

近似熵检测气候突变的研究^{*}

王启光^{1)†} 张增平^{1)‡}

1) 扬州大学物理科学与技术学院, 扬州 225009)

2) 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)

3) 国家气候中心, 中国气象局气候研究开放实验室, 北京 100081)

(2007 年 7 月 3 日收到, 2007 年 7 月 22 日收到修改稿)

引入近似熵方法应用于理想时间序列, 检验了该方法具有区别不同动力学结构和检测其突变的特性. 然后应用于中国气象局国家气候中心 740 站点 1960—2000 年的逐日气温和降水观测资料. 分析结果表明, 温度和降水分布存在明显的区域分布特征, 反映出气候系统动力学结构区域间的差异. 在 20 世纪 70 年代末到 80 年代初各站点具有不同程度的突变现象, 说明了突变时空分布的不均一性以及气候系统演化的复杂性.

关键词: 气候突变, 温度, 降水, 近似熵

PACC: 9260X

1. 引 言

气候系统是一个高度复杂的、耗散的、非绝热的非线性动力系统, 对于这样的系统采用非线性科学最新研究成果有助于更好地揭示其本质的特性和规律. 这正是近些年来非线性动力学分析方法在大气科学领域得到迅速推广应用和发展的重要原因之一. 事实上, 目前应用广泛的 Lorenz 混沌动力学理论模型最初就是由气象学者提出来的. 1963 年美国著名气象学家 Lorenz 在研究天气预报问题时, 发现确定方程中出现混沌现象, 得到天气具有 1—2 周的可预报性, 推动了 20 世纪 70 年代非线性科学尤其是混沌动力学的迅速发展.

中国幅员辽阔, 地形复杂, 气温和降水的空间分布具有较明显的区域差异, 因此温度和降水是十分重要的两大气象要素. 以往常用某些气象要素(比如温度和降水)的均值大小与分布来分析气候特征, 但由于气候是不断变化的, 因此有必要采用新的概念和方法来揭示这一特性, 采用能够反映气候变化不确定性特点的统计量来揭示区域气候特征. 同时, 自从 Lorenz 和 Charney 从理论上论证了气候突变存在的可能性^[1-3], 有关气候突变的研究广泛展开, 气候

突变的研究对认识气候变化以及气候预测都有着极其重要的意义. 在全球增暖的背景下, 许多学者对大气环流、气温和降水等不同地区的观测资料进行了研究^[4-6]. 中国在 20 世纪 70 年代末气象要素发生了不同程度的转折与突变, 其中表现最明显的是降水和温度的变化. 但两者变化的时空范围与尺度有待广泛而深入地研究.

20 世纪 90 年代, Pincus 从衡量时间序列复杂性的角度提出并发展了近似熵(approximate entropy)的概念^[7-9](以下简称为 ApEn). 因其在信号分析中所需数据量较短、有较好的抗噪抗干扰能力且不论是随机还是确定性的信号都可以使用等优点, 因此在医学、机械设备故障的诊断等方面得到广泛应用^[10-15]. 然而在气象学上的应用目前还没有涉及. 本文在介绍 ApEn 及其物理意义的基础上, 将其运用于理想时间序列, 研究该方法在区分不同的动力学结构, 检测其结构突变的有效性, 并且分别加入高斯白噪声和尖峰噪声之后, 检验该方法抗噪抗干扰的能力. 在此基础上应用于中国气象局国家气候中心 740 站 1960—2000 年逐日温度和降水资料.

2. ApEn 方法及其意义

ApEn 方法是一种基于边缘概率分布统计的量

^{*} 国家重点发展基础研究项目(批准号: 2006CB400503)和国家自然科学基金(批准号: 40675044)资助的课题.

