

中国气候系统复杂性及其地域差异

李双成 刘逢媛 赵志强

北京大学城市与环境学院, 北京 100871

摘 要 气候系统是一个非线性复杂系统, 对其复杂性进行度量有助于认识气候变化的内在本质。通过对全国 648 个台站 1961~1990 年 30 年日气温和日降水距平系列的近似熵分析, 发现该指数可以有效地度量气候系统的复杂性。气温日变化系列的近似熵高值带出现在大陆性过渡性气候和海洋性过渡性气候带, 而在海洋性气候或大陆性气候单一气候类型控制的区域近似熵值普遍较低, 这反映了两种或两种气候类型交互地带气候特征较为复杂; 降水日变化的高值区出现在海拔较高、地表起伏较大的区域, 反映了地形条件是塑造降水系列复杂性的重要因素。

关键词 气候系统 复杂性格局 近似熵 地域差异
文章编号 1006-9585 (2008) 01-0031-08 **中图分类号** P423.3 **文献标识码** A

Climate Complexity and Spatial Variation in China

LI Shuang-Cheng, LIU Feng-Yuan, and ZHAO Zhi-Qiang

College of Urban and Environmental Sciences, Peking University, Beijing 100871

Abstract Climate system is a complex nonlinear system. Estimation of the complexity is of great interest in climatic forecast and prediction. We proposed the use of approximate entropy (ApEn) to measure the complexity of daily temperature and precipitation series in China. The result indicated that ApEn successfully capture complex information in both climatic time series. The fact that higher ApEn values of temperature series in China appear in transition zone between monsoon and non-monsoon climatic types indicate the complexity of system emerges from the coupling between processes or sub-systems. The consistency of spatial pattern between ApEn values of precipitation series and terrain indicates that topographical factors play a significant role in precipitation.

Key words climate system, complexity pattern, approximate entropy, regional difference

1 引言

气候系统是一个外有强迫、内有非线性耗散的开放系统^[1~3]。Nicolis 等^[4]和 Rind^[5]认为, 诸如气温、降水等非平稳的一维过程数据中包含着远为丰富的信息, 它蕴藏着参与动态的全部其他变量的痕迹^[4,5]。通过对这些单变量时间系列数据进行分析处理, 可以展示出气候系统的非线性特

征。近年来, 非线性动力学的一些指数被用来表征气候系统的复杂性, 如关联维、李雅普诺夫指数、柯尔莫哥罗夫熵等^[6,7]。其中关联维是衡量系列最后收缩到的子空间的维度, 也表征了描述一个复杂系统所需的最少实质性状态变量; 李雅普诺夫指数用来度量相空间中轨迹的发散或收缩速率, 而柯尔莫哥罗夫熵则说明了系统状态信息的损失率。然而要通过这些指数准确地刻画气候系统, 需要较多的观测数据支撑。Wolf 等^[8]曾经指出要

收稿日期 2006-09-20 收到, 2007-08-24 收到修定稿
资助项目 国家重点基础研究发展规划项目 2003CB415101、教育部科学研究重大项目 306019 和国家自然科学基金 40771001
作者简介 李双成, 男, 1961 年出生, 博士, 副教授, 主要从事地表格局与过程复杂性计算与模拟研究。E-mail: scl@urban.pku.edu.cn

可靠地刻画一个 m 维吸引子所需数据点多达 30^m 个。寻找更加有效的动力学指数和算法来准确描述气候系统仍是具有挑战性的研究课题。从数据点要求及算法的适应性和稳健性上看, Pincus^[9] 在 1991 年提出的近似熵 (Approximate entropy, ApEn) 是一个值得考虑的指数。近似熵具有所需数据点较少、有较好的抗噪和抗干扰能力以及对随机信号、确定性信号或两种混合信号都能适应等特点。近似熵是用一种有效的统计方式——边缘概率的分布来区分各种过程。它用一个非负数来表示一个时间序列的复杂性, 反映一个时间序列在模式上的自相似程度的大小, 衡量维数变化时序列中产生新模式的概率的大小以及时间序列中新信息发生率, 越复杂的时间序列对应的近似熵越大^[10~13]。在已有的近似熵应用案例中, 集中在生物信号的处理上^[14~17], 用以度量气候系统复杂性的工作尚不多见。本文引入近似熵算法测定中国气候系统的复杂性, 以期丰富气候系统非线性动力学表征方法, 同时通过对中国各个气象台站气温降水数据的分析, 揭示气候系统复杂性的地域差异。

