

利用正交函数作气候要素场定量描述的试验*

施 永 年
(南京大学气象系)

提 要

在本文中,首先提出了用近似的解析表达式定量描述气候要素场的一般问题。作者认为:利用正交函数的线性组合作为近似表达式是具有许多优点的。提出了一个衡量准确度的计算公式。计算了两个实例,求出线性组合中各项的系数、用线性组合表示的定量描述方程、原始资料数据相对于描述方程的相关指数、标准误差和方差比,并且从各项系数对相关指数的贡献中比较了各项的相对重要性。

一、引 言

在气候学理论中,为了研究各种气候形成因子与气候要素场之间的相互关系,如何客观地对这些形成因子和要素的时空分布作一定量的描述,是一个极其重要的基本问题。

显然,各台站的原始资料和等值线图,虽然都是客观的,但是在理论上处理时很不方便,很难直接引用。而根据原始资料或等值线图求得的平均值或某些其他指数,又往往过于概略,不能全面代表分布特点,有些甚至完全失去了代表性。因此,最理想的办法是寻求一个数学上的函数解析表达式 $f_n(x, y, z; t)$, 根据我们所要求的准确度充分近似地定量描述客观存在的、但其解析表达式并不知道的气候要素场 $f(x, y, z; t)$ 。

这样一个问题,从气候统计学的观点来看,就是求回归的问题。但是,如果用解正规方程组的方法,即使是求一个自变数的曲线回归,已经相当麻烦。而且,根据要求的准确度发现计算结果不够满意而要增加项数时,所有的系数必须重新进行计算。至于求多维空间中的曲面回归,当然更是非常麻烦的事。

但是,如果利用数学上提供的正交函数组,则可以使我们避开或减小如上所述的种种缺点和麻烦。而且可以针对不同的问题,对正交函数组作最有利的选择(如使方程求解简化、收敛加快等)。正因为如此,目前在气象学中利用正交函数的线性组合定量描述要素场的例子已经很多。如:利用三角函数组作气温和土壤温度的谐波分析^[1,2],利用等距点正交多项式作气候变迁的趋势分析^[3],利用 Legendre 多项式表示要素或形成因子的经向分布^[4-6],都是属于一元正交函数组应用于描述气候要素的时间变化或一度空间分布的问题。而球函数^[7]、二元等距点正交多项式^[8-12]、三角函数与正交多项式的混合函数^[12]等等二元的正交函数组也已经应用于描述要素场或形成因子的水平分布。在 E. H. Блинова 的一系列论文中,更是经常利用球函数或三角函数与球函数的混合函数^[13]来描述气候要素或形成因子的空间分布及其时间变化。

* 本文于 1964 年 5 月 20 日收到。

二、平均最高温度的气候变迁趋势

据 S. K. Pramanik 等^[3]的意见, 要研究温度的气候变迁趋势, 最好是对最高温度与最低温度分别求它们的趋势曲线方程。为了确定年与年之间的变化究竟是纯粹偶然的抑或是其中包含了某种规律性的变化, 可以首先用转折点法^[4]对最高温度或最低温度的变化数列进行随机性测验。从分布在全国的、纪录年代较长的八个台站的年平均每日最高温度和年平均每日最低温度的数列, 进行转折点法随机性测验的结果来看, 仅有上海(1904—1958 共 55 年)的平均最高温度和沈阳(1906—1940 共 35 年)的平均最高温度两个连续数列是非随机的。也就是说, 对于其他数列虽可用数学方法配合最小二乘方意义下的趋势曲线方程, 但这种曲线并没有什么实际意义, 因为数列中每年平均最高温度或平均最低温度的升降很可能纯粹是偶然的。为此, 我们仅对非随机的两个数列求取其变迁趋势曲线。