[†] E-mail: qiguangw@sina.com

化时间序列复杂程度的方法,其具体算法扼要介绍如下^[7,12]:

假设长度为 N 的时间序列 $u(1), u(2), \dots, u(N)$ 对这个时间序列构造一组维数为 m 的向量 $X(1), X(2), \dots, X(N-m+1)$, 其中向量

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

定义向量 $X(i)$ 和 $X(j)$ 之间的距离 $d[X(i), X(j)]$ 为两者对应元素中差值最大的一个, 即

$$d[X(i), X(j)] = \max_k |u(i+k) - u(j+k)|, \quad k = 0, 1, \dots, m-1. \quad (2)$$

对于每一个 $i (1 \leq i \leq N-m+1)$ 定义

$$C_i^m(r) = \{d[X(i), X(j)] \leq r \text{ 的个数} \} / (N-m+1), \quad (3)$$

$C_i^m(r)$ 表示以 $X(i)$ 为中心, 在维数为 m , 容许偏差为 r 的情形下向量 $X(j)$ 与 $X(i)$ 的距离 $d[X(i), X(j)]$ 小于 r 的概率, 从而表示所有 $X(j)$ 与 $X(i)$ 之间相互近似的程度即关联程度.

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

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r). \quad (4)$$

将维数 m 加 1, 重复上述步骤, 得 $C_i^{m+1}(r)$ 和

$\phi^{m+1}(r)$. 理论上此序列的近似熵为

$$\text{ApEn}(m, r) = \lim_{N \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]. \quad (5)$$

一般而言, 此极限以概率 1 存在. 实际工作中 N 不可能为 ∞ . 当 N 为有限值时, 得到 ApEn 的估计值, 记为

$$\text{ApEn}(m, r) = \phi^m(r) - \phi^{m+1}(r). \quad (6)$$

显然 ApEn 的值与 m, r 的取值有关. Pincus 根据实践, 建议取 $m=2, r=0.1-0.2\sigma$ (σ 是原始数据的标准偏差)^[7-15]. 本文均取 $m=2, r=0.15\sigma$.

由以上推导过程易知, 近似熵方法反映序列在 m 维情况下两点组成的模式间的近似程度, 以及当维数变化时产生新模式的可能性大小. ApEn 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. ApEn 给出新模式发生率随维数而增减的情况, 因而反映序列在结构上的复杂性^[11]; 同时它并不是企图完全重构吸引子, 因此对各种非线性时间序列有很好的适用性.

3. ApEn 优劣性的检验

运用 ApEn 方法在区别动力学系统的异同性和反映结构突变方面做初步探讨, 验证其有效性. 首先构造如下—理想时间序列 $y(t)$:

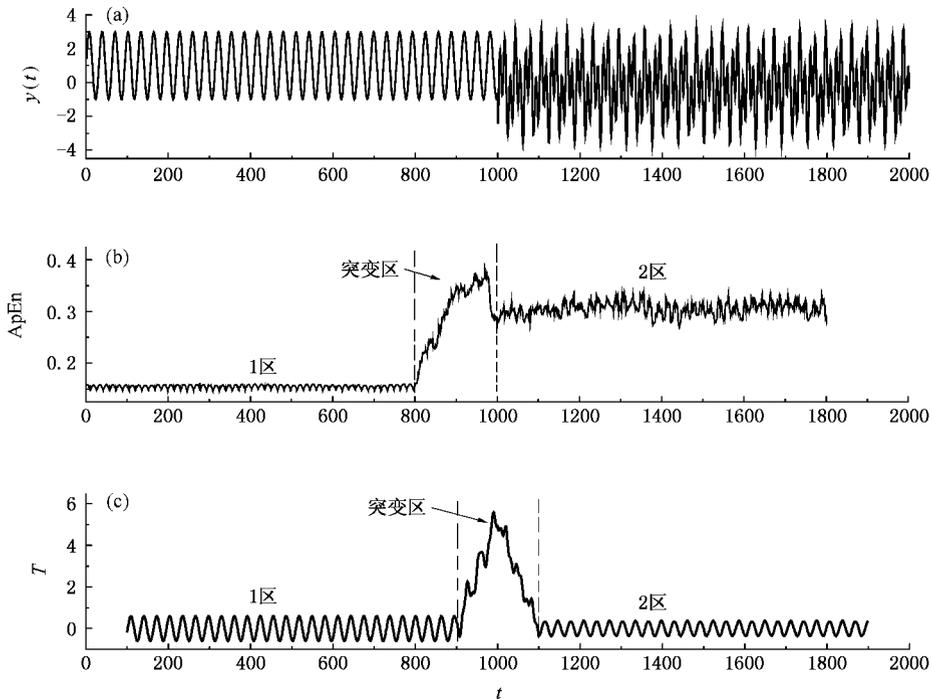


图 1 理想时间序列及其 ApEn 检测 (a)理想时间序列 (b)滑动窗口长度为 200 步长为 1 的 ApEn (c)滑动 t 检测

$$y(t) = \begin{cases} 2\sin(0.2t) + 1, & 1 \leq t \leq 1000, \\ 1.5\sin(0.2t) + 2\cos(0.5t) - 0.2, & 1000 < t \leq 2000. \end{cases} \quad (7)$$