2 近似熵算法

本质上, 近似熵属于非线性动力学参数, 因而利用它对一个时间序列进行分析的过程类似于计算关联维或李雅普诺夫指数的步骤, 下面概要介绍近似熵的算法^[9]。

给定一个单变量时间序列数据 $u(1), u(2), u(3), \dots, u(N)$, 其中 u 代表系列的变量, 在本研究中为气温或降水,

(1) 在 R^m 内构造一个矢量系列 $x(1), x(2), \dots, x(N-m+1)$, 其中:

$$x(i) = [u(i), u(i+1), \dots, u(i+m-1)], \quad (1 \leq i \leq N-m+1);$$

(2) 定义两个矢量 $x(i)$ 和 $x(j)$ 的最大分量差作为距离, 即

$$d[x(i), x(j)] = \max_{k=1, 2, \dots, m-1} (|u(i+k) - u(j+k)|);$$

(3) 设阈值 r 为一给定正数, 对每一 i 值统计 $d[x(i), x(j)]$ 小于 r 的数目及此数目与距离总数 $N-m+1$ 的比值, 记作 $C_i^m(r)$, 即

$$C^m(r) = \frac{\sum_{i=1}^{N-m+1} C_i^m(r)}{N-m+1}; \quad (1)$$

(4) 将 $C_i^m(r)$ 取对数, 再求其对所有 i 的均值, 记作 $\phi^m(r)$, 即

$$\phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N-m+1}; \quad (2)$$

(5) 增加维数, $m \rightarrow m+1$, 重复步骤 (1) ~ (4), 得到 $\phi^{m+1}(r)$;

(6) 理论上此单变量时间系列的近似熵为

$$E_A(m, r) = \lim_{N \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)]. \quad (3)$$

实际观测的数据点 N 不可能为无穷个, 当 N 为有限样本时, 通过上述步骤计算出的近似熵是公式 (3) 的无偏估计值。记作:

$$E_A(m, r, N) = \phi^m(r) - \phi^{m+1}(r). \quad (4)$$

式中, m 为相空间维数, r 为阈值型的正定常数, N 为系列数据点数。显然, 近似熵的计算值与这 3 个参数有关。与重构相空间方法计算关联维、李雅普诺夫指数和柯尔莫哥罗夫熵一样, 一些参数的取值需要一定的经验。Pincus^[9] 根据实践建议, $m=2$, $r=(0.1 \sim 0.2) D_x$ (D_x 为原始数据的标准差)。Pincus 还证实, N 的取值在 100~5 000 点就能取得近似熵的稳健估计值。

3 数据准备与近似熵计算

3.1 气候站点选择与数据处理

本研究的目的是通过计算中国各个气象台站气候要素值的近似熵, 阐明中国气候系统复杂性的地域差异。因而, 应当多选择气象台站, 并且在空间上应尽可能均匀。但由于各地建站早晚不同, 气温与降水观测数据的时间系列长短差异较大。通过权衡台站数目与时间系列长短之间关系, 选取 648 个气象台站的日气温和日降水 1961~1990 年 30 年系列数据作为近似熵计算的对象。西藏和新疆中西部地区由于建站时间过晚, 符合数据要求的台站较少 (图 1)。

