

滑动去趋势波动分析与近似熵在动力学结构 突变检测中的性能比较*

何文平^{1)†} 吴琼²⁾ 张文³⁾ 王启光⁴⁾ 张勇¹⁾

1) 国家气候中心, 北京 100081)

2) 中国气象科学研究院, 北京 100081)

3) 南京邮电大学通达学院, 南京 210003)

4) 扬州大学物理科学与技术学院, 扬州 225002)

(2008 年 7 月 16 日收到; 2008 年 9 月 12 日收到修改稿)

近似熵 (A_{pEn}) 被认为是一种有效的动力学结构突变检测方法. 将一种新的动力学结构检测方法——滑动去趋势波动分析 (MDFA) 与 A_{pEn} 的检测结果进行了比较, 检验了新方法的性能. 结果表明, 新方法的检测结果几乎不依赖于子序列的长度, 而 A_{pEn} 虽然能在一定程度上识别系统的动力学结构突变, 但其检测结果依赖于子序列长度, 且不能准确地检测出突变点的位置. 因此, 相对于 A_{pEn} 方法而言, MDFA 方法更适合于动力学结构突变检测, 其优越性是显而易见的.

关键词: 滑动去趋势波动分析, 近似熵, 动力学结构突变

PACC: 9260X

1. 引 言

气候突变现象及其理论的研究是近代气候学一个新兴的研究领域^[1], 传统的气候突变理论一般认为气候突变是指气候从一种稳定状态跳跃式地转变到另一种稳定状态的现象^[2]. 气候突变在相空间中的表现为系统的状态不再趋向于原来的吸引子, 而是趋向于新的吸引子. 两种情况可能会导致系统出现“突变”. 一是系统内部机理或关系发生演化, 旧的吸引子消失或变为不稳定吸引子, 分叉出新的吸引子; 二是对系统状态的扰动过大, 迫使系统离开原吸引子的范围, 转而趋向于新的吸引子附近, 系统的结构重新获得稳定. 简言之, 系统突变对应着演化状态趋向于新的吸引子的变化^[3]. 显然这两种可能的突变均与系统的动力学结构发生了变化有关, 然而, 即使系统的动力学结构不发生任何变化, 系统也可能会从一种稳定状态跳跃式地转变到另一种稳定状态. 例如, 由于气候系统变化的多时间尺度, 在不

同时间尺度层次上都存在冷暖相位的变化, 小尺度上的暖相位, 在更大尺度意义上却极有可能成为冷相位. 因此, 传统的气候突变理论对于气候突变的定义存在一定的局限性. 在本文中, 我们按照控制气候系统的动力学方程是否发生变化将气候突变分为动力学结构突变和气候状态变量在统计意义上发生了显著变化的突变, 如均值(相位)突变、趋势突变等. 气候状态变量在统计意义上的突变显然具有多时间尺度特征, 气候状态的跳跃可以发生在月、季、年、十年、百年, 甚至更长时间尺度上, 状态变量统计意义上的显著变化所导致的突变是一种普遍现象. 如何有效识别不同性质的突变点是一个亟待解决的问题.

传统的气候突变检测方法如滑动- t 检验、Cramer 法、Yamamoto 法以及 Mann-Kendall 法等^[4-6], 大多基于状态变量在统计意义上的显著变化来判断突变的发生, 其检测结果均具有多尺度特征. 此外, 还有学者采用 Fisher 方法^[7]、最大概率变点检验方法^[8]以及小波分析方法^[9,10]等研究气候突变, 但是

* 国家重点基础研究发展计划(批准号 2004CB418303, 2006CB400503)和国家科技支撑计划(批准号 2007BAC03A01, 2007BAC29B01)资助的课题.

† E-mail: wenping_he@163.com

这些方法的检测结果依然具有多尺度特征,然而系统的动力学结构突变发生的时间与特定的时间尺度没有必然的关联,因此基于系统状态变量在统计意义上的显著变化来判断突变的传统检测方法不能够对系统的动力学结构突变进行有效识别。

近年来,封国林等^[11-16]引进并发展了一系列新的气候突变检测方法,如条件熵、动力学相关因子指数(Q 指数)、启发式分割算法、复杂度、幂律指数、高阶矩等。应用结果表明,这类新方法不仅能够有效地检测统计意义的突变点,而且对于系统的动力学结构突变也具有一定的识别能力,这对传统的气候突变检测方法无疑是一个有益的补充。但是我们也注意到这类方法在对于动力学结构突变检测的过程中仍然存在很大的不确定性,不能够准确地检测到系统的动力学结构突变点。

近似熵($ApEn$)^[17-19]是一种衡量时间序列复杂性的方法,它对噪声具有较强的抗干扰性,对于随机信号、确定性信号或两种混合信号都适用,且其计算过程中所需数据的样本量较少。因此,该方法一经提出,便得到了广泛的使用。王启光等^[20]将其用于系统的动力学结构突变检测,发现其能够对线性系统的动力学结构突变进行准确检测,随后他们将其用于日温度和降水序列的动力学结构突变检测,初步确定 20 世纪 70 年代末和 80 年代初我国温度和降水的突变性质,得到了有意义的结论。鉴于 $ApEn$ 在动力学结构突变检测中的有效性^[20],本文将作者此前提出的一种新的动力学结构突变检测方法^[21]——滑动去趋势波动分析(MDFA)与之在理想序列和实际观测资料中的性能进行了系统的比较。

2. MDFA 方法及数据

Peng 等^[22-24]在研究核糖核酸行为机制的过程中提出了一种新的标度分析方法——去趋势波动分析(DFA),它能够很好地处理非平稳数据,通过 DFA 方法获得的标度指数较传统的方法更为可靠。因此,自 DFA 方法提出以来便在各个科学领域得到广泛应用^[25-29]。何文平等^[30]通过对一维至四维的四个典型混沌系统的标度分析发现,四种不同混沌系统均展示了标度行为,而且不同动力学系统的标度指数具有显著的差异。这表明标度指数能够作为一种表征系统动力学结构特征的物理量。随后,他们在 DFA 的基础上提出了一种新的动力学结构突变

检测方法——MDFA^[21]。已有的研究表明,自然界中的大量时间序列均具有标度行为^[25-40],因而 MDFA 方法非常适合不同领域内许多时间序列的动力学结构突变检测。

