

排列熵应用于气候复杂性度量

郝成元^{1,3}, 吴绍洪¹, 李双成²

(1 中国科学院地理科学与资源研究所, 北京 100101; 2 北京大学环境学院资源与环境地理系, 北京 100871;
3 中国科学院研究生院, 北京 100049)

摘要: 云南省南部地区由于多季风系统影响和大地形作用的多样性, 气候系统复杂多样。本文利用多年日气温和降水资料, 依据基于长时间序列的相邻数据对比的复杂度参数——排列熵, 对这种气候系统复杂性的度量效果是简单而有效的: 气温熵值的空间格局反映了区域大气环流的复杂性和哀牢山山地作用的显著性; 降水熵值的空间格局不仅指示了水汽来源的方向, 而且反映了降水空间递变强度的区域差异。并很好的诠释了当地气候的大气环流和局地地形的双重影响效果。

关键词: 排列熵; 气候复杂性; 大气环流; 地形因子

文章编号: 1000-0585(2007)01-0046-07

1 引言

多年来, 气象工作者对大气科学中的非线性与复杂性进行了多层次的研究^[1]。气候系统是一个非线性、无标度、多层次、强迫耗散十分复杂的巨系统^[2], 对其内涵的复杂性认识具有重要的理论意义和实用价值^[3]。即由于气候本身是变化的, 故而采用某一能够反映其变化的不确定性特点的统计量来分析某地气候特点, 更具有合理性, 而不是用某些要素的平均值大小与分布来分析^[4]。这个能够度量某些要素不确定性程度的指标就是熵。

熵原本是一个热力学概念, 是用来描述热力学系统混乱(无序)程度的。它反映一个系统宏观态所具有的微观态数目或热力学几率, 与热力学几率的对数成正比。由于热力学几率愈大表示系统处于愈混乱的状态, 因此熵可作为系统复杂性的度量^[5]。在信息论建立以后, 关于熵的概念和理论得到了迅速发展。

近年来, 随着非线性科学的发展, 不管在社会科学还是自然科学领域中, 对各种复杂现象复杂度的研究已成为各学科的前沿课题^[6,7]。我国学者在气候气象上的熵研究有, 汤懋苍在上世纪 80 年代中期建立了熵气候模式^[8]; 张学文于 1986 年提出“熵气象学”的概念, 并于 1992 年出版了《熵气象学》一书^[9]。

在众多熵值的计算模式中, 排列熵 (Permutati n Entr py) 是一个基于长时间序列的相邻数据对比的复杂度参数, 它的优点是易于计算、简单而实用^[10]。在医学上, 依据排列熵的计算, 基于心脏、脑活动数据的复杂性成功地诊断出疾病或是否健康以及健康程度^[11,12], 但在气候气象学的应用到目前为止还不多见, 因为气候系统复杂性受到包括大气环流、下垫面特征、太阳辐射等多种因素的影响, 尤其是地形地貌多样的山地高

收稿日期: 2006-03-18; 修订日期: 2006-07-17

基金项目: 国家重点基础研究发展计划 (973 项目) (2003CB415101) 资助

作者简介: 郝成元 (1969), 男, 山东曹县人, 博士, 讲师。主要从事自然地理综合研究。

E-mail: ha cy@ igsnr. ac. cn.

原区^[13]。

因此, 本文试图在大气环流多变、地形地貌复杂的典型地区——云南省南部, 利用气象物理量——多年日气温和日降水计算其排列熵, 对气候系统的两大最重要要素——气温和降水——的复杂性进行度量, 并探讨其地理学成因以及研究意义。