由(7)式知,所描述的时间序列代表两个不同的动力学结构:

- 1)当 $1 \leq t \leq 1000$ 时,子序列由振幅为 2,周期为 10π 的正弦函数组成;
- 2)当 $1000 < t \leq 2000$ 时,子序列由振幅为 1.5,周期为 10π 的正弦和振幅为 2,周期为 4π 的余弦函数叠加而成.

显然, $t = 1000$ 时为两动力学结构突变点,如图 1(a)所示.对该理想时间序列进行近似熵的计算,采用滑动窗口的方法,取窗口宽度 $l = 200$,滑动步长为 1.图 1(b)表明,不同的动力学结构的系统 A_{pEn} 值明显不同,且在系统动力学结构突变区 A_{pEn} 值会发生显著跃变,与理想时间序列的突变点相符合.采用不同窗口宽度和滑动步长进行比较,仍可得到类似结果(图略).分别比较图 1(a)和(b)中的 1 区

和 2 区可以发现,虽然滑动 t 检测方法也可以检测出相应的结构突变,但是不能反映出突变前后的动力学结构差异.而 A_{pEn} 方法中突变前后熵值明显不同,且与时间序列前后两段的动力学结构复杂程度很好地对应.

进一步,在(7)式中加入信噪比为 25 的高斯白噪声,计算结果如图 2(b)所示,前后两不同的动力学结构因噪声的加入,其复杂度均有所增加,两者 A_{pEn} 都有所增大.但不同的动力学结构所对应的 A_{pEn} 仍有明显差别,同时动力学结构突变点的 A_{pEn} 跃变仍然明显.

在(7)式构造的时间序列随机加入 6 个尖峰噪声,噪声大小在 2—7 之间变化(图 2(c)所示). A_{pEn} 计算结果如图 2(d),从图中可以看出, A_{pEn} 方法对系统中偶尔产生的瞬间强干扰也有较好承受能力.

由以上分析可知,当系统动力学结构发生改变时,对应的熵值会发生明显跃变,且动力学结构越复杂,熵值就越大.因此 A_{pEn} 方法可以有效地反映系统动力学结构变化和辨别动力学结构的差异以及动

