

# 基于条件熵长江三角洲温度的 非线性动力学特征分析\*

封国林<sup>1,2)†</sup> 侯 威<sup>1,2)</sup> 董文杰<sup>2,3)</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区域气候-环境重点实验室, 北京 100029)

3) 国家气候中心气候研究开放实验室, 北京 100081)

(2005 年 3 月 24 日收到, 2005 年 5 月 24 日收到修改稿)

通过粗粒化方法得到符号序列, 如果两个符号序列的条件熵在相对位移为零时达到最小值, 那么这两个时间序列可能属于同一个动力学系统. 进一步可以判断两个时间序列之间的耦合强度并提取出两个动力学耦合信号之间的内在滞后关系. 本文用条件熵算法分析了长江三角洲地区上海、南通、常州、南京和杭州的逐日气温资料, 判断这些地区的温度是否属于同一个动力学系统以及它们之间的动力学耦合关系, 并提取出它们之间的滞后信息.

关键词: 粗粒化, 条件熵, 动力学耦合, 滞后时间

PACC: 9260X

## 1. 引 言

短期气候预测是世界大气科学领域的前沿课题, 也是极其困难的跨学科难题, 气温预测作为气候预测的重要组成部分, 也是备受关注和亟待解决的难题之一. 由于短期气候预测时间跨度长并涉及五大圈层间的相互作用, 表现在时空上的多尺度性、非平稳性等, 从而增加了许多不确定性因素, 给预测带来了极大困难, 因此有必要从非线性的角度进一步研究我国气温的区域动力学行为<sup>[1-7]</sup>.

已有的研究表明在气候系统中可能存在着一些低维吸引子<sup>[8,9]</sup>, 在气候学等领域, 判断所研究的各种时间序列之间的动力学异同性甚至比研究时间序列中存在的吸引子的性质更加重要. 那么对于两个或者更多的时间序列, 有没有一种方法可以判断出这些时间序列是否同属一个潜在的动力学系统? 并且, 对于两个动力学耦合信号, 如何提取出它们之间的滞后信息? 基于这一问题<sup>[10,11]</sup>, 利用不同的观测资料, 通过粗粒化方法得到它们的符号序列, 并计算这两个符号序列之间的条件熵, 观察条件熵随两个符号序列之间相对位移的变化, 从而判断这两个时

间序列的动力学异同性, 这一方法还可以判断两个时间序列之间的耦合关系, 即对于两个动力学耦合的时间序列  $X$  和  $Z$ ,  $X$  对  $Z$  的依赖程度可能与  $Z$  对  $X$  的依赖程度完全不同, 这可以通过计算条件熵  $E(X/Z)$  和  $E(Z/X)$  得到; 此外, 这一方法也可以用来提取两个耦合信号之间的时间滞后关系, 而时间滞后因子是一个很重要的物理量, 它可以导致许多简单系统出现复杂的动力学行为.

本文研究我国长江三角洲上海、南通、常州、南京和杭州地区近 40 年来的逐日温度资料, 得到了这几个地区气温的动力学异同性以及它们之间的动力学耦合关系, 提取出它们之间的滞后信息, 这将有助于建立区域气候模型<sup>[12-17]</sup>.

## 2. 算法简介

### 2.1. 粗粒化方法

假设一时间序列为  $X(i) (i = 1, \dots, m)$ , 其所对应的符号序列为  $S(i) (i = 1, \dots, m)$ ;  $A(i) (i = 1, \dots, m)$  为一组自定义的含有  $m$  个符号的符号串;  $\alpha(i) (i = 0, 1, \dots, m)$  是时间序列  $X(i)$  的  $m+1$  个

\* 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 90411008 和 40325015)资助的课题.

