frac{16}{9}\pi\right)\right] & \frac{16}{9}\pi \leq x \leq 2\pi \end{cases}$$

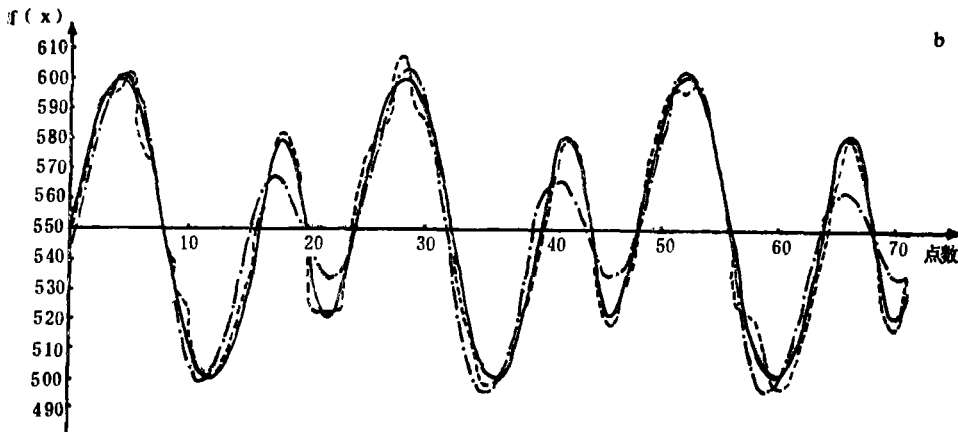
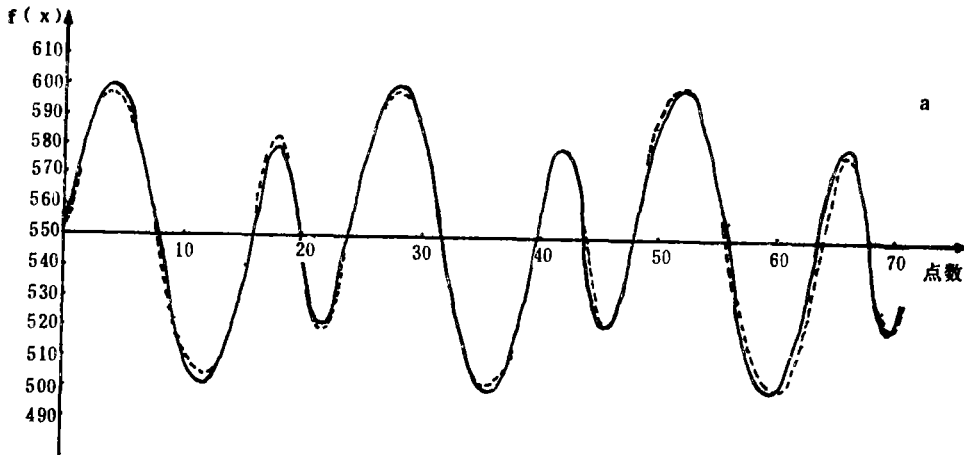


图 2 用理想函数时截取不同项数的拟合情况

(a) 粗实线代表真值, 虚线代表 n 取 12;

(b) 粗实线代表真值, 虚线代表 n 取 35, 点划线代表 n 取 8)

$f(x)$ 的变化曲线可见图 2 中的粗实线。将该函数离散化后, 取 72 个点上的值作为真值, 在其上迭加随机误差(为 STYR 程序库中子程序 GGNML 产生的 $(0, 1)$ 正态分布的随机

