气象统计应用讲座

第四讲　气候状态变化的持续性和周期性分析

黄嘉佑

（北京大学地球物理系，100871）

研究前期气象要素对后期的关系能反映前期外力因素及气象要素自身对后期气象要素的影响，这种影响称为要素持续性。气候状态的形成是多种具有不同尺度变化的大型气候因子影响的共同结果。因此，通过对气候状态变化的不同尺度结构的分析来了解外力的影响是十分必要的。气候序列变化的结构通常用不同周期振动来表现，即从频域来进行分析，亦即所谓谱分析。另方面，还可以用不同尺度的过滤器来放大或缩小某种尺度的变化，从而达到了解某种尺度的气候变化演变特征。这方面的分析工具是各种过滤器和近年来发展起来的小波分析。本讲将介绍这方面的应用。

1　气候变化过程的持续性

天气过程的持续性是大气过程的可预报性研究的重要方面。大多使用自相关系数和自回归模式进行研究。

1.1　自相关系数与大气环流的持续性

计算各个网格点的不同落后步长（天）的自相关系数，用它们来度量大气过程逐日持续性程度。显然，自相关系数为正值且很大时，表明大气过程有高持续性。自相关系数计算用如下公式：

$$r(\tau)=\sigma(\tau)/\sigma^2 \tag{1}$$

其中自协方差 $\sigma(\tau)$ 用下式估计：

$$\sigma(\tau)=1/n\sum_{i=1}^{n-\tau}x(i)x(i+\tau)-(\bar{x})^2 \tag{2}$$

其中 $\bar{x}$ 及 σ 为样本均值和标准差。

在研究高度场的持续性中，自相关系数的计算常在网格点中进行。而网格点资料易受各种随机因素影响，其稳定性不强，在此基础上的持续性研究也受到影响。为克服这一缺点，可以用模式相关来表征某一局地范围内环流或天气系统的持续性。定义区域第 t 天和第 $t+\tau$ 天的模式相关为

$$p(t,\tau)=\frac{\langle z(x,t)z(x,t+\tau)\rangle-\langle z(x,t)\rangle\langle z(x,t-\tau)\rangle}{s[z(x,t)]s[z(x,t+\tau)]} \tag{3}$$

其中

$$s^2[z(x,t)]=\langle z^2(x,t)\rangle-[\langle z(x,t)\rangle]^2$$

为区域样本方差，$\langle\ \rangle$ 表示区域权重平均，$z(x,t)$ 表示在 t 时刻区域内格点的标准化距平值。上式实际表征不同天两区域之间的空间相关系数，以此来表示环流模式持续性大小。若 $p(t,\tau)$ 接近于 $p(t+1,\tau)$，$\tau=1,\cdots,5$ 时，表示环流系统在 5 天内处于准定常状态，因此可用来描述天气系统的阻塞过程。

对雨季降水持续性研究也是十分重要的。但是研究对象是分隔一段时间的不连续资料，怎样对它们作持续性研究呢？例如对某地区的 J 年雨季（6 — 8 月（$n=92$ 天））逐日降水距平资料，使用下式计算不同落后步长自相关系数

$$r(\tau)=\frac{\sum_{1}^{J}\sum_{i=1}^{n-\tau}x(i)x(i+\tau)}{\{[\sum_{1}^{J}\sum_{i=1}^{n-\tau}x^2(i)][\sum_{1}^{J}\sum_{i=1}^{n-\tau}x^2(i+\tau)]\}^{1/2}} \quad (5)$$

用自相关系数表示持续性时，持续性的检验可以用对自相关系数的检验来代替。例如检验落后一个时刻的持续性时，在显著水平5%下，把计算值$r(1)$与下式临界值进行比较

$$r_c=\frac{-1+1.645(n-2)^{1/2}}{n-1} \quad (6)$$

若$r(1)>r_c$则认为有持续性存在。

为了估计气候序列总体的持续性情况，还需要求不同落后步长自相关系数的95%置信区间。因为自相关函数不遵从正态分布，用常规方法求置信区间有困难。可以使用Fisher的z变换，用如下公式把相关系数r转换为新变量z。

$$z=(1/2)\log[(1+r)/(1-r)] \quad (7)$$

新变量z的95%置信区间为

$$z\pm1.96[1/(n-3)]^{1/2}$$

然后利用式(7)可从z的95%置信区间转换为r的95%置信区间。n为样本容量，如果序列是随机独立样本的话，那么其自由度等于序列的样本容量，但有些变量逐日变化有持续性而不是独立的。因此需考虑用有效自由度，即使用下式计算有效自由度：

$$N_{eff}=n/T_0 \quad (8)$$

式中T_0为有效样本值之间的特征时间，即

$$T_o=1+\sum_{\tau=1}^{n}2(1-\tau/n)r(\tau) \quad (9)$$

1.2　自回归与持续性

在描述天气过程的变化中往往用自回归模式来描述。在方程中序列落后时刻的关系就反映其变化的持续性。最简单的自回归模式是一阶自回归模式，又称一阶马尔可夫过程，或称红噪音过程，即天气状态的变化仅与前一天状态有关。例如Trenberth(1985)对南半球1000和500hPa高度场作了逐日的持续性分析，发现在中纬度大部分地区用一阶自回归模式拟合得很好，但在极地和低纬地区用高阶自回归模式更为合适。