† E-mail: feng\_gl@sina.com

分割点.有

$$S(j) \equiv A(k), \text{若 } \alpha(k) < X(j) < \alpha(k+1), \\ k = 0, 1, \dots, m-1. \quad (1)$$

如将符号  $A(k)$  直接用数字  $k$  来表示,那么对于一个长度为  $L$  ( $L < n$ ) 的子符号序列  $X(i) (i=1, \dots, L)$ ,它所对应的符号序列为  $S(i) (i=1, \dots, L)$ ,可以很方便的定义其特征指数为

$$l_X(L, i) = \sum_{h=1}^L m^{L-h} S_{h+i-1}, \quad (2)$$

这里的  $i$  表示子符号序列在整个符号序列  $S(i)$  中的起始位置.一个子符号序列可以用其特征指数  $l_X(L, i)$  来唯一的标识.用  $P_{l_X}$  来表示一个具有某一特定值的  $l_X$  的出现概率,则信息熵可以定义为

$$E = -\frac{1}{L} \sum_{l_X} P_{l_X} \ln P_{l_X}. \quad (3)$$

对于一个时间序列  $X(i)$ ,其最佳分割点就是在所有可能的分割点之中,使信息熵  $E$  达到最大值的分割点.通常,分割点数目的增加将导致  $E$  的增加,但当  $E$  不再随分割点数目的增加而增加的时候,此时的分割点数目就是最佳的.

## 2.2. 条件熵算法

设有时间序列  $X(i) (i=1, \dots, n)$  和  $Z(i) (i=1, \dots, n)$  经过粗粒化之后,其所对应的符号序列分别为  $S(i) (i=1, \dots, n)$  和  $G(i) (i=1, \dots, n)$ ,它们之间的条件熵定义如下:

$$E(X/Z) = -\frac{1}{N_l} \sum_{l_X} \frac{1}{L} \sum_{l_Z} P(l_Z/l_X) \ln P(l_Z/l_X), \quad (4)$$

这  $l_Z$  里的定义同  $l_X$ ,  $P(l_Z/l_X)$  是当符号序列  $S(i)$  的

某一特征指数取值为  $l_X$  时,符号序列  $G(i)$  的与  $l_X$  处于相同位置的特征指数取值为  $l_Z$  的条件概率,  $N_l$  是所有的  $l_X$  中不同值的个数.

如果时间序列  $X(i)$  和  $Z(i)$  来自于同一动力学系统,那么  $l_X(L, i)$  和  $l_Z(L, i)$  就会具有相关性,但是,如果使两个时间序列彼此之间产生一定的相对位移,就会消除这种相关性,也就是计算  $l_X(L, i)$  和  $l_Z(L, i+n_0)$  的条件熵,  $n_0$  为相对位移.如果在  $n_0=0$  时或者  $n_0$  在零值附近,条件熵达到了最小值,那么这两个时间序列就来自同一动力学系统.

$E(Z/X)$  代表了当  $X(i)$  取一个特定值时,限定  $Z(i)$  所取的那些值的信息,并且  $E(Z/X)$  的值可以和  $E(X/Z)$  的值不一样;条件概率  $P(l_Z/l_X)$  和  $P(l_X/l_Z)$  的值也可以不同,这取决于两个时间序列彼此之间的相对依赖性.

如果在属于同一系统的两个时间序列之间存在着延迟耦合,即现在的  $Z(i)$  的值取决于先前的某个  $X(i)$  值,那么,当相对位移  $n_0$  同实际的延迟一致时,条件熵就达到了最小值.

## 3. 应用和结果

### 3.1. 条件熵在 Lorenz 映射中的应用

取 Lorenz 映射中  $X$  分量与  $Z$  分量的 3000 个数据点,分别计算条件熵  $E(X/Z)$  和  $E(Z/X)$ .图 1 显示了 Lorenz 映射的 (a)  $X$  分量和 (b)  $Z$  分量时间序列.从图 2 中可以看出,条件熵  $E(X/Z)$  和  $E(Z/X)$

