

基于启发式分割算法的气候突变检测研究*

封国林^{1)2)†} 龚志强¹⁾³⁾ 董文杰²⁾ 李建平³⁾

¹⁾ (扬州大学物理科学与技术学院, 扬州 225009)

²⁾ (国家气候中心气候研究开放实验室, 北京 100081)

³⁾ (中国科学院大气物理研究所东亚区域气候-环境重点实验室, 北京 100029)

(2005 年 3 月 23 日收到; 2005 年 4 月 25 日收到修改稿)

气候系统的非线性、多层次性和非平稳性对气候突变的检测方法提出了较高的要求。基于 t 检验将非平稳序列分割为多个不同尺度的自平稳子序列, Bernaola Galvan 提出的启发式分割算法 (BG 算法) 对非平稳时间序列的突变检测效果较好。在 BG 算法的基础上, 通过理想时间序列验证 BG 算法处理非平稳时间序列的有效性, 并对近 2000a 北半球树木年轮距平宽度序列基于不同层次的思想, 检测和分析其中包含的各种尺度的气候突变事件, 成功地区分不同尺度的突变。定义的新物理量——突变密度的分析表明, 自然因素作用的基础上, 人为因素影响的加剧可能导致近 1000a 来突变密集段和稀疏段分布失衡, 这可能是全球变化的重要表现之一。

关键词: 气候突变, 均值段, 突变密度, 人为因素, 全球变化

PACC: 9260X, 9260S

1. 引 言

气候系统是非线性、多层次的, 一般认为从一个层次跳跃到另一个层次就意味着气候发生了突变, 气候突变是气候系统的非线性的特殊表现形式之一。气候突变现象又称气候变化的不连续性、气候的跳跃^[1], 是普遍存在于气候系统中的一个重要现象^[2]。冰芯、石笋、孢粉、树轮、动物化石中保留着大量的长期变化的信息。树轮记录还具有分辨率高和记录环境气候信息量大的特点。通过对树轮记录的定性、定量分析和应用统计检验方法检测气候突变事件, 进而用数值模拟等方法找出其发生突变的机理, 已在古气候研究中取得了显著的成果。由于气候系统的复杂性, 必然导致这些信息的多尺度性、多层次性和非平稳性, 这无疑大大增加了分析和检测气候突变的难度, 对检测的方法也提出了较高的要求。地球上的气候变化是由多种因素导致的, 主要可以分为两大类: 1) 自然因素, 包括太阳活动异常, 火山爆发等; 2) 人为因素, 包括温室气体的排放、气溶胶和下垫面的改变、城市化等。因此选用何种方法检测

气候突变? 如何在一定程度上区分人为因素和自然因素对气候的影响? 这些问题的解决对人们认识过去、预测未来的气候变化将具有十分重要的现实意义。本文通过理想时间序列来验证 Bernaola Galvan 提出的启发式分割算法^[3] (BG 算法) 在处理非平稳时间序列中的有效性。对北半球树木年轮距平宽度序列 $S(t)$ (A. D. 1—1980) 基于高频和低频序列两部份, 分别用 BG 算法对其中包含的气候突变事件进行检测和分析, 试图区分不同尺度的突变。定义新物理量——突变密度, 分析近 2000a 来突变密集段和稀疏段的分布特点, 探讨近 1000a 来, 特别是工业革命以来人为因素影响的加剧可能导致的气候变化。

2. Bernaola-Galvan 启发式分割算法的统计意义

目前在气候变化中, 检测突变的方法主要有: 1) 滤波检测法; 2) 滑动 t , F 检测法; 3) Gramer 法; 4) Yamamoto 法, 以及张丕远等的最大概率变点检验方法^[4]; Yang 等的 Fisher 最优分割法^[5] 等。但将上述基于线性和平稳过程的方法应用于处理非线性、非平

* 国家重点发展基础研究项目 (批准号: 2004CB418300) 和国家自然科学基金 (批准号: 90411008, 40325015) 共同资助的课题。

† E-mail: feng. gl @sina. com

稳的气候资料时均存在一定的缺陷^[4-6]. Bernaola-Gálvan 在 2001 年提出的启发式分割算法是一种有别于上述理论的突变检测方法,其主要思想介绍如下^[3]:

对于一个由 N 个点构成的时间序列 $x(t)$, 从左到右分别计算每个点左边部分和右边部分的平均值 $\mu_1(i)$ 和 $\mu_2(i)$ 及标准偏差 $s_1(i)$ 和 $s_2(i)$, 则 i 点的合并偏差 $S_D(i)$ 为

$$S_D(i) = \left[\frac{(N_1 - 1) \times s_1(i)^2 + (N_2 - 1) \times s_2(i)^2}{N_1 + N_2 - 2} \right]^{1/2} \times \left[\frac{1}{N_1} + \frac{1}{N_2} \right], \quad (1)$$

其中, N_1, N_2 分别为 i 点左边和右边部分的点数. 我们用 t 检验的统计值 $T(i)$ 来量化表示 i 点左右两部份均值的差异:

$$T(i) = \left| \frac{\mu_1(i) - \mu_2(i)}{S_D(i)} \right|, \quad (2)$$

对 $x(t)$ 中的每一个点重复上述计算过程, 得到与 $x(t)$ 一一对应的检验统计值序列 $T(t)$, T 越大, 表示该点左右两部份的均值相差越大. 计算 $T(t)$ 中的最大值 T_{\max} 的统计显著性 $P(T_{\max})$:

$$P(T_{\max}) = Prob(T \leq T_{\max}), \quad (3)$$

$P(T_{\max})$ 表示在随机过程中取到 T 值小于等于 T_{\max} 的概率. 一般情况下 $P(T_{\max})$ 可近似表示为

$$P(T_{\max}) \approx (1 - I_{\nu/(v+T_{\max}^2)}(\nu, \nu)). \quad (4)$$

由蒙特卡洛模拟可以得到 $\ln P(T_{\max}) = 4.19 \ln N - 11.54$, $\nu = 0.40$, N 是时间序列 $x(t)$ 的长度, $\nu = N - 2$, $I_x(a, b)$ 为不完全 Γ 函数. 我们设定一个临界值 P_0 , 如果 $P(T_{\max}) > P_0$, 则于该点将 $x(t)$ 分割成两段均值有一定差异的子序列, 否则不分割.

对新得到的两个子序列分别重复上述操作, 如果子序列有 $P(T_{\max}) > P_0$, 并且子序列与其左、右相邻的子序列间均值的差异程度均满足上述条件, 则对子序列进行分割, 否则不分割. 如此重复直至所有的子序列都不可分割为止. 为确保统计的有效性, 当子序列的长度小于等于 l_0 (l_0 为最小分割尺度) 时不再对其进行分割. 通过上述操作, 我们将原序列分割为若干不同均值的子序列, 分割点即为均值突变点. 我们定义原时间序列被分割得到的每个子序列所对应的时间段为“均值段”. 一般而言, l_0 的取值不小于 25, P_0 可取 0.5—0.95.

$$x(t) = \begin{cases} 0.90 + \varepsilon_1 & 1 \leq t < 800, \\ 0.60 + \varepsilon_2 & 800 \leq t < 2800, \\ 0.80 + \varepsilon_3 & 2800 \leq t < 5000, \\ 0.50 + \varepsilon_4 & 5000 \leq t < 7000, \\ 0.85 + \varepsilon_5 & 7000 \leq t < 8000. \end{cases} \quad (5)$$

我们构建一理想时间序列 $x(t)$ (8000 个点), 并用 BG 算法对该序列进行突变检测, 取 $l_0 = 100$, $P_0 = 0.95$. 由(5)式可知理想时间序列 $x(t)$ 由一个常数 c 和一个等概率分布的随机扰动 ε_i 叠加而成, 分别在 t 为 799, 2799, 4999 和 6999 四点发生了均值突变(图 1(a)). 原序列的四个突变点的位置(图 1(b)), 恰好与 1—8000 的 $T(t)$ 有极大值的位置相对应, 故用 BG 算法进行检测时, 突变点一般出现在 T 有极大值的地方. $t = t_1$ 时有 $T_{\max 1}$, 此时 $P(T_{\max 1}) > 0.95$, 在 $t_1 = 799$ 处原序列被分割为两段; 对新得到的子序列 799—8000 继续用 BG 算法进行处理, 图 1(c) 为点 799—8000 的 T 值, $t = t_2$ 时 $P(T_{\max 2}) > 0.95$, 并且该段子序列与其左邻的子序列之间的 $P(T_{\max}) > 0.95$, 故在 $t_2 = 6999$ 处该子序列被二次分

