

东江流域降水空间分布模式识别^{*}

刘丙军^{1,2}, 陈晓宏²

- (1. 水资源与水电工程科学国家重点实验室, 湖北 武汉 430072
2. 中山大学地理科学与规划学院水资源与环境系, 广东 广州 510275)

摘 要: 一定区域范围内, 近地层气象因子往往是在统一的构造背景控制下形成的, 降水、气温、蒸发等气象因子具有自相似性特点。以广东省东江流域内 29 个雨量观测站点 1956–2005 年月平均降水资料为样本, 运用信息熵原理中有向信息传输指数研究了雨量观测站点之间的信息传输特征, 并结合聚类分析方法界定各观测站的所属关系, 提出了基于信息熵原理的流域降水空间分布模式识别方法, 从而为流域水文要素空间分布模式识别和水资源利用方案提供重要理论和实践依据。

关键词: 降水; 模式识别; 有向信息传输指数; 东江流域

中图分类号: P333 **文献标识码:** A **文章编号:** 0529-6579 (2009) 05-0148-06

Mode Identification of Spatial Rainfall Distribution in the Dongjiang Basin

LIU Bingjun^{1,2}, CHEN Xiaohong²

- (1. State Key Laboratory of Water Resources and Hydropower Engineering Science,
Wuhan 430072, China; 2. Department of Water Resources and Environment,
Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Since surface atmospheric meteorological factors are formed against the same tectonic background in a certain area, self-similarity is shared among meteorological factors such as rainfall, temperature, evaporation. The average rainfalls were collected from 29 rainfall observation stations in the Dongjiang Basin of Guangdong Province from 1956 to 2005 for studying the features of information transfer among different stations using transfer index of directional information entropy theory. Cluster analysis is applied to define the relationship between stations. Based on information entropy theory, a mode identification method of spatial rainfall distribution was proposed for providing theoretical and practical basis for mode identification of spatial distribution of hydrological factors and water resources utilization plans.

Key words: rainfall; mode identification; transfer index of directional information; Dongjiang Basin

由于受气候变异和剧烈人类活动的双重影响, 流域水文要素时空变异十分突出^[1], 科学构建流域降水时空分布模式识别方法对于深刻理解流域尺度的水文循环和水资源管理十分重要^[2]。刘春蓁等^[3]采用 Mann-Kendall 方法对近 50 年海河流域山区 20 个子流域的径流及降水的变化趋势和径流对

气候变化的敏感性进行分析, 提出了一个半定量分析的研究思路和方法; 陈晓宏等^[4]以广东省内 186 个站点的年降雨量为样本, 运用聚类分析理论方法识别了该地区的降雨空间分布特征; Hall^[5]和张静怡^[6]等应用自组织特征映射人工神经网络方法和模糊聚类法进行了水文要素分区识别; 陈英等^[7]

* 收稿日期: 2008-11-05

基金项目: 水资源与水电工程科学国家重点实验室开放基金资助项目; 国家自然科学基金资助项目 (50909106); 中国水利水电科学研究院开放研究基金资助项目

作者简介: 刘丙军 (1976 年生), 男, 博士, 讲师; E-mail: liubj@mail.sysu.edu.cn

利用分级随机典型分布法对降雨时空分布特征进行了探讨等等。上述研究均属统计学模式识别范畴,只要使用的水文特征适当和充分、样本足够全面和具有代表性,就可以得到更为客观的分类结果。其缺点是如果样本序列不足,或使用了不适当的典型模式,或者类别之间存在交集,模式分类系统将无法区分类别。

根据局部相似性理论^[8],一定空间尺度范围内,气象场是一个连续介质的热力-动力系统,近地层气象因子往往是在统一的构造背景控制下形成的,降水、气温、蒸发等气象因子具有自相似性特点。基于此,将信息熵理论引入水文要素时空分布模式识别中,通过计算观测站点之间的信息关联程度来测度各观测站的所属关系,从而为水文要素模式识别理论提供一种新的研究途径,相关研究已经运用于水文气象要素时空变异性测度等方面^[9]。

1 信息熵基本原理

1.1 观测站点信息网络特征分析

分析各水文气象观测站的特点,可将水文气象观测网络系统看成是一个以气象场为载体的信号通讯网络,各观测站均可以看成是水文气象因子如降水等动态变化信号的接受者,各观测站之间存在信息交流或信息传递,这种“通讯网络”具备信号传输的基本特点:

(1) 信号的可传递性。一定区域尺度范围内,气象场中任一观测站降水量的突变将会反映在其它观测点中,尽管这种反映的程度和时间可能出现差异,但仍能表现出气象场的信息传递的特征。

(2) 信号之间的相关性和差异性。在一定空间尺度范围内,降水量具有稳定性、连续性的特点,临近或相近的气象观测站之间必然表现出更加密切的联系,彼此之间信息的转换和传递更加密切,而当两观测站相距较远时,信息的转换和传递将出现较大的差异,而且差异随距离加大而增加。

(3) 信号传递的衰减性。在同一气象场内,受地形、云层等因素影响,蒸发等水文气象因子中心的强度远大于边缘地带的强度,可见从源到汇水文气象因子的变动或变化幅度呈现出由强到弱的特点,这种特点可以理解为一水文气象因子信号传递过程中出现的衰减性。

1.2 信息熵基本原理

1.2.1 基本概念 信息熵是信号通讯理论中用来评价随机信号中所含信息量大小的基本概念之一,它对随机信号中所含信息量的度量依据随机信号出

现的概率大小。Shannon 将熵的概念引入信息论,把熵作为一个随机变量的不确定性或信息量的度量,认为对于取有限个值的随机变量 X , $P\{X = x_i\} = p(x_i) (i = 1, 2, \dots, n)$, $p(x_i) \geq 0$,

$\sum_{i=1}^n p(x_i) = 1$, 则信息熵为:

$$H(X) = H(p(x_1), p(x_2), \dots, p(x_n)) \\ = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

式中, $p(x_i)$ 为发生事件 x_i 的概率, n 为可能发生的事件(状态)总数。对于完全确定的变量 X , $H(X) = 0$; 对于随机变量 X , $H(X) > 0$ (非负性), 且 $H(X)$ 的值随状态数 n 的增加而增大(递增性)。

设 Y 为另一离散型随机变量, 称 $H(X, Y)$ 为 X 和 Y 的联合熵, 称 $H(X|Y)$ 为在 Y 已知的条件下 X 的条件熵, 即:

$$H(X|Y) = - \sum_{i,j} p(y_j) p(x_i|y_j) \log p(x_i|y_j) \quad (2)$$

式中, $p(y_j)$ 为单独事件 y_j 的概率, $p(x_i|y_j)$ 为在发生事件 y_j 的条件下事件 x_i 发生的条件概率。定义 X 和 Y 的互信息 $I(X, Y)$ 为:

$$I(X, Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (3)$$

式(3)证明了互易性, 引入 X 和 Y 的联合概率信息 $H(X, Y)$, 即:

$$H(X, Y) = - \sum_{i,j} p(x_i, y_j) \log p(x_i, y_j) \quad (4)$$

式中, $p(x_i, y_j)$ 为 x_i 与 y_j 的联合概率, 即事件 x_i 与事件 y_j 同时发生的概率, 这样互信息 $I(X, Y)$ 为:

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (5)$$

1.2.2 信息传输 信息传输是指信息通过传输通道被系统所接收和利用的一个过程。对于各个测站而言, 信息传输是指一个测站对临近测站的“影响”, 文献[10]定义了判断这种能力的指标, 并称之为有向信息传输指数(Directional Information Transfer Index DIT), 即观测站 X 到测站 Y 的 DIT 为:

$$DIT_{XY} = I(X, Y)/H(X) \quad (6)$$

式(6)描述了站 X 对站 Y 传输信息的一种能力, 但是因 $I(X, Y) = I(Y, X)$, 致使熵较小的站点其 DIT 大, 而熵较大的站点其 DIT 小, 这与实际情况不相符合。文献[11]对 DIT 进行了一定的调整, 调整后的有向传输指数为 DITI:

