

# 基于近似熵的突变检测新方法\*

何文平<sup>1)†</sup> 何涛<sup>2)</sup> 成海英<sup>3)</sup> 张文<sup>4)</sup> 吴琼<sup>5)</sup>

1) (国家气候中心, 北京 100081)

2) (济南市环境保护监测站, 济南 250014)

3) (盐城工学院基础教学部, 盐城 224002)

4) (南京邮电大学通达学院, 南京 210003)

5) (国家卫星气象中心, 北京 100081)

(2010年6月6日收到; 2010年7月12日收到修改稿)

近似熵是一个有效的非线性动力学指数, 能够用于表征时间序列的复杂性, 通过滑动窗口技术, 近似熵对于一维时间序列的动力学结构突变具有一定的识别能力, 但其突变检测结果依赖于子序列长度的选择, 且不能准确定位突变点. 鉴于此, 本文提出了一种新的突变检测方法——滑动移除近似熵. 测试结果表明, 滑动移除近似熵具有检测结果稳定性好、准确性高等特点, 明显优于滑动近似熵和 Mann-Kendall 方法, 其在实际观测资料中的应用进一步证实了新方法的可靠性.

**关键词:** 近似熵, 滑动移除近似熵, 突变检测

**PACS:** 92.60.Wc

## 1. 引言

近年来, 随着非线性科学的不断发展, 非线性动力学领域中的诸多指数被用于描述系统的动力学特征, 如关联维数、李雅普诺夫指数、柯尔莫哥罗夫熵等<sup>[1-8]</sup>, 其中关联维数描述了一个复杂系统所需要的最少有效状态变量; 李雅普诺夫指数则被用于度量非线性耗散系统的两个运动特性: 收缩与发散, 表征了相空间中在不同方向上轨道的收缩或发散的速率, 是判断系统动力学行为是否为混沌的一个很好的物理指标; 由于每一个正的李雅普诺夫指数表示该方向上长度是伸长的, 因此该方向上初始信息不断损失. 所有正的李雅普诺夫指数之和代表了总的信息损失率, 此即为柯尔莫哥罗夫熵的定义, 它是表征系统产生信息量的多少、快慢的物理量.

然而, 要直接根据这些指数的定义来计算分维和李雅普诺夫指数并不容易, 实际中往往会面临单变量的一维时间序列, 很显然, 系统的非线性特征信息蕴藏在这些序列中, 如何从中提取这类信息成

为一个非常困难的问题<sup>[9]</sup>. 鉴于此, Packara 提出了通过时间序列相空间重构来计算李雅普诺夫指数和关联维数, 但是为了得到一个好的最大李雅普诺夫指数估计值, 必须要求: 1) 所给定的时间序列有足够的长度, Wolf 认为, 这个长度的估计范围为  $10^D - 30^D$  ( $D$  为吸引子的分维); 2) 延迟时间的选取; 3) 临近相空间点的选取. 但目前, 相空间重构中关于延迟时间的选取是一个难题, 而对初始相点的最邻近点的选择更为困难, 尤其是需要花费大量的机时不断增加嵌入空间的维数以便获得稳健的李雅普诺夫指数. 因此, 这些非线性动力学指数在实际应用中显得并不方便, 需要寻找更为有效、稳健且对数据长度要求较短的新方法.

近似熵<sup>[10-12]</sup> (approximate entropy, ApEn) 方法无疑是一个理想的非线性动力学指数, 它具有所需数据点较少、有较好的抗噪和抗干扰能力且适用于随机信号、确定性信号或两种混合信号等特点. 因此, ApEn 方法在医学、生物学、机械设备故障诊断、语音信号端点检测等方面得到了广泛应用<sup>[13-17]</sup>.

气候系统是一个复杂的非线性动力学系统, 如

\* 国家自然科学基金 (批准号: 40905034, 40875040, 40930952), 公益性行业 (气象) 科研专项基金 (批准号: GYHY200806005, GYHY200906019) 资助的课题.

† E-mail: wenping\_he@163.com

何从过去的气候记录中检测气候突变,对于预测未来的气候演变趋势有着重要的现实意义<sup>[18-26]</sup>. 已有学者对此做了大量卓有成效的工作<sup>[27-35]</sup>,例如基于标度理论先后发展了滑动去趋势波动分析方法、滑动移除去趋势波动分析方法、滑动移除重标极差方法<sup>[25-28]</sup>,这些方法的出现很大程度上丰富了突变检测的手段,并能够很好的对大量实际观测资料进行动力学结构突变检测,但是必须指出的是,这些方法也存在一个共同的缺陷,即应用这三种方法的前提是待分析时间序列具有分形特征,但并不是所有的时间序列都能够满足这一条件. ApEn 对于各种信号的广泛适用性,使得其成为一种检测突变的另一途径. 类似于滑动- $t$  方法,王启光等<sup>[36]</sup>将 ApEn 方法引入到气候系统的动力学结构突变检测中,利用滑动技术通过理想试验证实了 ApEn 具有区分不同动力学结构的能力,表明其能够用于动力学结构突变检测中,并将其应用于实际气候资料的突变检测,得到了较好的结果. 但是这种滑动技术本身存在着众所周知的先天不足:其一,检测结果依赖于子序列长度;其二,不能够准确的定位突变点的位置,只能给出一个大致突变区域,这显然不能够满足实际应用的需求.