1.3　相关系数

在考虑长期过程持续性的度量方面，也可以用相关系数来衡量。例如考虑6月的热力持续性，可以计算6月多年的序列与7月序列的相关系数。其持续性的检验用相关系数检验代替。但是，如果考查特殊年份与一般年份不同的持续性特征，用统一一个样本下计算的持续性相关系数就不好度量了。这时可以用分类相关系数来度量其持续性。例如研究秋季到冬季700hPa环流形势场的持续性时，用

$$p_1=\frac{1}{(N_1S_xS_y)}\sum_{i=1}^{N_1}(X_iY_i) \quad (10)$$

$$p_2=\frac{1}{(N_2S_xS_y)}\sum_{i=1}^{N_2}(X_iY_i) \quad (11)$$

表示特殊年和一般年份的持续性。式中X_i和Y_i分别为秋和冬季形势场高度距平，S_x和S_y分别为两样本标准差，N_1为特殊年(如ENSO年)样本，N_2为一般年(如非ENSO年)样本。当N_1或N_2有一个为0时就是相关系数，但一般来说p_1，p_2不是严格的相关系数，它们的值有可能超过1.0。它们的显著性检验可用Monte Carlo法作检验。

1.4　概率

对降水和气温的持续性也可用转移概率来度量。把气象要素某种状态用状态转移的观点进行研究。从马尔科夫链的观点来看，一种状态持续而不转移或一种状态转移到其他状态的概率很小时，定义此状态有持续性。显然，若将天气气候现象按其属性划分为两种或多种状态，例如晴雨、干湿日、高低温和旱涝年等。这样就可把天气气候序列看成状态离散、时间离散的马尔科夫链。设有两种状态A和B。令第i日为A状态，而第$i+n$日仍为

A 状态的转移概率记为 $P_n(A|A)$；第 i 日为 B 状态、第 $i+n$ 日仍为 B 状态的概率类似地有 $P_n(B|B)$。按照上述观点当状态的转移概率大于状态的气候概率时，认为序列有持续性。

事件重现的概率，例如为考察北京地区旱涝现象年持续性可用对应事件重现的概率进行分析。记前一年出现(旱)涝事件为 A，第二年仍为(旱)涝事件为 B，设它们出现是相互独立，则(旱)涝现象持续一年的概率(即两事件同时发生的概率)为

$$P(AB)=P(A)P(B)$$

同样，持续二年的(旱)涝事件的持续性度量可用概率

$$P(ABC)=P(A)P(B)P(C)$$

表示。其中 $P(C)$ 为第三年(旱)涝事件出现的概率。根据各年旱涝出现的概率(用 Gamma 分布求出)可计算出持续二年或三年的旱涝出现的概率。附表给出该地区不同气候时段(30 年)旱涝事件持续性比较。从表中可见，干旱的持续性随气候时段演变是增加的，涝的持续性却减少。

附表　不同气候时段北京旱涝事件持续性比较

时段	两年		三年	
	旱	涝	旱	涝
1870—1899	0.172	0.655	0.143	0.571
1900—1929	0.448	0.207	0.393	0.286
1930—1959	0.483	0.379	0.464	0.393
1960—1989	0.517	0.345	0.500	0.250

