

## 基于样本熵的降雨和径流时间序列突变检验

薛联青<sup>1,2,3</sup>, 刘远洪<sup>1</sup>, 张梦泽<sup>1</sup>, 王思琪<sup>1</sup>, 李 军<sup>4</sup>

(1. 河海大学 水文水资源学院, 江苏 南京 210098; 2. 石河子大学 水利建筑工程学院,  
新疆 石河子 832003; 3. 河海大学 文天学院, 安徽 马鞍山 243031;  
4. 泰安市水文局, 山东 泰安 271000)

**摘 要:**传统的突变检验方法 Mann-Kendall 法主要基于线性和概率统计理论, 对于高度复杂和非线性的水文时间序列检测能力不足。采用基于非线性动力学参数的样本熵方法, 结合滑动技术和滑动移除技术, 利用湘江流域 1961~2009 年多个气象站点日降雨资料和控制站湘潭站日径流资料, 对降雨和径流时间序列进行突变检验。结果表明: 日尺度降雨和径流时间序列样本熵值最小, 复杂度最低, 可预测性最高; 径流分别在 1966 年、1983 年、1992 年以及 2002 年发生突变, 降雨在 1966 年、1987 年以及 2002 年发生突变; 滑动样本熵与滑动移除样本熵相结合不仅能有效地识别突变, 更能发现突变前后时间序列的动力学变化特征。

**关键词:**工程水文学; 时间序列; 突变检验; Mann-Kendall 法; 样本熵; 复杂度; 湘江流域

**中图分类号:** P333; TV211.1

**文献标志码:** A

## Mutation Test on Rainfall and Runoff Time Series Based on Sample Entropy

XUE Lian-qing<sup>1,2,3</sup>, LIU Yuan-hong<sup>1</sup>, ZHANG Meng-ze<sup>1</sup>, WANG Si-qi<sup>1</sup>, LI Jun<sup>4</sup>

(1. School of Hydrology and Water Resources, Hohai University, Nanjing 210098, Jiangsu, China; 2. School of Water and Architectural Engineering, Shihezi University, Shihezi 832003, Xinjiang, China;  
3. School of Wentian, Hohai University, Ma'anshan 243031, Anhui, China;  
4. Tai'an Bureau of Hydrology, Tai'an 271000, Shandong, China)

**Abstract:** The traditional mutation test method including Mann-Kendall method, which is mainly based on the theories of linear, probability and statistic, is insufficient for testing the highly complex and nonlinear hydrological time series. Combined with moving and moving cut data technologies, sample entropy based on nonlinear dynamic parameter was applied for the mutation test on rainfall and runoff time series in Xiangjiang river basin. The test data included daily rainfall in several meteorological stations and daily runoff in Xiangtan control station of Xiangjiang river basin from 1961 to 2009. The results show that the sample entropy of daily rainfall and runoff time series is smallest, the complexity is lowest, and the predictability is highest; the mutations of runoff occur in 1966, 1983, 1992 and 2002, the mutations of rainfall occur in 1966, 1987 and 2002; the combination of moving sample entropy and moving cut data sample entropy can not only recognize the mutation effectively, but also discover the characteristics of dynamic variation before and after the mutation of time series.

**Key words:** engineering hydrology; time series; mutation test; Mann-Kendall method; sample

收稿日期: 2014-12-31

基金项目: 国家自然科学基金项目(41371052, U1203282, 51269026); 江苏省“青蓝工程”项目; 兵团科技创新团队项目

作者简介: 薛联青(1973-), 女, 新疆石河子人, 河海大学教授, 博士研究生导师, 工学博士, E-mail: lqxue@hhu.edu.cn.