在选取气温与降水系列后, 对每一台站的数据进行了标准化处理。标准化处理后的时间系列总体平均值为 0, 均方差为 1^[3]。



图1 中国气象台站分布
Fig. 1 Meteorological stations in China

3.2 近似熵的计算

依照（4）式，对中国 1961~1990 年 648 个台站的逐日气温和降水系列进行了近似熵计算。计算参数为 $m=2$ ， $r=0.18$ ， $N=10\ 957$ （含 7 个闰年）。

3.3 计算结果

表 1 是对气温与降水近似熵计算结果的统计分析。从表中可以看出，气温系列的近似熵值大多在 1.0~2.0，均值为 1.841；而降水系列的近似熵值在 0.27~0.68，均值为 0.46。气温系列的近似熵值普遍高于降水系列，且标准差较大，这是因为气温每日均有一个观测值，其系列构型复杂；而降水并非每日都有，在无降水或微降水情形下，系列中的数据值为零，特别是对于降水稀少的季节来说，降水系列中零值较多，降低了系列的复杂程度。

近似熵值反映了系列的复杂性大小，熵值越大说明系列复杂性越高，随机性越大，可预测性越低。反之，熵值越小说明系列复杂性越低，随机性越低，可预测性越高。在本研究中，作者采用 matlab 中的 randn 和 diric 函数分别生成与气温和降水系列等长的随机系列和周期正弦波系列，并用相同参数计算了这两个生成系列的近似熵。结果为：随机系列的近似熵为 2.26，而正弦波系列的近似熵值只有 0.12。由此表明，气温和降水系列的动力学特性介于随机系列和确定性系列之间。

4 中国气候系统复杂性地域差异

为了展示中国气候系统复杂性的地域差异，利用 ArcGIS8.3（ESRI Inc. 1999~2002 年）对 648 个气象台站的近似熵进行了克里格插值。

表 1 中国日气温和日降水系列近似熵统计结果
Table 1 Statistical results of approximate entropy of daily temperature and precipitation in China

系列	样本数	平均值	中值	最小值	最大值	四分位数	标准差	斜度	峰度
气温	648	1.841	1.872	1.063	2.041	0.157	0.118	-1.714	6.101
降水	648	0.460	0.455	0.265	0.684	0.084	0.063	0.260	0.414

4.1 气温系列复杂性地域差异

4.1.1 空间分布及统计特征

利用 ArcGIS 中的空间统计模块确定出气温系列近似熵的高值分布走向和范围（图 2）。从图 2 中可以看出，气温近似熵的高值区基本上呈东北西南带状分布，近似熵 1.90 高值中心区（图 2 中浅色椭圆形区）自东北向西南穿越富锦、通河、克什克腾旗、鄂托克旗、天祝、兴海、称多和那曲。近似熵 1.95 高值中心区（图 2 中深色椭圆形区）的方向略有变化，但也接近呈东北西南向。气温近似熵的低值区分布格局较为分散，典型的低值区是东北北部、华南沿海、成都平原以及新疆、西藏中西部等。

4.1.2 成因解析

通过近似熵值高低反映出来的气温系列复杂性大小的地域差异是显著的。何以形成这种复杂性的空间格局？按照非线性动力学理论，若一个系列由多个非线性过程相互作用而成，则这一系列必定是非平稳的系列，复杂性也较高。在一个复杂系统中，两相或多相物质体系相互作用区，往往是非线性表现最明显的区域。

通过与中国主要气候类型分布格局相比（图 3）^[18]，近似熵的高值区基本上分布在大陆性过渡

气候与海洋性过渡气候类型区内，而在西北内陆大陆性气候区和华南沿海以及成都平原海洋性气候区，近似熵空间格局呈现明显的低值连续区。另外，在西南季风作用的区域以及黑龙江北部也有小范围的低值区。这说明，两个或两个以上气候类型交互作用的区域比单一气候类型控制的区域，气温系列的复杂性要高。在东北西南的高值带上，有一个位于内蒙古东部、河北北部及辽宁西南部的高值中心，这个明显的高值中心的气温近似熵值均在 1.95 以上。该中心恰好位于我国温带大陆性气候和温带季风气候交互作用带，说明该区域海洋与陆地显著而强烈的相互作用塑造了气候的复杂性。