MDFA 的主要步骤如下:首先选择一定长度的子序列,采用 DFA 计算该子序列的标度指数,随后逐步移动该子序列,但保持其长度不变,再采用 DFA 计算新子序列的标度指数。如此重复操作,直至子序列移动至所分析序列的末端。根据已有的研究结果^[21,30]可知,如果时间序列中不存在动力学结构突变,通过 MDFA 分析得到的标度指数序列的变化较平缓,标度指数的波动幅度在一个较小的范围内,这种波动的原因主要在于子序列长度过短;若存在动力学结构突变,标度指数序列将在突变点附近剧烈波动,其幅度远大于由于子序列长度过短所引起的波动。在文献 [21,30] 中,我们所考虑的动力学结构突变主要包括以下三种情形:(1)动力学方程的参数发生了变化,记为 I 型动力学结构突变;(2)动力学方程组的形式完全发生了变化,记为 II 型动力学结构突变;(3)动力学方程组的部分方程形式发生了变化,记为 III 型动力学结构突变,主要是指在某一时段,系统受外部系统的影响,其演化规律发生了变化,如外源强迫或者系统之间的相互作用。在本文中,我们主要以 I 型动力学结构突变为例,对 MDFA 与 $ApEn$ 方法的优劣进行了比较。类似于 DFA 分析,在采用 MDFA 进行动力学结构突变检测过程中,若多项式拟合的阶数为 n ,我们则将这一分析过程简记为 MDFA $_n$ 。

在本文的比较试验中,线性理想序列同文献 [20],其长度为 2000,所采用的非线性理想序列由各种典型的混沌模型产生,其序列总长度均为 1000,且动力学结构突变均发生在时间序列的中部,即理想序列的前后两段分别来自不同的动力学系统。文中所用到的日地表极端最高温度资料来源于中国气象局国家气象信息中心气象资料室,资料长度为 46a,1a 即为 1 年(1960 年—2005 年)。

3. MDFA 与 $ApEn$ 在动力学结构突变检测中的性能比较

3.1. MDFA 与 $ApEn$ 在线性时间序列中的检测性能比较

为了对 MDFA 与 $ApEn$ 方法的性能进行全面的

比较,我们首先就 MDFA 与 $ApEn$ 在线性序列的动力学结构突变检测中的有效性进行了试验.理想序列 ISO 采用如下方程^[20]获取:

$$y(t) = \begin{cases} 2\sin(0.2t) + 1 & (1 \leq t \leq 1000), \\ 1.5\sin(0.2t) + 2\cos(0.5t) - 0.2 & (1000 < t \leq 2000). \end{cases} \quad (1)$$

由(1)式可知,在 $t = 1001$ 时,理想序列 ISO 发生了动力学结构突变,其控制方程由简单的正弦函数突变为正弦和余弦所组成的新方程,其时间序列已在图 1(a)中给出.

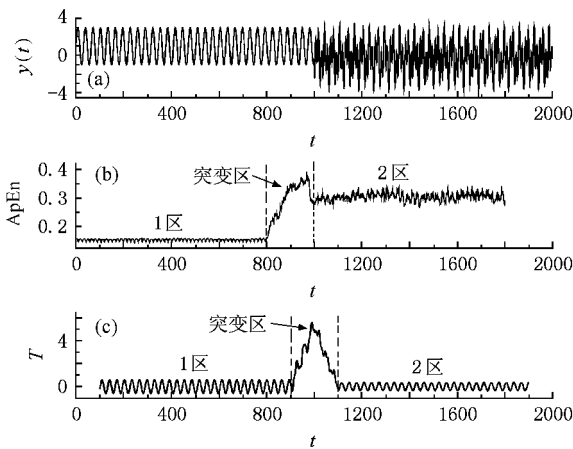


图 1 理想时间序列 ISO 及其 $ApEn$ 检测 (a)理想时间序列 ISO; (b)滑动窗口长度为 200,滑动步长为 1 的 $ApEn$ 检测 (c)滑动- t 检验,滑动窗口长度为 100

王启光等^[20]采用 $ApEn$ 和滑动- t 检验对理想序列 ISO 进行了突变检测,其检测结果已由图 1 给出.从图 1(b)中不难看出, $ApEn$ 能够根据时间序列的复杂性来很好区分线性动力学系统的异同,当子序列来源于同一个动力学系统时,随着子序列的滑动其 $ApEn$ 值的变化十分稳定,而当子序列中的数据来自于不同的动力学系统时,序列的复杂性显然较单一动力学系统更高,因而其 $ApEn$ 值也随之增大.基于此,可以很容易地判断出原时间序列约在 $t = 1000$ 附近发生了动力学结构突变.由 ISO 的滑动- t 检验结果可知,其检测到在 $t = 1001$ 时发生了显著的均值突变,这与系统的动力学结构突变完全重合,但与 $ApEn$ 检测结果不同,滑动- t 检验不能够明确地区分出突变点前后系统的动力学结构异同性(详见图 1(c)).

图 2 给出了理想序列 ISO 的 MDFA2 检测结果,从图 2(a)容易看出,在 $t = 1001$ 处,标度指数 γ 发生

了一次显著的均值突变,采用滑动- t 检验容易对此进行验证(图略).由图 2(b)可知,子序列长度的增加更有利于从 MDFA2 的检测结果中准确判断原序列的动力学结构突变点,子序列长度为其他值时的检测结果可以进一步验证这一结论(图略).标度指数在突变点前后的这种显著差异显然是由于原序列中的动力学结构突变所造成的,即不同动力学系统的标度指数具有显著的差异,可以据此来判别动力学系统的突变. ISO 的更高阶 MDFA 分析结果类似于 MDFA2,在此不再赘述.

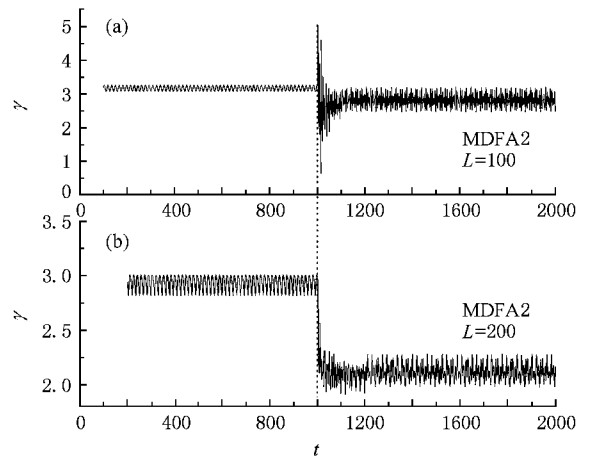


图 2 线性序列 ISO 的 MDFA2 突变检测 (a) $L = 100$ (b) $L = 200$

从 MDFA 与 $ApEn$ 对于理想序列 ISO 的分析结果看,二者都能够对其进行较为准确的检测,而且我们发现噪声对 MDFA 分析结果的影响不大^[21], $ApEn$ 也同样具有较强的抗噪能力^[20].因此,MDFA 与 $ApEn$ 在线性时间序列的动力学结构突变检测中的效果基本相当.实际上,自然界的演化是纷繁复杂的,大多数时间序列都是非线性的.因此,突变检测方法对于非线性时间序列的适用性问题显得尤为重要,而文献[20]仅仅考虑了 $ApEn$ 对于线性时间序列的检测效果,显然存在着一定的不足.为此,本文着重比较了 MDFA 与 $ApEn$ 方法在非线性时间序列的动力学结构突变检测中的优劣.