割: 对子序列 1—799 用 BG 算法进行处理, 图 1(d) 为该段时间序列的 T 值, 其 $P(T_{\max 3}) < 0.95$, 故这段子序列不能被继续分割. 如此重复, 我们得到 4 个分割点分别在 t 等于 799, 2799, 4999 和 6999 的位置, 与原序列的四个突变点的位置一致. 通过 BG 算法的处理, 我们能检测到 $x(t)$ 中的 4 个突变点, 并将 $x(t)$ 分割为 5 个不同均值的平稳子序列(均值段).

取 $P_0 = 0.95$, l_0 分别取 1000, 1100, 3000, 3100 再次对 $x(t)$ 进行检测(表 1). 由表 1 可以看出, l_0 取 1000 到 3100 的范围内检测结果相同, 这进一步说明了 BG 算法能有效地检测突变. 与传统的方法相比, 其特征主要表现在: 该方法基于 t 检验将非平稳序列分割为多个具有不同均值的平稳子序列(均值段), 各子序列表征了不同的物理背景, 分解得到的

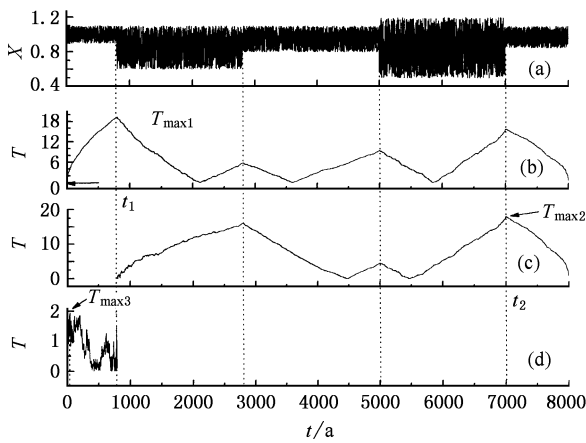


图1 理想时间序列BG算法处理结果

各均值段的尺度具有可变性,不受方法本身的限制,克服了以往的检测方法基于平稳和线性过程的问题;由于分割时采用多次迭代的一分为二的方法,大大减少了计算量,实用性较好;可以通过调整 l_0 和 P_0 的大小改变检测的尺度和精度,即能够检测不同尺度和幅度的突变;白噪声和尖峰噪声对该方法的影响较小,特别适合于处理类似气候资料的时间序列^[7].

表1 $x(t)$ 在不同 l_0 下的突变年份

l_0	1000	1100	3000	3100
突变点	799	799	799	799
	2799	2799	2799	2799
	4999	4999	4999	4999
	6999	6999	6999	6999
合计	4	4	4	4

3. 实际资料的突变检测

基于BG算法,对北半球树木年轮距平宽度序列 $S(t)$ (A.D. 1—1980) 进行分析. 树轮记录的形成和降水有着非常紧密的联系,因此树轮序列中包含着降水随年际变化的信息,对其进行研究,有助于更好地认识近 2000a 来北半球的气候变化,特别是降水的变化特点^[8-12].

$S(t)$ 是一个非线性、多层次的非平稳序列,其中包含着各种不同尺度的突变信息(图 2(a)). 直接用BG算法对 $S(t)$ 进行检测,对于得到的突变点,无法区分其属于小尺度还是大尺度的突变点(即真正意义上的气候变化重大事件). 本文使用的由 Zhang

等提出的滤波方法,是一种滤波频率带很狭窄的多级滤波器实现高、低频的分离. 这种滤波方法被 Rasmusson 等、Wang 等用于分离海平面温度 29 年以上的慢变的平均值,利用这种方法滤波后的序列的长度与原序列相同,程序由 Wang 等提供^[13-16]. 对 $S(t)$ 进行 100a 尺度的滤波,我们将 $S(t)$ 分解为低频和高频两个新序列. 低频率序列 $S_1(t)$ (图 2(b)) 和高频率序列 $S_2(t)$ (图 2(c)) 分别包含了不同尺度的气候变化信息. 基于BG算法,分别对低频率序列 $S_1(t)$ 和高频率序列 $S_2(t)$ 进行处理,我们可以分别得到大尺度和小尺度的突变点(图 3、图 4),即通过先滤波后分割的处理,有可能判别树轮记录中的各种不同尺度的突变信息.