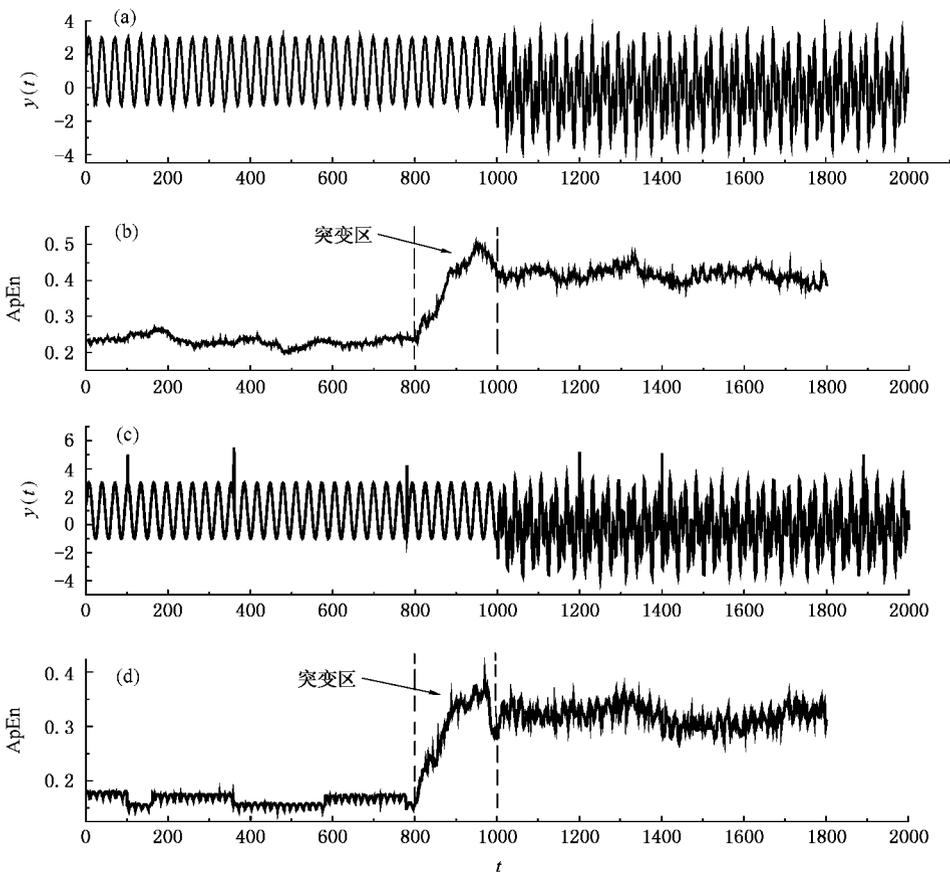


图 2 加噪后的理想时间序列及其 A_{pEn} (a)(b)分别为加入信噪比为 25 的高斯白噪声后的信号及其 A_{pEn} (c)(d)加入尖峰噪声后的理想时间序列及其 A_{pEn}