$$DITI_{XY} = I(X, Y)/H(Y) = \frac{H(Y) - H(Y|X)}{H(Y)} \quad (7)$$

$$DITI_{YX} = I(Y, X)/H(X) = \frac{H(X) - H(X|Y)}{H(X)} \quad (8)$$

在式(7) $DITI_{XY}$ 中, 称 X 为基点, 称 Y 为辅点。

从上述分析可知, DITI 能够反映系统的两个特征: 信息传输能力和站对之间的相关联程度。具体算法参见文献 [12]。

在某一空间范围内, 有 DITI 是一个站点对周围区域的“影响”能力。若两观测站点 $DITI_{XY}$ 与 $DITI_{YX}$ 均较大, 则表明两站点间信息的传递或交流较强, 其关联程度较大, 两个观测站点处于同一区间; 若两站点之间没有一个 DITI 有较高的数值, 则它们是相互独立的, 故应处于不同区间。根据 DITI 的这一特性, 利用聚类分析方法研究 DITI 分布矩阵, 通过分析不同阈值情况下聚类分析结果, 确定降水量的空间相似性区间, 从而揭示流域降水量的空间分布规律。

1.2.3 聚类分析方法 为了建立模糊相似矩阵 R , 引入相似系数 r_{ij} , 表示 2 个样本 x_i 与 x_j 之间的相似程度, R 为:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (9)$$

式中 r_{ij} 可以用数量积分法、夹角余弦法、距离法、主观评分法等确定。本文各站点之间的 DITI 表示为:

$$r_{ij} = r_{ji} = (DITI_{x_j x_i} + DITI_{x_i x_j})/2 \quad (10)$$

上述标定过程得到的模糊关系 R 一般只具有自反性和对称性, 不满足传递性, 是模糊相似矩阵, 必需对 R 进行改造。本文运用传递闭包方法对模糊集进行聚类分析, 在得到模糊等价关系 $t(R)$ 后, 就可以在适当的阈值 λ 上截取 $t(R)$, 得到普通等价关系 $[t(R)]_\lambda$, 从而确定降水量空间分布分类结果。

2 降水空间分布模式识别分析

2.1 降水序列时间熵空间分布特征分析

运用信息熵原理, 根据式 (9) 和式 (1), 求出了东江流域 29 个雨量站降水量的时间熵, 得出图 1 各个观测站的熵值与其所处经纬度的关系。由图 1 可见, 降水量的时间熵在经度和纬度方向上分布比较杂乱, 表明东江流域降水量与所处地域的经度、纬度相关性不强 (相关系数仅为 0.25 和 0.22)。根据熵原理可知, 各观测站点降水量时间熵值反映样本时间分布的不均匀程度, 各站点时间熵值不一致, 表明在同时刻流域降水量在空间分布比较分散。东江流域降水时间序列变差系数 C_v 等值线主要以某些分散雨量站为中心呈离散分布, 在经度方向或纬度方向上无明显规律 (图 2), 即降

水量在经度或纬度方向上不存在显著相关, 此结论与降水量的时间熵在经度或纬度方向分布不均匀的结论一致。

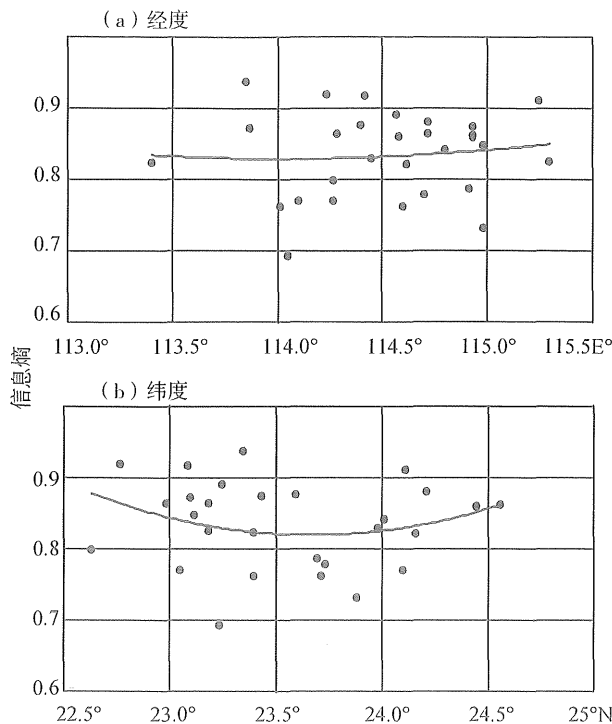


图 1 单站时间熵与经度 (a)、纬度 (b) 的关系

Fig. 1 The relationship between ET_0 entropy and latitude (a) and longitude (b)