4.2 降水系列复杂性地域差异

4.2.1 空间分布及统计特征

从图 4 可以看出，降水系列近似熵的高值区集中分布在 4 个区域：一是北疆的阿尔泰地区，二是东北山地（包括大小兴安岭、长白山以及内蒙古东部地区），三是青藏高原东南部，四是湘西及贵州高原大部。从 4 个区域平均状况看，近似熵的平均值为 0.5044，标准差为 0.0193（表 2）。

从出现频率上分析，4 个区域的近似熵集中

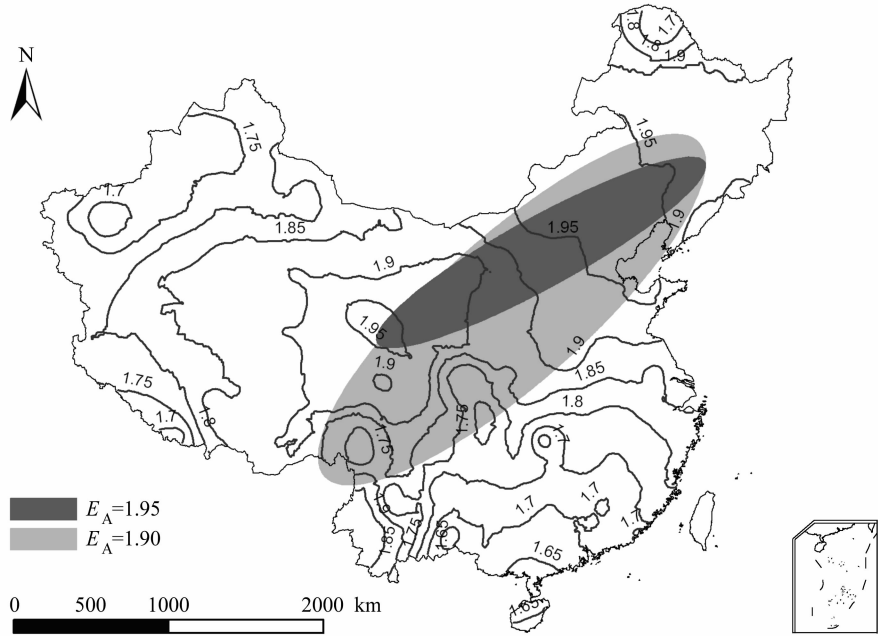


图 2 1961~1990 年中国气温日变化系列近似熵空间分布
Fig. 2 Spatial pattern of approximate entropy of daily temperature series during 1961—1990 in China

