

用 Bivariate-Test 方法分析温度 资料的不均一性*

王晓春 蔡雅萍 梁幼林

(中国科学院大气物理研究所资料中心, 北京 100080)

提 要

Bivariate-Test 方法(以下简称 B-T 方法)是一种基于二维正态分布随机矢量的假设检验方法,可以在一个分量的均值不改变的前提下,检验另一分量的均值有无改变。一些气象研究工作者最早用这种方法来分析年降水量的不均一性。

从温度资料不均一性产生的原因看,一般可分为两类:突变型不均一及渐变型不均一。对两种类型的不均一性,我们分别推导了理想情形下 B-T 方法的检验结果,并用随机试验的方法分析了检验结果对资料的敏感性。以上两方面的结果表明,渐变型及突变型不均一分别对应明显不同的 B-T 检验结果,在实际分析中这种差异可以用来定性判断不均一性的类型。

东北区 20 站 30 年年平均温度资料的试算结果表明,有 7 站资料不能通过信度 95% 的 B-T 检验,有 3 站资料不能通过信度 90% 的 B-T 检验。其中有 4 个站的检验结果与理想情形下渐变型不均一的检验结果相似,有 4 个站的检验结果与理想情形下突变型不均一的检验结果相似。有一个站存在个别点的资料错误,有一个站可能有一次以上的不均一性发生。

计算结果也表明,尽管 B-T 方法按照理论推导只是用来检测资料的不均一性,但对个别点上的资料质量问题也有一定的检测效果。

关键词: Bivariate-Test 方法;不均一性;渐变型;突变型。

一、引 言

温度资料的不均一性对温度的分析研究有很大影响。Hansen^[1] 及 Jones 等^[2] 在分析北半球地表温度的长期变化时,所用的原始资料绝大部分是一样的,得到的两条北半球温度变化曲线的相关系数高达 0.86。但 Elsner 及 Tsonis^[3] 的分析表明,这两条曲线有不同的平均值,呈现了不同的线性趋势,两条曲线并非来自同一个总体。

Hansen 及 Jones 等在得到北半球地表温度的长期趋势时,对于这一问题的两个关键部分: 1. 资料空间覆盖率不均匀, 2. 资料不均一性,分别采取了不同的方法,这应该是这两条曲线呈现差别的原因。

Jones 等利用与邻近站比较及台站历史信息分析了月平均温度资料的不均一性。首先将温度资料转变为相对某一标准时段的距平,与邻近站资料逐一相减,找出间断的地

1993年12月29日收到,1994年4月29日收到修改稿。

* 本研究由 KY85-10 项目支持。

方,最后设法用邻近站的数据进行订正。在所分析的 2666 站中,有 645 站(占 24.2%)由于没有邻近站进行比较而没有进行检查,资料不正确的有 236 站(8.9%),38 站(1.4%)有非气候的增暖,249 站(9.3%)经过均一性订正,即资料肯定有问题的站共占到 19.6%。

根据许多作者的结果(例如 Jones 等^[2]),造成温度资料不均一的原因主要有: 1) 仪器及观测方法的变更, 2) 站址变迁, 3) 观测时间及计算月平均温度方法的变更, 4) 台站周围环境的变化(如城市热岛效应)。与邻近站进行比较并利用台站历史信息是检测温度资料不均一性的基本方法。但这一方法在判断不均一性是否发生时缺乏一个客观的标准。

Bivariate-Test 方法(以下简称 B-T 方法)是一种基于二维正态分布随机矢量的假设检验方法。最早由 Maronna 和 Yohai^[4] 提出, Potter^[5] 用来分析台站年降水量序列的不均一性。在一定信度下,这一方法可以用来较客观地判断不均一性是否发生,并得到不均一性发生时间及大小的极大似然估计。

本文在介绍这一方法之后,分析了理想情形下两种不同类型不均一性的检验结果。并用随机试验的方法分析了检验结果对资料的敏感性。第四节用东北区 20 站 30 年年平均温度资料试验了 B-T 方法,结论在最后一节给出。

二 Bivariate-Test 方法

1. Bivariate-Test 方法

Bivariate-Test 方法最早由 Maronna 和 Yohai 引入。设序列 $\{x_i, y_i\} i = 1, 2, \dots, n$ 为相互独立的两维随机矢量,每一矢量来源于正态总体。在本文中,我们仅分析年平均温度的均一性,因而 x_i, y_i 可理解为年平均温度, i 为年。

原假设 $H_0: \{x_i, y_i\} i = 1, 2, \dots, n$ 具有同样的正态分布 $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ 。

H_1 假设: 对某一 $i_0, 0 < i_0 < n$, 当 $i \leq i_0$ 时, $\{x_i, y_i\}$ 的分布为 $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 当 $i > i_0$ 时, $\{x_i, y_i\}$ 的分布为 $N(\mu_x, \mu_y + d, \sigma_x^2, \sigma_y^2, \rho)$, 其中 $d \neq 0$ 。