力学系统演化的复杂程度。

4. ApEn 方法在实际气象观测资料中的应用

本文所用资料为中国气象局国家气候中心 740 站点 1960—2000 年(共 41 年)的逐日降水和温度观测资料。为了真实地反映降水的时空分布规律,对

740 站点资料进行了检测,缺测较多或不满 41 年连续观测的站点,本文没有考虑。经检测有 435 个站点满足上述要求,对个别缺测通过插值方法补足。

4.1. 温度和降水 ApEn 的分布

基于 ApEn 方法分别对各站逐日观测资料进行分析,得到各站点 41 年序列的 ApEn 值。其分布情况如图 3、图 4 所示。

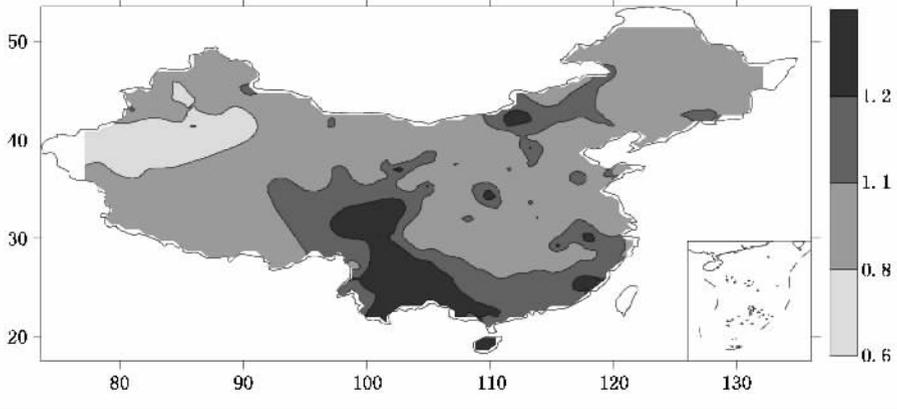


图 3 中国 435 站点 1960—2000 年逐日温度资料的 ApEn 分布情况

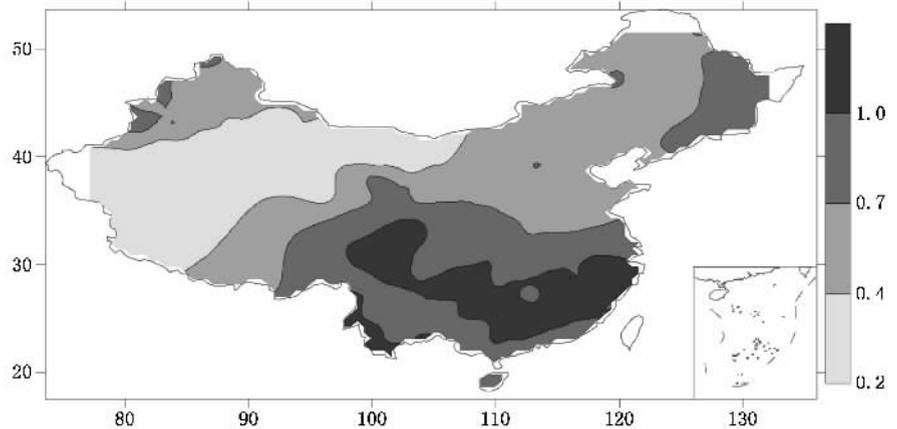


图 4 中国 435 站点 1960—2000 年逐日降水资料的 ApEn 分布情况

图 3 所示为各站点温度序列的 ApEn 值分布,总体呈现南大北小的趋势。其中海南、云南、贵州以及四川等西南地区各站点 ApEn 值普遍偏大,处于 1.2—1.5 之间;华南与华东大部分地区 ApEn 值处于 1.1—1.2 之间;长江以北的附近区域及华北和东北大部分地区 ApEn 值有所减小,基本处于 0.8—1.1 之间;新疆天山以南地区 ApEn 值偏小且内部差异不大,其中新疆地区 ApEn 值为中国最小值集聚区,在 0.6—0.8 之间。由此可见以上区域的气候系统的复杂程度是有明显区别的,反映了这些区域影

响气温的气候系统动力学结构很可能不同,可以应用于气候特征区的划分。

图 4 所示为中国降水序列的 ApEn 值分布,与中国的年降水量分布比较接近。ApEn 值具有很明显的区域分布特征,呈现东南大西北小趋势。其中华南大部分地区为 ApEn 值最大区,在 1.0—1.2 之间;长江以北区域明显减小,在 0.7—1.0 之间,并且在秦岭淮河一线(约 30°N)以北又明显减小,大部分在 0.4—0.7 之间;东北三江平原及东北南部 ApEn 值分布集中且较大,处于 0.7—1.0 之间;西北甘肃北

部和新疆南部 $ApEn$ 值为中国最小, 仅在 0.2—0.4 之间. 由 $ApEn$ 意义可知, 熵值越大, 反映系统的复杂程度越高, 可预测性也就越差. 从图 4 可以看出中国降水分布存在明显的层次差异, 长江以南地区熵值相对较高, 说明这一区域影响降水的因子较多, 相关的气候系统动力学结构复杂, 这与该地区易受台风影响、洪涝灾害频发且预报难度较大很好地对应. 比较图 3、图 4 可以发现中国南方地区温度和降水的 $ApEn$ 值都相对较大, 这很好地反映了南方地区气候系统的复杂性与较强的不确定性; 北方特别是西北地区温度和降水 $ApEn$ 值较小, 反映了影响该区域气候系统的因子可能相对较少.

中国地域广阔, 纬度差异及海拔落差均较大, 日照时间和强度分布不均, 不仅受东亚季风的显著影响, 而且还受印度季风的作用, 夏季东南沿海台风活动频繁. 同时温度和降水是各影响因子综合作用的结果. 基于以上地理、气候特点, 结合图 3、图 4 $ApEn$ 分布特征可以发现: 在中国西南地区多丘陵山脉, 并

且受印度季风的影响较大, 其温度变化规律性较差, 对应 $ApEn$ 值偏大; 东南地区受东亚季风和夏季台风的影响, 气候系统也较为复杂, 温度、降水 $ApEn$ 值较大. 中国西北部地区气候系统相对单一, 因而 $ApEn$ 值较小; 但新疆地区北部经常受冷锋影响, 而南部有天山阻挡, 所以温度及降水的 $ApEn$ 分布在新疆南北有所差异. 从中国温度和降水的 $ApEn$ 的分布可以了解中国气温与降水系统的不确定程度, 为区域气候研究、气候预测与评估提供参考.

4.2. 降水、温度的 $ApEn$ 逐年变化与气候突变的关系

气候状态的转折和突变源于气候系统的非线性、复杂性以及气候系统内部各种时空尺度的相互作用^[16—22], 而 $ApEn$ 可以反映出系统的动力学结构差异. 作为示例, 本文给出呼玛站(站点编号 50353; 东经 126.39 度, 北纬 51.43 度) 降水序列逐年 $ApEn$ 的变化情况, 如图 5 (a) 所示.