entropy; complexity; Xiangjiang river basin

## 0 引言

由于气象条件的变化和人类活动的影响,流域下垫面发生了变化,水文系统循环时空范围与尺度发生了改变,造成了水文时间序列的时空变异<sup>[1-3]</sup>,而降雨和径流作为水文系统的重要组成部分,其长期变化趋势及突变常与自然灾害相伴生,带来了一系列诸如水文极值事件、洪涝灾害、严重干旱频繁发生等新问题,破坏了水文时间序列的一致性,给人类生产和生活带来不利影响。因此,研究变化环境下的降雨和径流时间序列变异规律,对于研究水资源的演变规律以及治理与合理开发利用水资源具有重要的现实意义。水文时间序列具有高度复杂性、非稳定性、非线性等特征,而传统的突变检验方法主要从线性和概率统计方面进行研究<sup>[4-5]</sup>,不能有效识别非线性的时间序列特征和变异前后时间序列的状态趋势,检测能力也不足。随着非线性理论的发展,20世纪90年代初,Pincus首次提出用近似熵<sup>[6]</sup>计算时间序列复杂度,并在各领域得到广泛应用<sup>[7-10]</sup>。2000年,Richman等对近似熵进行修正,提出样本熵(Sample Entropy,简称 SampEn),旨在降低近似熵的误差,是一种与近似熵类似但精度更好的方法<sup>[11]</sup>;该方法被国内外学者广泛应用于降雨和径流时间序列的复杂度分析<sup>[12-17]</sup>。2011年,何文平等提出滑动移除近似熵进行突变检测,其结果优于滑动近似熵<sup>[18]</sup>。杜鹃等采用 Mann-Kendall 法(简称 M-K 法)对湘江流域主要水文站的径流和水位进行了趋势分析和突变检验,并从气候变化和人类活动两方面分析其造成洪水危险性增大的原因<sup>[19]</sup>。本文以湘江流域降雨和径流资料为研究对象,分别将滑动样本熵(Moving Sample Entropy,简称为 M-SampEn)和滑动移除样本熵(Moving Cut Data Sample Entropy,简称为 MC-SampEn)应用到降雨和径流时间序列的突变检验中,分析降雨和径流时间序列的复杂性,揭示降雨和径流时间序列的动力学特征。

## 1 研究方法

### 1.1 样本熵

样本熵是 Richman 等提出的一种新的时间序列复杂性测度方法,有别于近似熵不计入自身匹配的统计量,是对近似熵的改进。它是条件概率严格

的自然对数,可以用  $\text{SampEn}(m, r, N)$  来表示。 $N$  为样本时间序列长度; $m$  为维数,一般取 2; $r$  为相似容限,一般取  $(0.1 \sim 0.25)D$ ,  $D$  为时间序列标准差。样本熵表示非线性动力学系统产生新模式概率的大小,主要用来定量刻画系统的规则度及复杂度。样本熵值越低,时间序列自我相似性越高,产生新模式的概率越低,时间序列越简单;反之,样本熵值越高,时间序列自我相似性越低,产生新模式的概率越高,时间序列越复杂。对于长度为  $N$  的时间序列  $x(i)$ ,  $i=1, 2, \dots, n$ , 其算法如下:

(1) 将时间序列按照序号组成一组  $m$  维矢量  $\mathbf{X}(i) = (x(i), x(i+1), \dots, x(i+m-1))$ ,  $i=1, 2, \dots, N-m+1$ 。

(2) 定义  $\mathbf{X}(i)$  与  $\mathbf{X}(j)$  间的距离  $d(\mathbf{X}(i), \mathbf{X}(j))$  为两者对应元素中差值最大的一个,即

$$d(\mathbf{X}(i), \mathbf{X}(j)) = \max(|\mathbf{X}(i+k) - \mathbf{X}(j+k)|) \\ k=0, 1, 2, \dots, m-1 \quad (1)$$

对于每一个  $i$  值计算  $\mathbf{X}(i)$  与其余矢量  $\mathbf{X}(j)$  ( $j=1, 2, \dots, N-m+1$ , 且  $j \neq i$ ) 之间的  $d(\mathbf{X}(i), \mathbf{X}(j))$ 。