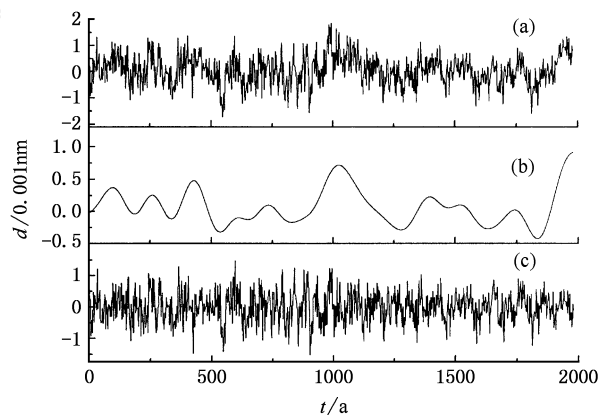


图2 $S(t)$ 的 100a 尺度滤波图像(横坐标为时间 t , 纵坐标为距平宽度 d)

对 $S_1(t)$ 用BG算法进行处理,取 $l_0 = 500, P_0 = 0.85$. 图 3 为低频率序列 $S_1(t)$ 在 $l_0 = 500$ 的突变图. 由图 3 可以看出, $S_1(t)$ 在 $l_0 = 500$ 时共检测到 7 个大尺度突变点,这 7 个突变点将 $S_1(t)$ 划分为 8 个相对平稳的均值段(表 2),即近 2000a 树轮记录 $S(t)$ 可能包含这 8 个大尺度均值段. 各均值段的宽度前 1000a 较后 1000a 略长,突变的频率后者略高,换言之,后 1000a 的气候变化可能要比前 1000a 的变化更复杂. 根据文献^[17,18],树轮序列大尺度突变点的分布与发生较大干湿变化的年份有一定的相似性.

图 3 中,900a 和 1600a 两个突变点前后气候突变的频率均有一定的变化,其原因可能有二:一是气候系统确实受到了某些外部系统的干扰,如太阳活动、火山或冰雪融化等;二是气候系统的动力结构发生了突变,换句话说,气候也许已从一个层次向另一

层次转化.以 900a,1600a 为分界点,近 2000a 的气候变化大致可以分为 1—900a,900—1600a 和 1600—1980a 三个阶段.

表 2 $S_1(t)$ 的 $l_0=500$ 的突变年份及各均值段

突变年份/a	子序列/a	均值/0.001mm	宽度/a
-	1—364	0.119	364
364	365—486	0.294	122
486	487—902	-0.108	416
902	903—1118	0.439	206
1118	1119—1349	-0.091	231
1349	1350—1573	0.105	224
1573	1574—1919	-0.150	346
1919	1920—1980	0.750	61

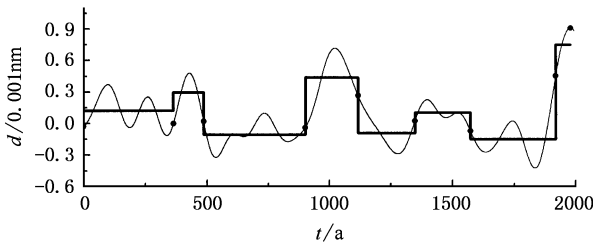


图 3 $S_1(t)$ 的 $l_0=500$ 的突变图(细线为 $S_1(t)$,粗线为均值线,黑点为突变点)