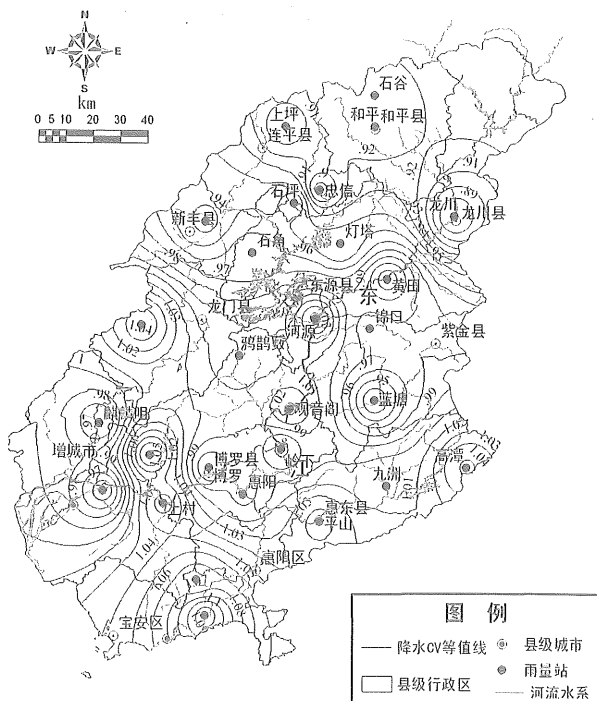


图 2 东江流域降水变差系数 C_v 值空间分布图

Fig. 2 Spatial distribution of rainfall C_v value in the Dongjiang Basin

2.2 有向信息传输指数特征分析

分析观测站有向信息传输指数与距离之间的相互关系,以博罗雨量观测站($114^{\circ}17'E$, $23^{\circ}11'N$)基点,从基点出发,分别沿东-西、南-北方向上选取若干个辅点,根据距离 d 和有向信息传输指数DITI构造而成的关系曲线,见图3所示。图3中DITI随着离基点的距离增加而减少,表明两观测站距离越远,则其信息传输能力越弱。博罗观测站点的这种特征在区域内普遍存在,而且图形的函数性质也相似,具有一般规律性。运用回归分析原理对图中曲线进行拟合,得出各拟合曲线的解析表达式,式中 d 表示辅助点与基点之间距离, R 为相关系数。

南北方向: $DITI = 0.827 - 0.002d - 4.911 \times 10^{-6}d^2$
 $R = 0.962$

东西方向: $DITI = 0.842 - 0.002d + 6.519 \times 10^{-6}d^2$
 $R = 0.825$

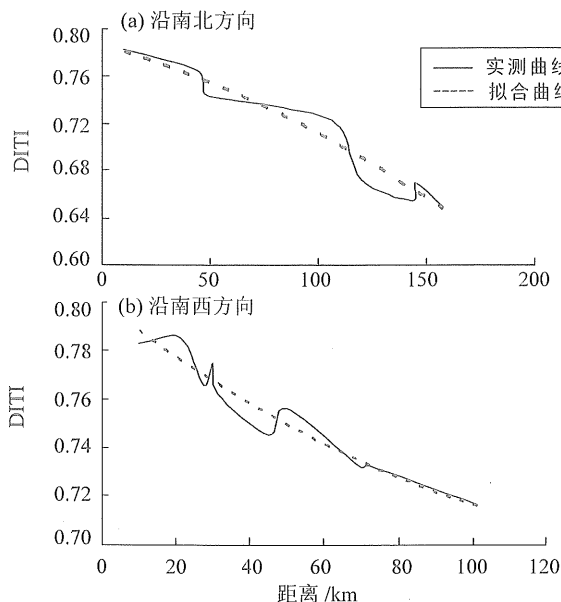


图3 博罗站有向信息传输指数曲线拟合图

Fig. 3 Curve-fitting chart of transfer index of directional information

2.3 聚类分析

运用模糊聚类分析方法对由有向信息传输指数构造得出的模糊相似矩阵 R 进行分析,得出 R 的传递闭包 $t(R)$ 。当阈值 λ 依次在区间 $[0, 1]$ 范围内变动时,不同的 λ 值对应着不同的区域划分。由聚类分析的原理可知,分在同一区域内的观测点降水特征具有较大的相似性。图4为 $\lambda \geq 0.88$ 情况下东江流域29雨量观测站降水量的聚类示意图,