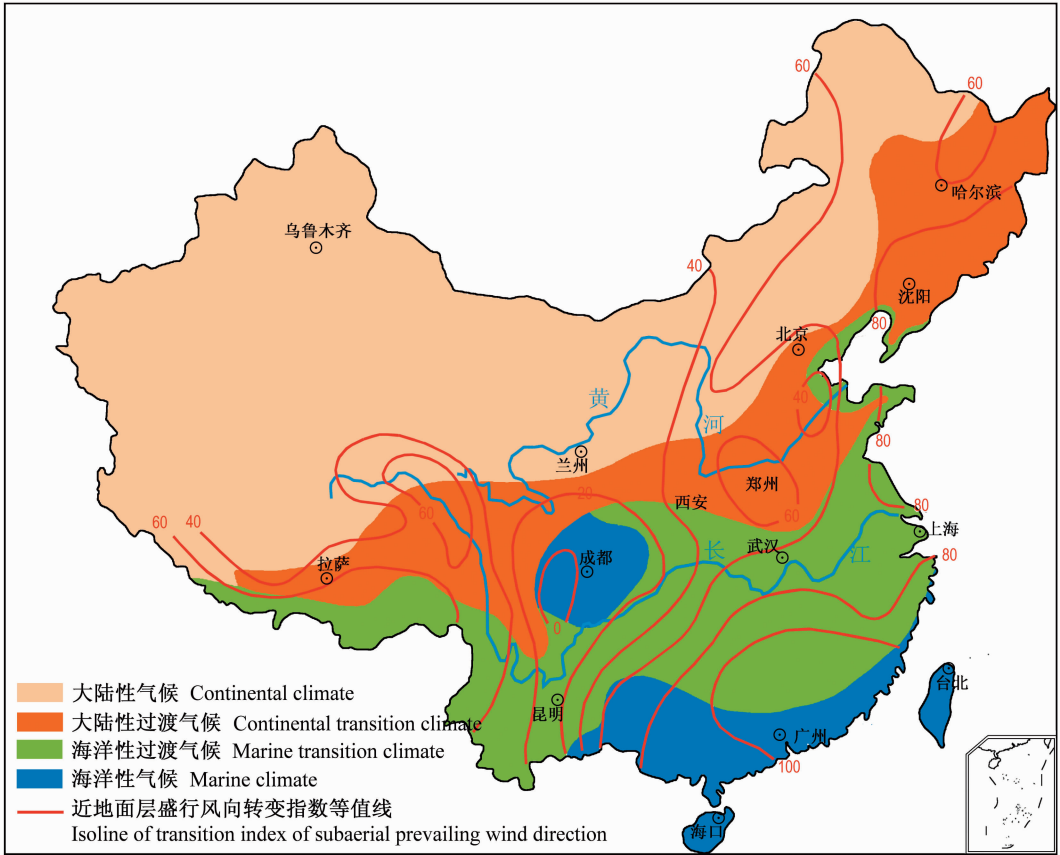


图 3 中国陆地和海洋性气候区划 (引自文献[18])
Fig. 3 The regionalization of continental climate and ocean climate in China (from reference [18])

在 0.49~0.55，两者占总体的 78.5%。

低值区分布贯通中国东西，尤以南疆、内蒙古中部、黄淮海平原等区域最为明显。其中，南疆地区近似熵最低值为 0.325，最高值为 0.387，均值为 0.356，标准差为 0.014；内蒙古中部最低值 0.379，最高值为 0.414，均值为 0.388，标准差为 0.009；黄淮海地区近似熵最低值为 0.389，最高值为 0.409，均值为 0.398，标准差为 0.004。图 5 和 6 是近似熵高值区和低值区各分段出现频率。

4.2.2 成因解析

从图 4 可以看出，降水系列的近似熵高低与地形要素有一定关系，对于大多数台站来说，地形越高，近似熵值越大。为了验证这种定性看法，对各个台站的近似熵和海拔高度进行了相关分析。结果表明，648 个台站降水系列近似熵与其海拔高度 (H) 的 Spearman 相关系数为 0.32，显著水平 $p < 0.05$ 。然而从相关分析结果看，降水系

列复杂性的成因比较难以解析，除地形外，可能有其他的气候或非气候因素在其中起重要作用。

5 结论与讨论

本文应用近似熵算法对全国 648 个台站 1961~1990 年 30 年日气温和日降水距平系列的复杂性进行了度量，结果发现：

- (1) 在度量系列复杂性方面，近似熵具有要求数据点少、抗噪声以及适应性广等优点，是一个值得考虑的非线性动力学参数。它比简单的统计参数如均值、方差等更能检视蕴含在系列顺序中的信息。一般来说，对于一个特定时间系列，近似熵值越高说明系列的随机性越大，可预测性就越低。
- (2) 近似熵能够较好地用在测定气候系列的统计特性上。本研究发现，气温和降水系列的动力学特性介于随机系列和确定性系列之间，即气

表 2 降水系列近似熵高值区统计特征值

Table 2 Statistical results of high approximate entropy of daily precipitation in China