3.2. MDFA 与 $ApEn$ 在非线性时间序列中的检测性能比较

我们首先构造了非线性时间序列 IS1,该序列来源于经典的 Lorenz 方程^[31],前 500 个数据为 Rayleigh 数取 29 时产生,其余数据由 Rayleigh 数突变为 28 时所得.由此可见,理想序列 IS1 在 $t = 501$ 处发生了 I 型动力学结构突变.图 3(a)展示了 IS1 随时间的

演化,当子序列长度 L 取为 100 时,相应的各阶 MDFA 检测结果已在图 3(b)(c)和(d)中给出.

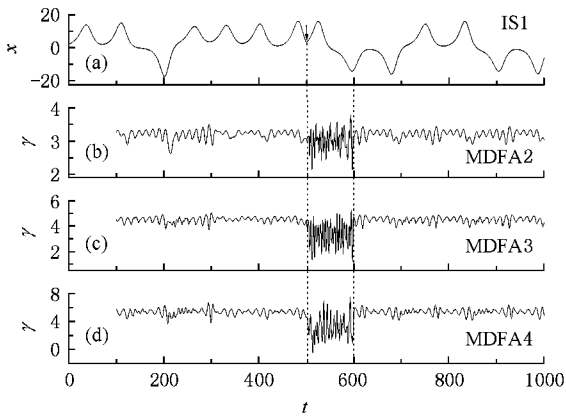


图3 理想时间序列 IS1 及相应的 MDFA 检测结果 子序列长度 $L=100$. (a)理想序列 IS1 (b)MDFA2 (c)MDFA3 (d)MDFA4

从图 3(b)(c)和(d)中不难发现,不同阶 MDFA 分析得到的标度指数随着子序列滑动的演化规律非常类似.在 $t \in [100, 500]$ 和 $t \in [601, 1000]$ 两个区间内,各阶 MDFA 分析所获取的标度指数波动均比较稳定,而且幅度较小,这种波动主要是由于所取子序列的长度较短造成的.而在 $t \in [501, 600]$ 的区间内,标度指数的波动幅度明显较另外两个区间大,这种异常波动的原因主要在于该区间内子序列中所包含的数据来自不同的动力学系统.因此,表征系统动力学无标度特征的标度指数相应地发生了异常变化,若所分析子序列中的数据不是来源于同一个系统,MDFA 方法所获取的标度指数将对此十分敏感,直观的表现是标度指数在突变点附近的剧烈变化,波动幅度显著大于正常情形.由于子序列长度过

短所引起的波动).进一步对比不同阶 MDFA 得到的标度指数序列可以发现,随着 MDFA 阶数的增加,获取的标度指数相应增大,究其原因主要在于原序列中的趋势没有完全去除,但这并不影响我们对序列中动力学结构突变点的准确检测. MDFA2, MDFA3 和 MDFA4 的分析结果表明,MDFA 方法能够十分精确地检测到理想序列 IS1 中所存在的动力学结构突变,而且 MDFA3 和 MDFA4 的结果较 MDFA2 更利于判断突变点的位置.

传统的突变检验方法如滑动 t -检验、Gramer 法、Mann-Kendall 法、Yamamoto 法等的检测结果均与所选取的子序列长度有关,不同长度的子序列得到的结果有时甚至大相径庭.因此,十分有必要考察 MDFA 的检测结果与子序列长度的相关性.为此我们仍然采用 MDFA2, MDFA3 和 MDFA4,研究了不同子序列长度对 MDFA 检测结果的影响,下面中主要以子序列长度 L 取 90 和 200 两种情形为例来研究这一问题.

图 4 给出了子序列长度 L 分别取为 90 和 200 时,采用 MDFA 方法对理想序列 IS1 分析得到的标度指数序列.由图 4(a)~(c)可知,当子序列长度 L 为 90 时,在 $t \in [500, 580]$ 时,各阶 MDFA 得到的标度指数在此区间的波动幅度相对于其他区域明显偏大,局部极大值和局部极小值之间的转换更加频繁.从图 4(a)可以发现,除在系统动力学结构突变点附近标度指数变化较大外,在系统动力学结构稳定的区域也出现了一些局部极小值.其原因一方面在于子序列长度过短,导致标度分析过程中最大去趋势波动分析的窗口长度较小,因而在对双对数曲线进行线性拟合时的点非常少,以至于得到的标度指

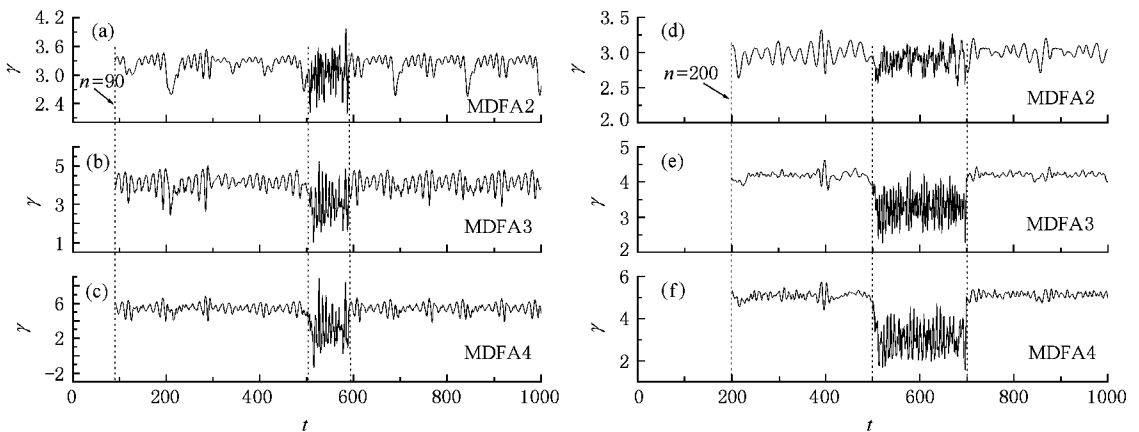


图4 理想时间序列 IS1 的 MDFA 分析结果 (a) $L=90$, MDFA2 (b) $L=90$, MDFA3 (c) $L=90$, MDFA4 (d) $L=200$, MDFA2 (e) $L=200$, MDFA3 (f) $L=200$, MDFA4