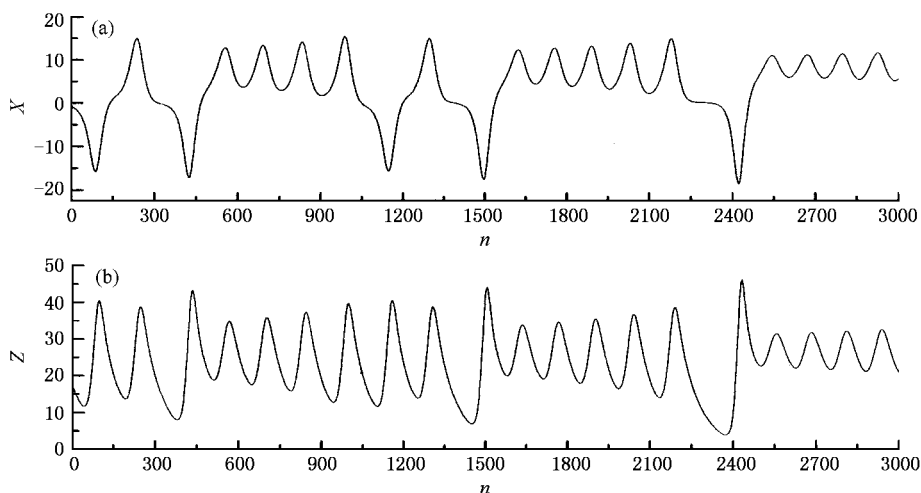


图 1 Lorenz 映射的 (a)  $X$  分量和 (b)  $Z$  分量的时间序列.  $x$  和  $z$  分别为  $X$  分量和  $Z$  分量的值,  $n$  为数据的样本量

分别在  $n_0 = 1$  和  $n_0 = -1$  时达到了最小值,说明时间序列  $X$  和  $Z$  来自于同一个动力学系统;并且

$E(X/Z) \neq E(Z/X)$  发生这样的差异是因为模型中不同分量彼此之间具有不同的相对耦合强度.

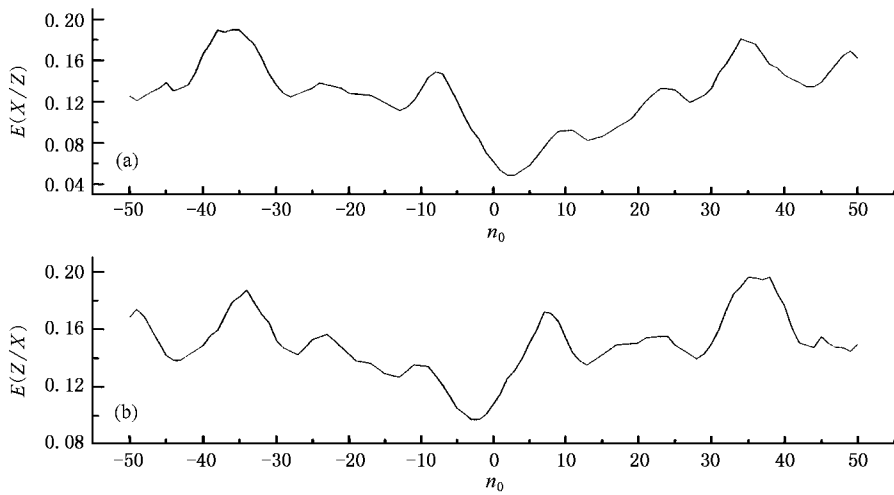


图2 Lorenz 映射的  $X$  分量和  $Z$  分量之间的条件熵  $E$  随  $n_0$  的变化 (a)为  $E(X/Z)$  (b)为  $E(Z/X)$ .  $n_0$  为相对位移

### 3.2. 资料的分析

为了研究长江三角洲各地区气温是否属于同一个动力学系统以及它们之间的动力学耦合关系,并提取出它们之间的滞后信息,从这一地区选取了五个典型站点,分别为南通、南京、常州、上海和杭州(见图3).利用各个站点记载的1960—2000年的逐日气温资料  $\{T\}$ ,分别计算每个站点与其他各个站点之间的条件熵.

时间序列  $X(i)$  的离散化方法如下:

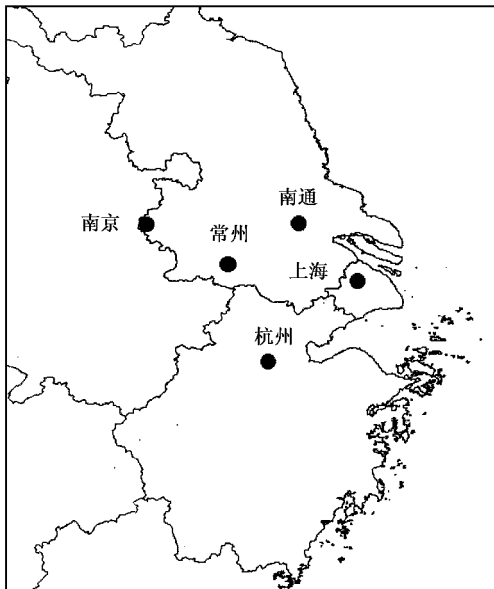


图3 站点选取示意图

$$X_n = X(n\tau) \quad (n = 0, 1, 2, \dots), \quad (5)$$

其中  $\tau$  值的选取要大于消相关时间.消相关时间被定义为 Nyquist 临界频率的倒数.在本文中,我们选取  $\tau = 5$ .

