

# 利用排列熵检测近 40 年华北地区气温突变的研究<sup>\*</sup>

侯 威<sup>1)2)†</sup> 封国林<sup>2)3)</sup> 董文杰<sup>2)3)</sup> 李建平<sup>4)</sup>

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中心, 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)

4)(中国科学院大气物理研究所 LASC, 北京 100029)

(2005 年 9 月 26 日收到; 2005 年 11 月 15 日收到修改稿)

运用一种新的动力学突变检测方法——排列熵(permutation entropy, PE)算法, 计算并分析了中国华北地区 52 个站点 1960 年—2000 年逐日平均气温资料的排列熵演化情况, 发现中国华北地区气温在 20 世纪 70 年代中期、80 年代初均发生了较大突变; 进一步用经验模态分解(empirical mode decomposition, EMD)方法对排列熵序列进行逐级平稳化处理, 结果发现这一地区的气温突变与准 10 年这一年代际时间尺度的周期变率密切相关, 其原因与太阳黑子活动有着密切联系.

关键词: 华北, 突变, 排列熵算法, 经验模态分解

PACC: 9260X

## 1. 引 言

自从 Lorenz<sup>[1,2]</sup> 和 Charney<sup>[3]</sup> 从理论上揭示了气候突变的可能性, 有关气候突变的研究得到了广泛的开展<sup>[4-8]</sup>. 在气象领域中, 精确描述气候突变发生的时间和地点对于提高气候诊断与预测的准确度有着极其重要的作用. 在过去 20 年中, 研究者们提出了一系列检测动力学突变的方法, 包括复原图<sup>[9]</sup> 及复原量化分析方法<sup>[10,11]</sup>、时空分离图以及与之相关的概率分布方法<sup>[12,13]</sup>、非线性相干预测分析法<sup>[14]</sup> 等, 这些方法都是基于将相空间中最近邻点的性质进行量化, 因而其计算过程非常复杂并且计算量十分巨大.

本文介绍一种新的动力学突变检测方法——排列熵(permutation entropy, PE)方法<sup>[15]</sup>, 这一方法具有概念简单、计算简捷的优点. 同以前的方法, 例如 Kolmogorov 熵和 Lyapunov 指数方法相比, PE 算法可以更好地检测出复杂系统的动力学突变. 本文采用 PE 算法计算并分析了中国华北地区 52 个站点 1960 年—2000 年逐日平均气温资料的排列熵, 得到了华北地区近 40 年来的温度突变情况; 并进一步用经验

模态分解(empirical mode decomposition, EMD)方法对计算结果进行逐级平稳化处理, 从中分离出不同周期的波动, 分析了不同时空尺度的周期变率在气温突变中的贡献.

## 2. 排列熵算法

设一离散时间序列为 $[x(i), i=1, 2, \dots, n]$ , 对其中任意一个元素 $x(i)$ 进行相空间重构, 得到

$$X(i) = [x(i), x(i+l), \dots, x(i+(m-1)l)], \quad (1)$$

式中 $m$ 和 $l$ 分别为嵌入维数和延迟时间. 将 $X(i)$ 的 $m$ 个重构分量 $[x(i), x(i+l), \dots, x(i+(m-1)l)]$ 按照升序重新进行排列, 即

$$[x(i+(j_1-1)l) \leq x(i+(j_2-1)l) \leq \dots \leq x(i+(j_m-1)l)]. \quad (2)$$

如果存在 $x(i+(j_1-1)l) = x(i+(j_2-1)l)$ , 此时就按 $j$ 值的大小来进行排序, 也就是当 $j_{i1} < j_{i2}$ 时, 有 $x(i+(j_{i1}-1)l) \leq x(i+(j_{i2}-1)l)$ . 所以, 任意一个向量 $X_i$ 都可以得到一组符号序列

$$A(g) = [j_1, j_2, \dots, j_m] \quad (3)$$

<sup>\*</sup> 国家自然科学基金(批准号: 90411008 和 40325015) 和国家重点发展基础研究项目(批准号: 2006CB400503)资助的课题.