(3) 给定  $r$ , 对每一个  $i$  值统计  $d(\mathbf{X}(i), \mathbf{X}(j)) < r$  的个数( $k$ ), 然后计算其与距离总数  $N-m$  的比值(称为模版匹配数), 记作  $C_i^m(r)$ , 即

$$C_i^m(r) = k / (N-m) \quad i=1, 2, \dots, N-m+1 \quad (2)$$

(4) 对  $C_i^m(r)$  求平均值  $C^m(r)$ , 即

$$C^m(r) = \sum_{i=1}^{N-m+1} C_i^m(r) / (N-m+1) \quad (3)$$

(5) 把维数加 1, 变成  $m+1$ , 重复步骤(1)~(4), 得到  $C^{m+1}(r)$ 。

(6) 所求的样本熵为

$$\text{SampEn}(m, r, N) = -\ln[C^{m+1}(r)/C^m(r)] \quad (4)$$

### 1.2 滑动移除样本熵

滑动移除样本熵是在样本熵的基础上通过滑动移除技术形成的一种突变检测方法。其具体计算过程为:①选择滑动移除数据的窗口尺度  $S$ ;②从待分析时间序列的第  $i$  个数据开始连续移除  $S$  个数据, 再将剩余  $N-S$  个数据直接连在一起得到一个新的时间序列, 其中  $i=1, 2, \dots, N-S+1$ ;③利用样本熵计算新时间序列的样本熵值;④保持移除数据的窗口尺度不变, 以步长  $L$  逐步移动窗口, 重复第②、③步操作, 直到原时间序列结束为止;⑤通过第①~④步操作可得到一个随着窗口移动的样本熵时间序

列;⑥基于不同动力学性质的数据复杂性不同,而具有相同动力学性质的数据复杂性差异不大的特点,结合第⑤步得到的样本熵时间序列判断突变点或突变区间。

## 2 研究区概况

湘江是湖南省最长、流量最大、流域面积最广的河流,流经湖南省永州、衡阳、株洲、湘潭、长沙等市区,在湘阴县濠河口分两支注入洞庭湖,为洞庭湖水系最大支流。其多年平均流入洞庭湖流量占有所有河流多年平均入洞庭湖总流量的 22.4%,流量变化对洞庭湖蓄水有着重要影响。其干流全长 856 km,流域面积为 94 660 km<sup>2</sup>,在湖南省境内流域面积为 85 383 km<sup>2</sup>。湘江流域属亚热带季风湿润气候,河流水量大多来自降雨,多年平均降水量约为 1 436 mm,河流流量受降雨影响很大;由于受季风气候和地貌条件影响,该流域呈现出河川径流年际和年内变化较大、地区分布不均匀以及气候类型多样等特点。

本文采用湘江下游控制站湘潭站 1961~2009 年逐日径流资料以及流域内道县、永州、常宁、郴州、衡阳、南岳、双峰、株洲等气象站点 1961~2009 年逐日降雨资料;径流数据来源于湖南省水文水资源勘测局和 2001~2011 年《长江泥沙公报》;气象数据来自于中国气象科学数据共享服务网(<http://cdc.nmic.cn/>);运用泰森多边形法计算面平均雨量,构成湘江流域 1961~2009 年逐日面平均雨量时间序列,其年际变化见图 1。

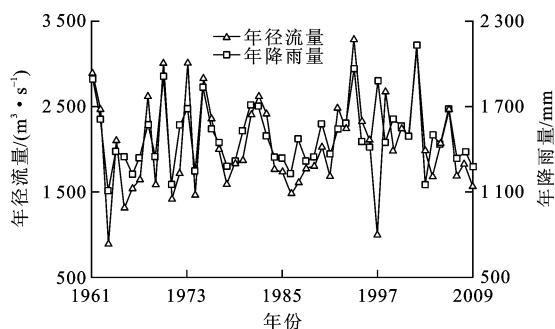


图1 湘江流域 1961~2009 年径流和降雨量年际变化  
Fig.1 Annual Variation of Runoff and Rainfall from 1961 to 2009 in Xiangjiang River Basin

