用最大熵谱方法分析气候序列的周期

The Periods of Climatic Time Series are Analysed by Using Maximum Entropy Spectral Method

涂方旭 Tu Fangxu

（广西气候中心 南宁市植物路50号 530021）
(Climate Centre of Guangxi, 50 Zhiwu Road, Nanning, Guangxi, 530021)

胡圣立 Hu Shengli

（广西气象局 南宁市植物路50号 530021）
(Meteorological Service of Guangxi, 50 Zhiwu Road, Nanning, Guangxi, 530021)

梁振海 Liang Zhenhai

（中国人民建设银行海口分行 海南省海口市 570003）
(Haikou Branch of the People's Construction Bank of China, Haikou, Hainan, 570003)

摘要 提出以正弦函数拟合序列的周期分量并进行周期分量显著性检验的方法。对南宁年平均气温序列和梧州年降水量序列进行了周期分析。

关键词 最大熵谱 气候序列 周期

Abstract The calculations of periodic components by sine function and significance level of periodic components are conducted in the analysis of the period of climatic time series by using maximun entropy spectrum. The periods of annual mean temperature time series in Nanning city and annual amount precipitation time series in Wuzhou city of Guangxi are also computed.

Key words maximum entropy spectrum, climatic time series, period

周期性是时间序列的一种重要特征。分析序列的周期有许多方法，而最大熵谱方法在周期分析中表现出公认的优良特性：对序列未作任何假定，能得到连续谱分布，精度高，并可能分析出超过样本长度的周期。近年来，国内外用最大熵谱方法讨论时间序列的可能周期已相当普遍，但目前主要限于分析序列可能存在的周期，没有给出各个周期的显著性水平，对提取周期分量的方法也未见讨论。国外还有人用计算机作过模拟计算，用 $\sin\lambda t$ 曲线，取其四分之一周期长度加白噪声，最大熵谱方法较准确地找出了周期点，但用此方法分析实际序列的长周期，确定长周期的显著性水平，提取长周期的周期分量，未见报道。因此，最大熵谱方法在实际业务工作中的应用受到限制。我们在对气候时间序列进行最大熵谱分析时，选用正弦函数确定周期分量，并给出了确定周期分量的显著性水平和周期分量的方法，对南宁年平均气温和梧州年降水量序列进行了周期分析。

1993 10 23收稿。1994 01 15修回。

1 统计方法

1.1 最大熵谱方法的主要计算公式

最大熵谱方法详见文献〔1〕〔2〕。其谱估计式可记为，

$$I(\omega)=\frac{\sigma_m^2\Delta}{\left|1-\sum_{k=1}^{m}b_k^{(m)}e^{-i\omega k}\right|^2} \tag{1}$$

式中 $\omega=2\pi f\Delta$，f 为频率，$f=\frac{1}{T}$，T 为周期长度。Δ 为取样间隔，由于取样为等间隔，故取 $\Delta=1$。$i=\sqrt{-1}$，为虚数单位。$b_k^{(m)}$ 为 m 阶自回归系数，采用 Burg 递推公式确定。m 为自回归的截止阶。σ_m^2 为 m 阶自回归预报误差的方差估计。截止阶 m 采用赤池提出的最终预报误差准则确定，即对于 $1\leqslant k\leqslant N-1$，计算，

$$\varphi_k=\sigma_k^2\frac{N+K}{N-K} \tag{2}$$

由 φ_k 的最小值或极小值对应的 k 为截止阶 m。

根据周期 T 与圆频率的关系，

$$\omega=2\pi f\Delta=\frac{2\pi\Delta}{T}$$

对于给定的周期 T，就可由(1)式计算出对应于周期 T 的最大熵谱 $I(\omega)$，记为 $I(T)$。若只需确定整数周期，可给定 $T=1,2,\cdots$，计算出 $I(1)$，$I(2)$，…，作出谱图。谱的峰值点对应序列的周期。一般，峰区越强，周期越显著。

1.2　周期分量及其显著性检验

最大熵谱方法本身未给出对应周期的显著性水平，也未给出确定周期分量的方法。为便于实际应用，我们采用曲线拟合方法确定周期分量，然后计算周期分量与原序列的相关系数，从而可以推断出所确定的周期的显著性水平。