<sup>†</sup> E-mail: hou\_w@sohu.com

其中  $g=1, 2, \dots, k$ , 且  $k \leq m$ !  $m$  个不同的符号  $[j_1, j_2, \dots, j_m]$  一共有  $m!$  种不同的排列, 也就是一共有  $m!$  种不同的符号序列, 符号序列  $A(g)$  是其中的一种. 计算每一种符号序列出现的概率  $P_1, P_2, \dots, P_k$ , 此时, 时间序列  $[x(i), i=1, 2, \dots, n]$  的  $k$  种不同符号序列的 PE 就可以按照 Shannon 信息熵的形式定义为

$$H_p(m) = - \sum_{v=1}^k P_v \ln P_v. \quad (4)$$

当  $P_v=1/m!$  时,  $H_p(m)$  就达到了最大值  $\ln(m!)$ , 为了方便起见, 通常将  $H_p(m)$  用  $\ln(m!)$  进行标准化处理, 即

$$0 \leq H_p = H_p(m) / \ln(m!) \leq 1. \quad (5)$$

$H_p$  值的大小表示时间序列  $[x(i), i=1, 2, \dots, n]$  的随机程度:  $H_p$  的值越小, 说明时间序列越规则, 反之则时间序列越接近随机;  $H_p$  的变化反映并放大了时间序列  $[x(i), i=1, 2, \dots, n]$  的微小的细节变化. 很显然, 如果  $m$  太小, 例如  $m=1$  或  $m=2$ , 那么 PE 算法就会失去有效性, 因为此时重构向量中仅仅包含很少几个状态; 一般在时间序列  $[x(i), i=1, 2, \dots, n]$  足够长的条件下可以选取较大的  $m$ . 但当使用 PE 算法来检测信号中的动力学突变,  $m$  的取值太大, 例如  $m=12$  或者  $m=15$  也不合适. 文献[15-17] 讨论了  $m=3, 4, \dots, 7$  时的计算结果, 但  $m=3$  或  $m=4$  仍然较小, 在本文中选取  $m=5$ .

用 PE 算法检测时间序列中的动力学突变可以描述如下: 将一个长时间序列分为若干个长度为  $w$  的子序列, 子序列样本的截止数据点为  $t$ . 这些子序列相互之间可以重叠也可以不重叠, 计算每个子序

列的  $H_p$ , 观察  $H_p$  随  $t$  的变化情况. 在本文中使用的最大重叠情形, 即将每个子序列向后移动一个数据点得到下一个子序列. 对于子序列的长度  $w$ , 如果  $w$  取值太小, 那么计算结果就会失去统计学意义, 但是为了精确检测突变信号,  $w$  取值又不能太大. 在文献[11] 中详细讨论了该算法在  $m, l, w$  取不同值时计算结果的有效性, 本文下面的计算和分析中均取  $m=5, l=1, w=2000$ .

### 3. PE 算法的运用

#### 3.1. PE 算法在 Logistic 模型中的运用

图 1 显示了用 PE 算法分析 Logistic 模型的结果. Logistic 映射随时间的演化是人们最熟知的非线性动力学系统. Logistic 映射可以写为  $x_{n+1} = ux_n(1-x_n)$  ( $n=0, 1, 2, \dots$ ),  $x_n \in [0, 1]$ ,  $u$  为控制参数. 如图 1(a) 所示, 当  $u \in [0, 1)$  时, 存在  $x_n \rightarrow 0$  的稳定不动点; 当  $x_n \in [1, 3)$  时, 存在  $x_n \rightarrow 0$  的不稳定不动点和  $x_n \rightarrow 1-1/u$  的稳定不动点, 周期 1 解; 当  $u=3$  时, 周期 1 解分叉为周期 2 解; 当  $u \in [3, u_\infty)$  时, 此时的  $u_\infty=3.569945672$ , 依次出现轨道周期变化  $1 \rightarrow 2 \rightarrow 4 \rightarrow 8 \rightarrow 16 \rightarrow \dots$ , 即倍周期分叉现象; 当  $u \in (u_\infty, 4)$  时出现混沌现象, 即时间序列依次经历了从稳定不动点  $\rightarrow$  不稳定不动点  $\rightarrow$  周期  $\rightarrow$  混沌四个不同的演化阶段. 在图 1(b) 中可以看到当时间序列从一种演化阶段跳跃到另一种演化阶段时, 排列熵均有明显的变化特征, 混沌区域中的周期窗口在排列熵的演