统计量 T_0 用如下方法计算:

$$T_0 = \max_{i < n} \{T_i\}, \quad (1)$$

$$T_i = [i(n-i)D_i^2 F_i] / (S_x S_y - S_{xy}^2), \quad (2)$$

其中

$$D_i = [S_x(\bar{Y} - Y_i) - S_{xy}(\bar{X} - X_i)]n / [(n-i)F_i], \quad (3)$$

$$F_i = S_x - (X_i - \bar{X})^2 ni / (n-i), \quad i < n \quad (4)$$

$$X_i = \frac{1}{i} \sum_{j=1}^i x_j, \quad Y_i = \frac{1}{i} \sum_{j=1}^i y_j, \quad (5)$$

$$\bar{X} = X_n, \quad \bar{Y} = Y_n, \quad (6)$$

$$S_x = \sum_{i=1}^n (x_i - \bar{X})^2, \quad S_y = \sum_{i=1}^n (y_i - \bar{Y})^2 \quad (7)$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y}). \quad (8)$$

对于一个实际的序列,可以求出 T_i $i < n$ 及 T_0 ,对信度 $1 - \alpha$,当 T_0 大于相应的置信限时,则拒绝原假设,即 y_i $i = 1, 2, \dots, n$ 是不均一的。而对应 T_0 之 i_0^* 及 D_0^* 则分别是 i_0 及 d 的极大似然估计。从 H_1 假设应该注意到, i_0 为不均一性发生的前一年。对于分布 $N(0, 0, 1, 1, \rho)$, Maronna、Yohai 及 Potter 用统计模拟的方法计算出了不同置信度下的置信限。为实际应用的方便,在本文中也直接给出了他们的计算结果(见表 1)。

表 1 Maronna, Yohai 及 Potter 等用随机模拟方法得出的不同样本数、不同信度下的置信限^[1]

α 样 本 数	0.25	0.10	0.05	0.01
10	4.7	6.0	6.8	7.9
15	4.9	6.5	7.4	9.3
20	5.0	6.7	7.8	9.8
30	5.3	7.0	8.2	10.7
40	5.4	7.3	8.7	11.6
70	5.9	7.9	9.3	12.2
100	6.0	7.9	9.3	12.5

2. B-T 方法的物理意义

用 x_i 构造的 y_i 线性回归方程为

$$\hat{y}_i = \bar{Y} + (S_{xy}/S_x)(x_i - \bar{X}), \quad (9)$$

可以定义

$$\hat{Y}_i = \frac{1}{i} \sum_{j=1}^i \hat{y}_j = \bar{Y} + (S_{xy}/S_x)(X_i - \bar{X}), \quad (10)$$

而(3)式可以改写为

$$[(n-i)F_i/nS_x]D_i = \bar{Y} + (S_{xy}/S_x)(X_i - \bar{X}) - Y_i, \quad (11)$$

将(10)式代入(11)式, D_i 的表达式可化简为

$$D_i = (\hat{Y}_i - Y_i)/[(n-i)F_i/nS_x], \quad (12)$$

因而在 D_i 的表达式中,分子为 y_i 的线性回归方程拟合程度的一种度量,当 x_i 本身均一时, $\hat{Y}_i - Y_i$ 中的 $(S_{xy}/S_x)(X_i - \bar{X})$ 是一个变化不大的量,若这时 y_i 也均一,则 $\hat{Y}_i - Y_i$ 亦为一变化不大的量;若 y_i 不均一,即 y_i 的均值有系统性的改变,则显然 $\hat{Y}_i - Y_i$ 在 y_i 均值系统性变化的前一年达到最大值。事实上, D_i 随 i 的变化主要由 $\bar{Y} - Y_i$ 项决定,而 T_i 与 D_i^2 成正比,因而 B-T 方法的物理意义在于考虑 $\hat{Y}_i$ 与 Y_i 的差别来检测 y_i 是否均一。这里 x_i 当然首先应该是均一的。在实际计算时,我们取 x_i 为该站周围若干站温度序列的算术平均。

三、理想情形下 B-T 方法的结果

从不均一性发生的原因,我们可以知道不均一性可以分成两种类型,突变型不均一及

渐变型不均一。第一种类型的不均一性可能由仪器及观测方法变更、站址变迁、观测时间及计算月平均温度方法变更等原因造成。而台站周围环境变化(譬如热岛效应)造成的资料不均一则很可能是渐变型的。

虽然渐变型的不均一不是 B-T 方法所能直接检测的问题,但由 Potter 的结果及 T_0 的计算过程可以推测,分析 T_i $i = 1, 2, \dots, n$ 随 i 的变化可以对该站资料的均一性有一个直观的了解。因而在下文中,我们对理想的突变型不均一及渐变型不均一都推导了相应的 T_i 演变。