借鉴于滑动移除去趋势波动分析方法、滑动移除重标极差方法中所采用滑动移除数据的方式,本文对 ApEn 在突变检测中的应用提出了一种改进的新方法——滑动移除近似熵 (moving cut data-approximate entropy, MC-ApEn). 试验结果验证了新方法的可靠性和准确性,其检测结果明显优于简单的滑动 ApEn 方法和 Mann-Kendall 方法,其在逐日降水资料中的应用进一步证实了新方法的可靠性,展示了其潜在的应用前景.

## 2. 方 法

### 2.1. ApEn 方法

ApEn 是一种基于边缘概率分布统计量化时间序列复杂程度的方法,其算法如下<sup>[10-12]</sup>:

1. 对时间序列  $\{u(i), i=1, 2, \dots, N\}$  进行相空间重构,重构维数为  $m$ ,据此可构造一组维数为  $m$  的新向量  $X(1), X(2), \dots, X(N-m+1)$ , 其中,

$$X(i) = \{u(i), u(i+1), \dots, u(i+m-1)\}, \quad i = 1, 2, \dots, N-m+1; \quad (1)$$

2. 计算任意向量  $X(i)$  与其余向量  $X(j)$  之间的相对欧式距离  $d[X(i), X(j)]$ ;

3. 设定容许偏差为  $r$ , 统计每个向量  $X(i)$  的  $d[X(i), X(j)]$  小于  $r$  的数目,求出该数目与向量总数  $N-m+1$  的比  $C_i^m(r)$ ;

4. 将  $C_i^m(r)$  取对数,再求其对所有  $i$  的平均值,记作  $\phi^m(r)$ ;

5. 维数  $m$  增加 1,重复 1—4 步,得  $C_i^{m+1}(r)$  和  $\phi^{m+1}(r)$ ;

6. 近似熵的估计值定义为  $\text{ApEn}(m, r) = \phi^m(r) - \phi^{m+1}(r)$ .

显然,ApEn 值与  $m, r$  的取值有关. Pincus 建议取  $m=2, r=0.1-0.2\sigma$  ( $\sigma$  是原始数据的标准偏差). 在本文中,取参数  $m=2, r=0.15\sigma$ .

近似熵方法反映了时间序列在模式上的自相似程度,以及当维数变化时,产生新模式的可能性大小. ApEn 值愈大,说明产生新模式的概率愈大,序列越复杂,系统可预测性越差. 它给出新模式发生率随维数而增减的情况,因而反映数据在结构上的复杂性. 同时它并不是企图完全重构吸引子,所以对各种非线性时间序列有很好的适用性<sup>[36]</sup>.

### 2.2. MC-ApEn 方法

类似于滑动移除去趋势波动分析方法、滑动移除重标极差方法,MC-ApEn 方法的计算过程如下:

1. 首先选择滑动移除数据的窗口长度  $L$ ;

2. 然后从待分析时间序列的第  $i$  ( $i=1, 2, \dots, N-L+1, N$  为时间序列中的总记录个数) 个数据开始连续移除  $L$  个数据,再将剩余  $N-L$  个数据直接连在一起得到一个新的时间序列;

3. 利用近似熵方法计算新序列的 ApEn 值;

4. 保持移除数据的窗口长度不变,以步长  $L$  逐步移动窗口,重复 2, 3 步操作,直到原序列结束为止;

5. 通过 1—4 步操作可得到一个随着窗口移动的 ApEn 序列;

6. 基于不同动力学性质的数据其复杂性大小不相同,而具有相同动力学性质的数据的复杂性差异不大这一特点,结合步骤 5 中得到的 ApEn 序列判断突变点或突变区间.

## 3. MC-ApEn 在突变检测中的性能测试

### 3.1. MC-ApEn 在线性时间序列中的性能测试

为便于对新方法与滑动近似熵的检测结果进

行比较,线性理想时间序列仍然由王启光等<sup>[36]</sup>所构造的方程获取,

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

图 1(a) 给出该理想时间序列的演化曲线,由 (2) 式可知,在  $t = 1001$  时,理想序列 IS1 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函数和余弦函数所组成的新方程。

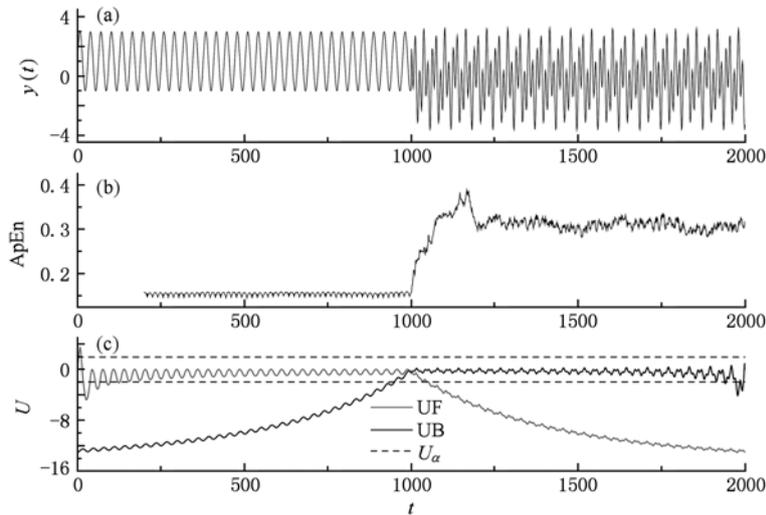


图 1 线性时间序列 IS1 及其突变检测 (a) 理想时间序列 IS1; (b) 滑动 ApEn 检测, 滑动窗口长度为  $S = 200$ , 滑动步长为 1; (c) Mann-Kendall 检测, UF 为时间序列顺序统计量, UB 为逆序统计量,  $U_\alpha$  线为置信区间, 显著性水平  $\alpha = 0.01$