数(即拟合斜率)在正常区域内的波动也会出现局部异常的现象.另一个重要的原因可能是经过 MDFA2 处理后的序列中仍然存在着较为严重的残留趋势,以至于得到的标度指数不能正确地反映原系统的标度特征.从图 4(a)还可以看出,在 $t = 210, 689, 843, 997$ 等四处出现了极小值,且在这四个点附近标度指数的波动幅度与系统动力学结构突变点附近的计算结果十分接近.因此,在实际应用中,如遇到这种情况很难将这些极值与真实的动力学结构突变点区分.值得庆幸的是,高阶 MDFA 方法能够很好解决这一问题.随着 MDFA 阶数的增加,序列中存在的各种趋势被逐步去除,MDFA 的检测结果越加清晰.由图 4(a)–(c)可以清晰地看到,高阶 MDFA 方法具有更强的去趋势能力,而 MDFA4 的突变检测结果明显好于 MDFA2,能够非常精确地检测出突变的位置.

由图 4(d)–(e)可知,在子序列长度 L 取为 200 时,根据标度指数波动幅度的大小可以将其大致的分成 $[200, 500]$ 、 $[501, 700]$ 及 $[701, 1000]$ 三个区间.在第一和第三区间内标度指数的变化情况类似,随着子序列的移动标度指数的波动幅度较小,在这两个区间标度指数的小幅度波动主要是由窗口数据长度过短所引起.在第二区间,MDFA2 得到的标度指数的波动幅度与第一、第三区间相差不大,明显的区别在于在此区间内标度指数的变率快于其他两个区间.但由图 2(e)和(f)可以发现,MDFA3 和 MDFA4 获取的标度指数的变化在三个区间具有十分鲜明的特征,其中在第二区间的波动幅度明显大于其他两个区间,能够很容易据此判断出突变点所在位置.从理想序列 IS1 在三种不同子序列长度下的试验看,MDFA 的检测结果几乎与子序列的选取没有必然的联系,当子序列选取其他值时,我们得到了类似的结果.

图 5 为理想序列 IS1 的 $ApEn$ 检测结果.从图 5(a)可以看出,在子序列长度 $L = 200$ 时,类似于 MDFA 的分析结果, $ApEn$ 值的演变情况也可以粗略地分为三个不同的区间,即 $[200, 390]$ 、 $[391, 600]$ 、 $[601, 1000]$.在第二区间与第一和第三区间交界处发生了两次显著的均值突变(滑动- t 检验,图略),且在第二区间内子序列的 $ApEn$ 值明显大于其他区间.当子序列长度增大至 300 后,发现 $ApEn$ 值仍然可以分为三个不同的区间,且在第二区间 $[484, 600]$ 内子序列的 $ApEn$ 值显著增大,但在 $t > 600$ 后 $ApEn$

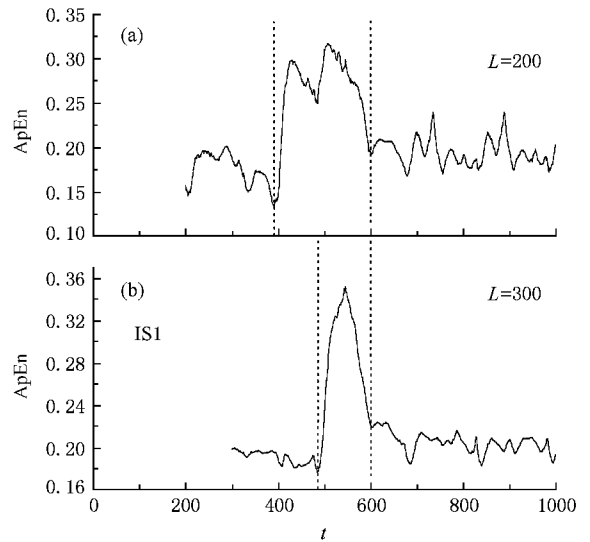


图 5 理想序列 IS1 的 $ApEn$ 检测结果 (a) $L = 200$ (b) $L = 300$

值又迅速减小,此后变化较稳定.由 $ApEn$ 值的物理意义可知,其值越大表明原时间序列的复杂性程度越高.因此,我们可以初步判断,在子序列长度 $L = 200, 300$ 时, $ApEn$ 在第二区间内所包含数据的复杂性程度较其他区域明显偏大,而且在第一和第三区间内,子序列的 $ApEn$ 值不仅偏小,而且其变化也相当稳定,故而在第一区间和第三区间系统的动力学特征比较稳定.从第一和第三区间内 $ApEn$ 的值来看,其平均值的差异十分显著,表明在这两个区间内的子序列的复杂性不同,因而其动力学方程也各异.第二区间子序列的 $ApEn$ 值的变化却非常剧烈,因此该区间应为一个突变区域,原时间序列在此区间内发生了动力学结构突变,导致在这一区域内子序列包含了不同的动力学系统的数据,因而其动力学复杂性程度显然要较单一动力学方程产生的数据的复杂性更大.从理想序列 IS1 的构造中可知,在 $t = 501$ 处, Lorenz 方程的 Rayleigh 数由 29 突然减小为 28,故对该时刻附近子序列的 $ApEn$ 值突然变大也就不难理解.由此可见, $ApEn$ 对于 IS1 的动力学结构突变检测具有一定的能力.同时我们也注意到, $ApEn$ 检测到的突变区间随着子序列长度的不同而不断变化着,通过 $ApEn$ 难以对时间序列中的突变点进行准确的检测(图 5).因此,在准确提取动力学结构突变点这一方面,MDFA 的优势是 $ApEn$ 难以媲美的.

在非线性时间序列 IS1 的动力学结构突变检测中,MDFA 检测结果较 $ApEn$ 更为准确,这种优势是普遍的现象,还是仅仅为一个特例?为此,类似于

IS1 的分析,我们分别采用 MDFA 和 ApEn 对理想序列 IS2 进行了动力学结构突变检测试验. 图 6(a)展示了由周期强迫下的 Brussel 振子^[32] x 分量构造的理想序列 IS2. 从直观上看,序列的演变规律较为简单,数据的连续性很好,无法判断何处发生了动力学结构突变,但当 $t = 501$ 时,Brussel 振子中表征外强迫强度的参数 a 由 0.05 突变变为 0.051,由文献 [32] 可知,在这两种不同参数情况下,Brussel 振子均进入混沌状态(已去除 40000 步积分暂态). 因此,理想序列 IS2 不仅仅是一个典型的非线性时间序列,而且在 $t = 501$ 时发生了 I 型动力学结构突变.