1. 突变型不均一性

对于温度序列 $x_i(i), y_i(i)$ $i = 1, 2, \dots, n$, 假定 $y_i(i)$ 有一理想的突变型不均一,即

$$y_i(i) = \begin{cases} x_i(i) + \Delta T, & i \geq i_0, \\ x_i(i), & i < i_0, \end{cases} \quad (13)$$

我们要分析这种情形下 T_i, D_i 及 T_0 的结果。

为简单起见并不失一般性,假定

$$\bar{x}_1 = \sum_{j=1}^n x_1(j) = 0, \quad (14)$$

则有

$$\sigma_x^2 = \frac{1}{n} \sum_{j=1}^n x_1^2(j), \quad (15)$$

$$\bar{y}_1 = \frac{1}{n} \sum_{j=1}^n y_1(j) = \frac{n - i_0 + 1}{n} \Delta T, \quad (16)$$

$$\begin{aligned} \sigma_y^2 &= \frac{1}{n} \sum_{j=1}^n (y_1(j) - \bar{y}_1)^2 = \sigma_x^2 - \frac{2}{n} \Delta T \sum_{j=1}^{i_0-1} x_1(j) \\ &\quad + \frac{(i_0 - 1)(n - i_0 + 1)}{n^2} \Delta T^2. \end{aligned} \quad (17)$$

注意到为应用 B-T 方法,要将原始资料标准化。将 $x_i(j), y_i(j)$ 标准化,并将标准化后的序列记为 $x(i), y(i)$, 即有

$$x(i) = x_1(i)/\sigma_x, \quad y(i) = (y_1(i) - \bar{y}_1)/\sigma_y. \quad (18)$$

由式(5)、(6)、(7)、(8),这时显然有

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x(j) = 0, \quad \bar{y} = \frac{1}{n} \sum_{j=1}^n y(j) = 0, \quad (19)$$

$$S_x = \sum_{j=1}^n (x(j) - \bar{x})^2 = n, \quad S_y = \sum_{j=1}^n (y(j) - \bar{y})^2 = n, \quad (20)$$

$$S_{xy} = \sum_{j=1}^n x(j)y(j) = n \frac{\sigma_x}{\sigma_y} - \frac{\Delta T}{\sigma_x \sigma_y} \sum_{j=1}^{i_0-1} x_1(j), \quad (21)$$

并且

$$X_i = \frac{1}{i} \sum_{j=1}^i x(j) = \frac{1}{i \sigma_x} \sum_{j=1}^i x_1(j), \quad (22)$$

$$\begin{aligned}
 Y_i &= \frac{1}{i} \sum_{j=1}^i y(j) = \frac{1}{i} \sum_{j=1}^i \frac{y_1(j) - \bar{Y}_1}{\sigma_y} \\
 &= \frac{1}{i\sigma_y} \sum_{j=1}^i y_1(j) - \frac{n-i_0+1}{\sigma_y n} \Delta T.
 \end{aligned} \quad (23)$$

为方便起见,我们引入记号

$$X_{1i} = \frac{1}{i} \sum_{j=1}^i x_1(j), \quad C_{10} = \sum_j^{i_0-1} x_1(j), \quad (24)$$

根据式(4),则有

$$F_i = S_x - X_i^2 n i / (n-i) = n - \frac{n i X_{1i}^2}{\sigma_x^2 (n-i)}. \quad (25)$$

由 $\bar{X} = 0$ 、 $\bar{Y} = 0$, (3)式中的 $S_x(\bar{Y} - Y_i) - S_{xy}(\bar{X} - X_i)$ 实际上为

$$S_x(\bar{Y} - Y_i) - S_{xy}(\bar{X} - X_i) = S_{xy}X_i - S_xY_i, \quad (26)$$

将 S_{xy} 、 X_i 、 S_x 、 Y_i 的表达式代入上式,易得

$$S_{xy}X_i - S_xY_i = \begin{cases} \frac{(i_0-1)(n-i)}{\sigma_y i} \Delta T - \frac{\Delta T}{\sigma_x^2 \sigma_y} C_{10} X_{1i}, & i \geq i_0, \\ \frac{n-i_0+1}{\sigma_y} \Delta T - \frac{\Delta T}{\sigma_x^2 \sigma_y} C_{10} X_{1i}, & i < i_0, \end{cases} \quad (27)$$