滑动 ApEn 对 IS1 的分析结果表明, ApEn 能够根据时间序列的复杂性来很好的区分线性动力系统的异同, 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(图 1(b)). 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, 显然若想进一步准确定位突变点则存在一定的困难。

从 Mann-kendall 方法对 IS1 的检测结果中可知(图 1(c)), 自  $t = 995$  开始, 理想序列呈现出明显的下降趋势突变, 且突变点通过了  $\alpha = 0.01$  的显著性检验. Mann-kendall 检测到的突变点 ( $t = 995$ ) 与真实突变点  $t = 1001$  非常接近, 这表明 Mann-kendall 在某些情况下具备一定的检测系统动力学结构突变的能力. 但必须指出的是, 这并不意味着 Mann-kendall 方法是一种检测时间序列动力学结构突变的好方法. 从本质上而言, Mann-kendall 方法仅仅是

一种检测趋势突变的方法, 若时间序列中不存在明显的趋势突变, 则其无法检测出时间序列中存在的动力学结构突变, 不再赘述。

从图 2 中可以发现, 四种不同窗口长度情形下的 MC-ApEn 检测结果有一个共同特点, 即在  $t < 1000$  时, 移除数据后得到的 ApEn 值明显大于  $t > 1000$  时的情形, 这表明以  $t = 1000$  为界, 可以将 ApEn 序列分为两个不同的演化阶段. 根据 ApEn 的物理含义, 值越大意味着序列的复杂性越大. 因此, 移除数据后相对较小的 ApEn 值意味着所移除的数据相对而言复杂性更大, 进而可以判断出  $t > 1000$  后数据的复杂性强于  $t < 1000$  前的数据. IS1 的 MC-ApEn 检测结果表明, 移除等长度的数据, 若具有相同动力学属性, 则移除数据对于时间序列的 ApEn 估算的影响大致相同, 尽管在移除数据尺度较小时, 得到的 ApEn 值会存在局部波动相对较大的现象(如图 2(a)), 但就统计意义而言, 不论在滑动移除数据的窗口长度较小时(如  $L = 10$  和  $20$ ), 还是窗口长度较大时(如  $L = 50$  和  $100$ ), 无需借助其

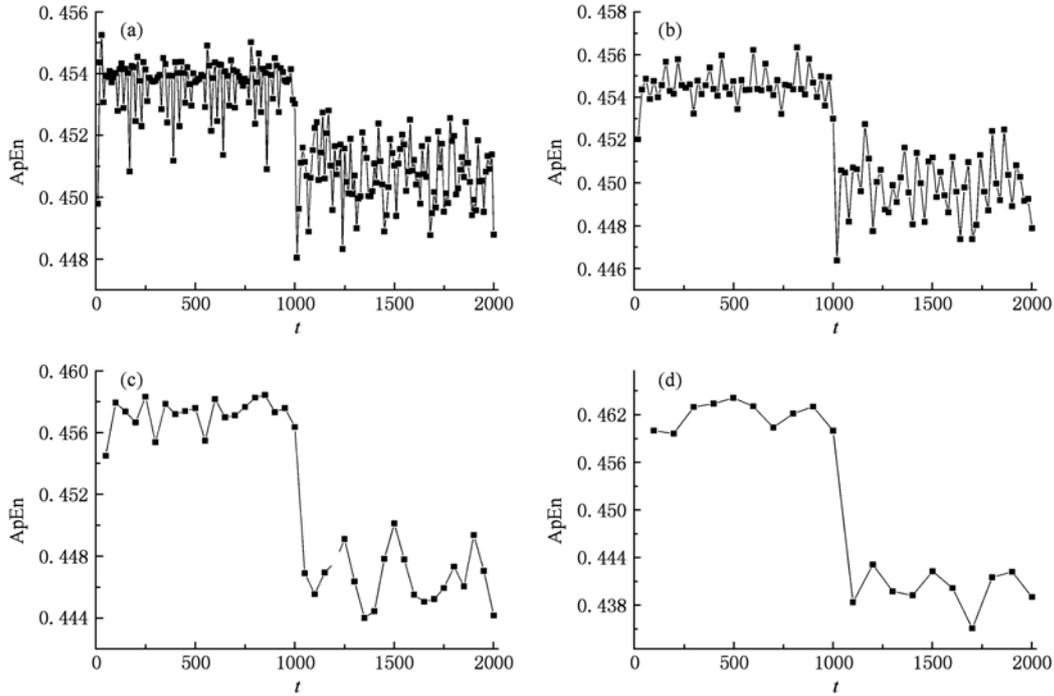


图2 线性时间序列 IS1 的 MC-ApEn 检测 (a) 滑动移除窗口长度  $L = 10$ ; (b)  $L = 20$ ; (c)  $L = 50$ ; (d)  $L = 100$

他统计工具就能够十分容易的将 MC-ApEn 得到的近似熵值分为明显变化的两个时段,进而能够大致的判断出约在  $t = 1000$  时发生了动力学结构突变. 由于图 2 中所选择的滑动移除窗口长度均属于一个特殊的长度,因此,得到的突变点完全一致,即在  $t = 1001$  处发生了突变. 由此引出如下问题:对于任意窗口尺度结果是否仍然完全一致? 结果显然是否定的,但不同窗口尺度检测到的突变点变化不大,即窗口尺度对检测结果影响较小.