数)以得到带有误差的“实测”资料。定义 $\sigma = \sqrt{\frac{1}{N}\sigma_1}$ ，由此计算出 n 从 1 到 35 项时所对应的不同 σ 值，根据 σ 的极小值 σ_{\min} 定出最佳截断项数 n_{op} 。当 σ_E 取 4.0 时，由计算知 σ_{\min} 为 2.4, n_{op} 为 12。参见图 2 a, n 取 12 项时展开式的值与真值的拟合较好，而当 n 取 8 ($\sigma_E = 7.9$) 或 n 取 35 ($\sigma_{35} = 4.2$) 时(图 2 b)，拟合得就差一些，而且还会有许多虚假的小波动。因此截取过多或过少的项都是有害的，这与定性结论一致。但由于实际气象资料的光滑程度远不及理想函数，所以我们还是着重于用实际资料进行实验。

2. 用实际资料作的实验

用美国气象局发布的 $0^\circ N, 15^\circ N, 30^\circ N, 45^\circ N, 60^\circ N, 75^\circ N$ 上的等压面高度和温度格点资料，每个纬圈上等距离分布着 72 个格点。此时 $N = 72, n' = 36$ 。虽然高度、温度的真值是无法得到的，但可以认为所选取的格点资料代表了某种天气形势下真实的高度、温度分布，而把迭加上误差值所得的高度、温度看作是带有误差的实际资料。把高度及温度沿各个纬圈展开，取 $\sigma = \sqrt{\frac{1}{N}\sigma_1}$ ，由此计算出不同纬度、不同资料误差情况下 n 从 1 到 36 项时不同的 σ 值，并根据 σ 的最小值 σ_{\min} 定出最佳截断项数 n_{op} 。其中 σ 表用实际资料计算的只截取前 n 项的有限 Fourier 展开式在格点上的均方误差。资料误差也用 STYR 库中子程序 GGNML 产生的 (0, 1) 正态分布的随机数。在以下的结果中高度的单位均为 dagpm，温度的单位均为 $^\circ C$ ，时间均为世界时。

表 1 为用冬季形势(1983年 12 月 1 日 12 时)下的高度和温度场进行计算的结果。可以看出 $\sigma_{\min} < \sigma_E$ ，这说明把实际资料作 Fourier 展开并截取前 n_{op} 项，可以使各个点上的误差有所减少，从 σ_E 取不同值所作的实验(见表 2)中亦可看到这一点，这就是最优化的优点

表 1 不同位势高度及 500hPa 温度的 σ_{\min} 和 n_{op} ($\sigma_E = 1.0$)

物理量	σ_{\min} 及 n_{op}	纬度					
		$0^\circ N$	$15^\circ N$	$30^\circ N$	$45^\circ N$	$60^\circ N$	$75^\circ N$
700hPa 高度	σ_{\min}	0.5	0.7	0.8	0.7	0.7	0.6
	n_{op}	1	12	22	22	15	8
500hPa 高度	σ_{\min}	0.7	0.7	0.8	0.7	0.7	0.5
	n_{op}	3	12	22	22	17	8
100hPa 高度	σ_{\min}	0.7	0.7	0.7	0.7	0.6	0.5
	n_{op}	14	12	21	14	12	5
500hPa 温度	σ_{\min}	0.7	0.7	0.8	0.7	0.7	0.5
	n_{op}	8	12	22	22	16	7

表 2 不同 σ_E 时的 σ_{\min} 及 n_{op}

σ_E	σ_{\min} 及 n_{op}	物理量					
		500hPa 高度			500hPa 温度		
		纬度					
		$15^\circ N$	$45^\circ N$	$75^\circ N$	$15^\circ N$	$45^\circ N$	$75^\circ N$
0.5	σ_{\min}	0.4	0.4	0.3	0.4	0.4	0.4
	n_{op}	21	22	8	22	22	12
1.0	σ_{\min}	0.7	0.7	0.5	0.7	0.7	0.5
	n_{op}	12	22	8	12	22	7
2.0	σ_{\min}	1.0	1.4	0.9	1.0	1.3	0.9
	n_{op}	5	14	7	4	12	7
4.0	σ_{\min}	1.5	2.4	1.7			
	n_{op}	3	14	7			