将(25)、(27)式代入 D_i 的表达式(3),并利用式(24)可以得到

$$D_i = \begin{cases} \frac{\frac{\Delta T}{\sigma_y} \left[\frac{(i_0-1)(n-i)}{i} - \frac{C_{10} X_{1i}}{\sigma_x^2} \right]}{(n-i) \left[1 - \frac{i X_{1i}^2}{\sigma_x^2 (n-i)} \right]}, & i \geq i_0, \\ \frac{\frac{\Delta T}{\sigma_y} \left(n - i_0 + 1 - \frac{C_{10} X_{1i}}{\sigma_x^2} \right)}{(n-i) \left[1 - \frac{i X_{1i}^2}{\sigma_x^2 (n-i)} \right]}, & i < i_0, \end{cases} \quad (28)$$

而(2)式中的 $S_x S_y - S_{xy}^2$ 项可以化简为

$$\begin{aligned}
 S_x S_y - S_{xy}^2 &= n^2 \left(1 - \frac{\sigma_x^2}{\sigma_y^2} \right) - \frac{\Delta T^2}{\sigma_x^2 \sigma_y^2} \\
 &\quad \cdot \left(\sum_{j=1}^{i_0-1} x_1(j) \right)^2 + \frac{2n\Delta T}{\sigma_y^2} \sum_{j=1}^{i_0-1} x_1(j),
 \end{aligned} \quad (29)$$

利用(26)式,关于 T_i 的表达式可以化简为

$$T_i = \frac{i n^2 (S_{xy} X_i - S_x Y_i)^2}{(n-i) F_i (S_x S_y - S_{xy}^2)}, \quad (30)$$

将(27)、(29)、(25)式代入式(30),并利用(24)式则可得

$$T_i = \begin{cases} \frac{in \left[\frac{(i_0 - 1)(n - i)}{i} - C_{10} X_{1i} / \sigma_x^2 \right]^2}{(n - i) \left[1 - \frac{i X_{1i}^2}{\sigma_x^2 (n - i)} \right] [(i_0 - 1)(n - i_0 + 1) - C_{10}^2 / \sigma_x^2]}, & i \geq i_0, \\ \frac{in(n - i_0 + 1 - C_{10} X_{1i} / \sigma_x^2)^2}{(n - i) \left[1 - \frac{i X_{1i}^2}{\sigma_x^2 (n - i)} \right] [(i_0 - 1)(n - i_0 + 1) - C_{10}^2 / \sigma_x^2]}, & i < i_0. \end{cases} \quad (31)$$

由这时 T_i 的表达式可以注意到, 在理想突变型不均一的情形下, T_i 随 i 的演变与 ΔT 及 σ_y 无关, 只与 i_0, i, n 及 $x_1(i) \ i = 1, 2, \dots, n$ 有关, 而凡与 $x_1(i)$ 有关的地方都有一个分母 σ_x , 即 $x_1(i)$ 本身的绝对大小不影响 T_i 随 i 的演变状况。我们的计算过程进一步表明, $C_{10} X_{1i} / \sigma_x^2$ 相对于 $(i_0 - 1)(n - i) / i$ 及 $n - i_0 + 1, i X_{1i}^2 / [\sigma_x^2 (n - i)]$ 相对于 1, C_{10}^2 / σ_x^2 相对于 $(i_0 - 1)(n - i_0 + 1)$ 都为小量, 因而在理想情形下, T_i 随 i 的演变与 $x_1(i)$ 几乎无关。

由 T_i 及 D_i 的表达式可以很容易得到, 当 $i = i_0 - 1$ 时, T_i 这时取得最大值 $T_{i_0-1} = n$, 且 $D_{i_0-1} = \Delta T / \sigma_y$ 。注意到这时 $y(i) = (y_1(i) - \bar{Y}) / \sigma_y$, 因而 B-T 方法可以精确得将这一不均一性发生的时间及大小检测出来。

2. 渐变型不均一性

类似地对于渐变型不均一性, 我们假定有温度序列 $x_1(i), y_1(i) \ i = 1, 2, \dots, n$, 且

$$y_1(i) = x_1(i) + \Delta T(i - 1), \quad (32)$$

不失一般性, 假设 $\bar{X}_1 = 0$, 则

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n x_1^2(i), \quad (33)$$

$$\bar{Y}_1 = \frac{1}{n} \sum_{i=1}^n y_1(i) = \frac{(n-1)}{2} \Delta T, \quad (34)$$

$$\sigma_y^2 = \sigma_x^2 + \frac{2\Delta T}{n} \sum_{i=1}^n i x_1(i) + \frac{\Delta T^2}{12} (n^2 - 1), \quad (35)$$

在 $\bar{Y}_1$ 及 σ_y^2 的推导中, 分别利用了数列求和公式

$$\sum_{k=1}^n k = \frac{n(n+1)}{2} \text{ 及 } \sum_{k=1}^n k^2 = \frac{1}{6} n(n+1)(2n+1).$$

将 $x_1(i), y_1(i) \ i = 1, 2, \dots, n$ 标准化, 并将标准化序列记为 $x(i), y(i)$, 即有

$$x(i) = x_1(i) / \sigma_x, \quad y(i) = (y_1(i) - \bar{Y}_1) / \sigma_y, \quad (36)$$