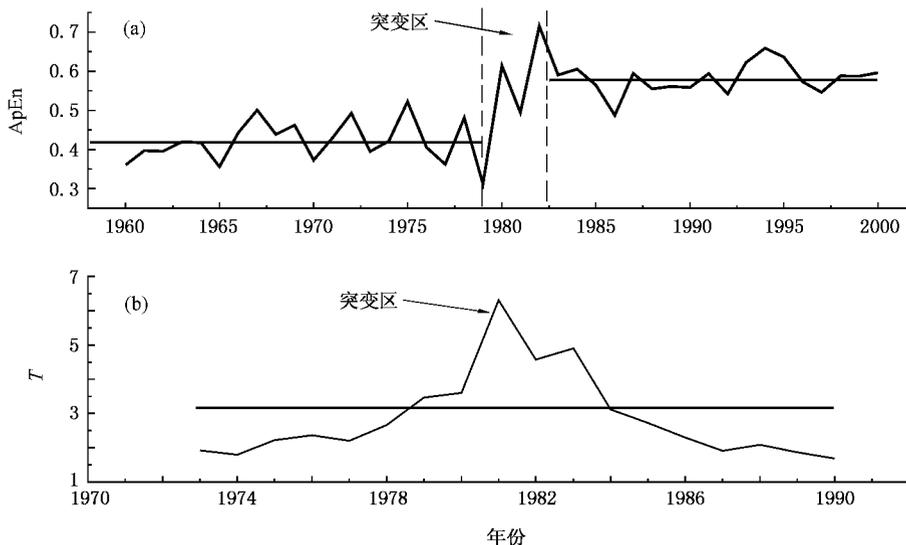


图 5 呼玛站逐年降水序列 $ApEn$ 变化情况 (a) 各年降水序列 $ApEn$ (b) 滑动 t 检验对 $ApEn$ 序列检测

从图中可以看出, 在 1981 年前后, 该站点的年降水序列的 $ApEn$ 发生明显的变化, 且与图 5 (b) 应用滑动 t 检验得到的检测结果一致. 这与该站点所在地区历史上的突变年份有很好地对应关系. 我们采用滑动 t 检验^[23] 对其他各站 $ApEn$ 逐年变化情况进行检验 (显著性水平为 0.05), 得到中国降水序列 $ApEn$ 突变年份分布情况, 如图 6 所示. 中国各个区域特别是北方在 1979—1985 年存在不同程度突变情况, 相对集中在 1979 年前后. 沿海地区各站点突变发生时间比较靠前, 基本集中在 1976—1979 年,

这可能与海陆相互作用有关. 而西北大部分地区相对中国而言突变较晚. 这与已有相关研究结果较好符合^[24].

中国 435 站点 1960—2000 年逐年温度资料 $ApEn$ 突变分布情况如图 7 所示. 温度资料 $ApEn$ 突变分布也有较明显的区域特征. 其中中部地区突变年份相对较早, 集中在 1976—1979 年, 其他地区各站点突变年份与降水资料类似, 也集中在 1979 年以后. 整体而言, 从发生突变的站点的年份来看, 降水和温度突变时间有相当好的对应关系.