此时, 所考虑的气候要素场仅是时间 t 的一元函数, 而且 t 的取值是离散、等距的。显然, 正交函数组应以选择一元等距点正交多项式最为有利。

根据 R. A. Fisher^[5] 的方法: 设 N 为等距点的数目(数列的年数), 当 N 为奇数时, 可取中间一年作为原点 ($t = 0$), 每递增或递减一年 t 亦增减 1。当 N 为偶数时, 可取中间两年分别为 $t = -\frac{1}{2}$ 和 $t = \frac{1}{2}$, 每递增或递减一年 t 亦增减 1。当 t 的坐标这样选取时, 规一化的一元等距点正交多项式 $\varphi_v(t)$ 可表示为:

$$\varphi_v(t) = \frac{P_v(t)}{\sqrt{\sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_v^2(t)}} \quad (1)$$

式中:

$$\left. \begin{aligned} P_0(t) &= 1, \\ P_1(t) &= t, \\ P_2(t) &= t^2 - \frac{N^2 - 1}{12}, \\ P_3(t) &= t^3 - \frac{3N^2 - 7}{20}t, \\ &\dots\dots\dots \end{aligned} \right\} \quad (2)$$

一般言之, 有递推公式:

$$P_{r+1}(t) = P_1(t)P_r(t) - \frac{r^2(N^2 - r^2)}{4 \times (4r^2 - 1)} P_{r-1}(t). \quad (3)$$

并且有公式:

$$\left. \begin{aligned} \sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_0^2(t) &= N, \\ \sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_1^2(t) &= \frac{N(N^2 - 1)}{12}, \end{aligned} \right\}$$

$$\left. \begin{aligned} \sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_2^2(t) &= \frac{N(N^2-1)(N^2-4)}{180}, \\ \sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_3^2(t) &= \frac{N(N^2-1)(N^2-4)(N^2-9)}{2800}, \\ &\dots\dots\dots \end{aligned} \right\} \quad (4)$$

一般言之,有:

$$\sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} P_r^2(t) = \frac{(N-1)^{(r)}(N+r)^{(r+1)}}{(2r+1) \binom{2r}{r}}, \quad (5)$$

式中:

$$\begin{aligned} (N-1)^{(r)} &= (N-1)(N-2)\cdots(N-r+1)(N-r), \\ (N+r)^{(r+1)} &= (N+r)(N+r-1)\cdots(N+2)(N+1)N, \\ \binom{2r}{r} &= \left[\frac{(2r)!}{r!r!} \right]. \end{aligned}$$

随 N, v, t 的值而变的 $P_v(t)$ 和 $\sum_i P_i^2(t)$ 的数值可以从表^[16]上查得。因此,系数 $C_v (v = 0, 1, 2, \dots)$ 的求得非常方便:

$$C_v = \sum_{t=\frac{1-N}{2}}^{\frac{N-1}{2}} T_M(t) \varphi_v(t). \quad (6)$$

其中 $T_M(t)$ 表示数列中每年的平均最高温度。

当 n 固定时,

$$\hat{T}_{Mn}(t) = \sum_{v=0}^n C_v \varphi_v(t). \quad (7)$$

就是最小二乘方意义下对最高温度数列 $T_M(t)$ 的最佳定量描述,它代表着最高温度气候变迁的抛物线趋势。

由下式定义的 ρ_n 可以作为用 $\hat{T}_{Mn}(t)$ 描述 $T_M(t)$ 的准确度:

$$\rho_n^2 = \frac{\sum_{v=1}^n C_v^2}{N\sigma^2}. \quad (8)$$

式中 σ^2 为数列 $T_M(t)$ 的方差。

从统计学的观点来看, ρ_n 实际上就是 $T_M(t)$ 相对于趋势曲线 $\hat{T}_{Mn}(t)$ 的相关指数。采用相关指数 ρ_n 作为准确度,其优点是意义明确,计算方便。根据已经求得的 $C_v (v = 1, 2, \dots, n)$ 代入(8)式很快就可求得 ρ_n , 当 $\hat{T}_{Mn}(t)$ 与 $T_M(t)$ 完全一致时, $\rho_n = 1$, $\hat{T}_{Mn}(t)$ 与 $T_M(t)$ 愈符合, ρ_n 就愈接近于 1。反之,若事先规定了必需满足的准确度 ρ_n , 则可逐步将求出的 C_1, C_2, \dots, C_n 代入(8)式直至符合要求为止, n 就是最恰当的