若 N 年的突变点数为 n ,我们定义为“突变密度”

$$= \frac{n}{N} \times 10, \quad (6)$$

设定密度常数 ρ_0 , ρ_0 时对应的时间段为突变密集段,否则为突变稀疏段.突变密度的定义,我们可以定量比较每个百年或更大尺度上突变点的密集程度,同时突变密集段和突变稀疏段分布特点的研究对认识气候的变化特点具有一定的现实意义.本文讨论每 100a 内突变点的疏密情况. $S_1(t)$ 的 $l_0=500$ 的 7 个突变点将 $S_2(t)$ 分为 8 个大尺度均值段.取 $l_0=30, P_0=0.85, \rho_0=0.5$,对 $S_2(t)$ 的前 7 个均值段,分别基于 BG 算法分析其每 100a 的突变密度分布(最后一个均值段的宽度较小,故不讨论).由表 3 可以看出,第 3,5,7 均值段,密集段和稀疏段交替出现.将各密集段和稀疏段对应的时段与张丕远等人根据 0 阶契比雪夫系数划分的旱涝时段^[4]比较发现,突变密集段大多为多涝和偏涝,稀疏段则大多为

多旱和偏旱(表 3).图 4 为 $S_2(t)$ 三个均值段的突变密度分布图.其他几个均值段,虽然疏密的分布没有表 3 中那么规则,但就整体而言,密集段和稀疏段的比例相当(图略).众所周知,6,12 和 20 世纪是三个典型的暖期^[19],由图 4 可以看出,6 和 12 世纪的突变点较稀疏,这两个世纪属于突变稀疏段,而 20 世纪则突变点较密集,属于突变密集段.Lemper-Ziv 复杂度对北京石花洞石笋微层厚度序列的处理结果表明,6,12,20 世纪均对应复杂度相对平稳的时段^[20].复杂度的类似和突变密度的不同,我们认为其原因可能在于自然因素和人为因素的共同作用,而 20 世纪人为因素的作用远大于 6 和 12 两个世纪,可能有与自然因素相当的作用.

表 3 $S_2(t)$ 三个均值段 $l_0=30$ 的突变的疏密分布及旱涝变化

均值段	时段/a	疏密	旱涝
3	487—586	0.40	疏 C
	587—686	0.50	密 B, C
	687—786	0.30	疏 C, D
	787—902	0.54	密 A, B, C
5	1119—1218	0.40	疏 D
	1219—1349	0.73	密 A
7	1574—1674	0.50	密 O
	1675—1774	0.30	疏 C
	1775—1919	0.63	密 A

注:A, B, C, D 分别表示多涝、偏涝、偏旱、多旱, O 表示旱涝持平.

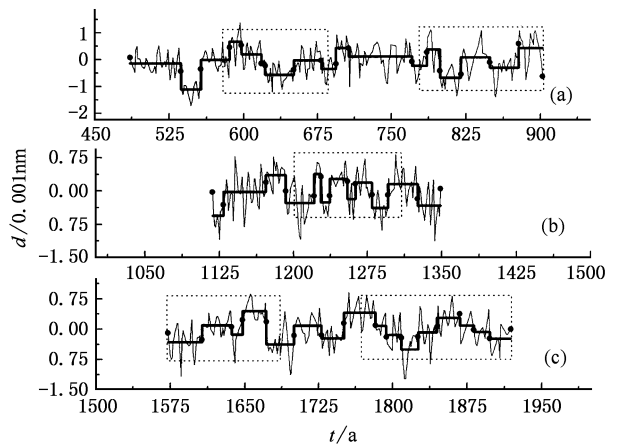


图 4 $S_2(t)$ 三个大尺度均值段 $l_0=30$ 的突变疏密分布图 ((a),(b),(c)分别为第 3,5,7 均值段突变的疏密分布图,虚线框为突变密集段)