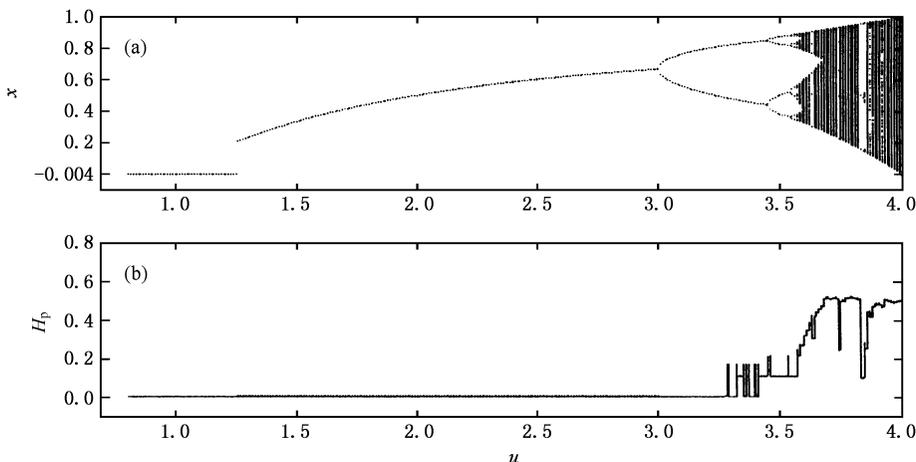


图 1 (a) Logistic 模型及其 (b) 排列熵随控制参数的演化.  $x$  为 Logistic 映射的值,  $H_p$  为其排列熵的值,  $u$  为控制参数

化中也有明显反映, 并且排列熵的变化与时间序列的演化特征步调一致. 由此可见, 排列熵可以提取并放大时间序列中微小的变化, 用来检测动力学突变.

将子序列长度和滑动步长以及  $m$  和  $l$  取不同的值进行组合并在原时间序列中加入噪声, 此时计算得到的排列熵的演变特征均类似于图 1(b).

### 3.2 用 PE 算法分析实际时间序列

本文所分析资料是将中国华北地区 52 个观测站点 1960 年—2000 年(共 40 年 14976 个数据点)的逐日气温观测资料  $x(i, j)$ , ( $i = 1, 2, \dots, 52; j = 1, 2, \dots, 14976$ ), 按(6)式进行平均得到华北地区逐日平均气温资料.

$$y(j) = \frac{\sum_{i=1}^n x(i, j)}{52} \quad (j = 1, 2, \dots, 14976). \quad (6)$$

将每个子序列计算得到的排列熵赋值给子序列的最后一个数据点, 且横坐标  $t$  可以按照(7)式来得到其所对应的年代.

$$\text{Year} = 1960 + [(t/365) - 1]. \quad (7)$$

图 2 中, 在 1974 年和 1980 年时发生了最大的两次突变, 说明此时华北地区气温发生了较大变化. 参照于已有的研究结果<sup>[18-20]</sup>, 发现华北地区在 20 世纪 70 年代中期经历了一次较大的气温突变, 出现了近 50a 来的第一次升温, 气候逐渐由湿润变为干旱; 80 年代初出现了第二次升温. 另外当  $t = 1966$  年,  $t = 1971$  年,  $t = 1972$  年,  $t = 1984$  年,  $t = 1989$  年和  $t = 1994$  年,  $t = 1997$  年,  $t = 1999$  年时也发生了突变. 为了进一步分析华北地区的温度变化, 用 EMD 方法对排列熵序列进行逐级分解, 以确定在何种时间尺度上的气候变率对华北地区的温度变化贡献最大.