区 域	最小值	最大值	范围	平均值	标准差
阿尔泰地区	0.4137	0.5353	0.1216	0.4868	0.0328
青藏高原东南部	0.4507	0.5721	0.1214	0.5157	0.0239
湘西及贵州高原	0.4765	0.5147	0.0382	0.4978	0.0056
东北山地	0.4719	0.5537	0.0818	0.5172	0.0150
平均	0.4532	0.5440	0.0908	0.5044	0.0193

候系列的复杂性高于确定性系列，而低于完全随机的系列。这一研究结果证实了气候系统是一个混沌系统，对其趋势进行预测应当采取审慎的

做法。

(3) 就气候系统来讲，气温系列的复杂性大于降水系列。这是由于气温是一个连续的观测值，

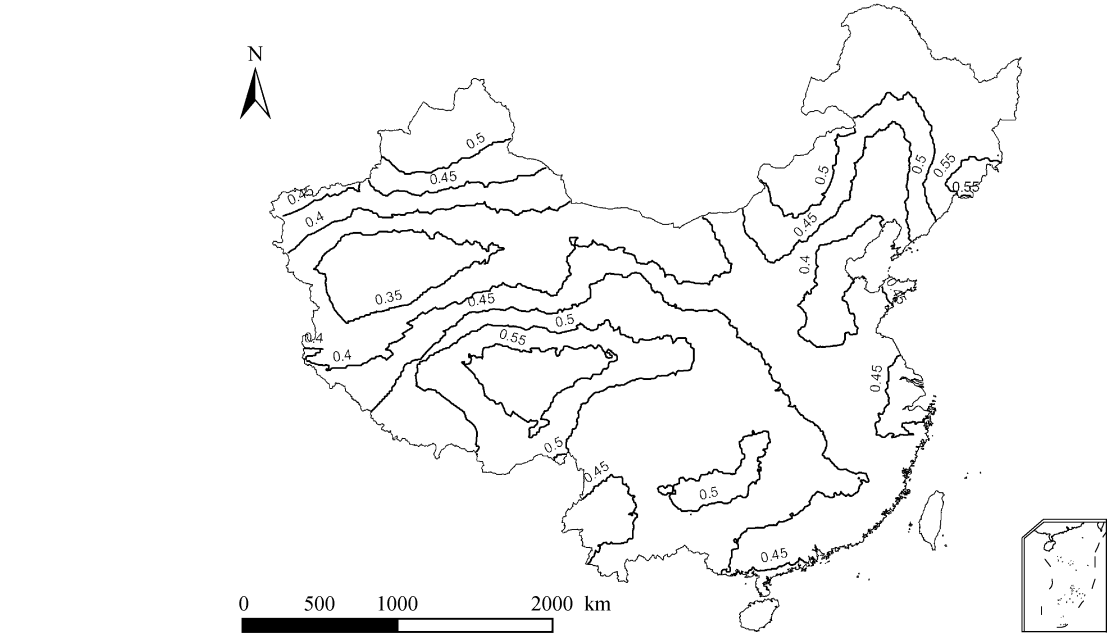


图 4 1961~1990 年中国降水日变化系列近似熵空间分布

Fig. 4 Spatial pattern of approximate entropy of daily precipitation series during 1961—1990 in China

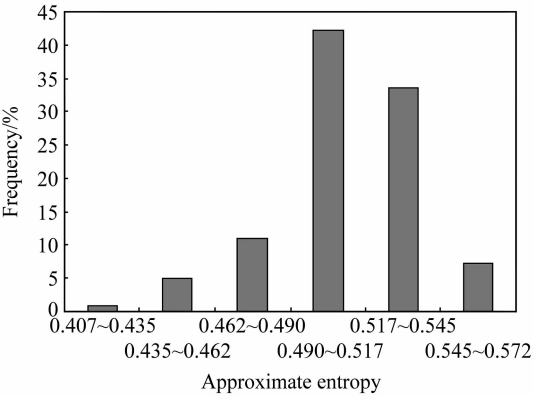


图 5 降水近似熵高值区域分布概率

Fig. 5 The distribution frequency of precipitation in high approximate entropy areas