所在。从表中还可以看到由于 100 hPa 上高度场以长波及超长波为主(即 M 较小)，因此较之 500 hPa, 700 hPa, 100 hPa 高度展开的 n_{op} 要小一些，这与定性结果一致，但这在 $0^\circ N$ 处不太明显。比较 500 hPa 高度和温度的 n_{op} ，由于 500 hPa 等压面上温度场与高度场配合得较好，即它们波动起伏的大小和波数较一致，因此在 500 hPa 上高度、温度场展开的

n_{op} 几乎相同, 尤其在中纬度更是如此。

表 2 列举了用冬季形势(1983 年 12 月 1 日 12 时)下的 500 hPa 位势高度和温度场进行计算的结果。可以看出, 无论高度还是温度, σ_E 越大, n_{op} 就越小, 这与定性结论一致。同时, 亦可以看出高度与温度展开的 n_{op} 比较接近。

从图 3 可以看出不同纬度上 σ 随项数的变化。由于在中纬度, 高度场的波动起伏变化比较大, 即 M 较大, 因此就要取更多的项数才能使 σ 达最小(亦可参见表 1 及表 2)。由理论估计式(15), 不等号右边第一项与 n 的平方成反比, 第二项与 n 的一次方成正比, 因此 n 较小时的误差比 n 较大时的误差要大, 这一点在图 3 以及图 4-6 中都可以看到。

让我们再来看看不同季节 σ 随项数的变化(图 4)。虽说中纬度地区冬、夏两季波数几乎相同, 但冬季波动的振幅显然比夏季大, 因此图 4 中的曲线(1)比之(2)先达到最佳项数, 两条曲线只有达到一定的项数后才比较接近。

图 5 表示的是在资料误差不同时 σ 随项数的变化。由图可见 σ_E 对 σ 的影响是很大的。当 σ_E 较小时, 截断项数达到一定值后 σ 增加得很慢, 即项数对 σ 的影响已不太大。而当 σ_E 较大时, σ 只在项数取某一范围的值如 10—20 时变化较小, 当项数取到一定的值后 σ 的增长较快, 此时如截取更多的项只会使误差进一步加大。

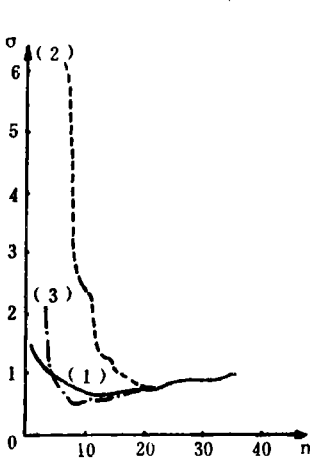


图 3 $\sigma_E=1.0$ 时不同纬度上 σ 随项数的变化
(1) 15°N , (2) 45°N , (3) 75°N ;
(1982 年 1 月 2 日 12 时)

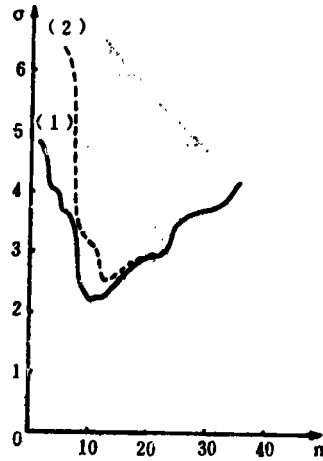


图 4 不同季节 σ 随项数的变化(45°N)
(1) 夏季, 1983.7.2.12 时 500 hPa 高度; (2) 冬季, 1982.1.2.12 时 500 hPa 高度)