为了更为准确地提取 IS1 中突变点的位置,在大致掌握 IS1 中突变点的位置后,在本节中利用 MC-ApEn 对 IS1 中部分时间序列进行突变检测,这里任取了 IS1 中的前 1200 个数据,图 3(a) 和 (b) 给出了滑动移除窗口长度  $L = 10$  和 20 时的 MC-ApEn 检测结果,相比于图 2,这样处理的优势在于可以降低整个待检测时间序列的复杂性,并将更易于区分不同复杂度的时间序列,有利于准确的判断突变点. 正如前文所言,窗口长度  $L = 10$  和 20 是两个比较特殊的尺度. 鉴于此,分别在滑动移除窗口尺度  $L$  为 1 和 2 时再次进行了 MC-ApEn 分析,结果表明,可以非常容易发现在  $t < 1001$  时,滑动移除数据对于所分析时间序列的 ApEn 的影响非常有规律,即 ApEn 随着滑动窗口的移动展现为一种准周期性的波动,且其波动幅度也大致相同(图 3(c) 和 (d)),

但自  $t = 1001$  后,这种波动规律明显不同于 ApEn 在此之前的演变,这意味着前后两段时间序列的复杂性不尽相同,因而数据的移除对于 ApEn 计算的影响各异. 基于此,可以准确地判断 IS1 在  $t = 1001$  处发生了动力学结构突变.

### 3.2. MC-ApEn 在非线性时间序列中的性能测试

为了测试 MC-ApEn 在非线性时间序列中的突变检测性能,构造了理想时间序列 IS2,序列总长为 2000,前 1000 个数据由 Logistic 映射<sup>[37]</sup>产生,其方程如下:

$$x_{n+1} = ux_n(1 - x_n), \quad x \in [0, 1], \quad (3)$$

上式中初值  $x_0 = 0.8$ , 参数  $u = 3.8$ . IS2 的后 1000 个数据为满足正态分布的随机数,其随时间的演化曲线已由图 4(a) 给出. 从其构造可知,IS2 在  $t = 1001$  时,系统的演化突然由确定性方程变为随机行为,由 ApEn 的物理意义可知,IS2 中后 1000 个数据的近似熵值要大于前 1000 个数据,即  $t = 1001$  后系统的复杂性突然增大了,可预测性变小了.

图 4 为理想时间序列 IS2 的 MC-ApEn 检测结果,从图 4(b) 和 (c) 中不难发现,在滑动移除窗口尺度  $L = 10$  和 5 时,ApEn 的演变趋势非常类似,在  $t = 1000$  前后,数据的移除对于 ApEn 值的计算影响各不相同,由于确定性方程 Logistic 模型产生的数据

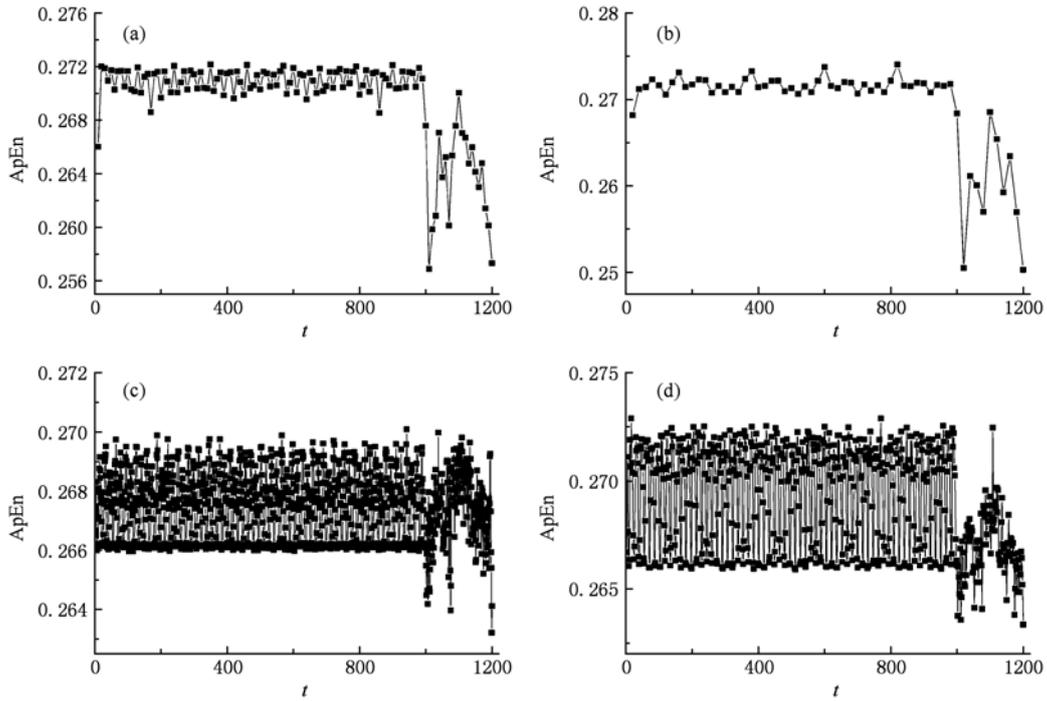


图3 在  $t < 1201$  时,对线性时间序列 IS1 中前 1200 个样本的 MC-ApEn 检测 (a) 滑动移除窗口长度  $L = 10$ ; (b)  $L = 20$ ; (c)  $L = 1$ ; (d)  $L = 2$