项数。

由 ρ_n^2 可以求出用 $\hat{T}_{Mn}(t)$ 定量描述 $T_M(t)$ 的标准误差 S ：

$$S = \sigma \sqrt{1 - \rho_n^2}. \quad (9)$$

此外,还可以用方差分析的方法测验所求得的定量描述解析表达式的显著性。这时,方差比 F 可用下式计算:

$$F = \frac{\frac{\rho_n^2}{n}}{\frac{1 - \rho_n^2}{N - n - 1}}. \quad (10)$$

而 F 的自由度为: $(n, N - n - 1)$ 。

计算结果:

(1) 沈阳(1906—1940)

年平均最高温度数列的方差: $\sigma^2 = (0.67^\circ\text{C})^2$ 。

系数 C_v 值

C_0	C_1	C_2	C_3
81.0	1.95	-0.428	0.063

将上表中 C_v 的值与(1)式中的 $\varphi_v(t)$ 代入(7)式,再合并同次项,即得趋势方程:

$$\hat{T}_{M3}(t) = 13.8 + 0.0132t - 0.000795t^2 + 0.0000159t^3. \quad (11)$$

按(8)式计算相关指数 ρ_3 以检查趋势方程与原数列的吻合程度:

$$\rho_3^2 = \frac{1.95^2 + 0.428^2 + 0.063^2}{35 \times 0.67^2} = 0.26.$$

即:

$$\rho_3 = 0.51$$

原数列相对于三次抛物线(11)式的相关指数为 0.51。明显看出,其中以 C_1 对相关指数的贡献最大,若考虑趋势方程为:

$$\hat{T}_{M1}(t) = C_0\varphi_0(t) + C_1\varphi_1(t) = 13.7 + 0.0326t. \quad (12)$$

则相关指数亦可达:

$$\rho_1 = \sqrt{\frac{1.95^2}{35 \times 0.67^2}} = 0.49.$$

据(19)式,可算出方程(11)描述原数列的标准误差:

$$S = 0.67 \times \sqrt{1 - 0.26} = 0.576^\circ\text{C}.$$

据(10)式,算得方差比:

$$F = \frac{\frac{0.26}{3}}{\frac{1 - 0.26}{35 - 3 - 1}} = 3.63.$$

如果取显著性水准为 5%, 则自由度为(3, 31)的 F 临界值: $F^* = 2.91^{[16]}$, 因此,至少有 95% 的把握可以认为方程(11)的配合不是随机的。

(2) 上海(1904—1958)

年平均最高温度数列的方差: $\sigma^2 = (0.74^\circ\text{C})^2$.

系数 C_r 值

C_0	C_1	C_2	C_3
154.9	0.735	-3.32	1.89

代入(7)式, 可得趋势方程:

$$\hat{T}_{M3}(t) = 21.4 + 0.0511t - 0.00199t^2 - 0.0000979t^3. \quad (13)$$

相关指数:

$$\rho_3 = \sqrt{\frac{0.735^2 + 3.32^2 + 1.89^2}{55 \times 0.74^2}} = 0.71.$$

因此, (13)式与原数列的吻合程度比沈阳的情况要好一些。但是, 在相关指数中主要的贡献是由二次和三次多项式的系数作出的, 所以上海的最高温度变迁趋势方程就不能不考虑二次项和三次项。若仅考虑一次项, 相关指数是:

$$\rho_1 = \sqrt{\frac{0.735^2}{55 \times 0.74^2}} = 0.13.$$

若仅考虑到二次项为止, 相关指数是:

$$\rho_2 = \sqrt{\frac{0.735^2 + 3.32^2}{55 \times 0.74^2}} = 0.62.$$

方程(13)的标准误差:

$$S = 0.74 \times \sqrt{1 - 0.71^2} = 0.52^\circ\text{C}.$$

方差比:

$$F = \frac{\frac{0.71^2}{3}}{\frac{1 - 0.71^2}{55 - 3 - 1}} = 17.2.$$

这一 F 值大大地超过了显著性水准为 0.1%、自由度为(3, 51)的临界值 $F^{*[16]}$, 因此, 可有 99.9% 以上的把握认为趋势方程(13)的配合不是随机的。