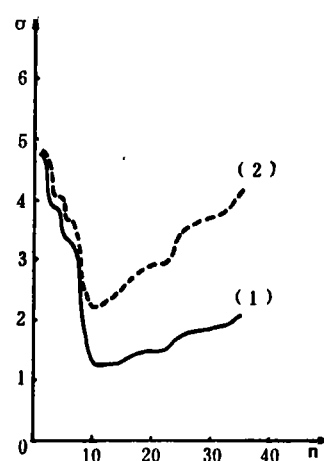


图 5 在 45°N 处对不同资料误差计算的 σ 随项数的变化
(1) $\sigma_E=2.0$, (2) $\sigma_E=4.0$; 1983 年 7 月 2 日 12 时, 500 hPa 高度)

图 6 给出了同一时次(1983 年 12 月 1 日 12 时)不同等压面上的 σ 随项数的变化。当项数较小时各等压面上的 σ 差别较大, 但当项数大于一定的值后各等压面上的 σ 值相差就不大了。这是因为当 n 取值较大时, 影响 σ 的主要是资料误差引起的项 $2n\sigma_E^2$ (见(15)式), 而此处 σ_E 相同。这也从另一个侧面说明了资料误差对展开式的误差是何等重要。

图 7 是用冬季形势下(1982 年 1 月 2 日 12 时) 45°N 处 500 hPa 高度值计算的 n 取不

同值时 Fourier 展开式的值与真值的拟合情况。如图所示,截取 35 项及 8 项的效果都不及截取 12 项的效果好。尤其是截取 35 项时,不仅计算量很大而且随机误差的影响相当大,计算出来的曲线上有许多虚假的小波动,这在数值预报中是有害的。因此,在资料误差较大时截取项数太多势必会事倍功半。当然在资料误差不太大的情况下项数取得多比取得少对计算结果的危害要小一些,这一点从估计式(15)及所给的图表中可以看到。

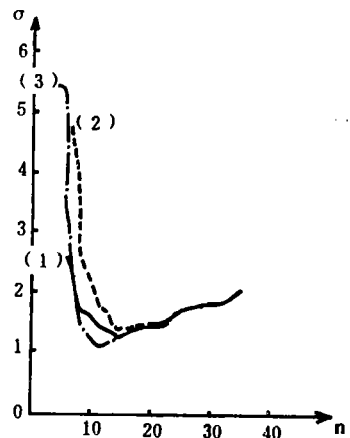


图 6 $\sigma_E=2.0$ 时不同高度上 σ 的变化
(1) 700hPa, (2) 500hPa, (3) 100hPa)

四、结论和建议

本文围绕着引言中提出的三个问题进行了一些尝试性的工作,得出以下结论:

1. 应用最小二乘法的原理和方法把用实际资料计位势高度

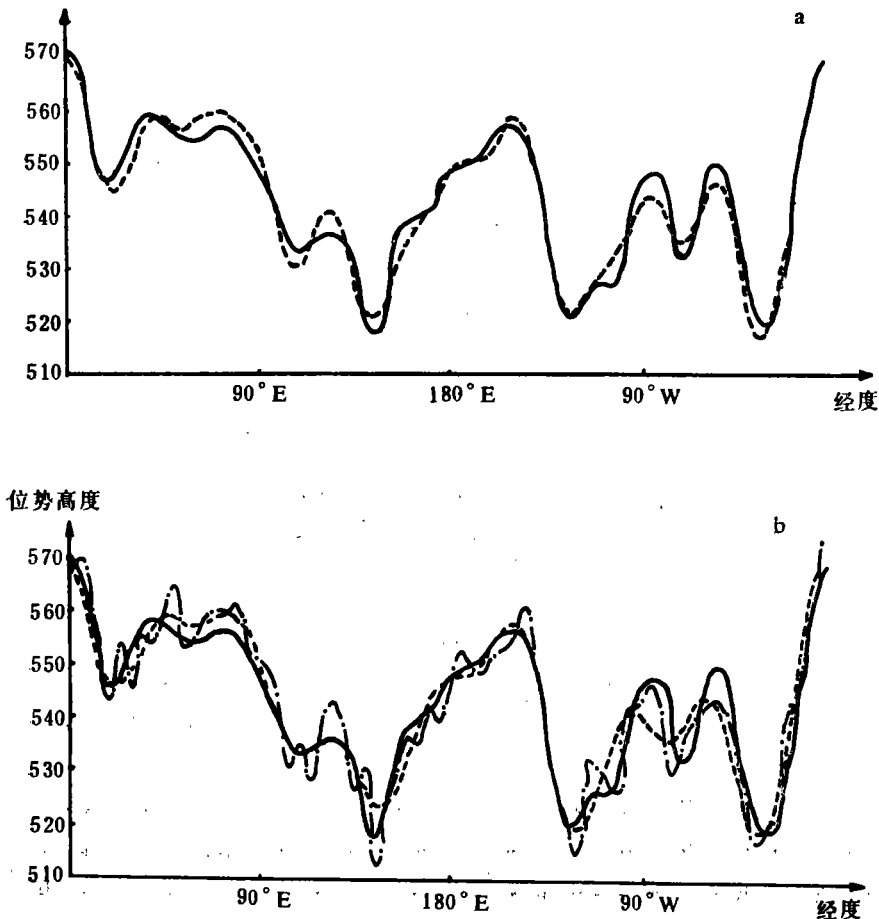


图 7 $\sigma_E=4.0$ 时截断项数 n 不同时展开式的值与真值的比较
(a 粗实线表真值,虚线表 n 取 $n_{opt}=12$ 项时的值;
b 粗实线表真值,虚线表 n 取 8 项,点划线表 n 取 35 项)

算的、只截取前 n 项的有限 Fourier 展开式的均方误差达最小作为确定最佳截断项数 n_{op} 的客观标准。

2. 如资料本身不含有误差, 只要计算条件允许, 在最大可能的项数范围内截取项数越多效果越好。但在资料含有误差时只有当 n 取 n_{op} 时才能达到最佳效果即使 $\sigma < \sigma_E$ 。若一味增加截断项数, 势必会事倍功半。从(15)式可以看到, n_{op} 与函数本身的变化有关系, 物理量的波动起伏越大就要取更多的波来逼近它。

3. 资料误差对展开式的值影响是很大的, 尤其是截取项数越多, 展开式中包含的资料误差就越多, 精度就越低。因此 n_{op} 与资料误差也有关系。资料误差的均方差越大, n_{op} 就越小。

4. 通过对某些天气形势下高度、温度场的数值实验, 分别从纬度变化、季节变化、不同等压面高度、不同物理量、不同资料误差等几个角度讨论了 σ 随项数的变化, 得到与定性结果一致的结论, 并为实际应用提供了参考。

上面的讨论和计算结果都是初步的。其中误差考虑得很简单, 也没有给出 σ^2 关于 n 的函数表达式; 而且, 只考虑了等压面位势高度和温度沿纬圈作 Fourier 展开时取最佳截断项数的问题, 没有涉及其他气象要素和多维的情况, 也没有涉及预报问题。我们的目的, 只是想从理论和计算上证明, 在资料有误差的情况下, 的确存在谱展开的最佳截断项数。显然, 在这样简单情况下得到的计算结果不能推广到一般的情况, 如预报的情况。这是不可不注意的。比如, 上面计算时所取的分辨率比较低, 得到的较大 n_{op} 只约为 20, 远比国外一些业务预报中心, 如 ECMWF 的 $63^{1)}$, 为小。

其实, 这个例子并不难理解。我们知道, 在 1983 年, ECMWF 预报中所取的经、纬度网格距为 1.875° , 线圈网格点数为 192, 在均方误差最小的情况, 最大可取波数为 96; 这样, 计算所取波数只约为最大可取波数的 $2/3$ 。在本文的计算中, $N=72$, 最大可取波数为 36, 计算所取波数, 即 n_{op} , 为 10—20, 即约为最大可取波数的 $1/3$ — $2/3$ 。这说明二者有一定的差别, 但差别还不是太悬殊的。我们认为, 造成这样的差别可能有如下原因:

1. ECMWF 等取波数不仅是针对位势高度和温度等比较光滑的场, 也针对分布比较零散或比较不太连续的地面海拔高度和水汽混合比等场, 而为了计算的方便, 又必须使各展开的对象的最大波数相等, 这就是其取高分辨和高波数的一个原因; 而如果只在纬圈展开位势高度和温度, 如果又不作预报的话, 波数看来是可以小些的。另一方面, 由于用资料的方便, 我们取的网格距较粗。如果我们也取和 ECMWF 同样的网格距的话, n_{op} 可能达到 50 左右, 这样, 其与 ECMWF 所取的就相差不多了。

2. 由于客观分析和预报的需要, 一般须作二维或更高维的谱展开, 又因有非线性项的存在, 情形比我们讨论的要复杂得多, 最佳截断项数的决定可能和一维的情形很不相同。

3. 在模式中有非绝热加热和耗散存在, 因它们在很大程度上和尺度较小的系统有关, 要想把它们的作用估计好, 有必要提高分辨率, 因而, 最佳截断项数也增加了。

所以, 对于二维以上或预报的情况, 还须要作专门的研究和试验。不过, 我们推测, 在

¹⁾ 指 1983 年的情况。

这样的情况下,仍然有小于最大可取波数的最佳的 n_o 存在;实际上, ECMWF 和国外一些用谱模式作预报的中心对截断波数的取法,一般比最大可取波数小,就可以间接说明这一点。

附 录

I. 离散形式的巴什瓦定理

由 $N = 2n' + 1$ 及文献[8] p153 上的公式 $\left(\frac{1}{2} + \sum_{k=1}^{n'} \cos kt\right) = \frac{\sin\left(n + \frac{1}{2}\right)t}{2 \sin \frac{t}{2}}$

便有

$$\begin{aligned} & N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^{n'} (a_m^2 + b_m^2) \right] \\ &= \frac{2}{N} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} F_j F_l \left[\frac{1}{2} + \sum_{m=1}^{n'} \cos(l-j) \frac{2\pi m}{N} \right] \\ &= \frac{2}{N} \sum_{j=0}^{N-1} F_j^2 \left(\frac{1}{2} + n' \right) + \frac{2}{N} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} F_j F_l \frac{\sin\left(n' + \frac{1}{2}\right)(l-j) \frac{2\pi m}{N}}{2 \sin(l-j) \frac{\pi m}{N}} \\ &= \sum_{j=0}^{N-1} F_j^2 + \frac{2}{N} \sum_{j=0}^{N-1} \sum_{\substack{l=0 \\ j \neq l}}^{N-1} F_j F_l \frac{\sin(l-j)m\pi}{2 \sin(l-j) \frac{\pi m}{N}} \\ &= \sum_{j=0}^{N-1} F_j^2; \\ \therefore \sum_{j=0}^{N-1} F_j^2 &= N \left[a_0^2 + \frac{1}{2} \sum_{m=1}^{n'} (a_m^2 + b_m^2) \right] \end{aligned}$$

上式即为离散形式的巴什瓦定理。

II. 复数形式 Fourier 展开式的 σ_1^2 的表达式

沿用文章中的符号便有

$$\begin{cases} F(x) = \sum_{m=-\infty}^{\infty} c_m e^{imx} \\ c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(x) e^{-imx} dx \\ \bar{F}(x) = \sum_{m=0}^{N-1} \bar{c}_m e^{imx} \\ \bar{c}_m = \frac{1}{N} \sum_{k=0}^{N-1} f_k e^{-i2\pi m k/N} \\ \bar{F}_n(x) = \sum_{m=0}^{n-1} \bar{c}_m e^{imx} \\ \tilde{F}(x) = \sum_{m=0}^{N-1} \bar{c}_m e^{imx} = \bar{F}(x) - \bar{E}(x) \\ \begin{cases} \bar{E}(x) = \sum_{m=0}^{N-1} \bar{c}_{E,m} e^{imx} \\ \bar{c}_{E,m} = \frac{1}{N} \sum_{k=0}^{N-1} \bar{E}_k e^{-i2\pi m k/N} \end{cases} \end{cases}$$

