

第三讲 气候状态变化趋势

与突变分析

黄嘉佑

(北京大学地球物理系, 100871)

本讲着重介绍怎样对气象要素时间演变序列作趋势与突变分析。

1 气候状态变化的趋势

天气过程长期演变趋势是气候变化或变迁研究的重要内容。所谓趋势是指气候要素大体的变化情况, 即描述很长时间尺度的演变过程。因而它能反映大尺度气候因子的影响。其研究方法大致有如下几种。

1.1 直线与曲线趋势

直线是最常用来表征气候变化的最大时间尺度的演变趋势, 直线的波长为无穷大, 它能反映最大的气候因素对局地气候变化的作用(即反映大型气候因子的综合作用)。利用气象要素时间序列, 用以时间为自变量, 以要素为因变量, 建立一元回归方程, 即直线方程, 其直线即为序列的直线变化趋势。表示为

$$x(t) = b_0 + b_1 t \quad (1)$$

$x_i (i = 1, 2, \dots, n)$ 为序列值, b_0 称为方程的截距, b_1 为斜率, 它们可以用最小二乘法定出, t 为样品序号, 当时间是用年份表示时, 方程中的截距和斜率值则有改变。如果令序列起始年份记为 a , 截距和斜率分别记为 b'_0 和 b'_1 , 容易导出

$$b'_0 = b_0 - b_1(a - 1)$$

$$b'_1 = b_1$$

往往在序列变化图上除绘出要素的变化

曲线外同时绘出其拟合直线, 从图中观察趋势演变是增加还是减少。直线的斜率的符号及大小可以度量其演变趋势是增加还是减少的程度。其显著程度还可对回归系数作显著性检验来确定。例如根据北京夏季 1951—1980 年降水量资料作趋势分析, 其方程为

$$x(t) = 564.6 - 5.286t$$

当起始年份定为 1951 时, 其方程为

$$x(t) = 10872.3 - 5.286t$$

从中可见, 地球气候因子的综合作用使北京地区在 30 年来夏季有变旱趋势。进一步根据概率划分级别的方法把 1870—1989 年北京各季降水量进行旱涝级别划分, 得到各季的旱涝级别序列。为研究不同季节旱涝多年变化趋势, 我们对各季旱涝级别序列作线性趋势分析。发现冬春与夏秋两季有相反的变化趋势, 前者为变涝, 后者为变旱, 其线性回归系数分别为 -0.004 , -0.003 , 0.006 和 0.002 。发现夏季的干旱趋势是较其它季节突出, 反映地球气候因子的综合作用使北京地区在近百年来夏季变旱。对长期变化趋势也可以用长尺度周期综合的曲线形式来表现, 即对序列用非线性多项式来拟合序列变化, 以表现气候的曲线变化趋势。例如, 用二次多项式

$$x(t) = b_0 + b_1 t + b_2 t^2 \quad (2)$$

来表现序列的长期变化趋势,反映长尺度周期变化气候因子的综合影响。同样,当时间是用年份表示时,方程中的系数亦有改变。如果令序列起始年份记为 a ,各系数分别记为 b'_0 、 b'_1 和 b'_2 ,容易导出

$$b'_0 = b_0 - b_1(a-1) + b_2(a-1)^2$$

$$b'_1 = b_1 - 2b_2(a-1)$$

$$b'_2 = b_2$$

对北京上述资料求得方程为

$$x(t) = 48495.1 - 43.5699t + 0.00973897t^2$$

从中可见,一次项系数为负,说明降水量的变化在大趋势上变旱;二次项系数为正,表明在近期有向偏多的趋势发展。从1980年至今的观测资料中的趋势变化似乎得以证实。

如果希望表现周期尺度更短些的气候因子的综合影响,可以用三次或四次多项式来表现。

1.2 低通滤波

它是把序列高频分量滤去以便突出长期或气候变化趋势的一种方法。常用的低通过滤方法是对序列作滑动平均,又称滑动平均法。其过滤后序列为:

$$\hat{x}(t) = \sum_{i=-L}^L w_i \cdot x(t-i) \quad (3)$$

$x(t-i)$ 表示过滤前序列,其权重为 w_i , $(2L+1)$ 称为滑动区间。权重可以相等也可以不等,但是必须满足所有权重之和等于1。平均的过程实际是压低小波动(即短周期振动)的影响,同时突出长周期波动的作用,因而也起到反映长周期气候因子的综合影响的作用。如对序列作5年等权重滑动平均时, L 取为2,有5个滑动权重,每个为0.2。滑动后序列能消除5年内短周期的影响,表现5年以上周期波动的综合影响。所取的滑动区间越大,过滤掉的短周期越长。但这种过滤会削弱过