## 3 结果分析

### 3.1 M-K 法突变检验

利用传统的突变检验方法 M-K 法对湘潭站 1961~2009 年年径流和湘江流域 1961~2009 年年降雨时间序列进行突变检验(图 2)。从图 2 可以看

出,年降雨和年径流时间序列的 M-K 法检验结果均未超过 95% 的显著性水平,呈现不显著的增加趋势。年径流时间序列 M-K 法统计量[图 2(a)]在 20 世纪 70 年代初有多次交点,在 1978 年和 1985 年也有 2 次交点;年降雨时间序列 M-K 法统计量[图 2(b)]在 20 世纪 70 年代也有多次交点,在 1980 年和 1989 年也有 2 次交点。M-K 法统计量表明降雨和径流在 20 世纪 70 年代发生多次突变,这与实际不符,且并未超过 95% 的显著水平,突变不显著,有必要采用样本熵法进行进一步验证。

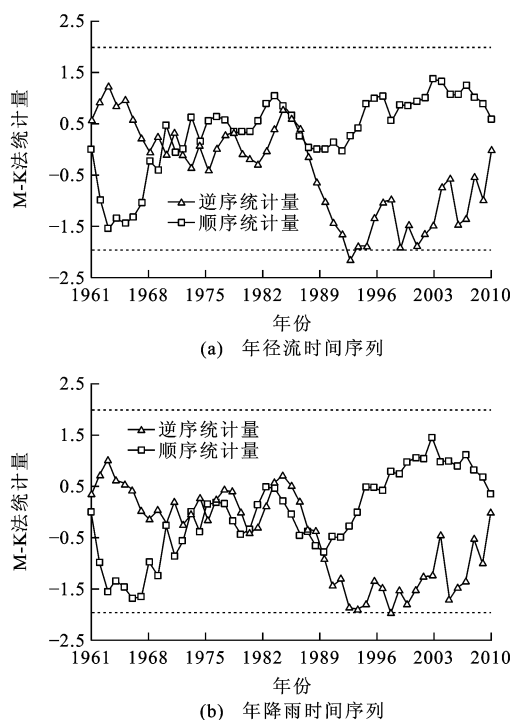


图2 1961~2009 年湘江流域年径流和  
年降雨时间序列 M-K 法统计量

Fig.2 Statistics of Annual Rainfall and Runoff Time Series by M-K Method from 1961 to 2009 in Xiangjiang River Basin

### 3.2 样本熵法突变检验

#### 3.2.1 滑动样本熵

结合滑动技术分别对湘潭站 1961~2009 年逐日降雨和径流进行样本熵计算,滑动窗口尺度  $S$  分别取 365、730、1 095、1 460 d,滑动步长  $L$  为 1 d。随着滑动窗口尺度的增大,得到的样本熵值时间序列越稳定,越易观察样本时间序列的动力学特征。由于不同动力学性质的数据复杂性不同,所以在时间序列点出现明显波动处可能发生突变。图 3 展示了滑动窗口尺度  $S$  为 730、1 460 d 的滑动样本熵计算结果。从图 3(a)、(b)可以看出:径流时间序列动态样本熵值波动较剧烈,整体呈增加趋势,分别在 1981 年、1986 年和