设气候序列为 x_t，$(t=1,2,\cdots,N)$，样本容量为 N。

周期分量用正弦函数(也可用余弦函数)确定。设周期分量的形式为，

$$\hat{x}_t = A + B\sin\frac{2\pi}{T}(t-t_0) \tag{3}$$

式中 $\hat{x}_t$ 为第 t 年原序列 x_t 的周期分量值。T 为周期长度。t 为时间，用公元年份。t_0 的作用相当于初位相，是一个待定参数。A、B 为系数，用最小二乘法确定。

较精确地确定 t_0 可采用数值解法。即对 t_0 给不同的值，计算原序列 x_t 与 $z_t=\sin\frac{2\pi}{T}(t-t_0)$ 的相关系数 R，

$$R=\frac{\Sigma Z_t x_t-\frac{1}{N}\Sigma Z_t\Sigma x_t}{\sqrt{[\Sigma Z_t^2-\frac{1}{N}(\Sigma Z_t)^2][\Sigma x_t^2-\frac{1}{N}(\Sigma x_t)^2]}} \tag{4}$$

取相关系数 R 最大值对应的 t_0。

确定 t_0 以后，用最小乘法确定系数 A 和 B。

$$B=\frac{\Sigma Z_t x_t-\frac{1}{N}\Sigma Z_t\Sigma x_t}{\Sigma Z_t^2-\frac{1}{N}(\Sigma Z_t)^2} \tag{5}$$

$$A=\frac{1}{N}\Sigma x_t-B\frac{1}{N}\Sigma Z_t \tag{6}$$

为了确定周期的显著性水平，需计算周期分量 $\hat{x}_t$ 与原序列 x_t 的相关系数 r，

$$r=\frac{\Sigma\hat{x}_t x_t-\frac{1}{N}\Sigma\hat{x}\Sigma x_t}{\sqrt{[\Sigma\hat{x}_t^2-\frac{1}{N}(\Sigma\hat{x})^2][\Sigma x_t^2-\frac{1}{N}(\Sigma x_t)^2]}} \tag{7}$$

将 $\hat{x}=A+B\sin\frac{2\pi}{T}(t-t_0)=A+BZ_t$　代入(7)式，可得 $r=R$。这里的 R 是 t_0 已确定条件下，由(4)式计算的 R 值。

求出 r，由给定的信度 α 查相关系数显著性检验表，得 r_α。若 $r>r_\alpha$，则说明在显著性水平 α 下，长度为 T 的周期显著。

1.3　拟合预测方程及其显著性检验

一般来讲，序列都存在几个显著周期。为了提高拟合预报的精度，可以逐次进行最大熵谱分析。即先确定一个最显著的周期，再用原序列减去已提取的显著周期分量，对剩余序列再进行最大熵谱分析，提取下一个显著周期分量，…，直到分析不出显著的周期为止。然后将各周期分量进行迭加，即可得到拟合预测方程。

设第 i 个周期分量为

$$\hat{x}_{it}=A_i+B_i\sin\frac{2\pi}{T_i}(t-t_i)$$

又设共有 s 个显著周期，则最后的拟合预测方程为

$$\begin{aligned}\hat{x}_t&=\sum_{i=1}^{s}\hat{x}_{it}\\&=\sum_{i=1}^{s}A_i+B_1\sin\frac{2\pi}{T_1}(t-t_1)+\cdots\\&\quad+B_s\sin\frac{2\pi}{T_s}(t-t_s)\end{aligned}$$

最后，参照(7)式，计算全部显著周期迭加拟合值 $\hat{x}_t$ 与原序列 x_t 的相关系数 $r(\hat{x},x_t)$，则可确定拟合预测方程的显著性水平。

2　实例

2.1　南宁年平均气温的周期分析

2.1.1　南宁年平均气温序列的整理

南宁从1922年起有月、年平均气温资料，但1940年、1945年及1930、1931、1939、1941、1943、1944、1949、1950年部分月份缺资料。同时，不同年代的资料存在观测时次、站址改变的差异，使序列均一性受到影响。