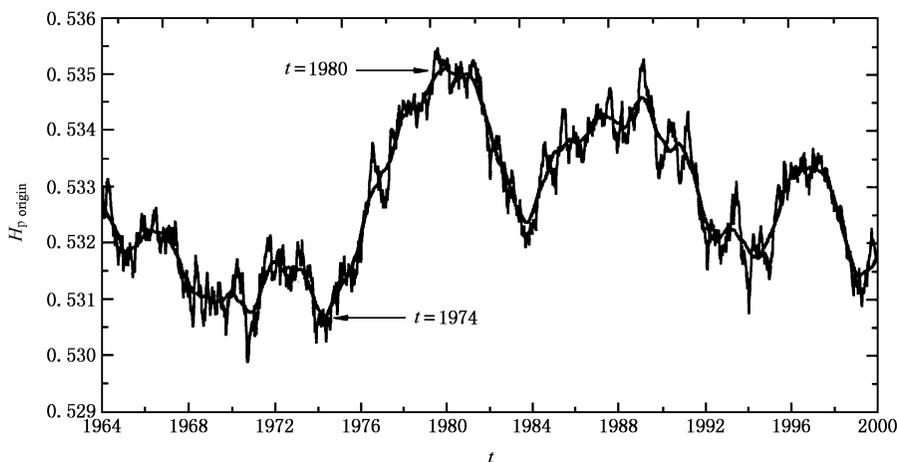


图2 华北地区平均温度的排列熵随时间的演化 ( $H_p$  为排列熵,  $t$  为演化时间)

表 1 各 IMF 分量的频率及其周期

分量	频率	周期
IMF1	0.1011	小于 1 年
IMF2	0.1027	小于 1 年
IMF3	0.0503	小于 1 年
IMF4	0.0236	小于 1 年
IMF5	0.0111	小于 1 年
IMF6	0.0054	小于 1 年
IMF7	0.0020	准 1 年
IMF8	0.0012	准 2 年
IMF9	6.1638e-004	准 4 年
IMF10	3.0821e-004	准 9 年
IMF11	7.7055e-005	准 35 年
IMF12	趋势分量	无

分解, 得到 12 个 IMF 分量(图 3), 每个 IMF 分量表示了在不同时间尺度上气候变率的变化情况. 分析这 12 个 IMF 分量, 各个 IMF 分量的频率和周期如表 1 所示. 这些 IMF 分量不仅包含了气候系统外在强迫的周期变化, 还包含了气候系统中的非线性反馈作用.

气候要素普遍存在的 4—6 年周期在 IMF9 中清晰地表示出来, 这是和 ENSO 相联系的特征周期. 而与太阳黑子活动密切相关的准 10 年周期也在 IMF10 中表现出来.

考察单个 IMF 分量, 即分析各种周期分量在气温变化中的可能作用, 本文重点分析 IMF10 分量. 从图 4 中可以看出 IMF10 分量随时间的演变和华北地

对华北地区平均温度的排列熵序列进行 EMD

区平均温度的排列熵随时间的演化特征几乎完全一致, 可以认为 IMF10 分量显示的是近 40a 华北地区温度变化的平均态. 这说明 20 世纪华北地区的气温

突变和 IMF10 分量这一时间尺度的气候变率有着密切联系, 可以认为 IMF10 分量即太阳黑子活动在图 2 中排列熵的波动变化中起主导作用, 也就是 IMF10