2 资料与计算方法

2.1 研究区概述

研究区位于云南省南部, 政区范围包括云南省 30 个县级行政区, 介于 $98^{\circ}40'53'' \sim 106^{\circ}11'33''\text{E}$ 和 $22^{\circ}26'34'' \sim 24^{\circ}27'35''\text{N}$ 之间 (图 1), 总面积 101900km^2 。研究区西部, 隶属滇西南山地, 纵向排列着老别山 (南定河)、邦马山 (澜沧江)、无量山 (把边江-阿墨江)、哀牢山 (元江-红河) 等纵向高大山地和深切河谷。其中哀牢山是纵贯研究区中部的西北~东南向的巨大山脉, 主峰海拔超过 3100m , 而且与西南季风气流几成正交; 哀牢山以东则隔元江与滇东高原相邻, 地势变化幅度相对较小, 虽受到元江、南盘江及其支流的切割, 但起伏相对和缓^[14]。研究区同时受两个热带海洋 (孟加拉湾和南海) 水汽源的影响, 特别是西部位于青藏高原的东南侧, 是西南季风进入中国的必经之地^[15]。冬半年 (11 月至次年 4 月), 受到高原冬季风及南支西风的控制, 部分地区受东亚冬季风影响, 天气相对干冷; 夏半年 (5 月至 10 月), 受西南季风和东南季风影响, 天气相对湿热。即研究区有明显的干季和雨季更替^[16]。由于这些原因, 使得研究区的气候具有显著的特殊性, 无论是冬季还是夏季, 一些天气气候特征为全国独有。世界三大气象前沿问题 (低纬度问题、大地形作用问题和热带海洋问题) 同时在这片地区显著地表现出来^[17]。



图 1 研究区位置图

Fig 1 Location of the study area

2.2 排列熵计算原理

本文排列熵计算是基于嵌入维的算法, 其计算原理如下。从实用角度出发, 先选择嵌入维 $n=3, \dots, 7^{[10]}$ 。

下面以 7 个数值的时间序列为例: $x=(4, 7, 9, 10, 6, 11, 3)$

根据邻近的两个数值相对大小, 可以组织六对, 其中 4 对 $x_i < x_{i+1}$, 而 2 对 $x_i > x_{i+1}$ 的数据对, 用 01 表示, 而 $x_i > x_{i+1}$ 用 10 表示。定义 2 阶的排列熵为 01 和 10 排列的概率测定, 因此有:

$$H(2) = - (4/6) \lg(4/6) - (2/6) \lg(2/6) \approx 0.918 \quad (1)$$

这里 \lg 是以 2 为底, 因此 H 的单位是比特。下面考虑连续三个数值组成的数据对, (4, 7, 9) 和 (7, 9, 10) 用 012 的排列表示, 因为它们是增加序, (9, 10, 6) 和 (6, 11, 3) 为 201 的排列, 因为 $x_{i+2} < x_i < x_{i+1}$ 。而 (10, 6, 11) 为 102 的排列, 因为 x_{i+1}

$< x_t < x_{t+2}$ 。

因此, 3 阶的排列熵定义为

$$H(3) = -2(2/5) \lg(2/5) - (1/5) \lg(1/5) \approx 1.522. \quad (2)$$

由上可知, 这里定义: 考虑时间序列 $\{x_t/t=1, \dots, T\}$, 我们研究所有 n 阶 $n!$ 个排列 π , 这里 n 为视为不同数值的可能阶。对于每一个排列 π , 我们定义相对频数, 即概率 (# 表示个数)

$$p(\pi) = \frac{\# \{t/t \leq T-n, (x_{t+1}, \dots, x_{t+n}) \text{ has type } \pi\}}{T-n+1} \quad (3)$$

上式对于估计有限时间序列的频数 π 是足够的。为了准确估计 $p(\pi)$, 我们必须假设一个无限的时间序列 $\{x_1, x_2, \dots\}$, 对上式取极限 $T \rightarrow \infty$ 。当内在的随机过程符合一个非常弱的平稳条件时, 这个极限存在概率为 1, 即对于 $k \leq n$, $x_t < x_{t+k}$ 的概率应当不依赖于 t 。

基于上述原理, 大于等于 2 阶排列熵的定义为:

$$H(n) = -\sum p(\pi) \lg p(\pi) \quad (4)$$

这里求和为所有 n 阶 $n!$ 个的排列 π 。它包含在时间序列的 n 个连续值比对的信息。很显然, $0 \leq H(n) \leq \lg n!$ 这里通过增加或减少序列的值得到下界, 而上界对于一个完全的随机系统, 所有的 $n!$ 可能排列具有同一概率。当 $H(n) < \lg n!$ 时, 时间序列表征出一些动态的分类特征。实际上, 在一个混沌时间序列中, $H(n)$ 确实随着 n 增大呈现线性增加。因此, 定义 n 阶每个符号的排列熵是有用的, 这里要被 $n-1$ 去除, 因为比对开始于第二个值。

$$h_n = H(n)/(n-1) \quad (5)$$

2.3 排列熵的计算

由熵定义可知, 排列熵作为一个反映系统性质的参数, 是动态复杂性的量度。应用到气候系统气象要素复杂性度量中, 在某些熵值大的地区就表示要素(气温和降水)的不确定性大, 反之, 熵值小的地方, 就具有较高的确定性。从而可根据熵值的分布来了解要素不确定性的地理分布特征, 进而分析此要素空间格局成因。