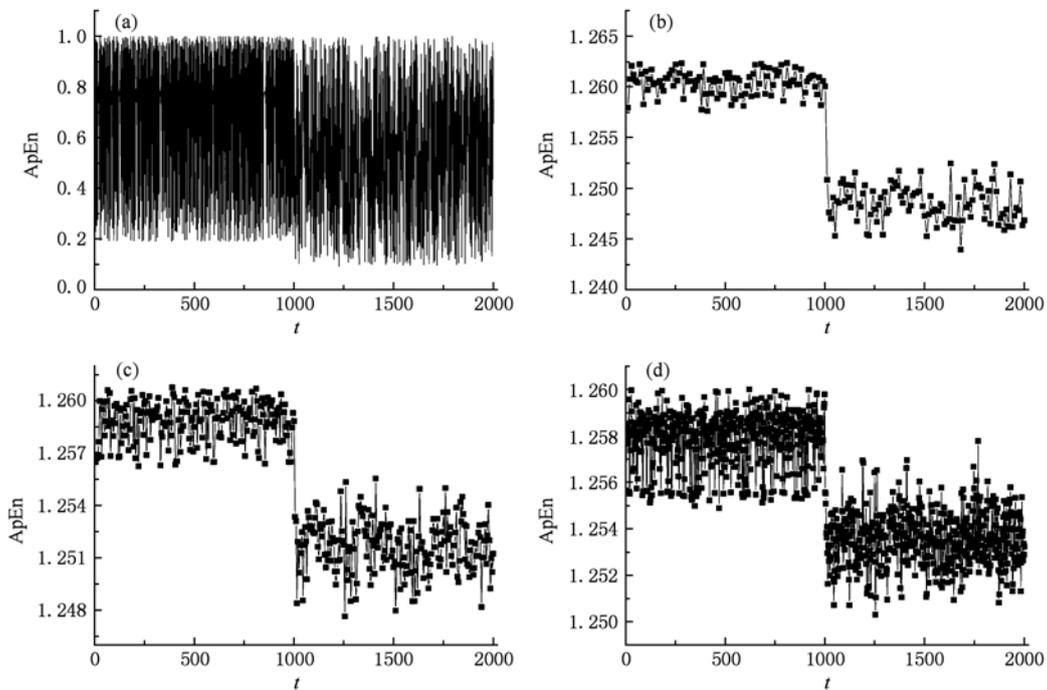


图4 非线性时间序列 IS2 及其 MC-ApEn 检测 (a) 理想序列 IS2, 长度为 2000; (b)  $L = 10$ ; (c)  $L = 5$ ; (d)  $L = 2$

其复杂性要明显低于随机数,因此,容易理解对于移除等长度的数据,移除随机数后所获取的 ApEn 值更小一些,即部分随机数的移除导致剩余数据的

复杂性减小.而且随着滑动移除窗口长度的增大,移除具有不同动力学性质的数据后计算得到的 ApEn 值之间的差异明显增大.随后考察了更小尺

度滑动移除窗口的情形,图 4(d) 给出了 IS2 在移除窗口长度  $L = 2$  时的 MC-ApEn 检测结果,与  $L = 10$  和 5 时的结果一致. 这表明,MC-ApEn 的检测结果对于窗口尺度的依赖性较小,该方法不仅适用于线

性时间序列的动力学结构突变检测,而且对于非线性时间序列的动力学结构突变也同样具有很强的识别能力.

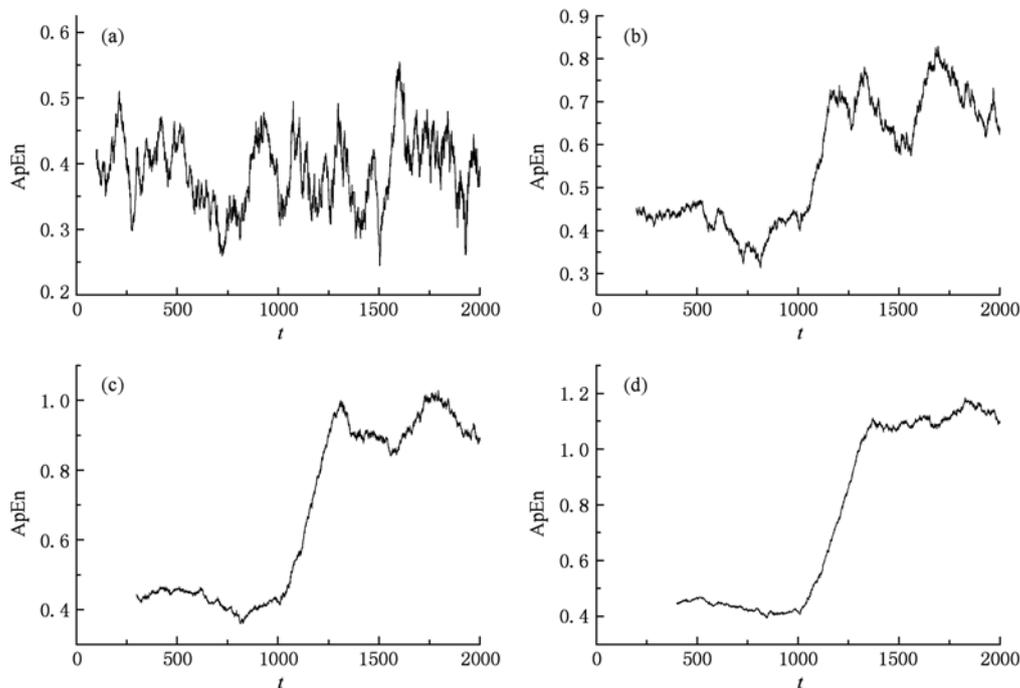


图 5 非线性时间序列 IS2 的滑动 ApEn 检测 (a) 滑动窗口长度  $S = 100$ ; (b)  $S = 200$ ; (c)  $S = 300$ ; (d)  $S = 400$