我们首先采用 MDFA 对理想序列 IS2 进行了动力学结构突变检测,图 6(b)(c)和(d)分别提供了子序列长度 $L = 100$ 时 MDFA2, MDFA3 及 MDFA4 的检测结果. 从图 6 容易发现,各阶 MDFA 的检测结果有以下一个共同的特点:当子序列滑动至 $t = 501$ 后,计算得到的标度指数的波动开始变得异常剧烈,这种情形一直持续到 $t = 600$,自 $t \geq 601$ 以后标度指数的波动幅度骤然变小. 进一步对比图 6(b)(c)和(d)可以发现,由 MDFA2 获取的标度指数平均值 (2.8582) 远大于 MDFA3 和 MDFA4 (其平均值分别为 0.87134, 1.00262). 为此,文献 [30] 中对样本量更大的 Brussel 振子 x 分量进行了研究,详细分析了在 DFA 处理过程中双对数曲线的拟合情况,发现当 $\lg s < 1.5$ (s 为去趋势波动分析过程中的窗口尺度)时,采用 DFA3 和 DFA4 所获取的双对数曲线存在一个标度指数较小的无标度区间,即高阶 DFA 分析得到的双对数曲线在此发生了转折(crossover)现象. 这一结果与 DFA2 分析不一致,其原因在于 DFA2 没有

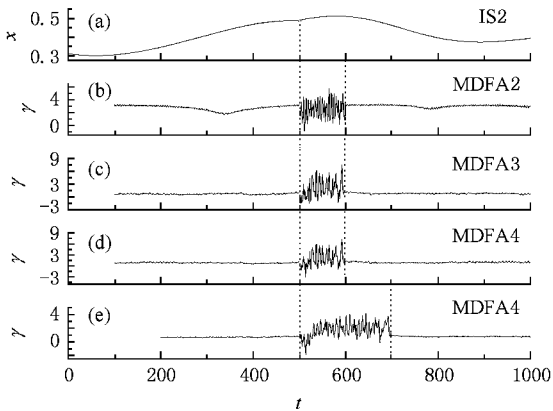


图 6 理想时间序列 IS2 及其动力学突变检测 (a)理想序列 IS2 在 $t = 501$ 时发生了 I 型动力学结构突变 (b) $L = 100$, MDFA2 检测 (c) $L = 100$, MDFA3 检测 (d) $L = 100$, MDFA4 检测; (e) $L = 200$, MDFA4 检测

完全去除序列中存在的趋势,掩盖了系统中真实的无标度特征. 但在 $\lg s > 1.5$ 后,各阶 DFA 分析得到的结果没有本质差别. 子序列长度 $L = 200$ 时的 MDFA4 检测结果与 $L = 100$ 时完全相同,其区别仅在于检测到突变区间的范围随之增大至与子序列长度 ($L = 200$) 相当(图 6(e)).

作为比较,理想序列 IS2 的 ApEn 检测结果已由图 7 给出. 由图 7(a)不难发现,在子序列长度为 200 时,通过 ApEn 方法检测到两个突变区间,分别约为 [525, 736] [946, 1000]. 显然,这两个区间均未包含真实的动力学结构突变点($t = 501$). 至于检测到的第一个突变区间,比较靠近序列中真实的突变点,但其 ApEn 的峰值约为 $t = 661$,与真实的突变点相差甚远. 在实际资料的动力学结构突变检测中,由于不可能事先获知系统在何时发生了动力学结构突变. 因此,在这种情况下根据 ApEn 的检测结果来判断突变点,得到的结论显然极不准确,而且 ApEn 所检测到的第二个区间明显是一个虚假的突变区. 当子序列长度 L 增大为 300 时,我们发现 ApEn 检测到的突变区间约为 [600, 800],虽然在子序列长度 L 取 200 时的虚假突变区间消失了,但是此时所检测到的突变区间仍然严重偏离了突变点的真实位置. 由此可见,ApEn 方法能够对理想序列 IS2 中的动力学结构突变点有所反映,但是其检测结果存在严重的漂移现象,而且所检测到的突变区间随着子序列长度 L 的不同而相差甚远,还存在虚假的突变区间. 因此,就 IS2 的动力学结构突变检测效果而言,MDFA 明显优于 ApEn. 本文对其他不同类型的动力

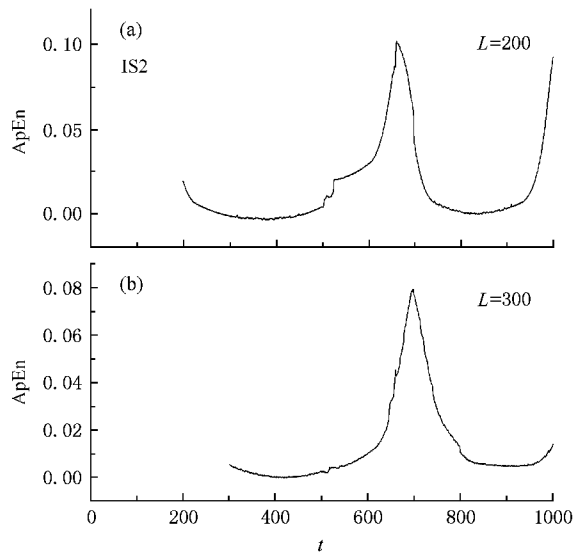


图 7 理想序列 IS2 的 ApEn 检测结果 (a) $L = 200$ (b) $L = 300$

学结构突变序列进行了类似的比较试验,得到了相同的结论,由于篇幅所限,在此不再赘述.

3.3. MDFA 与 ApEn 在实际资料中的应用比较

在本文中,我们以西沙站 1960 年—2005 年的日地表极端最高温度为例,比较研究了 MDFA 方法与 ApEn 方法在实际资料的动力学结构突变检测中的性能. 在对西沙站日地表极端最高温度的分析过程中, L 的长度取为 365 d,滑动步长为 1 a. 图 8 给出了西沙站日地表极端最高温度的 MDFA1 分析结果. 从图 8 容易发现,在子序列长度 L 分别为 8, 10, 12, 14 a 时,标度指数的演化规律基本类似,均在 1978 年处发生了均值突变,且通过了显著性水平为 0.01 的检验(滑动- t 检验,图略),子序列长度为其他值时得到了相同结论. 这表明西沙站的日地表极端最高温度在 1978 年发生了动力结构突变. 更高阶的 MDFA 分析结果与 MDFA1 类似,可见 MDFA 方法的检测结果不仅在理想试验中不依赖于子序列的长度,而且对于实际资料的检测结果与子序列长度仍然没有必然的联系,检测结果非常稳定. 因此,MDFA 方法在实际观测资料中的动力学结构突变检测中具有巨大的实用价值.