本研究依据研究区地理分布相对较为均匀的 30 个气象台站, 自 1971~2000 年日气温和日降水共 30 年资料, 应用 MATLAB6.1 (MathWorks Inc. Copyright 1984~2001) 语言采用 Statsoft 公司出品的神经网络设计分析软件 (Statistica Neural Networks v4.0e), 分别计算出它们的排列熵(包括气温和降水)。数值大小上, 排列熵(气温)数值在 0.9526~1.0245 之间, 数值较大, 变动范围仅 0.0719; 排列熵(降水)数值在 0.7060~0.9210 之间, 数值较小, 但变动范围较大为 0.2150, 见表 1。

3 研究结果

3.1 熵值空间分布特征

表 1 中的熵值(包括气温和降水)表现在空间上, 分别如图 2 和 3。由图 2 可知, 排列熵(气温)值较大的两级(熵值大于 1)均分布于哀牢山以西地区, 数值较小的两级(熵值小于 1)主要分布哀牢山以东地区; 且哀牢山以东地区的熵值有自北而南减小的过程。即全区以哀牢山为界线, 以西地区气温体系复杂度高而以东地区复杂度低; 哀牢山以东地区则北部气温体系复杂度高而南部复杂度低。由图 3, 得出研究区日降水排列熵的空间递变呈现

表 1 研究区气象站排列熵值

Tab 1 Permutation entropies of different weather stations in the study area

气象站	位置		排列熵		气象站	位置		排列熵	
	东经 (°)	北纬 (°)	气温	降水		东经 (°)	北纬 (°)	气温	降水
镇康	24 06	98 96	1.0245	0.8701	元阳	23 16	102 75	0.9594	0.8759
永德	24 15	99 23	1.0022	0.8297	建水	23 61	102 83	0.9937	0.7315
沧源	23 15	99 26	1.0196	0.8402	个旧	23 38	103 15	0.9767	0.8151
耿马	23 55	99 40	1.0143	0.8303	金平	22 78	103 23	0.9810	0.8798
双江	23 46	99 80	1.0166	0.8019	蒙自	23 38	103 38	0.9818	0.7656
临沧	23 95	100 21	1.0106	0.8129	屏边	22 98	103 60	0.9526	0.9210
景谷	23 50	100 70	1.0149	0.8344	河口	22 50	103 95	0.9578	0.8702
镇沅	23 88	100 88	1.0122	0.8000	邱北	24 05	104 18	0.9891	0.8362
普洱	23 03	101 28	1.0080	0.8194	文山	23 38	104 25	0.9706	0.7881
墨江	23 43	101 71	1.0074	0.8050	砚山	23 61	104 33	0.9675	0.8046
新平	24 06	101 96	0.9935	0.7450	马关	23 02	104 41	0.9701	0.8900
元江	23 43	101 98	0.9847	0.7060	西畴	23 45	104 68	0.9584	0.8682
绿春	23 00	102 40	0.9792	0.8457	麻栗坡	23 13	104 70	0.9585	0.8450
红河	23 36	102 43	0.9550	0.7268	广南	24 06	105 06	0.9737	0.8137
石屏	23 70	102 48	0.9883	0.7494	富宁	23 65	105 63	0.9745	0.8237