作为比较,图 5 给出了 IS2 的滑动 ApEn 检测结果,从图 5(a) 中可以发现,在滑动窗口尺度较小时,ApEn 的演变随着滑动窗口的移动其值不稳定,变化较大,这可能是 ApEn 算法自身的缺陷(即数据过短,得到的近似熵值不稳健)所造成. 因此,难以据此准确检测 IS2 中的动力学结构突变. 逐渐增大滑动窗口长度后发现,所得到的 ApEn 值渐渐变得比较平稳,而且对于 IS2 中的动力学结构突变也能够较好地反映出来. 从图 5(b), (c) 和 (d) 中可以发现三种不同窗口尺度下,均可以根据 ApEn 值的演变将其大致分为三个不同阶段,以滑动窗口长度  $S = 300$  时为例,可以粗略地分为:  $t \in [300, 1000]$ ,  $[1001, 1300]$ ,  $[1301, 2000]$  等三个演变阶段,在第一和第三区间 ApEn 值变化相对较为稳定,而且窗口长度越大,这种稳定性越好,而第二个区间为一个过渡阶段,这一特征主要是由于在滑动的过程中,窗口序列中所包含的数据由完全是确定性方程产生的数据变为随机信号与确定信号并存,这一过程直至窗口序列中的数据全部变为随机信号,因此,窗口序列的 ApEn 值也随之增大. 基于 ApEn 演变的这一特征,可以大致确定突变开始于

$t = 1000$  附近,但要想精确定位突变点还存在一定的困难. 王启光等<sup>[36]</sup> 提出采用滑动- $t$  检验法来进一步定位突变点,但由于滑动- $t$  检验的结果严重依赖于子序列长度的选取,故而未能从根本上解决此问题. 因此,这是影响滑动 ApEn 方法在突变检测中广泛应用的瓶颈之一.

当采用 Mann-Kendall 方法对非线性时间序列 IS2 进行突变检测时,发现顺序统计量 UF 和逆序统计量 UB 虽然在  $t = 993$  处存在一个交叉点,但其并未位于两条信度线之间,这意味着 IS2 中并不存在明显的趋势突变. 随后,对 IS2 中  $t \in [501, 1500]$  区间内的数据进行了 Mann-Kendall 检测,发现统计量 UF 和 UB 之间的交叉点为  $t = 1005$ ,且位于两条信度线之间,而 UF 在此处正好呈现显著的下降趋势,据此可以判断序列在  $t = 1005$  处存在一个下降的趋势突变. 对同一时间序列不同长度数据的 Mann-Kendall 分析却得到了完全不同的结论,表明了 Mann-Kendall 的分析结果严重依赖于时间序列的长度,显然该方法不适合用于检测时间序列的动力学结构突变.

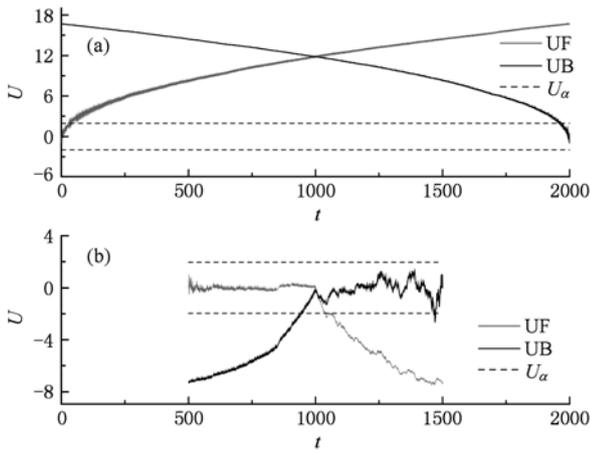


图6 非线性时间序列 IS2 的 Mann-Kendall 检测 (a) IS2 的 Mann-Kendall 检测; (b) 对 IS2 中  $t \in [501, 1500]$  区间内数据的 Mann-Kendall 检测

### 3.3. 应用实例

前文中所给出的两个例子均为理想化序列,而

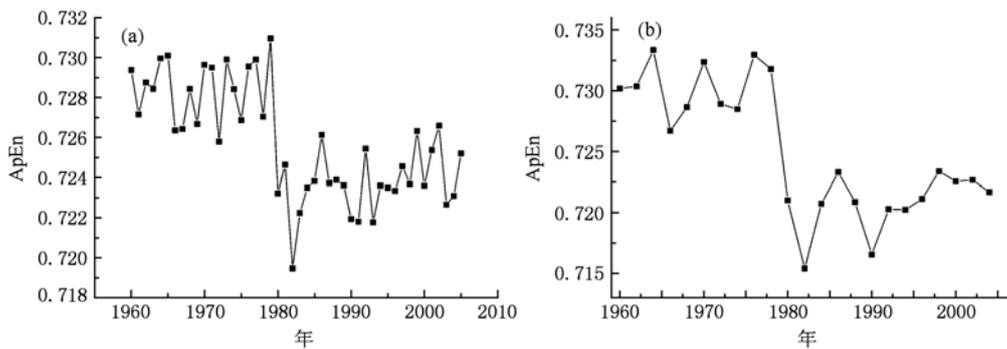


图7 呼玛气象观测站逐日降水资料的 MC-ApEn 检测,资料记录了 1960 年 1 月 1 日至 2006 年 12 月 31 日呼玛站的逐日降雨量 (a) 滑动移除窗口长度为 1 年,这里取  $L = 365$  天; (b)  $L = 730$  天