西沙站日地表极端最高温度相应的 ApEn 分析结果已在图 9 中给出. 从图 9 可以发现,当子序列

长度 $L = 8$ a 时,其 ApEn 演化曲线大致可以分为以下三个阶段:1960 年—1967 年、1968 年—1977 年、1978 年—1996 年,在第二个阶段 ApEn 处于一个低值时期,这表明该时段系统的动力学结构复杂性较低,因而在第二阶段系统的动力学结构可能发生了突变,但仅根据图 9(a)难以准确判断突变点. 类似于西沙站日地表极端最高温度的 MDFA 分析,我们逐步增大子序列的长度至 $L = 10, 12, 14$ a,期望能够结合不同子序列情况下的检测结果更为准确地提取系统中存在的动力学结构突变点. 从图 9(b)–(d)中不难发现,三种不同子序列长度的 ApEn 检测结果虽然都是在 1968 年附近 ApEn 的值为最小,但是自 1968 年以后,随着子序列的移动 ApEn 值逐步变大,这意味着系统的复杂性随着时间的增加而增大,系统一直处于不稳定的状态. 因此,难以根据增大子序列长度得到 ApEn 的演化来判断系统(指西沙站日地表极端最高温度的控制系统)的动力学结构突变. 鉴于此,我们采取减小子序列长度的方法来进行进一步利用 ApEn 对西沙站日地表极端最高温度进行动力学结构突变检测,图 10 展示出相应的检测结果. 在图 10(a)和(b)中,ApEn 曲线的演变情况基本相同,均在 1977 年发生了均值突变,且通过了显著性水平为 0.01 的检验(滑动- t 检验,图略). ApEn 值在 1977 年前后的显著差异表明系统的动力学结

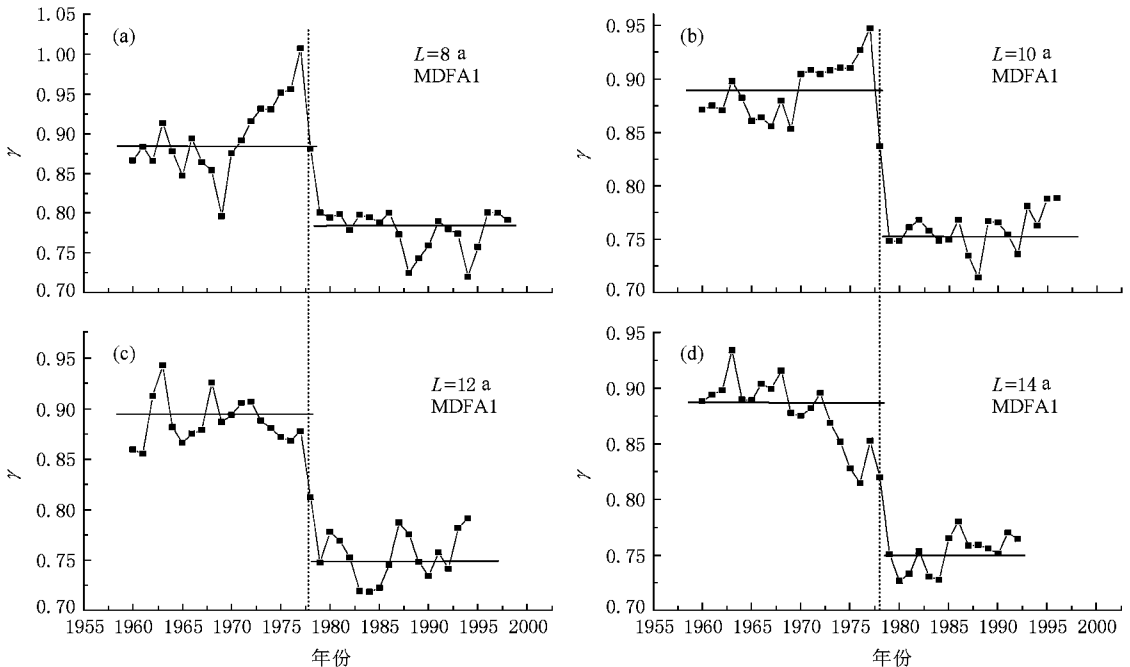


图 8 西沙站日地表极端最高温度(1960 年—2005 年)的 MDFA1 分析结果. 滑动步长为 1 a. (a) $L = 8$ a (b) $L = 10$ a (c) $L = 12$ a (d) $L = 14$ a. 水平实线分别为子序列滑动至 1978 年前后标度指数的平均值.