本文中取  $L = 5$  采用二次粗粒化方法对各个站点的温度时间序列进行粗粒化处理,也就是取  $m = 2$ .假设各站点温度资料的分割点为  $\{\alpha_0 = T_{\min}, \alpha_1 = T_1, \alpha_2 = T_{\max}\}$ ,  $T_{\min}$  和  $T_{\max}$  为温度资料  $\{T\}$  的最小值和最大值,采用 2.1 节中介绍的熵最大化方法来确定最佳分割点  $T_1$ ,经过计算得到南通  $T_1$  为 7.0999,南京为 21.5993,常州为 18.2996,上海为 8.7996,杭州为 20.4991.为了便于讨论,在下文的分析中统一采用各个站点的缩写形式,分别为 NT(南通),NJ(南京),CZ(常州),SH(上海)和 HZ(杭州).

### 3.3. 各站点相互之间的动力学分析

在图4中,  $E(CZ/NT)$  在  $n_0 = 0$  时达到了最小值,说明常州的温度受南通温度变化的影响,二者的耦合强度很高,属于同一动力系统;而  $E(HZ/NT)$ ,  $E(NJ/NT)$  均在  $n_0 = -1$  时达到了最小值,表明杭州、南京的温度虽然仍然和南通的温度属于同一动力系统,但是南京和杭州的温度变化均比南通滞后 -1,即南京或杭州的温度变化五天后会影响到南通的温度变化,并且南通温度变化的内在动力学机制和南京、杭州存在着较好的耦合关系.而  $E(SH/NT)$  则在  $n_0 = 0$  时没有最小值,说明上海和南通的温度

变化之间没有动力学耦合关系,不属于同一动力学系统.