首先经过相关比较分析，选用邻近、相关较好、插补年代均一性也较好的柳州沙塘、梧州和贵州罗甸的月平均气温资料，用回归分析方法对南宁缺测的月平均气温进行插补。插补月平均气温一共建立了32个直线回归方程，其相关系数＞0.9的14个，0.8～0.9的11个，0.7～0.8的5个，0.5～0.7的2个。最小的相关系数是罗甸与南宁8月平均气温的回归方程，相关系数为0.5103，样本容量为15，信度 α 接近0.05，用于插补1945年8月南宁月平均气温。其余31个方程的信度全部＜0.01，其中有29个方程的信度＜0.001。

经过插补得到南宁1922年1月至1992年12月逐月平均气温资料，由此再求得1922～1992年南宁逐年年平均气温。

其次，我们根据不同时次的对比，对1922～1935年的资料初步进行了时次订正。对1956～1965年的资料初步进行了高度订正。对1981年以后的资料初步进行了城市化订正。

这样，得到了一个用于进行气候分析的南宁年平均气温序列，样本容量 $N=71$。

2.1.2 南宁年平均气温的周期分析

用前述最大熵谱分析与正弦函数拟合的方法，对经过整理、订正的南宁年平均气温进行了周期分析。

首先计算南宁年平均气温序列 x_t 的最大熵谱。截止阶 $m=46$ 时的谱图见图 1。

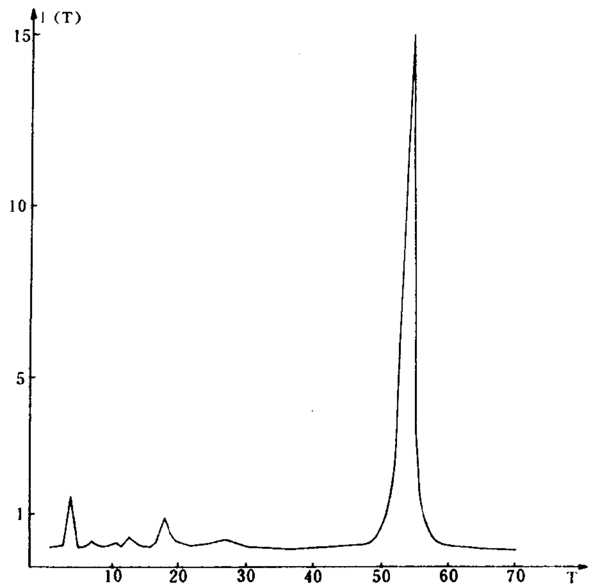

图 1 南宁年平均气温序列的最大熵谱($m=46$)

Fig. 1 The maximum entropy spectrum of annual mean temperature time series in Nanning

从图1可见，南宁年平均气温序列最显著的周期为54年。其它可能周期是4年、18年、13年、27年、11年和7年。为了使周期长度 T 更精确，我们采用了以最大熵谱峰值点为基础，以相关系数最大为原则，求其数值解，并精确到一位小数。求得第一周期 $T_1=55.6$ 年，其周期分量为：

$$\hat{x}_{1t}=21.74243+0.2471188\sin\frac{2\pi}{55.6}(t-1932.2)$$

相关系数 $r=0.4069$，信度 $\alpha<0.001$。

表1 南宁年平均气温的周期及其参数

Table 1 Periods and parameters of annual mean temperature time series in Nanning

i	T_i	A_i	B_i	t_i	r	α
1	55.6	21.74243	0.2471188	1932.2	0.4069	<0.001
2	8.3	−0.002477911	0.1630496	1920.2	0.3110	<0.01
3	4.0	−0.0006335762	0.1456062	1921.2	0.2933	<0.02
4	17.8	0.0001827195	0.1457203	1920.1	0.3056	<0.01
5	2.9	−0.0004795977	0.1401222	1922.1	0.3081	<0.01
6	7.2	0.0005265472	0.1092976	1920.6	0.2541	<0.05
7	25.9	0.005964419	0.08856881	1934.9	0.2091	<0.10

T_i = 周期长度 Length of period, A_i = 周期分量的常数项 Constant of periodic component, B_i = 周期分量的回归系数 Regression coefficient of periodic component, t_i = 位相特征 Phase characteristic, r = 相关系数 correlation coefficient, α = 显著性水平（信度）Significance level。