用转移概率可以构成某些统计量来度量持续性。例如在分析逐日降水持续性中假设降水出现过程为两状态(即有雨和无雨)，则降水持续性参数为

$$d=P_{11}-P_{01} \tag{12}$$

其中 P_{11} 和 P_{01} 分别为前一天有雨第二天有雨和前一天无雨第二天有雨的概率。显然，当 $d=0$ 时，无降水的持续性。

还有 Besson 持续系数，定义为

$$R_b=\frac{1-P}{1-P_{11}}-1 \tag{13}$$

式中 P 为雨日在样本中出现概率，P_{11} 为前一天有雨第二天仍有雨的概率。

还可以分别计算湿日 (W) 和干日 (D) 的期望持续天数，计算式为

$$E(W)=1/(1-P_{11}) \tag{14}$$

$$E(D)=1/(1-P_{00}) \tag{15}$$

式中 P_{00} 为前一天无雨第二天仍无雨的概率。一次天气过程是湿日和干日交替演变过程，因此天气过程周期变化天数还可以用

$$E(C)=E(W)+E(D) \tag{16}$$

来估计。

2　周期分析

在气象要素时间序列的演变规律中，周期性有很强的表现，如年变化现象。但是还存在其它的周期变化，对一时间序列 $x(t)$，若存在间隔一段时间 T 后有 $x(t+T)=x(t)$ 成立，则可以称该时间序列存在以 T 为周期的变化。如果其关系近似成立则可称为准周期波动。常见的周期变化是指序列呈规则的正弦波和余弦波动，但是也可以是不规则的波动，当然这种不规则波动也可以看成由若干个规则波动迭加形成。检测出不同周期波动是气候成因分析的重要方面，因为如果序列中含有的主要振动与大型气候因子(如太阳活动或太阳月亮共同作用或 ENSO 现象等)的振荡周期一致的话，则可认为它可能是局地气候变化的主要影响因子。使用的检测方法常用的有方差分析和谱分析方法。

2.1　方差分析方法可以检测出时间序列演变的不同周期的不规则波动分量。即序列 x_i ($i=1,n$)，用统计量

$$F=\frac{\sum_{i=1}^{k} n_i(\bar{x}_i-\bar{x})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2/(n-k)} \tag{17}$$

它遵从分子自由度 $(k-1)$ 分母自由度 $(n-k)$ 的 F 分布，式中 $\bar{x}_i$ 为第 i 组平均值，$\bar{x}$ 为序列平均值，n_i 为第 i 组样本容量，以不同周期长度作为分组数 k，计算 F 值，最大值对应的分组平均值作为不规则波动分量的估计。

2.2 谱分析方法可以检测出时间序列演变的不同周期分量。常用的有:(1)BT(Blackman—Tukey)谱,谱分析是从频域角度对时间序列进行分析的方法。它用序列的落后相关系数求不同频率波动的功率,即

$$S(k)=\frac{B_k}{m}\left[r(0)+\sum_{\tau=1}^{m-1}r(\tau)\left(1+\cos\frac{\pi\tau}{m}\right)\cos\frac{k\pi\tau}{m}\right]\quad (k=0,1,\cdots m)\tag{18}$$

式中 B_k 为1/2(当 k 取0和 m 时)和1(当 k 取其余值时),m 为取的最大落后步长。此式计算的功率谱称为标准化功率谱密度,以波数 k 为横轴,以标准化功率谱密度为纵轴作图,可得标准化功率谱密度图,简称为谱图或功率谱图。谱图中的横轴除使用波数外,一般还标出对应的周期或频率。由于标准化谱密度无单位,纵轴通常不标单位。若将式中的自相关函数改为自协方差,也得到相应的谱图,但是此谱图的纵轴应带有功率单位(即要素单位的平方)。

在谱图中除绘出各周期振动的功率值比较外,通常还绘出两种显著检验的曲线。一种是形如指数曲线,称为红噪声功率谱95%置信区间上限,它是在序列落后1时刻的自相关系数值显著大于0时使用。另一种是与横轴平行的直线,称为白噪声功率谱95%置信区间上限,它是序列落后1时刻的自相关系数值小于0时使用。

(2)FFT 谱,又称为傅立叶谱,或直接谱,也有称方差谱,它是比较各周期振动的方差贡献。序列中第 k 个波的振幅方差表示为

$$A_k^2=1/2(a_k^2+b_k^2)\quad (k=1,2\cdots,n/2)\tag{19}$$

$$a_k=2/n\sum_{t=1}^{n}x_t\cos\frac{2\pi k}{n}(t-1)$$

$$b_k=2/n\sum_{t=1}^{n}x_t\sin\frac{2\pi k}{n}(t-1)$$

式中 x_t 为序列 t 时刻距平值,n 为样本容量。它们可以用快速富里叶变换(FFT)计算,在计算时要求序列样本容量正好为2的正整数幂次,如不足,可以用0值补充。

(3)非整谱。在一般的功率谱分析方法(直接谱和间接谱)中不同周期功率的分辨只能在序列长度的最大周期范围内,而且分辨率高的谱段是在高频段。如对18个候的降水序列,仅能分辨18候以内的振荡,而且只有2—9候之间,特别是2—4候的振荡有较高的分辨率。如果我们希望研究低频振荡(周期在8—12候之间),那么高分辨率的频段很难分辨。非整波则可以解决这种困难,用它除了能在序列长度范围内有任意高的分辨率外,还可以分析长于序列长度的周期的谱,即能分析出大于18候的季节间振荡频谱。设某年主分量 y 序列长度为 N,则可把它看成由若干个非整正弦波和余弦波迭加而成,即表示为

$$y(t)=c+\sum a_k\cos(2\pi kt/N)+b_k\sin(2\pi kt/N)+e\tag{20}$$

其中 k 为非整正数,a_k 及 b_k 为系数,c 为常系数,e 为误差。上式可看成为一回归方程,系数可用最小二乘法定出。用偏回归方差贡献为纵轴以试验周期为横轴构成功率谱图。