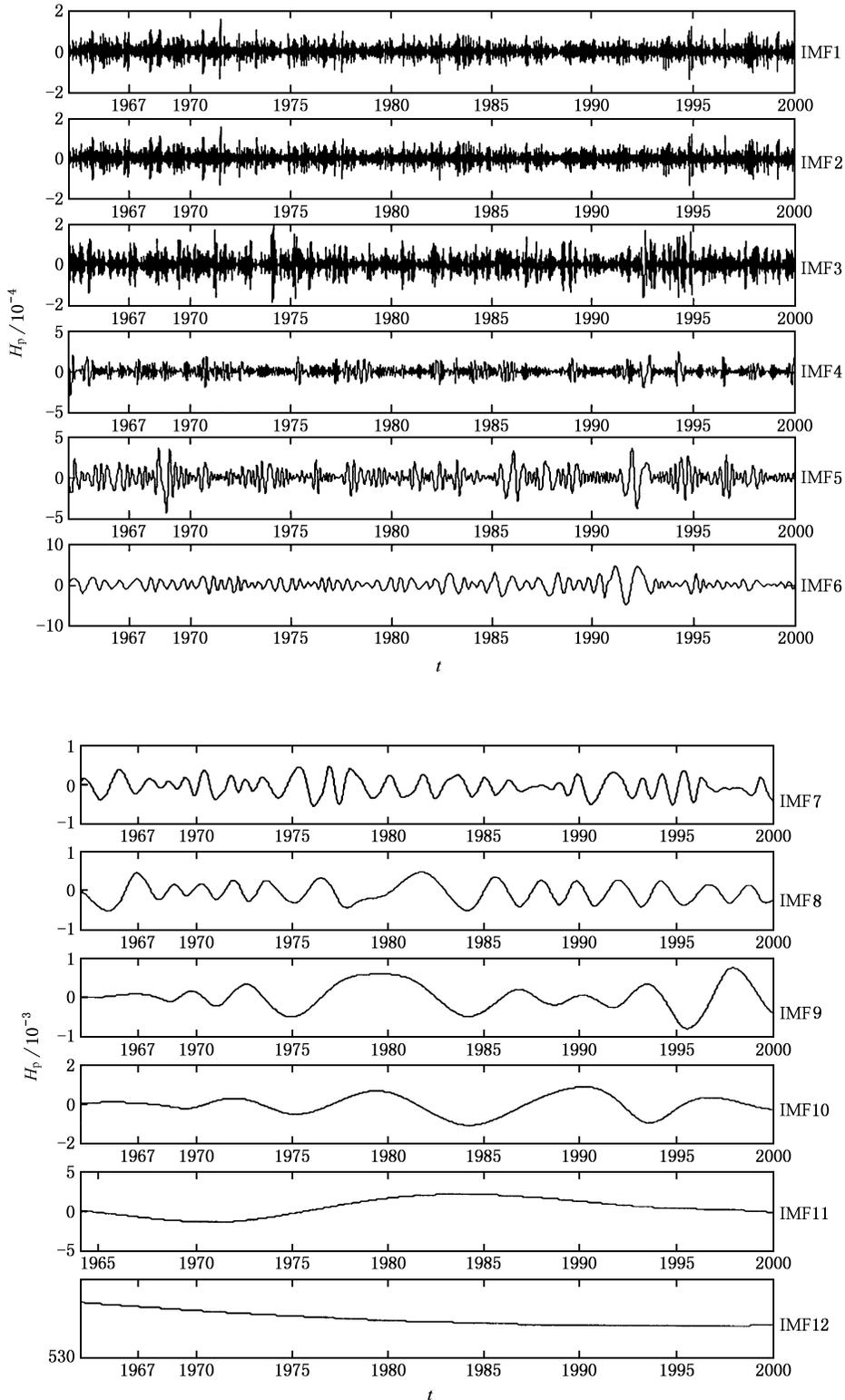


图 3 华北地区平均温度的排列熵经 EMD 分解后得到的 12 个 IMF 分量随时间的演化 ( $H_p$  为排列熵,  $t$  为演化时间)

分量对 20 世纪华北地区的气温变化贡献最大, 即气温变化主要受准 10 年周期气候变率的影响.