得出该流域降水空间分布具有如下结构特征:

(1) 受地形条件影响,东江流域降水在空间分布上存在相似性。由图4可知,当 $\lambda = 0.88$ 时,各观测站点可聚类划分成6个区域,其中除东江流域边界站点外,大部分站点集中划分在1、2类中。第1类区域中,各观测站点大部分所属东江流域灯塔盆地,盆地内各站点地形、气象因子比较相近,降水特征比较相近;第2类区域所属东江流域东部,该地区受粤东沿海海洋暖湿气流和莲花山脉的强烈抬升共同作用影响,形成降水的水汽、热力、动力条件均比较充分,降水特征与其他区域相比明显不同,具有降水强度大、量多等特点。将聚类分析结果与实际降水特征比较,特别是比较各站点之间的多年平均降雨量(图5)和年降雨量年际变化的变差系数 C_v 值在空间上的分布(图2),分类结果比较符合实际降水情况。

(2) 与常规模式识别方法比较,运用信息论中有向信息传输指数方法进行降水空间分布特征模式识别具有较大优越性。有向信息传输指数以各站点实际降水量为研究对象,综合流域内包括地形、气候等物理因素在内的绝大部分信息,在判断各个测站之间的降雨特性时不仅分析了水量上特征(多年平均降雨量),同时重点考虑了降水量时间序列上的变化特征(C_v 值),能更好反映降雨空间分布的实际情况。