根据式(5)、(6)、(7)、(8), 这时有

$$\bar{X} = 0, \quad \bar{Y} = 0, \quad (37)$$

$$S_x = n, \quad S_y = n, \quad (38)$$

$$X_i = \frac{1}{i} \sum_{j=1}^i x(j) = \frac{1}{i} \frac{1}{\sigma_x} \sum_{j=1}^i x_1(j), \quad (39)$$

$$\begin{aligned}
 Y_i &= \frac{1}{i} \sum_{j=1}^i y(j) = \frac{1}{i} \frac{1}{\sigma_y} \sum_{j=1}^i (y_1(j) - \bar{y}_1) \\
 &= \frac{1}{i\sigma_y} \sum_{j=1}^i x_1(j) - \frac{\Delta T}{2\sigma_y} (n-i),
 \end{aligned} \quad (40)$$

$$S_{xy} = \sum_{j=1}^n x(j)y(j) = n \frac{\sigma_x}{\sigma_y} + \frac{\Delta T}{\sigma_x \sigma_y} \sum_{j=1}^n jx_1(j). \quad (41)$$

为推导方便,我们引入记号

$$X_{1i} = \frac{1}{i} \sum_{j=1}^i x_1(j), \quad C'_{10} = \sum_{j=1}^n jx_1(j), \quad (42)$$

根据式(4),则可以直接得到

$$F_i = n - \frac{niX_{1i}^2}{\sigma_x^2(n-i)}, \quad (43)$$

利用式(38)、(41),则有

$$S_x S_y - S_{xy}^2 = \frac{\Delta T^2}{\sigma_y^2} \left[\frac{n^2}{12} (n^2 - 1) - \frac{C_{10}'^2}{\sigma_x^2} \right], \quad (44)$$

利用式(38)、(39)、(40)、(41),则有

$$S_{xy} X_i - S_x Y_i = \frac{\Delta T}{\sigma_x^2 \sigma_y} C'_{10} X_{1i} + n(n-i) \frac{\Delta T}{2\sigma_y}, \quad (45)$$

利用式(43)、(45),则有

$$\begin{aligned}
 D_i &= \frac{\frac{\Delta T}{\sigma_y} \left[\frac{(n-i)n}{2} + \frac{C'_{10} X_{1i}}{\sigma_x^2} \right]}{(n-i) \left[1 - \frac{iX_{1i}^2}{\sigma_x^2(n-i)} \right]},
 \end{aligned} \quad (46)$$

利用式(43)、(44)、(46),最后可得到理想情形下渐变型不均一 T_i 的表达式

$$\begin{aligned}
 T_i &= \frac{in \left[\frac{C'_{10} X_{1i}}{\sigma_x^2} + \frac{n}{2} (n-i) \right]^2}{(n-i) \left[1 - \frac{iX_{1i}^2}{\sigma_x^2(n-i)} \right] \left[\frac{n^2}{12} (n^2 - 1) - \frac{C_{10}'^2}{\sigma_x^2} \right]}.
 \end{aligned} \quad (47)$$

与突变型之 T_i 类似,渐变型不均一之 T_i 随时间的演变与 ΔT 也无关,与 $x_1(i)$ 有关的项也都有一个分母 σ_x 。计算结果也表明, T_i 的分布与 $x_1(i)$ 几乎无关。事实上这时若将上式中与 $x_1(i)$ 有关的项都略去,我们可以得到 T_i 的近似表达式为 $3in(n-i)/(n^2-1)$, 这是一条开口向下的抛物线,在 $i=n/2$ 时取得最大值 $3n^3/[4(n^2-1)]$ 。

2. 计算结果

我们用随机试验的方法,实际计算了理想的突变型不均一及渐变型不均一 T_i $i=1, 2, \dots, n$ 的演变。图 1a、b 分别为相应的结果。

图 1a 为突变型不均一所对应的 T_i 。图中也给出了 $x_1(i)$ 。在计算中,取 $n=30$, $i_0=16$ 。由图 1a 可以注意到, T_i $i=1, 2, \dots, n$ 与 $x_1(i)$ 几乎无关。 T_i 在 $i=$

$i_0 - 1$ 即 $i = 15$ 时, 取得最大值 30。由图 1 可以看出, 突变型不均一 T_i 的分布为一三角形, 而渐变型 T_i 的分布为一开口向下的曲线(图 1b), 在 $i = 11$ 至 21 的时段内, T_i 的极差只有 4.0 左右。由下文实际计算可以注意到, 两种类型不均一 T_i 分布的显著差异可以用来判断不均一的类型。