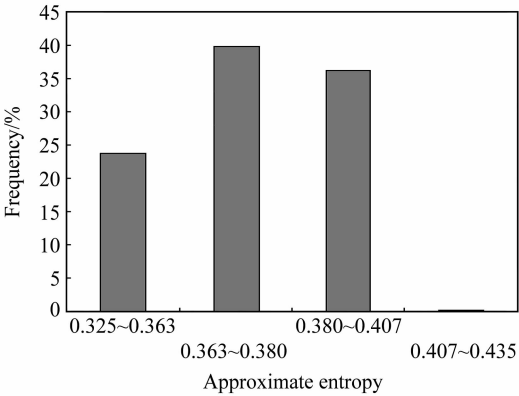


图 6 降水近似熵低值区域分布概率

Fig. 6 The distribution frequency of precipitation in low approximate entropy areas

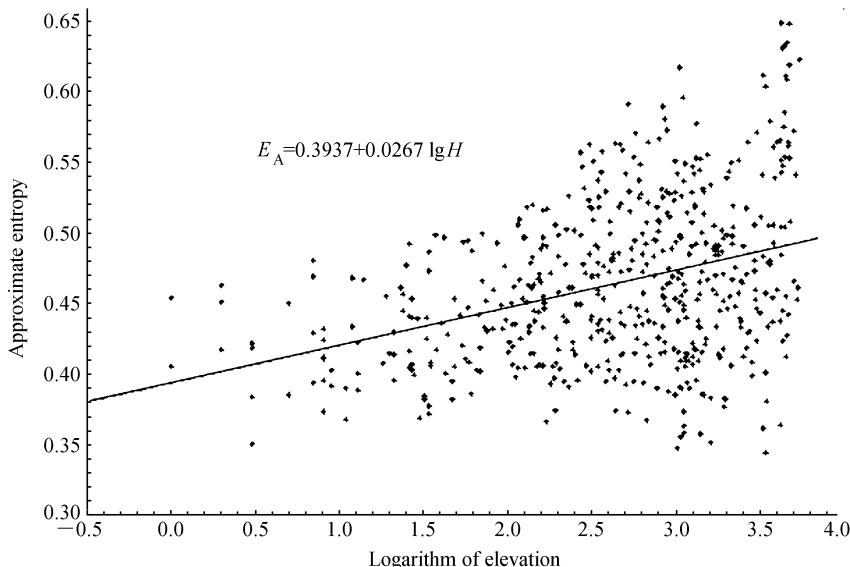


图7 各气象台站降水近似熵值与海拔高度的关系

Fig. 7 The relationship between approximate entropy of precipitation and elevation of the stations

系列构型复杂；而降水系列则由于无降水或微降水的出现充斥着数值为0的记录，特别是对于降水稀少的季节来说，降水系列中的很多数据段被0占据，降低了系列的复杂程度。尽管在数据前处理中，使用了距平方法，但对于系列的性质仍难加以改变。

(4) 通过对各个气象台站气温和降水近似熵的空间插值，可以清晰地反映出其空间结构和地域差异。气温系列近似熵的空间结构与中国主要气候类型的空间分布大致吻合。其近似熵高值区出现在大陆性季风气候与内陆性干旱气候过渡区域，即在两种或两种以上气候类型交互作用的区域，气温系列复杂性明显高于单一气候类型控制区域；降水系列的近似熵与地形因素具有一定关联，即在地形复杂、海拔较高的区域，降水系列复杂性明显高于其他地区。