用第一剩余序列 $x_t-\hat{x}_{1t}$，求第二周期，…，依次求出七个显著周期，各周期有关参数列于表 1。

2.1.3 南宁年平均气温的拟合与预测

根据周期分析结果，将各周期分量进行迭加，得到南宁年平均气温的预测方程：

$$\hat{X}_t=\sum_i[A_i+B_i\sin\frac{2\pi}{T_i}(t-t_i)]$$

$$\begin{aligned}=&21.7455126+0.2471188\sin\frac{2\pi}{55.6}(t-1932.2)\\&+0.1630496\sin\frac{2\pi}{8.3}(t-1920.2)\\&+0.1456062\sin\frac{2\pi}{4.0}(t-1921.2)\\&+0.1457203\sin\frac{2\pi}{17.8}(t-1920.1)\\&+0.1401222\sin\frac{2\pi}{2.9}(t-1922.1)\\&+0.1092976\sin\frac{2\pi}{7.2}(t-1920.6)\\&+0.08856881\sin\frac{2\pi}{25.9}(t-1934.9)\end{aligned}$$

对 1922～1992 年南宁年平均气温进行拟合，计算实测值 x_t 与预测值 $\hat{x}_t$ 的相关系数，得 $r(\hat{x}_t,x_t)=0.7033$，信度 $\alpha<0.001$。

预测方程对 1922～1992 年南宁年平均气温的拟合情况见图 2。

从显著性检验结果和图 2 看，拟合效果是比较好的。

应用预测方程对 1993～2000 年南宁年平均气温的计算结果（已初步滤去城市化影响）列于表 2。

表 2 1993～2000 年南宁年平均气温计算值

Table 2 Calculation of annual mean temperature for the period of 1993 to 2000 in Nanning

t	1993	1994	1995	1996	
$\hat{x}_t$ (℃)	21.93	22.11	22.40	22.17	
t	1997	1998	1999	2000	8 年平均 Average
$\hat{x}_t$ (℃)	22.07	22.38	22.02	21.66	22.1

由表 2 可知，计算的 1993～2000 年南宁年平均气温，多数年份在多年平均值 21.7℃以上波动。这 8 年计算值的平均值为 22.1℃，若考虑城市化影响，取较为保守的估计值 0.3℃，则 8 年平均值可达 22.4℃，可能与南宁平均气温较为温暖的 40 年代的平均值（22.1℃）较为接近。但 2000 年以后，气温将可能有回落的趋势。

2.2 梧州年降水量的周期分析

2.2.1 梧州年降水量序列的整理

梧州从 1898 年开始有降水量资料，但 1944 年 8 月～1945 年 10 月，1938 年 12 月，1898 年 1 月缺资

料。用直线回归方法，选用邻近、相关较好的昭平降水量资料插补1938年12月和1944年8月～1945年10月的月降水量。选用相关较好的北海降水量资料插补1898年1月降水量。共建立13个直线回归方程，相关系数为0.4630～0.8146，信度为<0.05到<0.001。

经过插补得到梧州1898年1月～1992年12月逐月降水量资料，由此求得梧州1898～1992年逐年降水量序列。样本容量为$N=95$。

2.2.2 梧州年降水量序列的周期分析

对1898～1992年梧州年降水量进行最大熵谱分析，当截止阶取$m=42$时，谱图（图略）反映的周期点按谱的强弱依次是103、32、2、8、10、12、17、5年，共8个可能周期。当取截止阶$m=77$时，还有一个46年的可能周期。