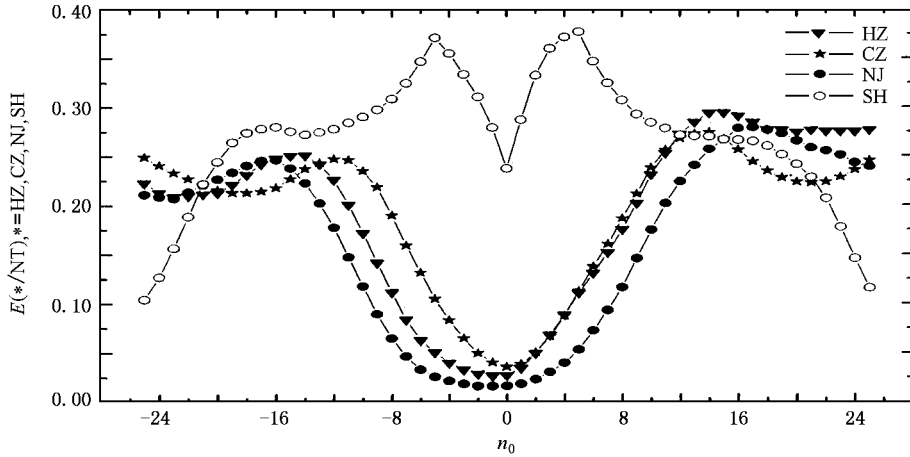


图4 各站点以南通为参考点时条件熵  $E$  随相对位移  $n_0$  的变化

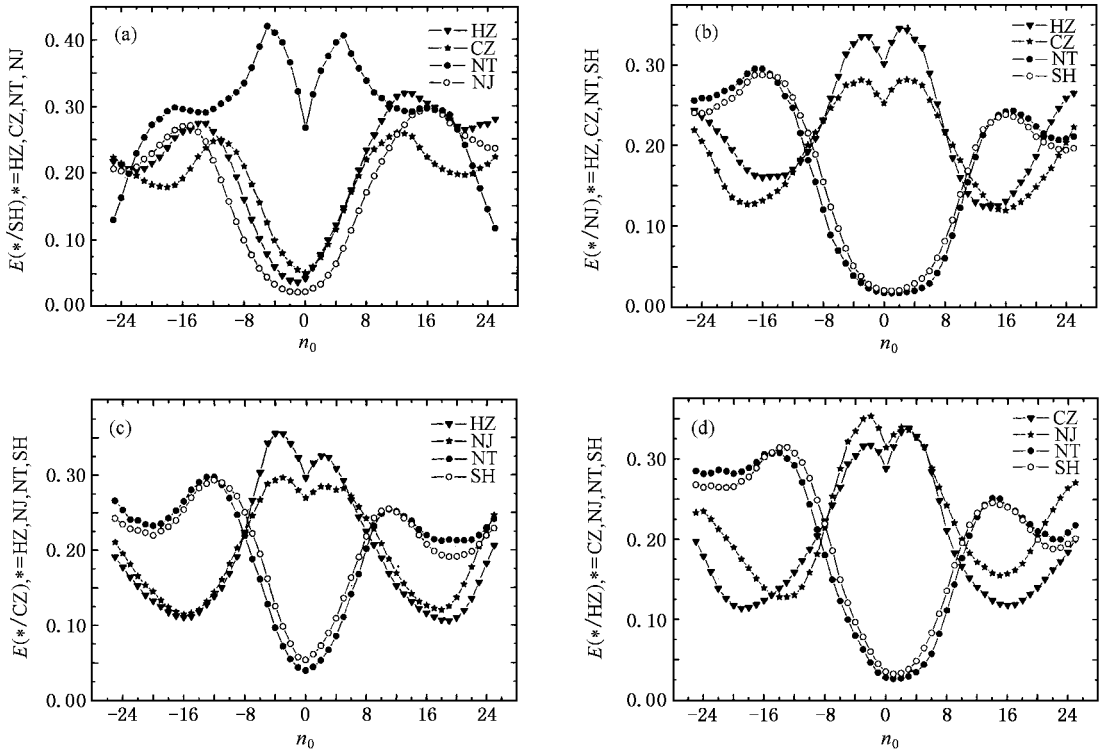


图5 各站点分别以(a)上海(b)南京(c)常州(d)杭州为参考点时条件熵  $E$  随相对位移  $n_0$  的变化

表 1 列出了图 4 和图 5 中分别以各站点为参考点时,条件熵达到最小值时  $n_0$  的取值情况.其中,当以上海为参考点时, $E(CZ/SH)$ 在  $n_0 = 0$  时达到最小值, $E(HZ/SH)$ , $E(NJ/SH)$ 均在  $n_0 = -1$  时达到最小值,而  $E(NT/SH)$ 没有在  $n_0 = 0$  时出现最小值;当以常州为参考点时,只有  $E(NT/CZ)$ , $E(SH/CZ)$ 在  $n_0 = 0$  时出现最小值;以杭州为参考点时, $E(NT/HZ)$ ,

$E(SH/HZ)$ 在  $n_0 = 1$  时具有最小值;当以南京为参考点时,也只有  $E(NT/NJ)$ , $E(SH/NJ)$ 在  $n_0 = 1$  时达到最小值.

综上所述,南通和常州的温度变化属于同一动力系统,具有很高的耦合强度,并且它们的耦合和相互依赖关系是双向的;常州和上海之间的温度变化具有类似情况,杭州和南京的温度变化比南通滞后

-1,即杭州或南京的温度变化五天后会影响到南通的温度,杭州和南京的温度同样也比上海滞后-1;而当分别以杭州和南京为参考点时,计算结果同样显示出南通及上海的温度变化比杭州和南京滞后

1,即落后五天,它们的耦合和相互依赖关系也是双向的.而其他站点之间则没有动力学耦合关系,由不同的动力系统控制.