图 1 和图 2 中的曲线就是根据趋势方程(11)和(13)作出的。



图 1 沈阳年平均最高温度变迁趋势

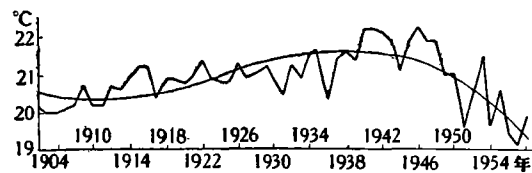


图 2 上海年平均最高温度变迁趋势

图上的折线代表沈阳和上海的年平均最高温度实际数列。由图可见, 沈阳的趋势曲线很接近于一条直线, 而上海的趋势曲线显然具有三次抛物线的特征。从相关指数的计算中,

我们早已肯定了这些事实。

由这两条曲线来看,从 1905 年至 1940 年左右,无论沈阳或上海,都具有升温的变迁趋势。而从上海的趋势曲线来看,从 40 年代开始,有降温的变迁趋势。趋势曲线的这些特征是和王鹏飞^[17]用其他方法研究我国近百年来来的气候变迁所得结论是一致的。

三、我国沿海温度场的定量描述

在 Albers 等积投影的地图上,选取如图 3 的矩形区域作为描述的区域。这个区域可以用 15×21 个格点分为 14×20 个正方形。每个格点的坐标是 (x, y) , x 的取值由 $0, 1, 2, \dots$ 直至 $14, y$ 的取值由 $0, 1, 2, \dots$ 直至 20 (见图 3)。

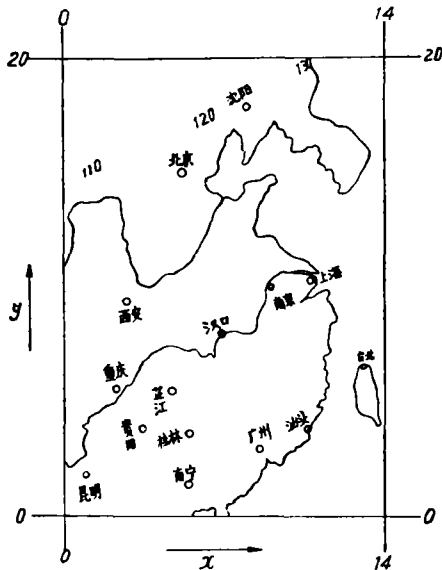


图 3 定量描述温度场的区域和坐标

根据资料^[19]作出了该区域上 1 月份的等温线图和 7 月份的等温线图。在每个格点上可以读出相应月份的温度数值 $T(x, y)$ 。

设 $X_r(x)$ 为 15 个点的规一化等距点 r 次正交多项式, $Y_s(y)$ 为 21 个点的规一化等距点正交多项式:

$$\left. \begin{aligned} X_r(x) &= \sum_{i=0}^r (-1)^i \binom{r}{i} \binom{r+i}{i} \frac{(x)^{(i)}}{(14)^{(i)}} \\ &= \frac{(14+r+1)^{(r+1)}}{(2r+1)(14)^{(r)}}, \\ Y_s(y) &= \sum_{i=0}^s (-1)^i \binom{s}{i} \binom{s+i}{i} \frac{(y)^{(i)}}{(20)^{(i)}} \\ &= \frac{(20+s+1)^{(s+1)}}{(2s+1)(20)^{(s)}}. \end{aligned} \right\} (14)$$

其中

$$\begin{aligned} (x)^{(i)} &= x(x-1)\cdots(x-i+1), \\ (14)^{(i)} &= 14 \times 13 \times \cdots \times (14-i+1), \\ (14+r+1)^{(r+1)} &= (14+r+1)(14+r)\cdots(14+1), \\ (14)^{(r)} &= 14 \times 13 \times \cdots \times (14-r+1). \end{aligned}$$