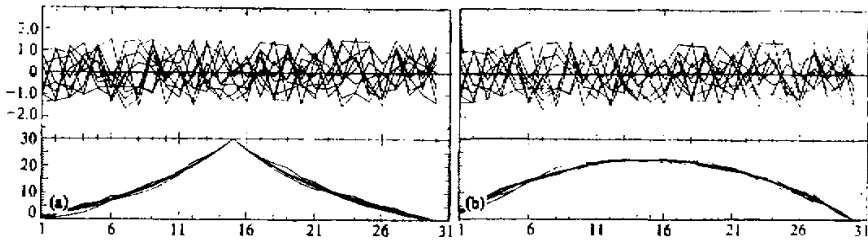


图 1 理想情形下突变型不均一及渐变型不均一所对应的 T_i 对资料 $x(i)$ 的敏感性
上半部分为 $x(i)$ 、下半部分为 T_i 的计算结果
(a) 突变型不均一, (b) 渐变型不均一。

四、中国东北区 20 站的试算结果

我们用中国东北区 20 站年平均温度资料试验了 B-T 方法, 并将计算结果同仅有的 5 站台站历史信息进行了比较。试算时段为 1959—1988 年。图 2 为台站的分布状况, 图中数字为台站站号(台站区号都为 50)。

1. 计算过程

首先将各站年平均温度资料距平化, 即将逐年平均温度减去多年平均值。第一轮计算时, 将任一站资料当做序列 $y(i)$, $x(i)$ 为其它所有站资料的算术平均, 序列 $x(i)$ 、 $y(i)$ 分别进行标准化后, 计算 T_i 及统计量 T_0 。对所有台站进行上述计算后, 将 95% 信度下不均一的台站进行剔除, 即在第二轮计算中, 这些站不参加 $x(i)$ 的计算。依此方法逐次进行, 直至参加 $x(i)$ 计算的所有台站都为均一的。

2. 计算结果

表 2 给出了计算结果。为与实际资料可直接比较, 表 2 给出了 $i_0 + 1$ 及用标准化以前的距平资料计算的 D_{i_0} 。表 2 表明, 第一轮计算中, 有 9 个站在 95% 信度下是不均一的。另有两个站, 虽然其 T_0 小于 95% 的置信限, 但由于这两站的 T_i 明显偏高, 而且其 T_0 大于 90% 的置信限, 因而对这两个站也进行了剔除, 即在第二轮计算中, 共有 11 个站

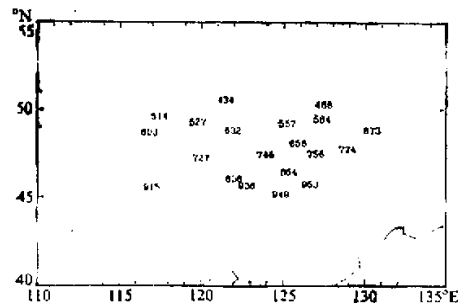


图 2 B-T 试算所用东北区 20 站的分布状况
图中数字为站号, 这些站的区号为 50。

不参加 $x(i)$ 的计算。由表 2 还可以看出,除科尔沁(50838)及嘉荫(50673)这两个站外,在第一、第二轮计算中,各站 $i_0 + 1$ (不均一发生时间的极大似然估计)及 D_{i_0} (不均一大小的极大似然估计)的改变都不大。 $i_0 + 1$ 最大的改变为 5 年(安达 50854), D_{i_0} 最大的改变为 0.09°C (爱辉 50468)。

表 2 用 Bivariate-Test 方法对东北 20 站不均一性的分析结果
(只给了第一轮计算中 T_i 超过 90% 置信限的站)

区站号	站名	第一轮计算				第二轮计算				不均一性类型
		$i_0 + 1$	D_{i_0}	T_{i_0}	信度	$i_0 + 1$	D_{i_0}	T_{i_0}	信度	
50434	图里河	1962	0.52	7.80	C	1962	0.51	6.62	D	
50468	爱辉	1985	0.59	9.67	B	1983	0.50	8.86	B	突变型
50564	孙吴	1984	0.42	9.51	B	1983	0.41	7.90	C	突变型
50632	博克图	1986	-0.99	11.31	A	1986	-1.00	10.76	A	资料错误
50658	克山	1986	-0.52	11.69	A	1986	-0.52	11.34	A	突变型
50673	嘉荫	1974	0.40	9.45	B	1984	0.53	8.12	C	渐变型
50727	阿尔山	1967	-0.46	11.49	A	1965	-0.50	12.42	A	渐变型
50745	齐齐哈尔	1963	-0.56	16.17	A	1964	-0.50	14.33	A	突变型
50774	伊春	1980	0.41	17.30	A	1980	0.43	14.14	A	渐变型
50838	科尔沁	1964	0.32	8.08	C	1977	0.23	7.38	C	
50854	安达	1972	-0.57	16.25	A	1977	-0.55	15.66	A	渐变型