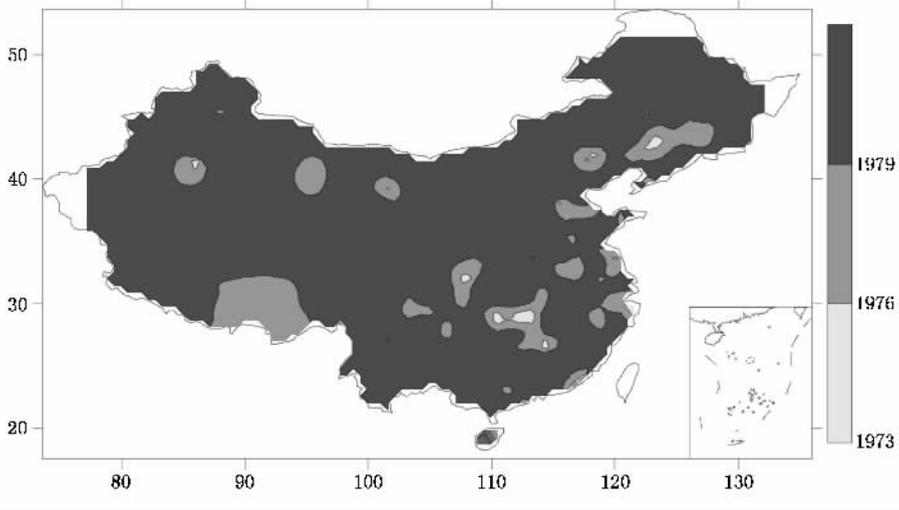


图6 中国435站点1960—2000年逐年降水资料ApEn突变分布情况

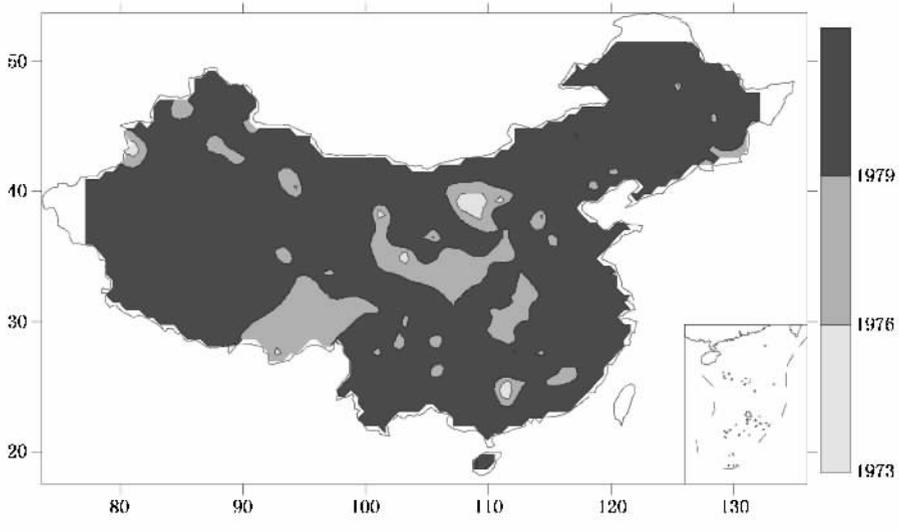


图7 中国435站点1960—2000年逐年温度资料ApEn突变分布情况

5. 结论和讨论

本文引入了一种用于分析气象资料的ApEn方法. 首先将该法用于分析构造的理想时间序列, 展示了很好的区分不同动力学结构和显示突变的能力, 并且具有良好的抗噪性能. 然后应用于中国435站点1960—2000年逐日气温和降水观测资料, 发现其ApEn值存在明显的区域差异, 这些差异与季风、地形地貌等因素有较好的对应关系, 为区域气候研究提供了一定的参考. 同时利用该方法结合滑动 t 检

验, 对各站逐年降水和温度序列进行处理, 得到近50年中国降水和温度的ApEn突变年份分布情况. 表明中国各站点降水和温度序列突变年份相对集中在1979年前后, 尤其是从温度和降水两个不同的气象要素得到较为一致的结果, 这是有别于传统分析方法的, 充分说明了ApEn能揭示一个气候系统的演化本质规律, 尤其降水和温度是复杂气候系统的子系统, 通过ApEn能从不同的气象要素反映系统的本质属性. 有关于ApEn在其他气象资料中的应用有待进一步探讨和研究.