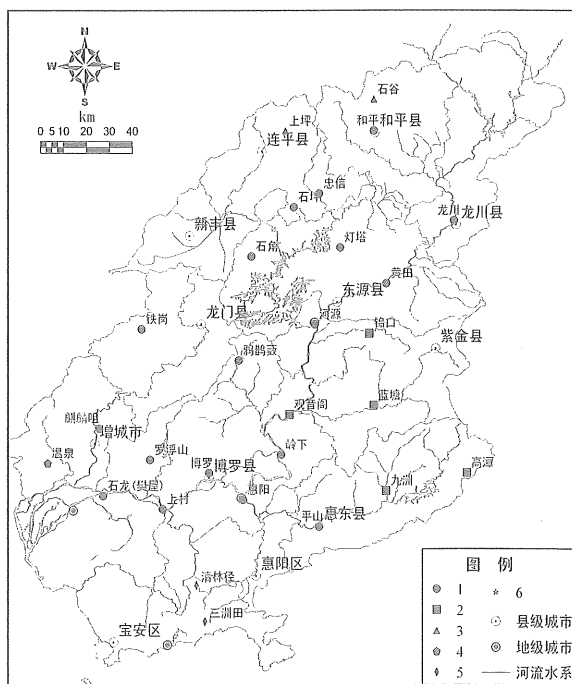


图4 东江流域降水聚类结果图

Fig. 4 Cluster analysis results of rainfall

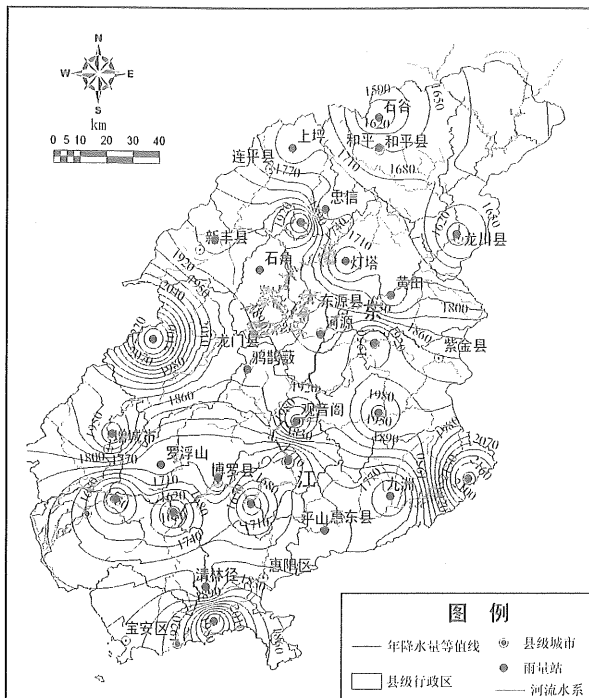


图5 东江流域年平均降水量等值线图

Fig. 5 Isolines map of the average rainfall

3 结论

利用信息熵理论中有向信息传输指数的概念,根据模糊聚类分析方法对东江流域29个雨量站50年月平均降水空间序列进行聚类分析。研究表明,由于降水系统内各站点存在信息传输,且各站点之间的有向信息传输指数在经度或纬度方向传输不一,可以根据站点间的有向信息传输指数大小判断各站点之间的关联程度,并结合模糊聚类分析方法界定各气象观测站的所属关系,从而研究区域降雨空间分布模式特征,从而为无资料地区水资源的空间分布模式和水资源利用方案重点理论依据。

参考文献:

- [1] VöRöSMARTY C, PAMELA G, JOSEPH S, et al. Global water resources: Vulnerability from climate change and population growth[J]. Science, 2000(288): 284 - 288.
- [2] LAKHTAKIA M N, YARNAL B, et al. Simulating the river basin response to atmospheric forcing by linking a mesoscale meteorological model and hydrologic model system[J]. Journal of Hydrology, 1999, 218: 72 - 91.
- [3] 刘春蓁, 刘志雨, 谢正辉. 近50年海河流域径流的变化趋势研究[J]. 应用气象学报, 2004, 15(4): 385 - 393.
LIU Chunzhen, LIU Zhiyu, XIE Zhenghui. Study of trends in runoff for Haihe river basin in recent 50 years. Journal of Applied Meteorological Science, 2004, 15(4): 385 - 393.
- [4] 陈晓宏, 刘德地, 王兆礼. 降雨空间分布模式识别[J]. 水利学报, 2006, 37(6): 711 - 716.
CHEN Xiaohong, LIU Dedi, WANG Zhaoli. Recognition method for spatial distribution pattern of precipitation. Journal of Hydraulic Engineering, 2006, 37(6): 711 - 716.
- [5] HALL M J, MINNS A W. The classification of hydrologically homogeneous regions[J]. Hydrol Sci J, 1999, 44(5): 693 - 704.
- [6] ZHANG Jingyi, HALL M J. Regional flood frequency analysis for the Gan-Ming River basin in China[J]. J Hydrol, 2004, 296: 98 - 117.
- [7] 陈英, 刘新仁. 淮河流域降水蒸发随机模型研究[J]. 河海大学学报, 1996, 24(3): 43 - 48.
CHEN Ying, LIU Xinren. Study on random model of precipitation and evapotranspiration in Huaihe River Basin [J]. Journal of Hohai University, 1996, 24(3): 43 - 48.
- [8] 胡隐樵, 张强. 论大气边界层局部相似性[J]. 大气科学, 1993, 17(1): 10 - 20.
HU Yinqiao, ZHANG Qiang. On local similarity of the atmospheric boundary layer[J]. Scientia Atmospheric Science, 1993, 17(1): 10 - 20.
- [9] 夏军, 欧春平, HUANG G H, 等. 基于GIS和差异信息测度的海河流域水文气象要素时空变异性分析[J]. 自然资源学报, 2007, 22(3): 409 - 414.
XIA Jun, OU Chunping, HUANG G H. The analysis of haihe river basin hydrometeorological spatio-temporal variability based on GIS and information difference measure [J]. Journal of Natural Resources, 2007, 22(3): 409 - 414.
- [10] YANG Y, BURN D H. An entropy approach to data collection network design[J]. J of Hydrology, 1994, 157: 307 - 324.
- [11] 张继国, 刘新仁. 降水时空分布不均匀性的信息熵分析—(I) 基本概念与数据分析[J]. 水科学进展, 2000, 11(2): 133 - 137.
ZHANG Jiguo, LIU Xinren. Information entropy analysis on nonuniformity of precipitation distribution in time-space I [J]. Advance in Water Science, 2000, 11(2): 133 - 137.
- [12] 杨志安. 用等间距分格子法集算互信息函数确定延迟时间[J]. 计算物理, 1995, 12(4): 442 - 448.
YANG Zhian. Determination of delay time by calculating mutual information with equally distant space cells[J]. Chinese Journal of Computational Physics, 1995, 12(4): 442 - 448.