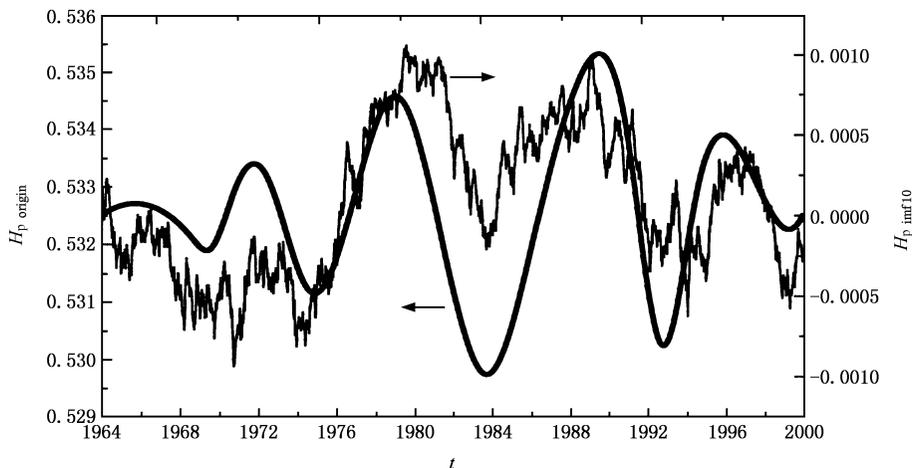


图 4 华北地区平均温度的排列熵与其 IMF10 分量随时间演化的对比 ( $H_{p\_origin}$  为平均温度的排列熵,  $H_{p\_IMF10}$  为其 IMF10 分量,  $t$  为演化时间)

进一步研究发现, 当  $w$  取不同值时仍然可以得出相同的结论. 当  $w = 1000$  时, 如图 5 所示, 华北地区气温发生突变的年份 (按 (7) 式计算得到) 与  $w = 2000$  时极其相似, 并且排列熵随时间的演变特征和

经过 EMD 分解得到的 IMF10 分量随时间的演变特征也几乎完全一致, 此时的 IMF10 分量为准 8—9 年周期, 从而得出同图 4 相同的结论. 这也说明了 PE 算法的有效性.

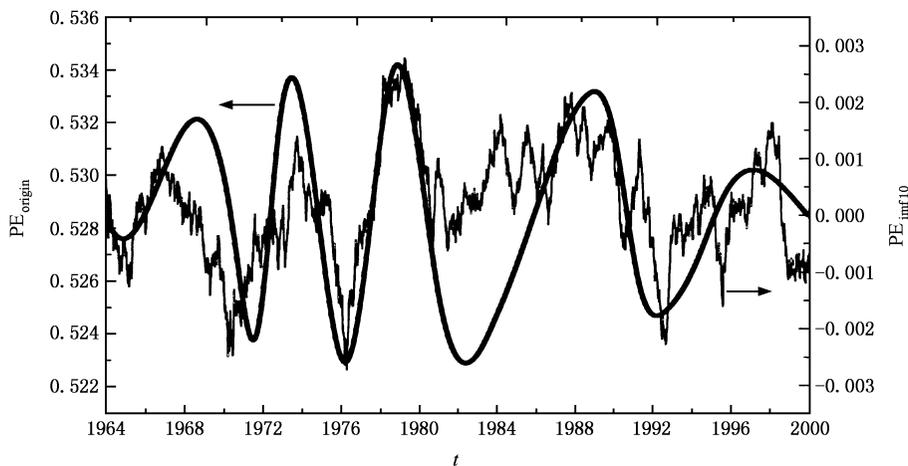


图 5 当  $w$  取 1000 时华北地区平均温度的排列熵与其 IMF10 分量随时间演化的对比 ( $H_{p\_origin}$  为平均温度的排列熵,  $H_{p\_IMF10}$  为其 IMF10 分量,  $t$  为演化时间)

## 4. 结 论

本文介绍了一种新的动力学突变检测方法——PE 算法, 并运用这一算法计算分析了中国华北地区 52 个站点 1960 年—2000 年的逐日平均气温资料, 得到这一地区其近 40 年来的气候突变情况, 发现在 20 世纪 70 年代中期, 80 年代初均发生了较大的气