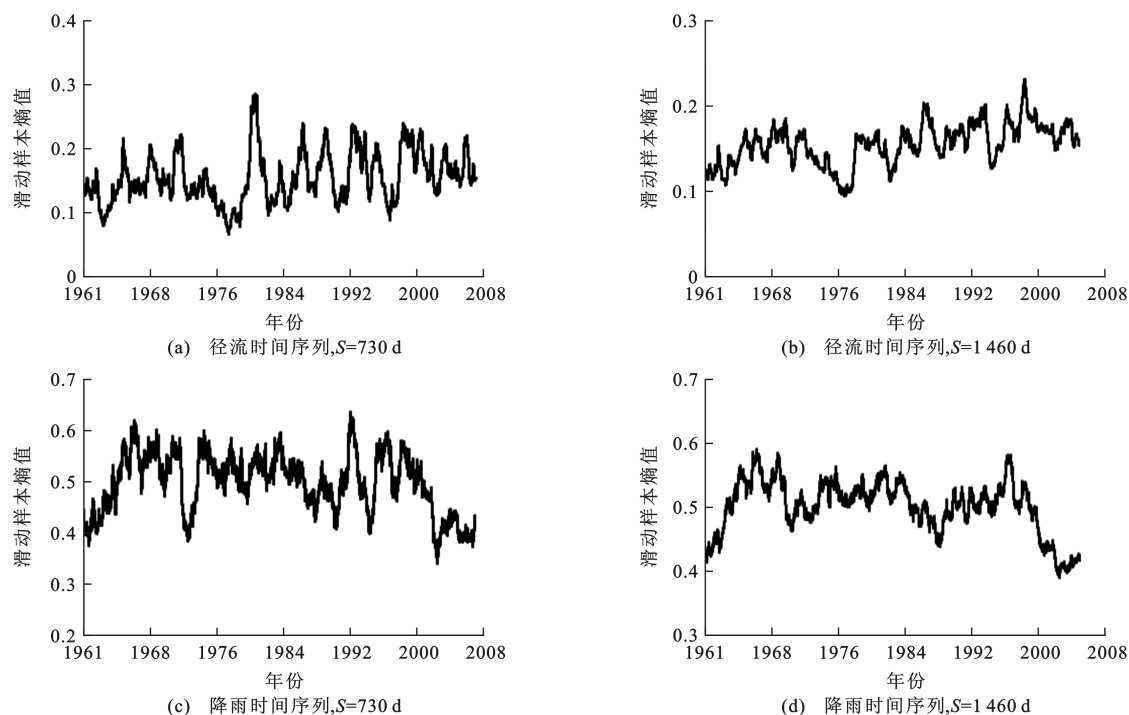


图3 不同滑动尺度的降雨和径流时间序列滑动样本熵

Fig. 3 Moving Sample Entropy of Rainfall and Runoff Time Series in Different Moving Scales

2000 年前后出现 3 次波谷,在 1991 年前后出现波峰;而降雨时间序列动态样本熵值整体呈减小趋势,波动相对较平稳[图 3(c)、(d)],在 1970 年和 1990 年前后出现波谷,在 2000 年前后出现波峰。

### 3.2.2 滑动移除样本熵

滑动样本熵能够较好地检测样本时间序列的突变,并体现其突变前后的动力学变化特征,但其结果依赖于滑动窗口尺度的大小,并且不能准确定位突变发生的起始点。因此,引进滑动移除技术,将滑动移除窗口尺度  $S$  取 365 d,滑动步长  $L$  取 365 d,依次移除 1 年的数据,将剩余年份的数据直接连在一起得到一个新的时间序列,并计算新生成时间序列的样本熵,得到该移除年份的滑动移除样本熵。对比各年份的滑动移除样本熵,当移除年份前后的滑动移除样本熵发生明显变化时,表明该年份发生突变[图 4(a)、(b)]。采用滑动  $t$  检验方法对图 4(a)、(b)的滑动移除样本熵时间序列进行突变检验,分别取滑动步长 6 年和 10 年,选取显著性水平 0.05 和 0.01,结果见图 4(c)、(d)。