## 4. 结果与讨论

本文基于一种非线性动力学指数——ApEn 能够反映不同动力学系统的复杂性特征这一特性,借鉴于滑动移除技术,提出了一种新的突变检测方法——MC-ApEn 方法. 理想试验结果表明,滑动 ApEn 方法对线性和非线性时间序列的动力学结构突变具有一定的识别能力,但是存在不能够准确定位突变点的问题,甚至会出现滑动窗口长度较小时难以检测突变的现象,因此这一缺陷制约了该方法在实际中的应用;而 MC-ApEn 适合线性和非线性时间序列的动力学结构突变检测,其检测结果对于

实际观测资料要比理想序列复杂得多. 鉴于此,本文测试了 MC-ApEn 在降水观测资料中的突变检测性能. 采用的降水资料来源于中国气象局国家气象信息中心资料室,对于资料中少数缺测记录视为无降水. 图 7 给出了呼玛气象观测站 1960—2006 年逐日降水资料的 MC-ApEn 检测结果,与前文理想试验中类似的情况是,逐日降水资料的 ApEn 序列可以分为明显不同的两个稳定变化阶段,1979 年以前的 ApEn 值普遍大于 1980 年以后情形,这意味着自 1980 年后至 2006 年期间内,降水的复杂性突然增大了,其可预测性相对于 1960—1979 年而言减小了. 基于此,可以判断 1980 年呼玛气象观测站发生了气候突变,这与已有的研究结果基本一致<sup>[36,38]</sup>. 而图 7 中滑动移除窗口长度为 1 年和 2 年时检测到的突变点完全一样,则进一步表明本文提出的新方法其检测结果对于窗口长度的依赖性较小.

滑动移除窗口尺度的依赖性较小,并能够更为精确的检测突变开始的时间,新方法检测突变的性能明显优于滑动近似熵方法. Mann-Kendall 方法虽然能够对本文中的线性理想时间序列中的动力学结构突变进行检测,但其本质上是一种针对时间序列趋势突变的检测算法,如果时间序列中不存在明显的趋势变化,则不可能有效的检测序列中的动力学结构突变. 而且 Mann-Kendall 方法的检测结果依赖于待分析时间序列的长度,因而它不是一种有效的动力学结构突变检测方法. MC-ApEn 在实际降水观测资料中的应用则进一步证实了新方法的有效性,展示了其潜在的应用前景.