余仿此。则它们的乘积 $X_r(x)Y_s(y)$ 就是在 15×21 个格点上正交的一个 $(r+s)$ 次规一化正交多项式。由此我们可以得到定量描述相应月份温度分布的方程:

$$\hat{T}(x, y) = \sum_{r=0}^l \sum_{s=0}^m a_{rs} X_r(x) Y_s(y). \quad (15)$$

其中每项的系数就是:

$$a_{rs} = \sum_{x=0}^{14} \sum_{y=0}^{20} T(x, y) X_r(x) Y_s(y). \quad (16)$$

相关指数 ρ_n 的计算公式此时化为下式:

$$\rho_n^2 = \frac{\sum_{r=0}^l \sum_{s=0}^m a_{rs}^2 - a_{00}^2}{\sum_{x=0}^{14} \sum_{y=0}^{20} T^2(x, y) - \frac{1}{15 \times 21} \left(\sum_{x=0}^{14} \sum_{y=0}^{20} T(x, y) \right)^2}. \quad (17)$$

从等值线图上读得 $T(x, y)$ 在各格点上的数值, 根据(16)式, 并利用数表^[16], 对1月份和7月份分别求得 a_{rs} 值, 定量描述方程, 原始资料对于方程的相关指数 ρ_n , 标准误差 S , 方差比 F 等如下:

(1) 1月份

a_{rs} 值

a_{00}	a_{10}	a_{01}	a_{20}	a_{11}
31.08	-105.24	162.94	17.64	22.67
a_{02}	a_{20}	a_{21}	a_{12}	a_{03}
3.48	-6.25	3.90	11.25	10.18

定量描述方程式:

$$\hat{T}(x, y) = 1.94 + 0.17x - 0.14y - 0.000567x^2 + 0.000666xy + 0.0000114y^2 + 0.0000605x^3 - 0.0000024x^2y + 0.0000034xy^2 + 0.0000117y^3. \quad (18)$$

原始资料相对于(18)式的相关指数, 据(17)式应为:

$$\rho_9 = \sqrt{\frac{105.24^2 + 162.94^2 + \dots + 10.18^2}{40925.76}} = 0.97.$$

标准误差:

$$S = \sqrt{\frac{40925.76^2 \times (1 - 0.97^2)}{15 \times 21}} = 2.6^\circ\text{C}.$$

方差比:

$$F = \frac{\frac{0.97^2}{9}}{\frac{1 - 0.97^2}{15 \times 21 - 9 - 1}} = 591.1.$$

(2) 7月份

a_{rs} 值

a_{00}	a_{10}	a_{01}	a_{20}	a_{11}
456.75	-21.10	26.33	-20.95	-9.34
a_{02}	a_{20}	a_{21}	a_{12}	a_{03}
-20.17	-7.96	-5.92	-4.35	-1.87

定量描述方程式:

$$\hat{T}(x, y) = 25.65 + 0.0529x - 0.0197y - 0.00251x^2 - 0.000390xy - 0.000256y^2 + 0.0000799x^3 + 0.00000392x^2y + 0.00000113xy^2 + 0.00000214y^3. \quad (19)$$

原始资料对于(19)式的相关指数:

$$\rho_9 = \sqrt{\frac{21.10^2 + 26.33^2 + \cdots + 1.87^2}{1938.80}} = 0.86.$$

标准误差:

$$S = \sqrt{\frac{1938.80 \times (1 - 0.86^2)}{15 \times 21}} = 1.54^\circ\text{C}.$$

方差比:

$$F = \frac{\frac{0.86^2}{9}}{\frac{1 - 0.86^2}{15 \times 21 - 9 - 1}} = 99.1.$$

由以上求得的 F 值与显著性水准为 0.1%, 自由度为(9,305)的临界值 $F^{*[16]}$ 相比较, 发现: 无论在 1 月份或 7 月份, F 都大大地超过了 F^* , 因此可以认为(18), (19)式的求得并不是随机的。