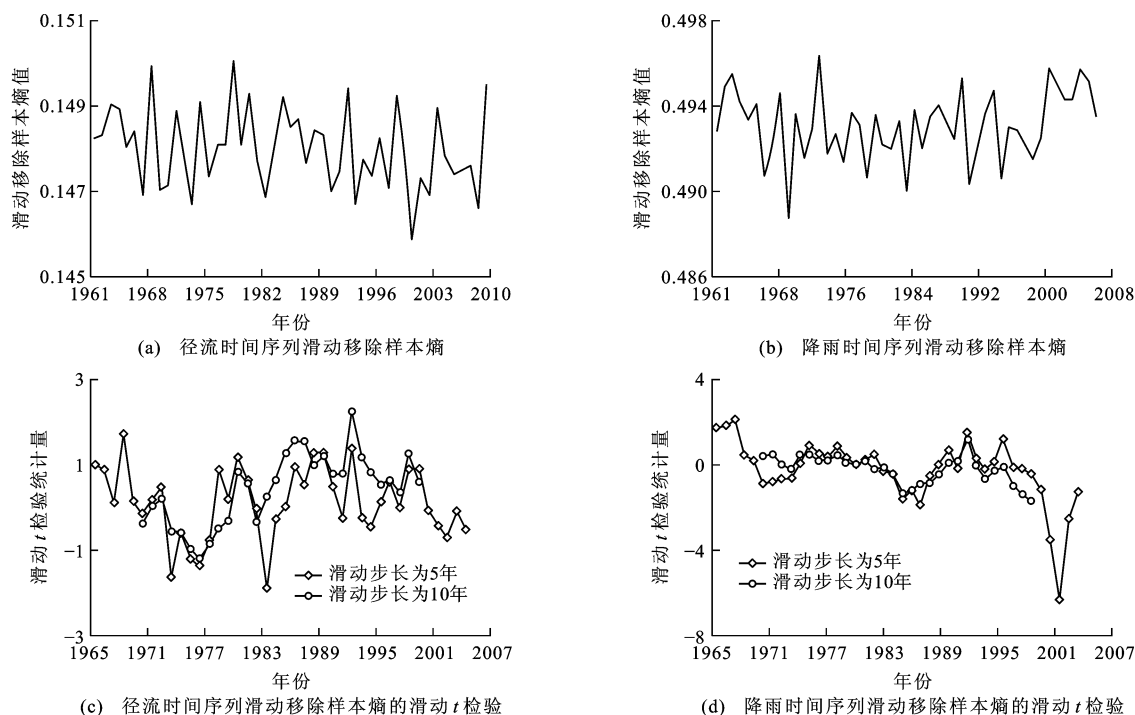
由图 4(c)可以看出:径流时间序列在 1966 年、1983 年、1992 年以及 2002 年发生多次显著突变,且 1966 年、1992 年以及 2002 年分别通过置信度为 95%的置信检验,1983 年通过置信度为 99%的置信检验;而降雨序列则是在 1966 年、1987 年以及 2002

年发生 3 次显著突变[图 4(d)],其中 1966 年和 1987 年通过置信度为 95%的置信检验,2002 年通过置信度为 99%的置信检验。

## 4 讨论

由滑动移除样本熵检测到降雨和径流均在 1966 年发生突变,两者在时间上较一致,这是由于在 20 世纪 50 年代初期,河川径流基本处于天然状态,径流主要受到气候变化的影响,所以降雨的突变直接导致径流的突变,并且径流时间序列的复杂度随着降雨时间序列复杂度的增大而增大。径流在 1983 年发生突变,与径流滑动样本熵时间序列[图 3(b)]的第 1 次波谷相对应,相对的降雨滑动样本熵时间序列[图 3(d)]并无明显变化,且该时间与湘江干流第 1 座水电站湘江水电站的建成时间吻合,表明此次突变与降雨无关,而是受到水利工程的影响。降雨时间序列在 1987 年发生第 2 次突变,与降雨时间序列滑动样本熵的第 2 次波谷(1990 年左右)相对应,而在第 1 次波谷处(1970 年左右)未检测到突变,表明 1987 年气象条件发生变化,但未引起降雨的突变;径流时间序列滑动样本熵在 1987 年左右出现第 2 次波谷,但径流时间序列滑动移除样本熵并未检测到突变,表明降雨的第 2 次突变造成径流的复杂度发生变化,且在 20 世纪 90 年代初,湘江干流



图4 降雨和径流时间序列滑动移除样本熵及其滑动  $t$  检验Fig. 4 Moving Cut Data Sample Entropy of Rainfall and Runoff Time Series and Their Moving  $t$ -test

第2座水电站宋家洲水电站开始修建,共同导致了径流在1992年发生突变。总之,人类活动的影响导致径流时间序列的复杂度增加。

降雨和径流均在2002年发生突变,但降雨时间序列在突变后复杂度呈减小趋势,而径流时间序列复杂度呈增大趋势。近年来,随着社会经济快速发展,城市化进程不断加快,人类活动明显增强,导致径流时间序列的复杂度不断增大。由中国科学院地理科学与资源研究所完成的1:100 000土地利用/覆盖数据可以看出,1986~2005年湘江流域的平原旱地减少271.4 km<sup>2</sup>,城镇用地面积增加661.7 km<sup>2</sup>,增量占1986年此类土地利用类型的35%左右<sup>[13]</sup>,改变了流域的产流条件,且随着水利工程的兴建,人类活动对径流的影响不断增强,相应的气象条件变化对径流的影响相对减弱。