滤后振动的振幅和改变其位相。不少人设计各种滤波器以便克服这一缺点,常用的是不等权重滑动平均。一般用权重分布形状来设计过滤权重。其中有二项系数过滤,此过滤器权重值的分布遵从二项分布,突出滑动中心点的作用。常用的有三点平滑的二项系数过滤器,滑动区间中 L 取为1,其权重分布值如表1所示。

表1 三点二项系数过滤器权重

i	i	$i \pm 1$
w_i	0.50	0.25

在序列两个端点由于没有前和后数据,可用端点后一点和前一点的数据代替,使得在过滤后的序列样品数与过滤前一样。显然,如果序列是年变化序列的话,过滤掉的是3年以内的短周期波动。滑动区间越大,平滑掉的波动周期越长。如取过滤器为九点二项系数(具体数值如表2所示),可以滤去9年内的波动周期。过滤时,为使过滤后序列样本容量个数不至于损失,其端点前后要插补4个资料,它们均用序列平均值代替,或以端点为中心将有资料的数据对称地外延到无资料的相应位置上来代替。

表2 九点二项系数过滤器权重

	i	$i \pm 1$	$i \pm 2$	$i \pm 3$	$i \pm 4$
二项权重	0.22	0.20	0.12	0.05	0.02
正态权重	0.24	0.20	0.12	0.05	0.01

除权重分布遵从二项分布外,还可以用遵从正态分布的权重。其权重分布值按正态分布概率密度曲线形状分布(如表3所示)。从表3可见,它与二项分布权重相差不大,而二项分布权重容易求。所以常用二项分布权重作低通过滤。

还可以用其它近似二项分布的权重,权重分布较二项分布更平缓。如对序列 $x(t)$ 使用五点过滤器,过滤器权重分布如表3。

表3 五点过滤器权重

	i	$i \pm 1$	$i \pm 2$
近似权重	0.330	0.250	0.083
二项权重	0.374	0.250	0.063

1.3 差分过滤

将序列后期值减去前期值即得差分序列:

$$\begin{aligned} dx(t) &= x(t+1) - x(t) \\ (t &= 2, 3, \dots, n) \end{aligned} \quad (4)$$

由于差分是一种高通滤波器,它能过滤出含高频振动的综合序列。那么将原序列减差分序列令

$$\begin{aligned} Lx(t) &= x(t) - dx(t) \\ (t &= 2, 3, \dots, n) \end{aligned} \quad (5)$$

得一低频段综合序列。当然,为了保持序列样本容量不变,也可以使用平均值外延序列。

1.4 累积距平曲线

用累积距平曲线的变化也可作气候趋势分析。对一距平值序列 $x_{dt}(t=1, n)$ 。在样本中某一时刻 t 的累积距平值表示为

$$I(t) = \sum_{i=1}^t x_{di} \quad (6)$$

把 $I(t)$ 随时间变化绘成曲线称为累积距平曲线。在累积距平曲线变化中,上升表示要素距平值增加,下降则减少。曲线上的微小变化所反映的是要素短期距平值的变化,而长时期的曲线演变则可反映要素长期演变趋势。在式(6)中也可使用标准化距平值来代替,相应的称为标准化累积距平曲线。

1.5 用最优分割法作分段趋势分析

例如对我国东北夏季气温的趋势变化分析,对该地区 1881—1970 年气温资料提取主要因子序列,然后用这种方法对第一因子序列作分割,结果分成 3 段:1881—1917 年,1918—1951 年和 1952—1970 年。第 1 时段

序列平均值为负距平值,第 2 段为正距平值,第 3 段接近于 0。说明在 1917 年以前气温是下降的,其后为上升阶段。从其累积距平曲线也发现在 1917 年为最小值。说明这一趋势分析结果是一致的。