4. 关于人为因素对气候影响的探讨

公元 1000 年以前,人为因素对气候的影响可能比较小,气候突变的原因可能是火山喷发、太阳活动异常等自然因素的作用.近 1000a 来特别是第一次工业革命以来,人为因素对气候的影响不断加剧,人为因素很可能已经成为继自然因素之后造成气候突变的重要原因之一.取 $l_0 = 30, P_0 = 0.85, \rho_0 = 0.5$, 我们对比分析 $S_2(t)$ 公元 1000 年前后两个千年的突变情况,并就人为因素对气候的影响进行初步的探讨.图 5 为 $S_2(t)$ 前后千年 $l_0 = 30$ 的突变图.可以看出前 1000a 发生了 41 次突变,后 980a 发生了 44 次突变,突变的频率较前者略高.表 4 为 $S_2(t)$ 在前后千年每 100a 内的突变密度分布.可以看出,前 1000a 突变密集段和稀疏段基本上是交替出现,两者的比例相当.后 1000a,密集段远多于稀疏段,疏密不交替出现.我们认为导致后 1000a,密集段远多于稀疏段,疏密不交替出现的可能原因与人为因素的影响有关.1800—1980a 间存在两个连续的突变密集段,其原因可能是工业革命以来,人为因素对气候的影响不断加剧,气候的全球变化导致突变频率的增大.我们认为气候突变的频率增大和气候突变疏密分布失衡是自然因素和人为因素共同作用,但人为因素的影响大大加强的结果.

表 4 高频序列 $S_2(t)$ 前后千年的每 100a 突变疏密分布

时段	1	2	3	4	5	6	7	8	9	10
	0.5	0.2	0.4	0.5	0.5	0.4	0.6	0.4	0.3	0.5
疏密	密	疏	疏	密	密	疏	密	疏	疏	密
时段	11	12	13	14	15	16	17	18	19	20
	0.3	0.5	0.6	0.3	0.6	0.5	0.5	0.4	0.5	0.5
疏密	疏	密	密	疏	密	密	密	疏	密	密

注:时段一行:1 表示 1—100a,2 表示 101—200a, ...,20 表示 1901—1980a.

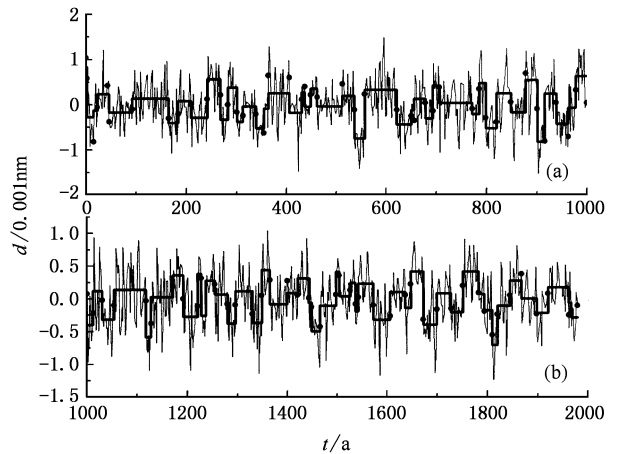


图 5 $S_2(t)$ 的前后千年 $l_0 = 30$ 的突变图((a),(b)分别为 1—1000a 和 1001—1980a 突变图)

5. 结 论

1. BG 算法是一种突变检测的有效方法,该方法可以将非平稳序列分割为多尺度的自平稳子序列;检测的尺度和精度具有可变性,能够检测不同尺度和不同幅度的突变;白噪声和尖峰噪声的对该方法的影响较小,特别适合于处理类似气候资料的时间序列.
2. 树轮序列中包含的信息,各层次的尺度各不相同,经过滤波,可以将大尺度和小尺度突变的信息分离,分别对滤波得到的高、低频率序列进行突变检测可以得到大尺度和小尺度的均值段.900a,1600a 前后的突变频率发生了一定的变化,近 2000a 的气候变化大致可以分为 1—900a,900—1600a 和 1600—1980a 三个阶段.
3. 根据低频率序列 $l_0 = 500$ 的突变将近 2000a 树轮序列划分为 8 个均值段.突变密度的定义和分析表明,前后 1000a 突变密集段和稀疏段在分布和比例上均存在一定的差异,人为因素影响可能导致近 1000a 突变密集段和稀疏段分布失衡,这可能是全球变化的重要表现之一.