## 5 结 语

(1)以样本熵为基础,分别与滑动方法和滑动移除方法相结合,分析了湘江流域降雨和径流时间序列的复杂度变化,并进行突变检验。基于非线性理论的样本熵能够较好地识别水文非线性系统动态演变特征。研究结果克服了传统的突变检验方法Mann-Kendall法不能有效识别非线性时间序列特征,检测突变能力不足的缺陷。

(2)滑动移除样本熵检测到湘江流域径流在1966年、1983年、1992年以及2002年发生多次突变,降雨在1966年、1987年以及2002年发生突变。结合滑动样本熵和滑动移除样本熵的结果发现:径流时间序列在突变后复杂度呈增加趋势,降雨时间序列在突变后呈减小趋势;在人类活动影响较小时,径流时间序列复杂度随着降雨时间序列复杂度增大而增大,而随着人类活动对径流的影响不断增大,径流时间序列复杂度不断增大。

## 参 考 文 献 :

### References :

- [1] 宋晓猛,张建云,占车生,等.气候变化和人类活动对水文循环影响研究进展[J].水利学报,2013,44(7):779-790.  
SONG Xiao-meng, ZHANG Jian-yun, ZHAN Che-sheng, et al. Review for Impacts of Climate Change and Human Activities on Water Cycle[J]. Journal of Hydraulic Engineering, 2013, 44(7): 779-790.
- [2] 孙永东,黄强,王义民,等.滑动近似熵在径流序列突变性分析中的应用[J].水力发电学报,2014,33(4):1-6.  
SUN Yong-dong, HUANG Qiang, WANG Yi-min, et al. Application of Moving Approximate Entropy to Mutation Analysis of Runoff Time Series[J]. Journal of Hydroelectric Engineering, 2014, 33(4): 1-6.
- [3] 白乐,李怀恩,何宏谋.窟野河径流变化检测及归因