* A、B、C、D 分别表示 T_i 超过了 99%、95%、90%、75% 信度的置信限

为了解这种改变发生的原因,我们给出第二轮计算中科尔沁站及嘉荫站的计算结果(图 3、图 4)。图 3 中上部实线为其它 9 个台站的年温度距平,虚线为科尔沁站年温度距平的逐年演变。下部为 T_i 的计算结果。由图 3 可以注意到, T_i 在 1963 年、1976 年及 1980 年分别有一个极大值,这可能意味着该站有一次以上不均一性发生,但因为台站历史信息缺乏,无法确认这一点。

图 4 为嘉荫站的计算结果。图中表明,从 1973 至 1983 年其 T_i 一直维持在 7.0 左右,因而计算结果便敏感地依赖于某些随机因素,当参加 $x(i)$ 计算的台站改变时,计算结果改变较大。

由理想情形下 T_i 分布的显著差异可以定性推断不均一性的类型。对于 10 个 95%

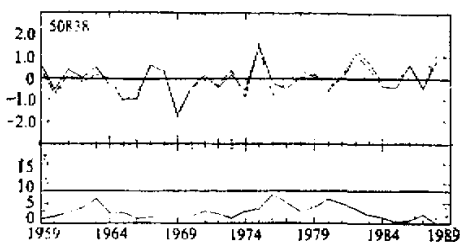


图 3 第二轮计算中科尔沁站(50838)的计算结果

图中上部实线为其他 9 站年温度距平的算术平均,虚线为该站的年温度距平。下部实线为 T_i 的计算结果。

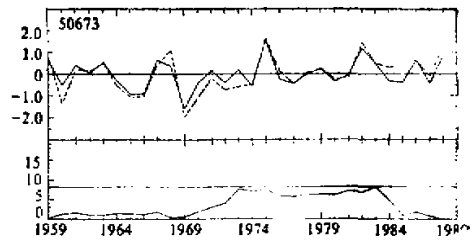


图 4 同图 3, 但为嘉荫站(50673)

信度下不均一的台站, 表 2 给出了不均一性分型的结果。有 4 个台站 T_i 的分布为平缓的曲线, 有 4 个台站 T_i 的分布为三角形。科尔沁站有 3 个极大值, 博克图站(50632)虽然 T_i 的分布为一三角形, 但进一步的分析表明, 这一 T_i 的极大值为个别点上资料的质量问题造成。

图 5 给出了齐齐哈尔站(50745)的计算结果。由图表明, 这时 T_i 随时间的演变为一三角形, 在 1963 年取得极大值。根据台站历史信息, 这一站在 1963 年 12 月有一次站址变迁。注意到 i_0 为不均一性发生的前一年, 因而 B-T 方法正确地描述了这一事件。图 6 与图 5 形成了鲜明的对照。图 6

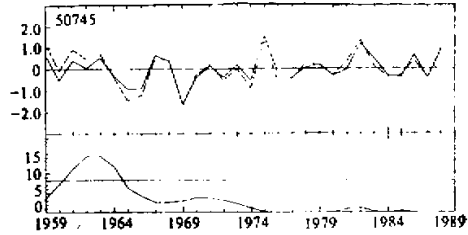


图 5 第二轮计算中齐齐哈尔站(50745)的计算结果
该站在 1963 年 12 月有一次站址迁移, 其他同图 3。

为伊春站(50774)的计算结果。图 6 表明, 伊春站的温度距平从 70 年代末以来比其周围站温度距平的算术平均一直系统性地偏高, 而 T_i 随时间的演变与渐变型不均一 T_i 的演变极为相似。这一站的不均一性很可能由台站周围环境变化(例如热岛效应)造成, 可惜没有台站历史信息, 无法对这一问题再做进一步分析。

图 7 给出了博克图站的计算结果。图中表明, 博克图站 1986 年年温度距平为

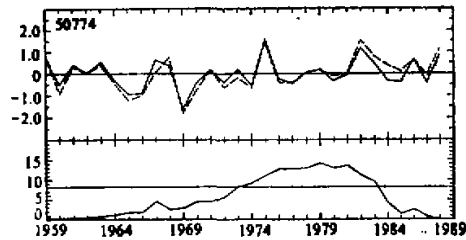


图 6 同图 3, 但为伊春站(50774)