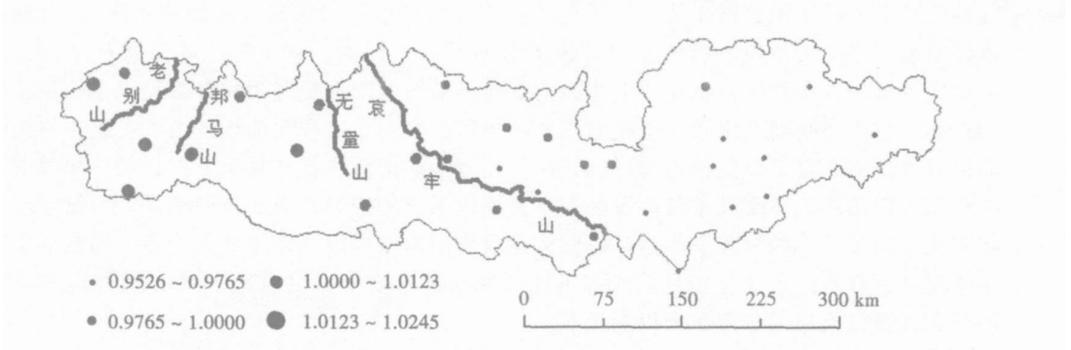


图 2 排列熵 (气温) 的空间分布

Fig 2 Spatial distribution of permutation entropies of daily temperature series

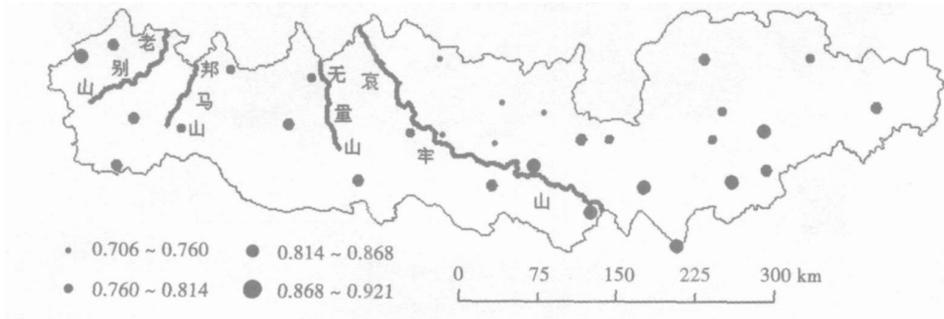


图 3 排列熵 (降水) 的空间分布

Fig 3 Spatial distribution of permutation entropies of daily precipitation series

出两个方向上的变化过程：整个研究区自西而东递减和哀牢山以东地区自西南向北和东减小，即全区整体规律是西部和中南部降水体系复杂度高，而东部和北部降水体系复杂度低。

3.2 格局解析

整个研究区的熵值分布规律，无论排列熵(气温)还是排列熵(降水)都是自西而东减小，它表明了影响气温和降水的大气环流系统的运动方向应该都是自西而东。正如前已论述，由于研究区所处的特殊地理位置，它受到的季风气流系统有其特殊性。冬季主要受到南支西风带的大陆干热气团和青藏高原干冷冬季风的影响，而夏季主要受湿热的南亚西南季风影响，三者均为偏西风，但水热性质迥异^[18, 19]。因此，西部由于其受影响较深而熵值大，东部则受影响程度弱而熵值小；而且由于纵向山地尤其是哀牢山对气流的强烈“阻隔”作用及对水汽的截留作用造成熵值在空间的递变梯度与哀牢山山地的位置有显著相关性，尤其是排列熵(气温)。

在熵值的区域分异上，哀牢山西、东地区对比显著。哀牢山以西地区，排列熵(气温)数值虽然都很大，但自西而东空间上的变化不突出，却有随山地位置自西而东增大减小、再一次增大减小的规律；排列熵(降水)也有类似变动过程。这种熵值空间分布规律显然与老别山、邦马山、无量山、哀牢山等纵向系列山地的分布密不可分，即排列熵与山地位置的相关性反映了高大山地对冷空气的截留效果。哀牢山以东地区，排列熵(气温)是自北而南递减，排列熵(降水)则是自北而南递增。这种特殊的熵值空间变化规律，其原因也与大气环流及大地形作用显著相关，特别是气势宏大的哀牢山对东来气流的“阻隔”，以及本地区所表现出的大面积平坦地势和谷狭沟深的元江-红河、富宁等地的强烈对比(图4)。冬季的哀牢山以东地区，不仅受上述的南支西风带的大陆干热气团和青藏高原干冷冬季风的控制，另外还受到北来强冷空气即东亚冬季风的影响，这样北部地区由于受此影响较强而显得气温体系复杂致使排列熵(气温)高，相对而言南部较低；夏季不仅受湿热深厚的西南季风气团影响，还受到来自南海的东亚夏季风系统的控制，故此排列熵(降水)值空间分布有从南向北减小的过程。当然，也同样是哀牢山山地的阻挡作用，无论冬季的东亚干冷冬季风还是夏季的东亚湿热夏季风都不能顺畅地进入哀牢山以西地区^[20]，因此这种熵值的分布规律仅仅局限于哀牢山以东地区。