$$\tilde{F}_n(x) = \sum_{m=0}^{n-1} \tilde{c}_m e^{imx}$$

同样定义 $\sigma_1 = \sum_{l=0}^{N-1} [F_l - \tilde{F}_n(x_l)]^2$

$$\sigma_1 = \sum_{l=0}^{N-1} \{ [F_l - \bar{F}_n(x_l)]^2 + 2[F_l - \bar{F}_n(x_l)] \cdot \bar{E}_n(x_l) + [\bar{E}_n(x_l)]^2 \}$$

利用指数函数正交关系式^[1]和对实函数有

$$\bar{F}_n(x_l) = \sum_{m=0}^{n-1} \tilde{c}_m^* e^{-i2\pi ml/N}$$

则 $\sum_{l=0}^{N-1} [F_l - \bar{F}_n(x_l)] \cdot \bar{E}_n(x_l) = 0$

因此 $\sigma_1 = \sum_{l=0}^{N-1} \{ [F_l - \bar{F}_n(x_l)]^2 + \bar{E}_n(x_l)^2 \}$

利用相对于指数函数的离散形式巴什瓦定理^[2]有

$$\sum_{l=0}^{N-1} [F_l - \bar{F}_n(x_l)]^2 = N \sum_{m=n}^{N-1} |\tilde{c}_m|^2$$

由正交关系式有 $\sum_{l=0}^{N-1} [\bar{E}_n(x_l)]^2 = N \sum_{m=0}^{n-1} |\tilde{c}_{Em}|^2$

因而 $\sigma_1 = N \sum_{m=n}^{N-1} |\tilde{c}_m|^2 + N \sum_{m=0}^{n-1} |\tilde{c}_{Em}|^2$

参 考 文 献

- [1] 华罗庚、王元, 数值积分及其应用, 科学出版社, 1963。
 [2] 马淑芬, 插值多项式最优化在客观分析中的应用, 气象学报, 38, 1, 73—79, 1980。
 [3] Simmons, A. J., Dynamical Prediction: Some results from operational forecasting and research experiments at ECMWF, WMO, PWPR, Long-rang forecasting research Pub. Ser. No. 1, 1983。
 [4] Manabe, S., D. G. Hahn and J. L. Holloway, Jr., Climate simulations with GFDL spectral models of the atmosphere: effect of spectral truncation, WMO GARP Pub. Ser. Vol. 1, No. 22, 1979。
 [5] 北京大学、吉林大学、南京大学计算数学教研室编, 计算方法, 人民教育出版社, 1961。
 [6] E. O. 布赖姆, 快速富里叶变换, 上海科学技术出版社, 1979。
 [7] Alaka, M. A., and F. Lewis, Numerical Experiments Leading to the Design of Optimum Global Meteorological Networks. Tech. Memo. WBTM TDL-7, ESSA, Weather Bureau, Washington, D. C., 1970。
 [8] [美] E. W. 切尼, 逼近论导引, 上海科学技术出版社, 1981。

THE DETERMINATION OF THE OPTIMUM NUMBER OF TRUNCATED TERMS OF FOURIER EXPANSION AND NUMERICAL EXPERIMENTS

Xiu Aijun Liao Dongxian

(National Meteorological Center, State Meteorological Administration)

Abstract

With data errors the optimum number of truncated terms of Fourier expansion is determined. According to the theoretical estimate the optimum number is related to data variation and their errors. The computations of the numerical experiments are in agreement with the theoretical results.