此外, 从 1 月份的 a_{rr} 值可以看出线性项是占绝对优势的。事实上, 如果仅考虑线性项:

$$\begin{aligned} \hat{T}(x, y) &= a_{00}X_0(x)Y_0(y) + a_{10}X_1(x)Y_0(y) + a_{01}X_0(x)Y_1(y) = \\ &= 1.90 + 0.182x - 0.145y. \end{aligned} \quad (20)$$

则原始资料相对于此描述方程(20)的相关指数为:

$$\rho_2 = \sqrt{\frac{105.24^2 + 162.94^2}{40925.76}} = 0.95.$$

因此可以认为 1 月份的温度分布基本上是水平坐标 (x, y) 的线性函数。

但 7 月份却是另外一种情形, 由 a_{rr} 值可以看出, 二次项对相关指数的贡献与一次项的贡献达同一量级, 甚至连三次项也有相当重要的贡献。这说明 7 月份我国沿海地区温度分布的形势要比 1 月份复杂得多。

为了验证定量描述方程(19), (20)式的正确性, 曾根据方程分别算出各点上的温度值, 绘制 1 月份和 7 月份的等温线图, 其结果是与原来的等温线图非常近似的。

* * *

上面所举的仅是利用一元和二元等距点正交多项式进行定量描述的两个实例。对于不同的目的, 利用各种正交函数组定量描述气候要素场的试验我们还在继续进行, 其结果待以后发表。

致谢: 本文承导师么枕生教授启发与鼓励, 得以完成; 又文中的两个实例, 一部分是在我系蒋燕霞(1962), 曹永华(1963)两位同学毕业论文中数据的基础上进行计算而得的, 在此一并致谢。

参 考 文 献

- [1] 么枕生, 地理学报, 18(1944), 126—128.
- [2] 么枕生, 气象学报, 21(1950), 69—78.
- [3] Pramanik, S. K., and P. Jagannathan, *Scientific Proceedings of the International Association of Meteorology*, 1956, 86—138.
- [4] Добрышман, Е. М., *Мет. и Гид.* 1956, 12, 18—25.

- [5] Ракипова, Л. Р., *Труды ГГО* вып. 47 (103), (1953).
- [6] Ракипова, Л. Р., *Тепловой режим атмосферы*, Гидромет. Изд. (1957).
- [7] 中央气象局气象科学研究所数值预报组, *气象学报*, **30** (1959), 405—413.
- [8] 张家诚、周家斌、黄文杰、马维华, *气象学报*, **33** (1963), 231—243.
- [9] Багров, Н. А., *Труды ЦИПа.*, **64** (1957), 3—25.
- [10] Зверев, Н. И., *Труды ЦИПа.*, **85** (1959), 27—39.
- [11] Sehak, K., *Arch. für Geophys. Met. und Bioq.*, S. A., **12** (1960), 40—61.
- [12] Hare, F. K., *Polar Atmosphere Symposium*, Part I. Meteorology Section (1958), 137—150.
- [13] Блинова, Е. Н., *Труды Ин-та Физ. Атм. АН СССР* №. 2 (1958), 5—22.
- [14] Kendall, M. G., *Advanced Theory of Statistics*, (II), (1946), Challes Griffin.
- [15] Fisher, R. A., *Statistical Methods for Research Workers* (11th ed.), (1950), Oliver and Boyd.
- [16] Fisher, R. A., and F. Yates, *Statistical Tables for Bio., Agri. and Med. Research* (5th ed.) (1957), Oliver and Boyd.
- [17] 王鹏飞, *天气月刊*, No. 3 (1958), 25—29.
- [18] *Climatic Atlas of Japan and Her Neighbouring countries*, Cen. Met. Obs., Tokyo, (1943).

AN EXPERIMENT ON THE QUANTITATIVE DESCRIPTION OF CLIMATIC ELEMENT FIELD BY ORTHOGONAL FUNCTIONS

SHI YONG-NIAN

(Department of Meteorology, Nanking University)

ABSTRACT

In this paper, the author firstly proposes the generalized problem of the quantitative description of climatic field by means of approximate analytic expression. It is submitted that the application of the linear combination of orthogonal functions in approximate expression has many advantages. A formula to measure the precision of the calculation is also proposed. Two calculated examples are given.