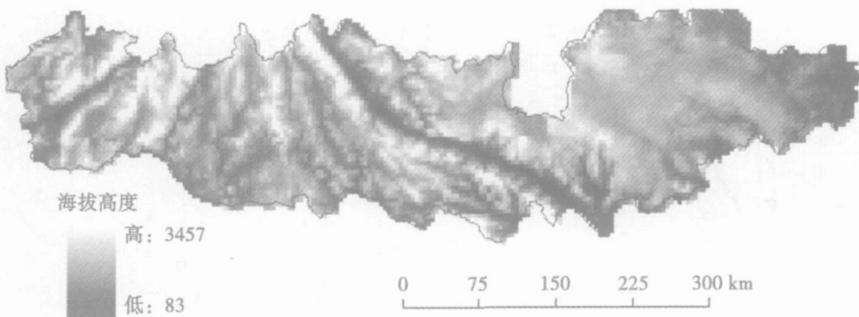


图 4 海拔高度空间格局

Fig 4 Spatial pattern of elevation

在两者的空间递变过程对比上，排列熵(气温)突出表现在数值自西而东有减小的趋势，尤其在哀牢山两侧突变较为显著；排列熵(降水)空间递变过程主要体现在哀牢山以东

地区, 自西南角向北有减小的过程, 而且以向西北方向减小幅度大于向东北方向的减小幅度。究其原因, 则是年内气温变化主要与年内气流路径的转换显著性有关, 由于整个研究区内年多受偏西风的南支西风、高原季风、西南季风控制, 尤其是哀牢山以西地区。因此, 排列熵(气温)数值递变过程有自西而东减小的趋势。而年内降水变化主要与水汽源在研究区的方位有关, 研究区大气降水水源地有两个: 一为孟加拉湾, 一为南海^[15]。其中前者的水汽从研究区西南角渐次扫过各个纵向山地, 由于它们对水汽的阻隔截留作用使得排列熵(降水)变化趋势不显著, 而到了哀牢山以东, 数值变化幅度明显加强, 特别是顺着哀牢山走向自东南向西北渐次减小, 充分体现了哀牢山大地形对来自南海的水汽截留作用的效果。

正是由于这种多气流系统和大地形作用的综合影响, 引起了研究区的多年日气温和日降水的排列熵呈现出了这种独特的时空分布格局。总之, 应用长时间序列的气候数据计算出的排列熵能够反映出区域气候气象要素的时空复杂性, 并能很好地解释当地气候的大气环流和局地地形的双重影响效果。

4 结语与讨论

(1) 排列熵(气温)值的空间分布格局反映了区域大气环流的复杂性以及大地形作用的显著性。熵值的自西而东的减小趋势及强度指示了研究区纵向山地对冷空气的阻隔效应; 哀牢山以东地区熵值由北向南的减小过程既反映了哀牢山大地形的作用, 也显现了研究区冬季受东亚季风的影响范围及程度。

(2) 排列熵(降水)值的空间分布格局不仅指示了水汽通道的方向, 而且反映了降水量空间递变强度。排列熵(降水)值自西而东的递减过程反映了纵向山地的地形效应以及水汽自西而东的来源方向; 研究区东部受南海水汽源的影响明显, 自南而北递变强度逐渐减弱, 多与滇东高原地势相对平坦有关。

(3) 基于长时间序列气象数据的排列熵能够反映区域气候系统气象要素的时空复杂度; 而且更为重要地是, 排列熵的空间分布格局得到了很好的地理诠释。

(4) 通过计算长时间序列的气候物理参数的排列熵, 成功地进行气候系统复杂度的度量和解释, 发现很多有趣问题需要探究。例如, 本研究区是一个气候系统相对比较复杂、地形效应显著的地区, 使得熵值在空间的递变过程明显, 但大气环流较为单一、地势较为平坦地区的熵值空间分布又会是一种怎样的情形? 还有, 本研究使用的熵值是依据1971~2000年30年数据而计算出来的, 但时间序列的长度显然影响熵值的大小, 这就存在一个合适的时间序列问题; 再者, 日气象数据显得原始数据臃肿、偶然性大, 由旬数据来替代又如何? 这些问题都是下一步工作值得商榷的。