从地理位置来看,南京四周低山盘曲,长江自西

表 1 以各站点为参考点时的条件熵达到最小值时  $n_0$  取值情况

|     | 站点名         | * = NT    | * = SH    | * = CZ    | * = HZ     | * = NJ     |
|-----|-------------|-----------|-----------|-----------|------------|------------|
| 条件熵 | $E(* / NT)$ | 无         | 无         | $n_0 = 0$ | $n_0 = -1$ | $n_0 = -1$ |
|     | $E(* / SH)$ | 无         | 无         | $n_0 = 0$ | $n_0 = -1$ | $n_0 = -1$ |
|     | $E(* / CZ)$ | $n_0 = 0$ | $n_0 = 0$ | 无         | 无          | 无          |
|     | $E(* / HZ)$ | $n_0 = 1$ | $n_0 = 1$ | 无         | 无          | 无          |
|     | $E(* / NJ)$ | $n_0 = 1$ | $n_0 = 1$ | 无         | 无          | 无          |

向东横穿市区,常州位于江南平原,北倚长江,东濒太湖,而杭州地处长江三角洲南翼,杭州湾西端,各地独特的地理位置形成了各自不同的区域气候子系统.由于各个站点之间的地域远近不同,所以南京、常州、杭州的温度变化对南通和上海温度变化的影响也有不同的滞后时间.虽然南通、上海的温度均受南京、常州和杭州温度变化的影响,但是由于上海地区的“热岛效应”较南通更为显著,所以南通和上海的气候特征也有较大差异.

### 3.4. 单个站点内部年代际之间的动力学分析

将各个站点的温度资料分别分成 1960—1969 年(1960s)、1970—1979 年(1970s)、1980—1989 年(1980s)和 1990—1999 年(1990s)四个年代,计算每两个年代之间的条件熵并判断这两个年代之间的动力学异同性.仍然取  $L = 5, m = 2$ ,采用 2.1 节中介绍的熵最大化方法来确定最佳分割点.

小值,说明南通的气温在 60 年代和 70 年代属于同一动力系统,具有很高的耦合强度;而  $E(80/70)$  和  $E(90/80)$  没有在  $n_0 = 0$  时出现最小值,说明南通的气温在 70—80 年代和 80—90 年代均不属于同一动力系统.因此在 70 年代末,南通的气候系统可能经历了一次突变,导致了气温的异常.

表 2 各站点内部年代间温度的动力学异同性

|    | 年代 | 60—70           | 70—80 | 80—90           |
|----|----|-----------------|-------|-----------------|
| 站点 | NT | 同( $n_0 = 1$ )  | 不同(无) | 不同(无)           |
|    | CZ | 同( $n_0 = 1$ )  | 不同(无) | 不同(无)           |
|    | HZ | 同( $n_0 = 3$ )  | 不同(无) | 不同(无)           |
|    | NJ | 同( $n_0 = -4$ ) | 不同(无) | 不同(无)           |
|    | SH | 不同(无)           | 不同(无) | 同( $n_0 = -4$ ) |

表 2 给出了图 6 和图 7 中各站点内部年代间温度的动力学异同性,各站点的气温在不同年代间的动力学差异呈现出一定的规律性.其中,南通、常州、杭州和南京都是 60 年代和 70 年代属于同一系统,而 70 与 80 年代和 80 与 90 年代都不属于同一系统;上海为 60 与 70 年代和 70 与 80 年代不属于同一系统,而 80—90 年代属于同一系统.综上所述,各站点 70—80 年代均属“不同”,表明长江三角洲地区在 70 年代末期可能经历了一次气候突变.

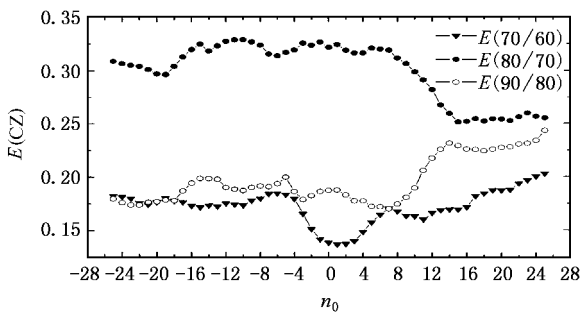


图 6 南通站各年代际之间的条件熵  $E(CZ)$  随相对位移  $n_0$  的变化

## 4. 结 论

本文介绍了一种判断不同时间序列之间动力学异同性的新方法:条件熵算法,讨论了其物理意义,并进一步用 Lorenz 映射的  $X$  分量和  $Z$  分量为资料,