1.6 趋势的统计量及其检验

1.6.1 差分平均值。对差分序列求其平均值,记为 mdx 。显然,若它大于 0 时,表明序列后面的值比前面的值普遍高,即有增加趋势,反之亦然。它等于 0 时,表明序列无趋势。对差分序列平均值还可作检验,使用统计量为

$$t = mdx(n-1)^{1/2}/s \quad (7)$$

式中, n 为差分序列的样本容量, s 为序列的标准差。该统计量遵从自由度为 $(n-1)$ 的 t 分布。

1.6.2 秩统计量。对给定的序列 $x_i (i=1, 2, \dots, n)$, 构造一秩统计量 Z , 定义为

$$Z = \frac{4 \sum_{i=1}^{n-1} N(i)}{n(n-1)} - 1 \quad (8)$$

其中, $N(i)$ 为序列 $x(t) (t=1, 2, \dots, n)$ 中大于 $x(i)$ 值样品个数。若 Z 值在 $-1 \sim 1$ 间变化,对递增直线或曲线,其 $N(i)$ 序列为 $n-1, n-2, \dots, 1$ 则

$$\begin{aligned} 4 \sum_{i=1}^{n-1} N(i) &= 4(n-1)(1+n-1)/2 \\ &= 2n(n-1) \end{aligned}$$

即, Z 为 1, 对递减和无变化趋势序列, Z 为 -1 。因此 0 值可以作为有递增或递减趋势的分界线。若 Z 的绝对值大于如下判据值

$$Z_c = 1.96 \left[\frac{4n+10}{9n(n-1)} \right]^{1/2} \quad (9)$$

则表明序列有显著的变化趋势(显著水平为 5%)。

2 气候突变

在天气过程的变化中存在着某种不连续现象,这种不连续性称为突变。根据突变原因

可划分为两类。其一称简单突变,它常发生在天气系统内部,即不考虑系统边界外力影响下或外界气候系统没有大的变化情况下出现的突变,如初霜冻和季风爆发等。这类突变常有周期变化,例如年变化周期。另一类为由于外力的突变使天气过程发生突变。这类突变现象不是有规律性的。近年来,气候突变成为长期天气过程变化研究的重要课题。另一方面,突变还常发生在台站的迁站址后的气象要素资料中,通过对突变现象的分析可以判断历史上何时发生迁站。因此,在判断气象要素出现突变时,还需要判断是自然气候变化的突变,还是由于人为的迁站的结果,或者是由于城市热岛效应所带来的影响。序列突变分析大致有如下方法。

2.1 信噪比

天气过程的突变常用气候状态分布特征量(例如平均值或方差)随不同时段的变化来衡量,Yamamoto 等人(1986)提出把几十年作为气候阶段,以相邻气候阶段的平均值的显著差异定义为气候突变。方法是滑动地选择某个参考年,分别计算两气候阶段内的平均值之差和标准差(σ_1 和 σ_2)。定义信噪比

$$S/N = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_1 + \sigma_2} \quad (10)$$

作为相邻两气候阶段差异的统计量,当这一信噪比数值大于 1.0 时认为存在气候突变。

2.2 t 统计量

还可以用 t 分布检验两相邻时段序列平均值有无显著差异来判断序列变化趋势的不同(参见第二讲)。例如 Cooper 等人对热带地区 Pitcairn 和 Tahiti 站月气压近 100 年序列的趋势进行 Mann-Kendall 分析,发现,气压变化趋势是增加的,近 25 年(1950—1975)气压值上升了 0.75—1.5hPa。与前一时段用平

均值比较其 t 值分别为 2.23 和 2.40,已达 5% 显著水平。说明近年来气压增加是非随机性的。

2.3 Mann-Kendall 秩统计量

定义为

$$Q = \sum_{i < j} h_{ij} \quad (11)$$

其中

$$h_{ij} = \begin{cases} +1 & \text{当 } x_i > x_j \text{ 时} \\ 0 & \text{否则} \end{cases}$$