参考文献:

- [1] 丑纪范. 大气科学中非线性与复杂性研究进展. 中国科学院院刊, 1997, (5): 325~ 329
- [2] 郭燕娟, 杨修群. 全球海气系统年际和年代际变化的时空特征分析. 气象科学, 2002, 22(2): 127~ 139
- [3] 栗珂. 气候系统态行为复杂性的探索. 气象科学, 2004, 24(1): 95~ 103
- [4] 俞佚名, 田英. 熵分析在气象上的应用. 贵州气象, 1997, 21(5): 42~ 44
- [5] 乔生炳. “熵”概念的拓展和应用. 现代物理知识, 2001, (3): 38~ 39
- [6] 谢惠民. 复杂性与动力系统. 上海: 上海科技教育出版社, 1994 186~ 209
- [7] 郝柏林. 复杂性的刻画与“复杂性科学”. 物理, 2001, 51(3): 1~ 11

- [8] 汤懋苍. 理论气候学概论. 北京: 气象出版社, 1989
- [9] 张学文, 马力. 熵气象学. 北京: 气象出版社, 1992 203~ 204
- [10] Bandt C, Pompe B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 2002, 88(174102): 1~ 4
- [11] Schwartz T U, Walczak R, Biel G. Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor subunit. *Protein Science*, 2004, 13: 2814~ 2818
- [12] Pard E, Tvar F J R. MAXENPER: a program for maximum entropy spectral estimation with assessment of statistical significance by the permutation test. *Computers & Geosciences*, 2005, 31: 555~ 567
- [13] 毛政旦. 论山地气候带和气候型. *地理研究*, 1989, 8(3): 21~ 29
- [14] 王声跃. 云南地理. 昆明: 云南民族出版社, 2002 46~ 48
- [15] 陈隆勋, 朱乾根, 罗会邦, 等. 东亚季风. 北京: 气象出版社, 1991 49~ 60
- [16] 强学民, 琚建华, 张浩瀚. 云南夏季风演变诊断分析. *云南大学学报(自然科学版)*, 1998, 20(1): 75~ 79
- [17] 秦剑, 琚建华, 解明恩. 低纬高原天气气候. 北京: 气象出版社, 1997 1~ 2
- [18] 郑度. 青藏高原形成环境与发展. 石家庄: 河北科学技术出版社, 2003 124
- [19] 马林, 李锡福, 张青梅, 等. 青藏高原东部牧区冬季雪灾天气的形成及其预报. *高原气象*, 2001, 20(3): 325~ 331
- [20] 张克映, 马友鑫, 李佑荣, 等. 哀牢山过山气流的气候效应. *地理研究*, 1992, 11(3): 65~ 70

Measurement of climate complexity using permutation entropy

HAO Cheng-yuan^{1,3}, WU Sha-hong¹, LI Shuang-cheng²

(1 Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2 College of Environmental Sciences, Peking University; The Key Laboratory for Earth Surface Processes, Ministry of Education, Beijing 100871, China;

3 Graduate School of the Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Permutation entropy is a complex parameter for time series based on comparison of neighboring values, which can measure spati-temporal complexity of weather factors, and more importantly, whose spatial pattern can get geographical explanation. Based on the climate physics elements for time series including daily temperature and daily precipitation from 1971 to 2000, this article takes southwest of Yunnan Province as a case study. The characteristic of varied atmospheric circulations and complex landforms, measure the climate systematic complexity and probe into their geographic causes. It is concluded that it is simple and effectual to measure the systematic complexity of climate by permutation entropy. On the one hand, the spatial pattern of permutation entropy (based on daily temperature) can reflect the complexity of regional atmospheric circulation and the significance of Aila Mountain barrier function. On the other hand, the spatial pattern of permutation entropy (based on daily precipitation) can not only indicate the direction of the water vapour sources, but also reflect regional difference of spatial precipitation variation intensity. In a word, there is a good geographical explanation for both atmospheric circulations and localized topography to climate elements.

Key words: permutation entropy; climate complexity; atmosphere circulation; topographic factor