从图 6 可以看出,  $E(70/60)$  在  $n_0 = 1$  时达到最

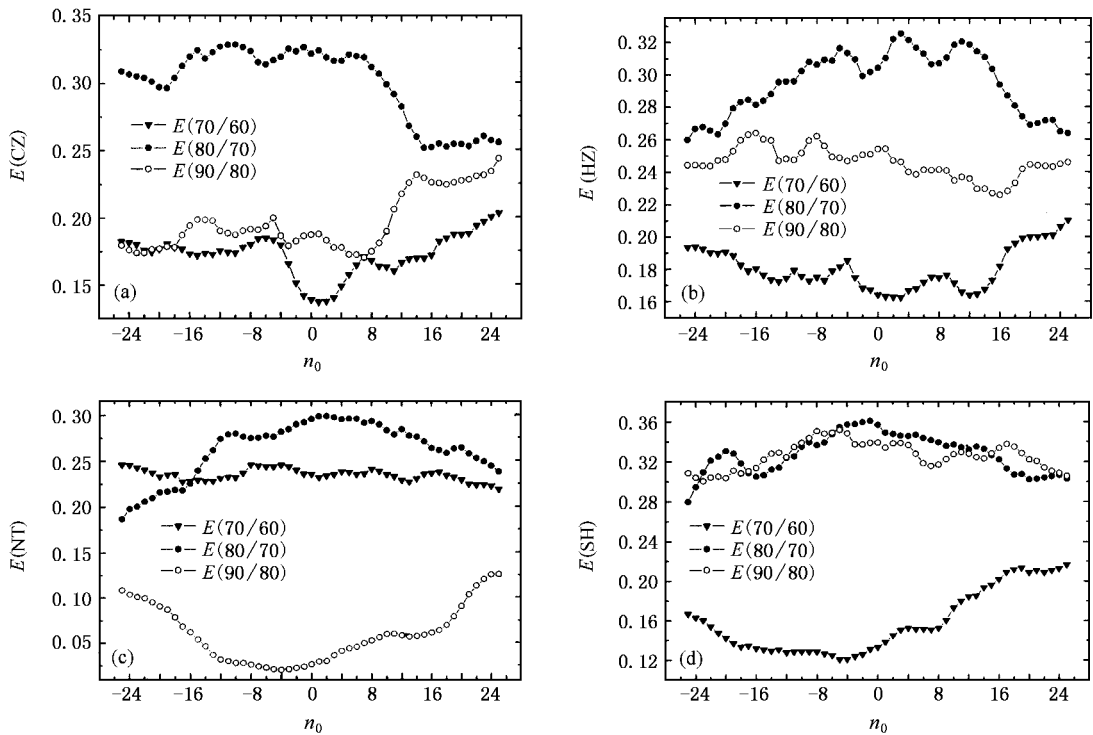


图 7 (a)常州(b)杭州(c)南京(d)上海各年代际之间的条件熵  $E$  随相对位移  $n_0$  的变化

计算了条件熵  $E(X/Z)$  和  $E(Z/X)$ , 验证了这一算法的有效性和可靠性. 利用长江三角洲地区上海、南通、常州、南京和杭州的逐日温度资料, 分别以每个站点为参考点, 计算并分析了其余各个站点相对于参考站点的条件熵. 结果显示, 南通和常州的温度变化属于同一动力系统, 具有很高的耦合强度, 并且它们的耦合和相互依赖关系是双向的; 常州和上海之间的温度变化具有类似情况; 杭州和南京的温度变化比南通滞后 -1, 即杭州或南京的温度变化五天后会影响到南通的温度. 杭州和南京的温度同样也比上海滞后 -1, 而当分别以杭州和南京为参考点时, 计算结果同样显示出南通及上海的温度变化比杭州和南京滞后 1, 即落后五天, 它们的耦合和相互依赖关系也是双向的. 而其他站点之间则没有动力学耦合关系, 由不同的动力系统控制. 另外发现不属于同

一动力学系统的站点之间的条件熵虽然在  $n_0 = 0$  时没有达到最小值, 但是却都出现了极小值. 这些极小值产生的原因以及它们是否具有特定的物理意义, 还有待于进一步的研究.