- [1] Wang G R, Cheng S G, Hao B L 1984 *Acta Phys. Sin.* **33** 1246 (in Chinese) [王光瑞、陈式刚、郝柏林 1984 物理学报 **33** 1246]
- [2] Cheng S G 1991 *Acta Phys. Sin.* **40** 584 (in Chinese) [陈式刚 1991 物理学报 **40** 584]
- [3] Wang A L, Yang C X 2002 *Acta Phys. Sin.* **51** 2719 (in Chinese) [王安良、杨春信 2002 物理学报 **51** 2719]
- [4] Wan 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]
- [5] 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]
- [6] Song A J, Han L 2007 *Acta Phys. Sin.* **56** 3820 (in Chinese) [宋爱军、韩雷 2007 物理学报 **56** 3820]
- [7] Yu S M 2008 *Acta Phys. Sin.* **57** 6859 (in Chinese) [禹思敏 2008 物理学报 **57** 6859]
- [8] Shi N 2009 *Meteorology statistical forecasting* (Beijing: Metrological Press) p116 (in Chinese) [施能 2009 气象统计预报(北京:气象出版社)第116页]
- [9] Lin Z S 2003 *Nonlinear science and its application in Geosciences* (Beijing: Metrological Press) p113 (in Chinese) [林振山 2003 非线性科学及其在地球学中的应用(北京:气象出版社)第113页]
- [10] Pincus S M 1991 *Proc. Natl. Acad. Sci. USA* **88** 2297
- [11] Pincus S M, Goldberger A L 1994 *Am. J. Physiol.* **266** H1643
- [12] Pincus S M 1995 *Chaos* **5** 110
- [13] Lei G X, Zeng Y C, Li L 2007 *Technical Acoustics* **26** 121 (in Chinese) [雷国雄、曾以成、李凌 2007 声学技术 **26** 121]
- [14] Li Q, Wang T Y, Leng Y G, He G Y, He H L 2007 *Acta Phys. Sin.* **56** 6803 (in Chinese) [李强、王太勇、冷永刚、何改云、何慧龙 2007 物理学报 **56** 6803]
- [15] Li X F, Yu W L 2008 *Acta Phys. Sin.* **57** 2587 (in Chinese) [刘小峰、俞文莉 2007 物理学报 **57** 2587]
- [16] Xie Y, Xu J X, Yang H J, Hu S J 2002 *Acta Phys. Sin.* **51** 205 (in Chinese) [谢勇、徐健学、杨红军、胡三觉 2002 物理学报 **51** 205]
- [17] Cao B, Lv X Q, Zeng M, Wang Z M, Huang S S 2006 *Acta Phys. Sin.* **55** 1696 (in Chinese) [曹彪、吕小青、曾敏、王振民、黄石生 2006 物理学报 **55** 1696]
- [18] Feng G L, Dong W J, Li J P 2004 *Chin. Phys.* **13** 1582
- [19] Zeng H L, Gao X Q, Zhang W 2005 *Chin. Phys.* **14** 1265
- [20] Dai X G, Fu C B, Wang P 2005 *Chin. Phys.* **14** 850
- [21] Gao X Q, Zhang W 2005 *Chin. Phys.* **14** 2370
- [22] Shi N, Gu J Q, Yi Y M, Lin Z N 2005 *Chin. Phys.* **14** 844
- [23] Shi N, Yi Y M, Gu J Q, Xia D D 2006 *Chin. Phys.* **15** 2180
- [24] Feng G L, He W P 2007 *Chin. Phys.* **16** 2825
- [25] Zhang W, Wan S Q 2008 *Chin. Phys. B* **17** 2311
- [26] Feng G L, Gong Z Q, Zhi R, Zhang D Q 2008 *Chin. Phys. B* **17** 2745
- [27] Feng G L, Dai X G, Wang A H 2001 *Acta Phys. Sin.* **50** 606 (in Chinese) [封国林、戴新刚、王爱慧 2001 物理学报 **50** 606]
- [28] Feng G L, Dong W J, Jia X J 2002 *Acta Phys. Sin.* **51** 1181 (in Chinese) [封国林、董文杰、贾晓静 2002 物理学报 **51** 1181]
- [29] Feng G L, Dong W J, Gong Z Q, Hou W, Wan S Q, Zhi R 2006 *Nonlinear theories and methods on spatial-temporal distribution of the obserbational data* (Beijing: Metrological Press) p27—56 (in Chinese) [封国林、董文杰、龚志强、侯威、万仕全、支蓉 2006 观测数据非线性时空分布理论和方法(北京:气象出版社)第27—56页]
- [30] Zhi R, Lian Y, Feng G L 2007 *Acta Phys. Sin.* **56** 1837 (in Chinese) [支蓉、廉毅、封国林 2007 物理学报 **56** 1837]
- [31] Feng G L, Wang Q G, Hou W, Gong Z Q, Zhi R 2009 *Acta Phys. Sin.* **58** 2853 (in Chinese) [封国林、王启光、侯威、龚志强、支蓉 2009 物理学报 **58** 2853]
- [32] He W P 2008 *Ph. D. Dissertation* (Lanzhou: Lanzhou University) p23 (in Chinese) [何文平 2008 博士学位论文(兰州:兰州大学)第23页]
- [33] He W P, Feng G L, Wu Q, Wan S Q, Chou J F 2008 *Non. Proc. Geophys.* **15** 601
- [34] He W P, Wu Q, Zheng W, Wang Q G, Zhang Y 2009 *Acta Phys. Sin.* **58** 2862 (in Chinese) [何文平、吴琼、张文、王启光、张勇 2009 物理学报 **58** 2862]
- [35] He W P, Deng B S, Wu Q, Zheng W, Cheng H Y 2010 *Acta Phys. Sin.* **59** 2864 (in Chinese) [何文平、邓北胜、吴琼、张文、成海英 2010 物理学报 **59** 2864]
- [36] 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]
- [37] May R 1976 *Nature* **261** 459
- [38] Zhi R, Gong Z Q 2008 *Acta Phys. Sin.* **57** 4629 (in Chinese) [支蓉、龚志强 2008 物理学报 **57** 4629]

## A new method to detect abrupt change based on approximate entropy \*

He Wen-Ping<sup>1)†</sup> He Tao<sup>2)</sup> Cheng Hai-Ying<sup>3)</sup> Zhang Wen<sup>4)</sup> Wu Qiong<sup>5)</sup>

1) (*National Climate Center, China Meteorological Administration, Beijing 100081, China*)

2) (*JiNan Environment Protection Monitoring Station, Jinan 250014, China*)

3) (*Department of Basic Course, Yancheng Institute of Technology, Yancheng 224002, China*)

4) (*TongDa College, Nanjing University of post & telecommunications, Nanjing 210003, China*)

5) (*National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081, China*)

(Received 6 June 2010; revised manuscript received 12 July 2010)

### Abstract

Approximate entropy (ApEn) is valid index which can be used to quantitatively reflect dynamic characteristics and complexity of a time series. The ApEn has been developed to detect an abrupt change in one-dimension time series by sliding a fixed widow, which can be identified with an abrupt dynamic change to some extent, but the sliding ApEn results depend on the window scale, and cannot accurately position the time-instant of an abrupt change. Based on this, a new method is proposed in the present paper, i. e. , moving cut data-approximate entropy (MC-ApEn), which can be used to detect an abrupt dynamic change in time series. Tests on model time series indicate that the detection results from the present method show relatively good stability and high accuracy, obviously better than those from the sliding ApEn method and the Mann-Kendall method. The applications in daily precipitation records further verify the validity of the present method.

**Keywords:** approximate entropy, moving cut data-approximate entropy, abrupt change detection

**PACS:** 92.60.Wc

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 40905034, 40875040, 40930952), the Special Scientific Research Fund of Meteorological Public Welfare Profession of China (Grant Nos. GYHY200806005, GYHY200906019).

† E-mail: wenping\_he@163.com.