候突变, 其中在 70 年代初期前后出现了近 50a 来的第一次升温, 气候逐渐由湿润变为干旱; 80 年代中期出现了第二次升温. 进一步用 EMD 方法对计算得到的排列熵序列进行逐级平稳化处理, 从中分离出不同周期的波动成分. 通过对比分析, 发现这一地区的气温突变主要受准 10—12 年周期的气候变率的影响. 因此可以认为在准 10 年这一年代际时间尺度上的周期气候波动直接导致了这一地区的气温突

变, 而准 10 年的周期与太阳黑子活动有着密切联系, 它对全球和我国的气候变化均有较大影响, 特别

对华北地区的干旱化趋势有着很大的影响.

- [1] Lorenz E N 1963 *J. Atmos. Sci.* **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] Shi N, Gu J Q, Yi Y M *et al* 2005 *Chin. Phys.* **14** 844
- [5] Rial J A 2004 *Global and Planetary Change* **41** 95
- [6] Cury R, Dickson B, Yashayaev L 2003 *Nature* **426** 826
- [7] Feng G L, Dong W J 2003 *Chin. Phys.* **12** 1076
- [8] Wang P, Dai X G 2004 *Chin. Phys.* **13** 1770
- [9] Edkmann J P, Kamphorst S O, Ruelle D 1987 *Europhys. Lett.* **4** 973
- [10] Trulla L L, Giulant A, Zbilut J P *et al* 1996 *Phys. Lett. A* **223** 255
- [11] Gao J B, Cai H Q 2000 *Phys. Lett. A* **270** 75
- [12] Provenzale A, Smith L A, Vio R *et al* 1992 *Physica D* **58** 31
- [13] Yu D J, Lu W P, Harrison R G 1999 *Chaos* **9** 865
- [14] Schreiber T 1994 *Phys. Rev. Lett.* **78** 843
- [15] Cao Y H, Tung W W, Gao J B *et al* 2004 *Phys. Rev. E* **70** 217
- [16] Bandt C, Pompe B 2002 *Phys. Rev. Lett.* **88** 102
- [17] Bandt C, Keller G, Pompe B 2002 *Nonlinearity* **15** 1595
- [18] Shi L, Wang C H, Li X *et al* 2003 *Scientia Meteorologica Sinica* **23** 300 (in Chinese) [史 岚, 王翠花, 李 雄等 气象科学 **23** 300]
- [19] Gao X Q, Feng G L, Dong W J *et al* 2003 *Acta Mechanica Sinica* **19** 134
- [20] Dai X G, Wang P, Chou J F 2003 *Chinese Science Bulletin* **48** 2483 (in Chinese) [戴新刚, 汪 萍, 丑纪范 2003 科学通报 **48** 2483]

## A technique for distinguishing dynamical species in the temperature time series of north China<sup>\*</sup>

Hou Wei<sup>1)2)†</sup> Feng Guo-Ling<sup>2)3)</sup> Dong Wen-Jie<sup>2)3)</sup> Li Jian-Ping<sup>4)</sup>

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

2) (Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China)

3) (Key Laboratory of National Meteorological Bureau Climate Research Center, Beijing 100081, China)

4) (Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China)

(Received 26 September 2005; revised manuscript received 15 November 2005)

### Abstract

A new method named "permutation entropy" for detecting dynamical changes in time series was used to analysis the day-to-day temperature time series of north China from 1960a to 2000a. We detected three abrupt climate changes in the mid-1970s and early 1980s. Further analysis using Empirical Mode Decomposition to decompose the pemutation entropy shows that the approximate 10a cycle is most important. So we can draw the conclusion that north China suffered three abrupt climate changes in the mid-1970s and early 1980s, and moreover, the the approximate 10a cycle induced the climate changes directly.

**Keywords:** north China, abrupt cliamate change, pemutation entropy, empirical mode decomposition

**PACC:** 9260X

<sup>\*</sup> Project supported by the National Natural Science Foundation of China (Grant Nos. 90411008 and 40325015) and the State Key Development Program for Basic Research (Grant No. 2006CB400503).

† E-mail: hou\_w@sohu.com