(4)非线性谱。常用的最大熵谱(MEM),是一种非线性谱,它不同于普通的线性谱分析方法,它避免了数据周期性延伸的假定,也不需假设所得的记录长度以外的数据为零(如FFT方法)。由它分析得到的谱有较高的分辨率,它可以用于记录长度较短的序列。

熵是信息论中用来衡量随机事件不肯定性程度的量。当随机变量的方差较大时,有较大的熵值。当振动的熵值越大,表明其方差贡献越大。最大熵的功率谱为

$$S_h(l)=\frac{\sigma_{k_0}{}^2}{\left[1-\sum_{k=1}^{k_0}b_k\cos(\pi lk/m)\right]^2+\left[\sum_{k=1}^{k_0}b_k\sin(\pi lk/m)\right]^2} \tag{21}$$

式中 σ_{k_0} 为预报误差均方差，k_0 为最小预报误差自回归方程阶数，b_k 为满足向前和向后预报误差最小的自回归方程系数，l 为波数，m 为取的最大落后步长，有人建议 m 取为 $2n/\ln(2n)$，n 为序列样本容量。

(5)动力谱。不同时期的资料序列，其谱分析结果是不同的。因此有人建议为了较好地作谱估计，用滑动样本来作谱估计。一方面可以研究不同时期序列谱结构的变化。另方面可以将滑动后某频段的谱值作平均来估计序列的谱结构。

(6)单序列不规则周期模式。一般的方差分析可以诊断出主要的不规则周期，它们用周期平均序列来表现。但是，不同的周期段实际周期变化是不同的。黄嘉佑(1990)把不同周期的振动排列成二维矩阵，然后用EOF分析寻找存在于序列中的真实周期波型模式，取得比传统的方差分析好的效果。

(7)奇异谱分析。它是近来新发展的谱分析方法。对一时间序列 $x(t)$，$t=1,2,\cdots,n$，经不同排列可转化为一二维矩阵为

$$\boldsymbol{X}_{[(n-m+1)\times m]}=\begin{Bmatrix} x(1) & x(2) & \cdots & x(m) \\ x(2) & x(3) & \cdots & x(m-1) \\ x(n-m-1) & x(n-m-2) & \cdots & x(n) \end{Bmatrix}$$

对此新资料阵EOF分析则可求得相应的特征向量和时间函数。不同特征向量反映在 m 点时间间隔的不同落后时刻的变化特征，取不同的 m 值进行分析则可得到不同周期振荡方差贡献，比较它们在EOF中的方差贡献大小即得谱图，从中可发现主要周期振荡。不过此周期不同于经典谱分析的周期波动，是不规则波动。

3　小波滤波

频率滤波，把气候变化中某些频率振动的波动过滤出来，并把它们转变为时间序列，以便表现具有某种频率振荡的外力对局地气候变化的影响。除常用的低通、高通、带通过滤器外，还可以使用近年来发展的小波分析新方法。小波又称子波(wavelet)，小波变换实质上是一种数据平滑过程，因而也是一种过滤。在形式上类似傅立叶变换，只不过基底函数不是指数函数而改用小波基底函数。对实数域上的一个可测的平方可积函数 $x(t)$，它的小波变换可表示为

$$W(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)g\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{22}$$

式中 a 是伸缩尺度，b 是平移因子，有称为窗口。由于 b 是平移因子，对一时间函数 $x(t)$ 取为某一时刻 t_0，则在该时刻具有尺度 a 的墨西哥帽小波变换表示为

$$W(a,t_0)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)g\left(\frac{t-t_0}{a}\right)\mathrm{d}t \tag{23}$$

$g(t,a,b)$ 是一个基本小波函数，可根据研究的问题的特点来选择。常用有墨西哥帽、Morlet 小波和框架帽等。墨西哥帽的函数形式为

$$g(x)=(1-x^2)\exp(-x^2/2) \tag{24}$$

对一时间序列 $x_t(t=1,2,\cdots n)$，可把积分换成求和形式。将 t_0 取值遍历所有样本，这相当于对序列作平滑处理，其权重函数的分布形如墨西哥帽。此权重函数与正态分布很相似，尺度 a 越大相当于散布越大，低频分量过滤出来的就越多，反之亦然。