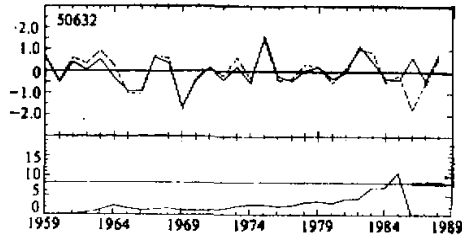


图 7 第二轮计算中博克图站(50632)的计算结果
该站 1986 年 11 月的平均温度存在错误, 其他同图 3。

-1.78°C , 而其它 19 站平均的温度距平为 0.56 。 T_i 在 1985 年出现一个超过 99% 信度置信限的最大值。并且 T_i 随时间的演变为一三角形。进一步的分析表明, 其它 19 站 1986 年最强的负温度距平(-0.22°C)出现在与博克图并不直接相邻的克山站(50658)。博克图站历史资料的分析表明, 这一站 1986 年 11 月的月平均温度为 -35.2°C , 而该站 1951 至 1980 年 30 年平均的 11 月平均温度为 -10.9°C 。因而基本上可以确定这一站 1986 年 11 月的温度存在错误。由这一例子我们可以知道, 尽管 B-T 方法按照理论推导是用来检测资料的不均一性, 但对个别点上的资料质量问题也有一定的检测效果。

五、结 论

(1) 理想情形下, Bivariate-Test 方法对突变型及渐变型不均一性有明显不同的检测结果, 这种检测结果的显著差异可以用来定性判断不均一性的类型。

(2) 实际计算表明, Bivariate-Test 方法对个别点上的资料质量问题也有一定的检测效果。

(3) 对东北区 20 站 30 年年平均温度资料的分析表明, 有 7 个站不能通过信度 95% 的 B-T 检验, 有 3 个站不能通过信度 90% 的 B-T 检验。其中有 4 个站的检验结果与理想情形下突变型不均一的检验结果相似, 有 4 个站的检验结果与理想情形下渐变型不均一的检验结果相似。有一个站存在个别点上的资料错误, 有一个站可能有一次以上的不均一性发生。

致谢: 承蒙曾昭美提供部分温度资料及台站历史信息, 董东风在调用资料时热情协助, 张玉秋完成手稿的汉字录入, 特此表示衷心感谢。

参 考 文 献

- [1] Hansen, J., S. Lebedeff, 1987, Global Trends of Measured Surface Air Temperature, *Journal of Geophysical Research*, **92**, No. D11, 13345—13372.
- [2] Jones, P. D., S. C. B. Raper, B. S. Bradley, H. F. Diaz, P. M. Kelly, T. M. L. Wigley, 1986, Northern Hemisphere Surface Air Temperature Variation: 1851—1984, *Journal of Climate and Applied Meteorology*, **25**, 161—179.
- [3] Elsner, J. B., A. A. Tsonis, 1991, Comparisons of Observed Northern Hemisphere Surface Air Temperature Records, *Geophysical Research Letters*, **18**, No. 7, 1229—1232.
- [4] Maronna, R., V. J. Yohai, 1978, A Bivariate-test for the Detection of a Systematic Change in Mean, *Journal of American Statistical Association*, **73** No. 363, 640—645.
- [5] Potter, K. W., 1981, Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series, *Mon. Wea. Rev.*, **109**, 2040—2045.
- [6] Tao S., C. Fu, Z. Zeng, Q. Zhang, 1991, Two Long-term Instrumental Climatic Data Bases of the People's Republic of China, ORNL/CDIAC-47, NDP-039.

To Detection of the Inhomogeneity of Yearly-mean Surface Temperature Data Using the Bivariate-Test Method

Wang Xiaochun, Cai Yaping and Liang Youling

(*Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100080*)

Abstract

Bivariate-Test method (B-T method) is a statistical method based on the hypothesis of normal distribution of two-dimensional random vector. The B-T method can be used to detect a systematic change of mean in one variable under the hypothesis that the mean of the other variable does not vary. Researchers of meteorology have already used this method to detect the inhomogeneity of yearly total precipitation.

The inhomogeneity of temperature data may fall into two classes: abrupt inhomogeneity and gradual inhomogeneity. The abrupt inhomogeneity may be caused by station move, change of the method used to compute the monthly-mean temperature, etc. The gradual inhomogeneity, on the other hand, may be caused by the change of environment around the station, such as urban heat-island effect, and so on.

In this paper, we analysed the results of B-T method of the ideal abrupt inhomogeneity and the ideal gradual inhomogeneity. Our results show that there is significant difference between the B-T method result of the ideal abrupt inhomogeneity and that of the ideal gradual inhomogeneity, and the difference can be used to fix the reasons of temperature inhomogeneity.

The yearly-mean surface temperature data of 20 stations in Northeast China over 30 years were used to analyse their inhomogeneity. The results show that there are 7 stations that are inhomogeneous on a significance level of 95 percent, 3 stations that are inhomogeneous on a significance level of 90 percent. The B-T method results of 4 stations are very similar to that of an ideal abrupt inhomogeneity. And the results of 4 stations are very similar to that of an ideal gradual inhomogeneity. There exists one station with data error. One station may have several inhomogeneities.

Our results also show that though the B-T method is theoretically used to detect the systematic change of mean, it is also effective in the detection of data error.

Key words: Bivariate-Test method; inhomogeneity; gradual; abrupt.