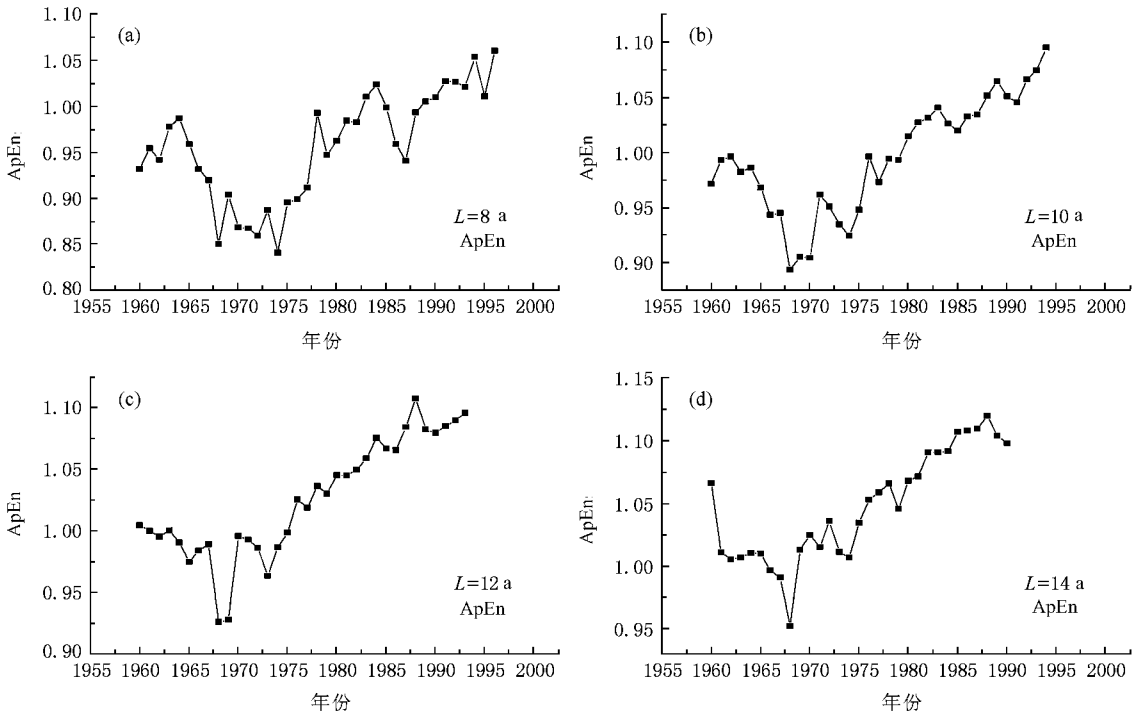


图 9 西沙站日地表极端最高温度(1960年—2005年)的 ApEn 分析 (a) $L=8$ a (b) $L=10$ a (c) $L=12$ a (d) $L=14$ a

构复杂性在此年份发生了突变,而且此后系统的 ApEn 曲线的演变非常稳定,表明系统自 1977 年进入了一个稳定变化时期.子序列长度 $L=8$ a 时的 ApEn 分析结果(图 9(a))实际上也在 1977 年发生了显著的均值突变,这可以通过滑动- t 检验很容易证实,但我们发现 $L=10, 12, 14$ a 时的 ApEn 值发生均值突变的时间分别为 1975 年、1975 年、1979 年,这

与 $L=5, 6, 8$ a 时的结果迥异.由以上所述可知,西沙站日地表极端最高温度的动力学结构突变检测中,相对于 ApEn, MDFA 的优势是非常明显的,ApEn 检测结果仅在子序列长度较小时(如 $L=5, 6, 8$ a)基本上与 MDFA 相当.因此,实际观测资料的 ApEn 检测结果仍然与子序列的选取有关,在实际应用 ApEn 的过程中对这一问题应予以重视.

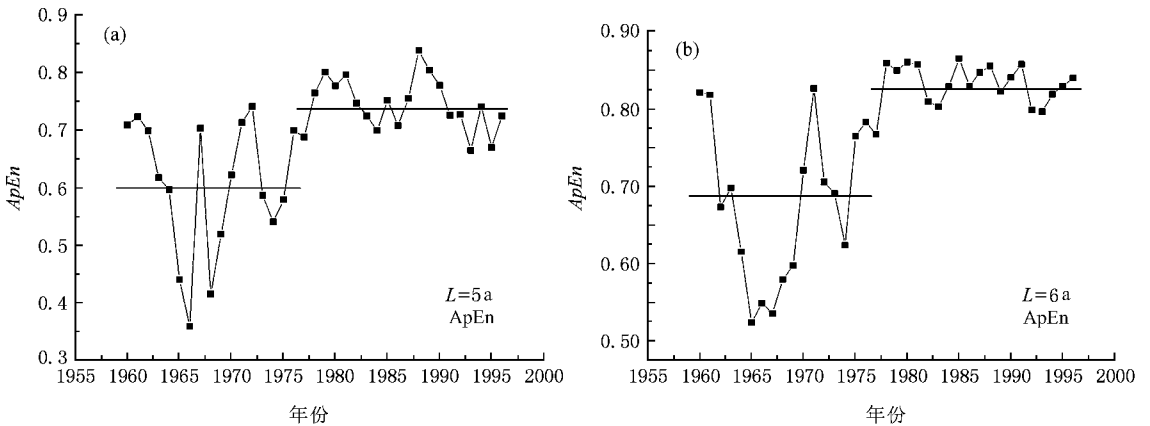


图 10 西沙站日地表极端最高温度的 ApEn 分析 (a) $L=5$ a (b) $L=6$ a



4. 结 论

本文将一种新的动力学结构突变检测方法 (MDFA) 与 A_{pEn} 方法进行了比较. 结果表明, 对于线性时间序列, MDFA 与 A_{pEn} 均能对序列中的动力学结构突变信号进行有效检测, 而且均能够根据相应的指标值(如标度指数和表征时间序列复杂性的 A_{pEn} 值)区分不同的动力学系统, 检测结果也基本相当. 就非线性时间序列而言, A_{pEn} 方法虽然能够对理想序列中的动力学结构突变有所识别, 但不能准确地给出突变时间, 且检测到的突变区间严重依赖于子序列的选择, 在某些情况下甚至会检测到虚假的突变区间. 而 MDFA 方法则能够非常准确地提取出时间序列中的动力学结构突变点, 虽然检测

到的突变区间与子序列的长度基本相当, 但不同子序列检测到的突变时间基本上相同, 检测到的突变点与子序列的长度没有必然的联系, 且检测结果中不存在虚假的突变区间. 显然, 在非线性时间序列的动力学结构突变检测中, MDFA 方法较 A_{pEn} 具有更为明显的优势. 根据已有的研究结果^[21]可知, MDFA 具有较强的抗干扰能力. MDFA 方法与 A_{pEn} 在中国西沙站近 46 a 日地表极端最高温度资料的动力学结构突变检测中的应用, 进一步显示了 MDFA 分析在实际观测资料中的实用性, 充分展示了这一方法的优越性能. 由于自然界中的许多时间序列具有非线性特征, 而且展现出各种标度行为, 因此 MDFA 方法在时间序列的动力学结构突变检测中无疑具有广阔的应用前景和重大的实用价值.