(5) 本文用近似熵算法度量了中国气温和降水系列的复杂性及其空间格局，并初步解释了这种格局的成因，然而仍有一些需要改进的地方。首先，近似熵算法的一些参数设置如嵌入维和阈值需要一定的经验，寻求更加客观的参数设置方法需要进一步探索；其次，欲准确客观地反映中国气候系统的复杂性，仅用近似熵单一指数是不够的，应当结合其他非线性动力学参数如关联维、李雅普诺夫指数、柯尔莫哥罗夫熵以及定量递归

分析的一些特征值等。这些不同方法的指数之间相互校验，可以获得更准确的气候系列复杂性及其空间格局的计算结果。

参考文献 (References)

- [1] 李建平, 丑纪范. 非线性大气动力学的进展. 大气科学, 2003, **27** (4): 653~673
Li Jianping, Chou Jifan. Advances in nonlinear atmospheric dynamics. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2003, **27** (4): 653~673
- [2] 穆穆, 李建平, 丑纪范, 等. 气候系统可预报理论研究. 气候与环境研究, 2002, **7** (2): 227~235
Mu Mu, Li Jianping, Chou Jifan, et al. Theoretical research on the predictability of climate system. *Climatic and Environmental Research* (in Chinese), 2002, **7** (2): 227~235
- [3] 辛国君. 气候噪声和气候系统的分维. 应用气象学报, 1997, **8** (1): 85~91
Xin Guojun. Climate noise and estimating the dimensions of climate attractor. *Quarterly Journal of Applied Meteorology* (in Chinese), 1997, **8** (1): 85~91
- [4] Nicolis C, Nicolis G. Is there a climate attractor? *Nature*, 1983, **311**: 529~632
- [5] Rind D. Complexity and climate. *Science*, 1999, **284**: 105~107
- [6] Torres M E, Gamero L G. Relative complexity changes in time series using information measures. *Physica A*, 2000,

- 286:** 457~473
- [7] Li S C, Zhou Q F, Wu S H, et al. Measurement of climate complexity using sample entropy. *Int. J. Climatol.*, 2006, **26**: 2131~2139
- [8] Wolf A, Swift J B, Swinney H L. Determining Lyapunov exponents from a time series. *Physica D*, 1985, **16**: 285~317
- [9] Pincus S M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences (USA)*, 1991, **88**: 2297~2301
- [10] Pincus S M, Viscarello R R. Approximate entropy: A regularity measure for fetal heart rate analysis. *Obstetrics and Gynecology*, 1992, **79**: 249~255
- [11] Pincus S M, Goldberger A L. Physiological time-series analysis: What does regularity quantify? *American Journal of Physiology*, 1994, **266**: H1643~H1656
- [12] Pincus S M. Greater signal regularity may indicate increased system isolation. *Mathematical Biosciences*, 1994, **122**: 161~181
- [13] Pincus S M. Approximate entropy (ApEn) as a complexity measure. *Chaos*, 1995, **5**: 110~117
- [14] Dawes G S, Moulden M, Sheil O, et al. Approximate entropy, a statistic of regularity, applied to fetal heart rate data before and during labor. *Obstetrics and Gynecology*, 1992, **80**: 763~768
- [15] Fleisher L A, Pincus S M, Rosenbaum S H. Approximate entropy of heart rate as a correlate of postoperative ventricular dysfunction. *Anesthesiology*, 1993, **78**: 683~692
- [16] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology*, 2000, **278**: H2039~H2049
- [17] Souza G M, Ribeiro R V, Pincus S M. Approximate Entropy as a tool to measure complexity in time series of plant physiological data. In: 1st International Conference on Bioinformatics and Computational Biology, Proceedings of the 1st International Conference on Bioinformatics and Computational Biology. 2003, Ribeirão Preto
- [18] 中国自然地理图集编辑委员会. 中国自然地理图集 (第二版). 北京: 地图出版社, 1998
The Editorial Committee of Chinese Physical Geography Atlas. *Chinese Physical Geography Atlas* (2nd ed.). Beijing: Map Publishing Houses, 1998