- 研究[J]. 水利发电学报, 2015, 34(2): 15-22.
- BAI Le, LI Huai-en, HE Hong-mou. Analysis on Detection and Attribution of Runoff Change in Kuye River Basin[J]. Journal of Hydroelectric Engineering, 2015, 34(2): 15-22.
- [4] 周念清, 杨 硕, 沈新平, 等. 洞庭湖流域降雨序列突变性和多尺度特征分析[J]. 同济大学学报: 自然科学版, 2014, 42(6): 867-872.
- ZHOU Nian-qing, YANG Shuo, SHEN Xin-ping, et al. Mutation and Multi-scale Characteristics Analysis of Rainfall Series in Dongting Lake Watershed[J]. Journal of Tongji University: Natural Science, 2014, 42(6): 867-872.
- [5] 王国庆, 张建云, 李雪梅, 等. 黄土高原泾水河流域径流变化及其对环境要素的响应[J]. 水土保持研究, 2014, 21(3): 192-196.
- WANG Guo-qing, ZHANG Jian-yun, LI Xue-mei, et al. Runoff Variation of the Xinsuihe River Catchment and Its Response to Environment Change on the Loess Plateau[J]. Research of Soil and Water Conservation, 2014, 21(3): 192-196.
- [6] PINCUS S M. Approximate Entropy as a Measure of System Complexity [J]. PNAS, 1991, 88(6): 2297-2301.
- [7] HORNERO R, ABOY M, ABASOLO D, et al. Interpretation of Approximate Entropy: Analysis of Intracranial Pressure Approximate Entropy During Acute Intracranial Hypertension[J]. IEEE Transactions on Biomedical Engineering, 2005, 52(10): 1671-1680.
- [8] 黄宁波, 王义民, 苏保林. 汉江上游洪水特性复杂度分析[J]. 南水北调与水利科技, 2012, 10(1): 45-48.
- HUANG Ning-bo, WANG Yi-min, SU Bao-lin. Complexity Analysis of Flood Characteristics in the Upper Hanjiang River [J]. South-to-north Water Diversion and Water Science and Technology, 2012, 10(1): 45-48.
- [9] 王启光, 张增平. 近似熵检测气候突变的研究[J]. 物理学报, 2008, 57(3): 1976-1980.
- WANG Qi-guang, ZHANG Zeng-ping. The Research of Detecting Abrupt Climate Change with Approximate Entropy[J]. Acta Physica Sinica, 2008, 57(3): 1976-1980.
- [10] 张亮亮, 贾元华, 牛忠海, 等. 基于近似熵的交通流序列趋势变化检测[J]. 北京交通大学学报, 2014, 38(6): 7-11.
- ZHANG Liang-liang, JIA Yuan-hua, NIU Zhong-hai, et al. Trend Change Detection of Traffic Flow Based on Approximate Entropy[J]. Journal of Beijing Jiaotong University, 2014, 38(6): 7-11.
- [11] RICHMAN J S, MOORMAN J R. Physiological Time-series Analysis Using Approximate Entropy and Sample Entropy[J]. American Journal of Physiology, Heart and Circulatory Physiology, 2000, 278(6): H2039-H2049.
- [12] 彭 涛, 陈晓宏, 庄承彬. 基于样本熵的东江月径流序列复杂性分析[J]. 生态环境学报, 2009, 18(4): 1379-1382.
- PENG Tao, CHEN Xiao-hong, ZHUANG Cheng-bin. Analysis on Complexity of Monthly Runoff Series Based on Sample Entropy in the Dongjiang River[J]. Ecology and Environmental Sciences, 2009, 18(4): 1379-1382.
- [13] 黄生志, 黄 强, 王义民, 等. 基于启发式分割和近似熵法的径流序列变异诊断[J]. 中山大学学报: 自然科学版, 2014, 53(4): 154-160.
- HUANG Sheng-zhi, HUANG Qiang, WANG Yi-min, et al. The Study on the Runoff Change Based on the Heuristic Segmentation Algorithm and Approximate Entropy[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2014, 53(4): 154-160.
- [14] 原继东, 王志海. 时间序列的表示与分类算法综述[J]. 计算机科学, 2015, 42(3): 1-7.
- YUAN Ji-dong, WANG Zhi-hai. Review of Time Series Representation and Classification Technique[J]. Computer Science, 2015, 42(3): 1-7.
- [15] CHOU C M. Complexity Analysis of Rainfall and Runoff Time Series Based on Sample Entropy in Different Temporal Scales [J]. Stochastic Environmental Research and Risk Assessment, 2014, 28(6): 1401-1408.
- [16] ZHOU Y, ZHANG Q, LI K, et al. Hydrological Effects of Water Reservoirs on Hydrological Processes in the East River(China) Basin: Complexity Evaluations Based on the Multi-scale Entropy Analysis[J]. Hydrological Processes, 2012, 26(21): 3253-3262.
- [17] CHOU C M. Applying Multiscale Entropy to the Complexity Analysis of Rainfall-runoff Relationships[J]. Entropy, 2012, 14(5): 945-957.
- [18] 何文平, 何 涛, 成海英, 等. 基于近似熵的突变检测新方法[J]. 物理学报, 2011, 60(4): 820-828.
- HE Wen-ping, HE Tao, CHENG Hai-ying, et al. A New Method to Detect Abrupt Change Based on Approximate Entropy[J]. Acta Physica Sinica, 2011, 60(4): 820-828.
- [19] 杜 鵬, 徐 伟, 赵智国, 等. 湘江流域近 30 年径流量与水位的长期变化规律研究[J]. 应用基础与工程科学学报, 2010, 18(3): 369-378.
- DU Juan, XU Wei, ZHAO Zhi-guo, et al. Trends of Runoff and Water Level over the Past Thirty Years in Xiang River Basin[J]. Journal of Basic Science and Engineering, 2010, 18(3): 369-378.