- [1] Fu C B, Wang Q 1992 *Chin. J. Atmos. Sci.* **16** 483 (in Chinese) [符淙斌、王 强 1992 大气科学 **16** 483]
- [2] Fu C B 1994 *Chin. J. Atmos. Sci.* **18** 373 (in Chinese) [符淙斌 1994 大气科学 **18** 373]
- [3] Yang M X, Yao T D 1999 *Discov. Nat.* **18** 29 (in Chinese) [杨梅学、姚檀栋 1999 大自然探索 **18** 29]
- [4] Mann H B 1945 *Econometric* **13** 245
- [5] Kendall M G, Charles G 1975 *Rank Correlation Methods* (New York: Oxford University Press) p202
- [6] Yamamoto R, Iwashima T, Sanga N K 1985 *J. Meteor. Soc. Jpn.* **63** 1157
- [7] Herron M M, Langway C C 1985 *Greenland Ice Core: Geophys. Geochem. Envir.* **33** 77
- [8] Zhang P Y, Ge Q S 1997 *Earth Sci. Frontiers* **4** 122 (in Chinese) [张丕远、葛全胜 1997 地学前缘 **4** 122]
- [9] Liu S D, Chen J, Liu S K 1999 *J. Appl. Meteor. Sci.* (Suppl.) **10** 10 (in Chinese) [刘式达、陈 炯、刘式适 1999 应用气象学报(增刊) **10** 10]
- [10] Liu S D, Rong P P, Chen J 2000 *Acta Meteor. Sin.* **58** 111 (in Chinese) [刘式达、荣平平、陈 炯 2000 气象学报 **58** 111]
- [11] Wang S Q, Feng G L, Dong W J, Li J P 2005 *Acta Phys. Sin.* **54** 5487 (in Chinese) [万仕全、封国林、董文杰、李建平 2005 物理学报 **54** 5487]
- [12] Feng G L, Gong Z Q, Dong W J, Li J P 2005 *Acta Phys. Sin.* **54** 5494 (in Chinese) [封国林、龚志强、董文杰、李建平 2005 物理学报 **54** 5494]
- [13] Gong Z Q, Feng G L, Wan S Q, Li J P 2006 *Acta Phys. Sin.* **55** 477 (in Chinese) [龚志强、封国林、万仕全、李建平 2006 物理学报 **55** 477]
- [14] Feng G L, Hou W, Dong W J 2006 *Acta Phys. Sin.* **55** 962 (in Chinese) [封国林、侯 威、董文杰 2006 物理学报 **55** 962]
- [15] Zhang W, Gao X Q, Dong W J, Li J P 2006 *Acta Phys. Sin.* **56** 2657 (in Chinese) [张 文、高新全、董文杰、李建平 2006 物理学报 **55** 2657]
- [16] Gong Z Q, Feng G L, Dong W J, Li J P 2006 *Acta Phys. Sin.* **55** 3180 (in Chinese) [龚志强、封国林、董文杰、李建平 2006 物理学报 **55** 3180]
- [17] Pincus S M 1991 *Proc. Natl. Acad. Sci. USA* **88** 2297
- [18] Pincus S M, Goldberger A L 1994 *Am. J. Physiol.* **266** H 1643
- [19] Pincus S M 1995 *Chaos* **5** 110
- [20] Wang Q G, Zhang Z P 2008 *Acta Phys. Sin.* **57** 1996 (in Chinese) [王启光、张增平 2008 物理学报 **57** 1996]
- [21] He W P, Feng G L, Wu Q, Wan S Q, Chou J F 2009 *Non. Proc. Geophys.* **15** 601
- [22] Peng C K, Havlin S, Schwartz M, Stanley H E 1991 *Phys Rev. A* **44** R2239
- [23] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E, Goldberger A L 1994 *Phys Rev. E* **49** 1685
- [24] Peng C K, Havlin S, Stanley H E, Goldberger A L 1995 *Chaos* **5** 82
- [25] Bunde E K, Bunde A, Havlin S, Goldreich Y 1996 *Physica A* **231** 393
- [26] Bunde E K, Bunde A, Havlin S, Roman H E, Goldreich Y, Schellnhuber H J 1998 *Phys. Rev. Lett.* **81** 729
- [27] Bunde A, Havlin S, Kantelhardt J W, Penzel T T, Peter J H, Voigt K 2000 *Phys. Rev. Lett.* **85** 3736
- [28] Koutsoyiannis D 2002 *Hydro. Sci. J.* **47** 573
- [29] Koutsoyiannis D 2003 *Hydro. Sci. J.* **48** 3
- [30] He W P 2008 *Ph. D. Dissertation* (Lanzhou: Lanzhou University) p13 (in Chinese) [何文平 2008 博士学位论文(第 13 页)]
- [31] Lorenz E N 1963 *J. Atmos. Sci.* **20** 130

- [32] Qiu Q , He W P 2006 *J. Yangtze Univ. (Nat. Sci. ed)* **3** 22 (in Chinese] 邱 庆、何文平 2006 长江大学学报 (自然科学版) **3** 22]
- [33] Hastings H M , Kissells R 1998 *Nat. Res. Model.* **11** 83
- [34] Talkner P , Weber R O 2000 *Phys. Rev. E* **6** 150
- [35] Govindan R B , Vjushin D , Brenner S , Bunde A , Havlin S , Schellnhuber H J 2001 *Physica A* **294** 239
- [36] Syroka J , Toumi R 2001 *Geo. Res. Lett.* **28** 3255
- [37] Caballero R , Jewson S , Brix A 2002 *Clim. Res.* **21** 127
- [38] Eichner J F , Koscieny-Bunde E , Bunde A 2003 *Phys. Rev. E* **8** 046133
- [39] Fraedrich K , Blender R 2003 *Phys. Rev. Lett.* **90** 108501
- [40] Eichner J F , Kantelhardt J W , Bunde A , Havlin S 2007 *Phys. Rev. E* **75** 011128

Comparison of characteristics of moving detrended fluctuation analysis with that of approximate entropy method in detecting abrupt dynamic change ^{*}

He Wen-Ping^{1)†} Wu Qiong²⁾ Zhang Wen³⁾ Wang Qi-Guang⁴⁾ Zhang Yong¹⁾

1)(National Climate Center , Beijing 100081 , China)

2)(Chinese Academy of Meteorological Science , Beijing 100081 , China)

3)(Tongda College , Nanjing University of Post and Telecommunications , Nanjing 210003 , China)

4)(College of Physics Science and Technology , Yangzhou University , Yangzhou 225002 , China)

(Received 16 July 2008 ; revised manuscript received 12 September 2008)

Abstract

Approximate entropy (ApEn) is well known to be an effective abrupt change detection method in dynamic structure. Based on this , we compare the performances between the moving detrended fluctuation analysis (MDFA) and ApEn in detecting abrupt dynamic change. The results show that the MDFA results almost do not depend on length of subseries , and while the ApEn could identify dynamic structure to some extent but still depends on the length of subseries. At the same time , there exists a huge drift for the ApEn results which means the actual time-instants of abrupt change don't coincide with the detected one. Therefore , compared with ApEn , MDFA is more suitable to be used to detect abrupt dynamic change , and the advantage of MDFA is obvious.

Keywords : moving detrended fluctuation analysis , approximate entropy , abrupt dynamic change

PACC : 9260X

^{*} Project supported by the State Key Development Program for Basic Research of China (Grant Nos. 2004CB418303 , 2006CB400503) and the State Key Program of Science and Technology of China (Grant Nos. 2007BAC29B01 , 2007BAC03A01).

[†] E-mail : wenping_he @ 163.com