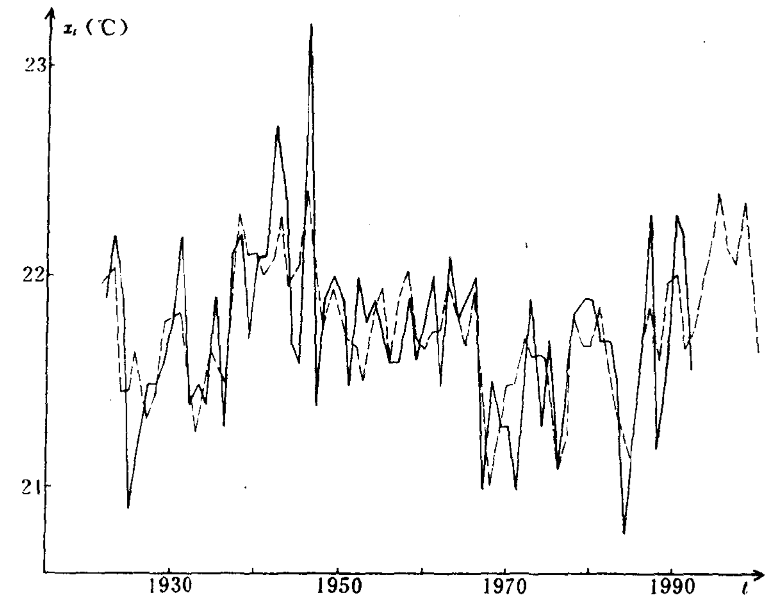

图2 南宁年平均气温实测值（实线）与计算值（虚线）曲线图

Fig.2 Curve of observation (Solid line) and calculation (Dashed line) of annual mean temperature in Nanning

按前述方法，对梧州年降水量序列进行周期分析，得$\alpha<0.05$的周期10个（约等于样本容量的1/10），各周期有关参数列于表3。

值得一提的是：梧州年降水量最显著的周期是102.4年，这个周期长度超过了样本容量（$N=95$），而且谱图也反映得比较精确（$m=42$时，谱图最大峰值点$T=103$），显示了用最大熵谱方法分析气候序列超过样本长度的周期的能力。

2.2.3 梧州年降水量的拟合与预测

根据表3，即可得到梧州年降水量的预测方程：

$$\hat{x}_t = 1384.834319 + 139.456\sin\frac{2\pi}{102.4}(t-1938.9) + \cdots + 52.73582\sin\frac{2\pi}{10.7}(t-1905.0)$$

对1898～1992年梧州年降水量进行拟合，计算实测值x_t与计算值$\hat{x}_t$的相关系数，得

$r(\hat{x}_t, xt) = 0.6925$

经显著性检验，信度$\alpha < 0.001$。

应用预测方程对1993～2000年梧州年降水量的计算结果列于表4。

由表4可知，计算的1993～2000年梧州年降水量的8年平均值为1239 mm，比梧州年降水量的多年平均值1385 mm偏少．比本世纪梧州降水量较少的30年代的平均值1174 mm偏多。

表3 梧州年降水量的周期及其参数

Table 3 Periods and parameters of annual amount precipitation time series in Wuzhou

i	T_i	A_i	B_i	t_i	r	α
1	102.4	1381.462	139.456	1938.9	0.4042	<0.001
2	2.9	0.6217166	91.67076	1899.6	0.2823	<0.01
3	31.0	1.340578	88.12884	1907.5	0.2860	<0.01
4	9.8	−1.305327	79.65746	1899.9	0.2680	<0.01
5	8.7	0.4607251	64.86475	1904.1	0.2254	<0.05
6	7.7	0.5693171	68.48439	1902.2	0.2439	<0.02
7	45.7	1.190759	62.09185	1903.2	0.2270	<0.05
8	19.4	−1.070731	55.32769	1902.5	0.2080	<0.05
9	2.0	0.4064142	38.60941	1898.5	0.2114	<0.05
10	10.7	0.7125466	52.73582	1905.0	0.2075	<0.05

表中符号意义同表1 Meaning of symbols in table 3 same as table 1.

表4 1993～2000年梧州年降水量计算值

Table 4 Calculation of annual amount precipitation for the period of 1993 to 2000 in Wuzhou

t	1993	1994	1995	1996	
$\hat{x}_t$ (mm)	1354.6	1222.3	1157.4	1219.7	
t	1997	1989	1999	2000	8年平均 Average
$\hat{x}_t$ (mm)	1229.8	1077.6	1367.0	1283.9	1239

实例分析说明，用最大熵谱分析与正弦函数拟合相结合，分析气候序列的周期的方法是可行的，对气候序列和其他时间序列的周期分析具有实际意义。

参考文献

1 罗乔林．介绍最大熵谱估计．数学的实践与认识．1980.（1）：74～80.

2 黄嘉佑，李黄．气象中的谱分析．北京：气象出版社，1984．79～96.

（责任编辑：蒋汉明 梁积全）