- [1] Lorenz E N 1963 *J. Atmos.* **20** 130
- [2] Lorenz E N 1976 *Quat. Res.* **6** 495
- [3] Charney J G ,Devore J G 1979 *J. Atmos. Sci.* **36** 1205
- [4] 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]
- [5] Shi N 2005 *Chin. Phys.* **14** 844
- [6] Dai X G ,Fu C B ,Wang P 2005 *Chin. Phys.* **14** 850
- [7] Pincus S M 1991 *Proc. Natl. Acad. Sci. USA* **88** 2297
- [8] Pincus S M 1995 *Chaos* **5** 110
- [9] Pincus S M ,Goldberger A L 1994 *Am. J. Physiol.* **266** 1643
- [10] Dong G Y ,Wu Q Y 1999 *Chinese Journal of Medical Instrumentation* **23** 311 (in Chinese)[董国亚、吴祈耀 1999 中国医疗器械杂志 **23** 311]
- [11] Yang F S ,Liao W C 1997 *Chinese Journal of Medical Instrumentation* **21** 283 (in Chinese)[杨福生、廖旺才 1997 中国医疗器械杂志 **21** 283]
- [12] Jang C H ,Feng H Q ,Liu D L *et al* 2005 *J. Biomed. Eng.* **22** 649 (in Chinese)[江朝晖、冯焕清、刘大路等 2005 生物医学工程学杂志 **22** 649]
- [13] Zhang C ,Zheng C X ,Yu X L *et al* 2006 *Space Medicine & Medical Engineering* **19** 459 (in Chinese)[张 崇、郑崇勋、于晓琳等 2006 航天医学与医学工程 **19** 459]
- [14] You R Y ,Chen Z 2004 *Chinese Journal of Computational Physics* **21** 34 (in Chinese)[游荣义、陈 忠 2004 计算物理 **21** 34]
- [15] Jang C H ,Feng H Q ,Xie D 2004 *Chinese Journal of Medical Instrumentation* **28** 82 (in Chinese)[江朝晖、冯焕清、谢 丹 2004 中国医疗器械杂志 **28** 82]
- [16] Yang P C ,Bian J C ,Wang G L *et al* 2003 *Chinese Science Bulletin* **48** 1470 (in Chinese)[杨培才、卞建春、王革丽等 2003 科学通报 **48** 1470]
- [17] Yang M X ,Yao T D 1999 *Exploration of Nature* **18** 29 (in Chinese) [杨梅学、姚檀栋 1999 大自然探索 **18** 29]
- [18] Feng G L ,Dong W J 2003 *Chin. Phys.* **12** 1076
- [19] Hou W ,Feng G L ,Dong W J ,Li J P 2006 *Acta. Phys. Sin.* **55** 2663 (in Chinese)[侯 威、封国林、董文杰、李建平 2006 物理学报 **55** 2663]
- [20] Zhi R ,Gong Z Q ,Wang D Y ,Feng G L 2006 *Acta. Phys. Sin.* **55** 6185 (in Chinese)[支 蓉、龚志强、王德英、封国林 2006 物理学报 **55** 6185]
- [21] Feng G L ,Dong W J ,Li J P 2004 *Chin. Phys.* **13** 1582
- [22] Feng G L ,Dong W J ,Jia X J 2004 *Chin Phys.* **13** 413
- [23] Wei F Y 1999 *The technologies of statistics diagnosis and forecast in modern climate* (Beijing :the weather publishing company) p63 (in Chinese)[魏凤英 1999 现代气候统计诊断与预测技术 (北京 :气象出版社) 第 63 页]
- [24] Yu S Q ,Lin X C ,X X D 2003 *Climatic and Environmental Research* **8** 9 (in Chinese)[于淑秋、林学椿、徐祥德 2003 气候与环境研究 **8** 9]

The research of detecting abrupt climate change with approximate entropy^{*}

Wang Qi-Guang^{1,2)†} Zhang Zeng-Ping^{1,3)}

¹ *College of Physics Science and Technology, Yangzhou University, Yangzhou 225009, China*

² *Key Laboratory of Regional Climate-Environment Research for Temperature East Asia, The Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

³ *Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing 100081, China*

(Received 3 July 2007 ; revised manuscript received 22 July 2007)

Abstract

The approximate entropy method is applied to an ideal time series. It is indicated that the method is capable of distinguishing between different structural dynamics and detecting structural dynamics mutation. And then this method is applied to the observation data of the daily temperature and precipitation of the 740 sites between 1960 and 2000. The results showed that the approximate entropy distribution of the temperature and precipitation has a characteristic of obvious district distribution, and displays the territorial difference of dynamic structures of climate system. Abrupt climate changes have been detected by the approximate entropy of the annual precipitation and temperature between 1970s and 1980s. The characteristics of the inhomogeneity of temporal and spatial distribution of abrupt climate changes and the complexity of climate systems have been demonstrated.

Keywords : abrupt climate change, temperature, precipitation, approximate entropy

PACC : 9260X

^{*} Project supported by the National Key Program Development for Basic Research (Grant No. 2006CB400503) and the National Natural Science Foundation of China (Grant No. 40675044).

[†] Corresponding author. E-mail : qiguangw@sina.com