在年代际的时间尺度上分析了单个站点温度资料各个年代之间的动力学异同性, 将各个站点的气温资料分别分为 1960—1969 年、1970—1979 年、1980—1989 年和 1990—1999 年四个年代, 计算各个年代之间的条件熵. 结果表明, 在 70 年代末, 长江三角洲各地区可能普遍经历了一次气候突变. 其中, 南通、常州、杭州和南京都是 60 年代和 70 年代属于同一系统, 而 70 年代与 80 年代和 80 年代与 90 年代都不属于同一系统; 上海是 60 年代与 70 年代以及 70 年代与 80 年代不属于同一系统, 而 80 年代与 90 年代则属于同一系统.

- [1] Dai X G, Fu Z B, Wang P 2005 *Chin. Phys.* **14** 850  
 [2] Wang P, Dai X G, Feng G L et al 2004 *Chin. Phys.* **13** 1770  
 [3] Gao X Q, Feng G L, Dong W J et al 2003 *Acta Mecha. Sin.* **19** 134  
 [4] Feng G L, Dong W J 2004 *Chin. Phys.* **13** 413  
 [5] Feng G L, Dong W J, Chou J F 2004 *Acta Phys. Sin.* **53** 2389 (in

- Chinese) [封国林、董文杰、丑纪范 2004 物理学报 **53** 2389]  
 [6] Hou W, Feng G L, Gao X Q et al 2005 *Acta Phys. Sin.* **54** (in Chinese) [侯威、封国林、高新全等 2005 物理学报 **54**]  
 [7] Shi N 2005 *Chin. Phys.* **14** 844  
 [8] Grassberger P 1986 *Nature* **323** 609  
 [9] Nicolis C, Nicolis G 1984 *Nature* **311** 529

- [ 10 ] Rajeev K Azad , Rao Subba J , Ramaswamy Ramakrishna 2004 *Nonlinear Analysis :Real World Applications* **5** 487 2483 ( in Chinese ) [ 戴新刚、汪 萍、丑纪范 2003 科学通报 **48** 2483 ]
- [ 11 ] Lehrman M , Rechester A B 1997 *Phys. Rev. Lett.* **78** 54 [ 15 ] Wang P , Dai X G 2005 *Acta Phys. Sin.* **54** 4961 ( in Chinese ) [ 汪 萍、戴新刚 2005 物理学报 **54** 4961 ]
- [ 12 ] Fu Z T , Liu S S , Liu S D 2004 *Acta Phys. Sin.* **53** 343 ( in Chinese ) [ 付遵涛、刘式适、刘式达 2004 物理学报 **53** 343 ] [ 16 ] Zeng H L , Gao X Q , Zhang W 2005 *Chin. Phys.* **14**
- [ 13 ] He W P , Feng G L , Dong W J *et al* 2005 *Chin. Phys.* **14** 21 [ 17 ] Feng G L , Chou J M , Dong W J 2004 *Chin. Phys.* **13** 1582
- [ 14 ] Dai X G , Wang P , Chou J F 2003 *Chinese Science Bulletin.* **48**

## A technique for distinguishing dynamical species in the temperature time series of Yangtze River delta \*

Feng Guo-Lin<sup>1,2,†</sup> Hou Wei<sup>1,2</sup> Dong Wen-Jie<sup>2,3</sup>

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

2 *The Institute of Atmospheric Physics , Chinese Academy of Sciences , Beijing 100029 , China*

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

( Received 24 March 2005 ; revised manuscript received 24 May 2005 )

### Abstract

The conditional entropy of two symbolic sequences which are encoded faithfully from two different time series is minimized at the zero relative shift of the two time series if the two time series have their origins in the same underlying dynamics. This technique is also useful in studying the delayed coupling between different time series. Applications are made to the temperature data derived from five meteorological stations in China , namely Shanghai , Nantong , Changzhou , Nanjing and Hangzhou in order to test the commonality of their underlying dynamics , obtain the relative strength of the dynamics coupling between different variables , and extract the implicit delayed information in the dynamics.

**Keywords** : coarse-graining , conditional entropy , dynamical coupling , time-delays

**PACC** : 9260X

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

† E-mail : feng\_gl@sina.com