[1] Dai X G, Fu Z B and Wang P 2005 *Chin. Phys.* **14** 850
 [2] Shi N 2005 *Chin. Phys.* **14** 844
 [3] Bernaola-Galvan P 2001 *Phys. Rev. Lett.* **87** 168105
 [4] Shi Y F and Zhang P Y 1996 *The variation trend and its effect of the climate in China and sea level change—the historical climatic*

change in China (Jinan: Shan Dong Science and Technology Press) 383—393 (in Chinese) [施雅风、张丕远 1996 中国气候与海面变化及其趋势和影响中国历史气候变化(济南:山东科学技术出版社)第 383—393 页]

- [5] Yang W F, Li Z Y and Li X M 1997 *Quarterly Journal of Applied Meteorology* **8** 119
- [6] Yan Z W, Li Z Y and Wang X C 1993 *Scientia Atmospherica Sinica* **17** 663
- [7] Fukuda Kensuke *et al* 2004 *Phys. Rev. Lett.* **69** 021108
- [8] Hou A M, Peng S L and Zhou G Y 1999 *Ecologic Science* **18** 16 (in Chinese) [侯爱敏、彭少麟、周国逸 1999 生态科学 **18** 16]
- [9] Zhang Z F, Peng Z C and He J F *et al* 2001 *Geology Geochemistry* **29** 78 (in Chinese) [张兆峰、彭子成、贺剑峰等 2001 地质地球化学 **29** 78]
- [10] Yao T D, Yang X M and Kang X C 2001 *Quaternary Sciences* **21** 514 (in Chinese) [姚檀栋、杨学梅、康兴成 2001 第四纪研究 **21** 514]
- [11] Fu Z B and Wang Q 1992 *Scientia Atmospherica Sinica* **6** 482
- [12] Fu Z B 1994 *Scientia Atmospherica Sinica* **18** 373
- [13] Zhang D W and Dong D N 1986 *Acta. Astron. Sin.* **27** 368
- [14] Shi N 2000 *J. Trop. Meteor.* **16** 131 (in Chinese) [施能 2000 热带气象学报 **16** 131]
- [15] Rasmusson E M, Wang X L and Ropelewski C F 1994 *Secular variability of ENSO cycle* (San Dong: Academic Press)
- [16] Wang X L and Ropelewski C F 1995 *J. Climate* **8** 1584
- [17] Yao T D, Yang X M and Kang X C 2001 *Quaternary Sciences* **21** 514 (in Chinese) [姚檀栋、杨学梅、康兴成 2001 第四纪研究 **21** 514]
- [18] Zhang D E, Liu C Z and Jiang J M 1997 *Quaternary Sciences* **1** 1
- [19] Qing X G, Liu D S, Tan M and Wang X F *et al* 2000 *Science in China (D)* **30** 239
- [20] Hou W, Feng G L and Gao X Q *et al* 2005 *Acta. Phys. Sin.* **54** 2441 (in Chinese) [侯威、封国林、高兴全等 2005 物理学报 **54** 2441]

Abrupt climate change detection based on heuristic segmentation algorithm *

Feng Guo-Ling¹⁾²⁾ Gong Zhi-Qiang¹⁾³⁾ Dong WerrJie²⁾ Li Jian-Pin³⁾

¹⁾ (Department of Physics, Yangzhou University, Yangzhou 225009, China)

²⁾ (Laboratory for Climate Studies of National Meteorological Bureau, Climate Research Center, Beijing 100081, China)

³⁾ (Key Laboratory of Regional Climate-Environment Research for Temperature East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China)

(Received 23 March 2005; revised manuscript received 25 April 2005)

Abstract

Climate system is nonlinear, non-stationary and hierarchical, which makes even harder to detect and analyze abrupt climate changes. Based on Student's t -test, Bernola Calvin recently proposed a heuristic segmentation algorithm to segment the time series into several subsets with different scales, which is more effective in detecting the abrupt changes of nonlinear time series. In this paper, we try to verify the effectiveness of heuristic segmentation algorithm in dealing with nonlinear time series by an ideal time series. Through detecting and analyzing the information of abrupt climate changes contained in recent 2000a's tree annual growth ring, we succeeded in distinguishing abrupt changes with different scales. The research based on the newly defined parameter of abrupt change density shows that human activities might have lead to the recent 1000a's unbalanced distribution of serial and spares segments of abrupt climate changes, which may be one of the manifestations of global temperature change.

Keywords: abrupt climate change, mean segment, abrupt change density, human activities, global temperature change

PACC: 9260X, 9260S

* Project supported by the State Key Development Program for Basic Research (Grant No. 2004CB418300) and the National Natural Science Foundation of China (Grant Nos. 90411008, 40325015).