这一统计量即是考查序列中各时间点(j)前面的要素值大于该时刻的要素值的个数的总体情况。如该序列是递降序列 Q 将有较大的正值,其值趋于 n 。如为递升序列则其数值趋于 0。利用 Mann-Kendall 检验作趋势分析有广泛的应用。在变量任一时刻均遵从同一分布的假定下,统计量

$$U = [Q - E(Q)] / [\text{Var}(Q)] \quad (12)$$

渐近遵从标准正态分布。式中

$$E(Q) = n(n+1)/4$$

$$\text{Var}(Q) = n(n-1)(2n+5)/72$$

为 Q 的数学期望和方差。如 U 的绝对值大于 1.96 时,则表明序列有非随机性变化趋势(显著水平为 5%)。统计量 Q 是由时间序列按时间顺序产生的。也可以按时间逆序产生。即对原序列按 $j = 1, 2, \dots, n$ 取不同样本容量计算一系列 U 值,在图上点出形成向前 U 曲线。另外按 $j = n, n-1, \dots, 1$ 的样本容量计算可得后向 U 曲线。如果在某一时间点上有相交,且其绝对值小于 1.96 时,则可定为趋势突变点。

2.4 K 统计量

使用

$$U(t) = \sum_{i=1}^t \sum_{j=i+1}^N \text{sgn}(x_i - x_j) \quad (13)$$

式中 $\text{sgn}(x_i - x_j)$ 表示取序列(下转封三)

(上接第 57 页)

$x_i (i = 1, 2, \dots, N)$ 的第 i 与第 j 个值之差的符号, 若正值记为 1, 若负值记为 -1, 若相等记为 0。然后取其中最大者记为

$$K = \max_{1 \leq i \leq N} |U(i)| \quad (14)$$

则它所对应的年份 t 为可能突变年份。其突变出现概率近似为

$$P = 1 - \exp\left[-\frac{6K^2}{(N^2 + N^2)}\right] \quad (15)$$

还可以使用

$$K' = \max_{1 \leq i \leq N} \frac{|U(i)|}{(N_i - t^2)^{1/2}} \quad (16)$$

代替 K 作为检验突变年份的统计量, 式中 N_i 为 t 取不同时刻的样本容量。

2.5 累积距平统计量

即使用

$$C_d(t) = \sum_{i=1}^t x_{di} \quad (17)$$

或

$$C_z(t) = \sum_{i=1}^t x_{zi} \quad (18)$$

式中 x_{di} 和 x_{zi} 分别为序列距平和标准化距平值。若在某年份的累积距平值中, 绝对值出现极大值时, 所对应的年份为可能是突变年份。

2.6 双相回归分析

用它也可以找出序列的突变点。对时间序列 $x_i (i = 1, n)$, 其双向回归模型表为

$$x_i = \begin{cases} a_0 + b_{0i} + e_i, & i = 1, \dots, c \\ a_1 + b_{1i} + e_i, & i = c + 1, \dots, n \end{cases} \quad (19)$$

式中 e_i 为误差。设它是独立的且遵从平均值为 0 的正态分布。上式中的第一式表示突变点 c 以前的时段的拟合直线, 它反映该时段的线性趋势。第二式则反映后一时段的线性

趋势。需要求 c 的最大或然估计值, 其最大或然函数为

$$x_i = a_0 + b_{0i} + b(i - c)\text{IND}_c(i) + e_i \quad (20)$$

其中

$$b = b_1 - b_0 \quad (21)$$

$$\text{IND}_c(i) = \begin{cases} 0 & \text{当 } i \leq c \\ 1 & \text{否则} \end{cases}$$

对固定的 c 值上式为标准二元线性回归模型, 其因子为 i 和 $(i - c)\text{IND}_c(i)$ 。用最小二乘法容易求出回归系数 a_0, b_0 和 b 。根据式(21)可求出另一时段斜率 b_1 。Solow (1987) 用这一模型分析北半球气温 1858-1985 年序列的突变, 定出最大的突变点为 1887 年。对这一突变趋势是否显著还可以作检验。其统计量为

$$U = [(S_0 - S)/3]/[S/(n - 4)] \quad (22)$$

它遵从分子自由度为 3, 分母自由度为 $n - 4$ 的 F 分布。式中 S_0 为所有资料的直线回归方程的残差平方和, 即式(20)中 $c = i$ 时的方程的残差平方和。 S 为式(20)对应突变点的残差平方和。Solow 对北半球气温序列资料求出 U 值为 1.88, 在 5% 显著水平下是不显著的。但是若假定突变点在序列中抽样是独立情况下统计量

$$U' = (S_0 - S)/[S/(n - 3)] \approx 3U \quad (23)$$

遵从分子自由度为 1、分母自由度为 $n - 3$ 的 F 分布。在这种假定下 U' 值为 5.70。结果表明 1887 年的突变点是显著的。