

Complexity Analysis of Sediment Transport Quantity in the Mainstream of Huangshui River Based on Multiple Entropy Algorithms

Zhijin Ma¹ Bo Zhang²

1. Yellow River Hydrological and Water Resources Science Research Institute, Zhengzhou, Henan, 450004, China

2. Yellow River Conservancy Commission Yellow River Institute of Hydraulic Research, Zhengzhou, Henan, 450003, China

Abstract

This paper investigates the complexity evolution of sediment transport in the main stream of the Huangshui River, with a focus on employing three algorithms: Approximate Entropy, Sample Entropy, and Fuzzy Entropy. The study reveals significant spatial and temporal variations in sediment transport complexity, with the Xining station exhibiting greater complexity than the Minhe station. The interannual changes at both stations are concentrated in three periods: 1964—1969, 1980—1990, and 2000—2016, largely influenced by factors such as precipitation distribution, large-scale land reclamation, and reforestation efforts. Among the three entropy algorithms, the analysis shows that Fuzzy Entropy demonstrates higher sensitivity and computational stability in assessing sediment transport complexity, making it less susceptible to data anomalies. This makes Fuzzy Entropy more suitable for analyzing the complexity evolution across different time periods and spatial scales.

Keywords

sand sediment; complexity; approximate entropy; sample moisture; fuzzy entropy; Huangshui main stream

基于多种熵算法的湟水干流输沙量复杂性分析

马志瑾¹ 章博²

1. 黄河水文水资源科学研究院, 中国·河南 郑州 450004

2. 黄河水利委员会黄河水利科学研究院, 中国·河南 郑州 450003

摘要

论文探讨了湟水干流输沙量序列的复杂性演化过程, 重点采用了近似熵、样本熵和模糊熵三种算法进行分析。研究表明, 湟水干流输沙量的时空变化呈现出明显的差异性, 西宁站的复杂性显著高于民和站, 且两站的年际变化集中在1964—1969年、1980—1990年和2000—2016年三个阶段, 主要受降水分布、大规模开垦及退耕还林等因素影响。分析表明, 在三种熵值算法的对比中, 模糊熵相比其他熵值算法在输沙量复杂性分析中具有更高的敏感性和计算稳定性, 不易受到数据奇异值的影响, 适用于不同时间段和空间尺度上的复杂性演变研究。

关键词

输沙量; 复杂性; 近似熵; 样本熵; 模糊熵; 湟水干流

1 引言

水文系统是一个开放、动态和非线性的复杂系统, 其演变过程具有高度的复杂性^[1]。受动力因素的影响, 水文系统往往具备一定的规律性, 若失去动力驱动或者收到外界动力的扰动, 即在气候变化和人类活动的干扰下, 原系统序列会变得扰动“不安”, 这时其信号波开始变得混乱并丧失其稳定性^[1]。近年来, 伴随非线性理论的发展和完善, 熵的应用领域也在不断拓展, 从热力学到生物学、物理学, 以及在

时间序列分析上都有应用^[2]。熵这个概念最早源于热力学, 它的物理意义是体系中混乱程度或者复杂程度的度量^[3]。Pincus 提出的近似熵概念^[4], Richman 对近似熵的改进算法样本熵, 熵原理被广泛地应用于径流序列复杂性分析, 并取得了一些成果。

在河流系统中来水来沙条件是最为重要的影响因素, 对河流体系诸多关系因素的演变起到决定性作用。因此在流域治理开发规划及防洪、河道整治等都必须了解、掌握水沙变化的基本情况和基本规律。

在泥沙方面熵原理最初应用于河流冲刷或淤积概率沿程分布、含沙量垂线分布以及悬沙浓度分布方面, 并未用于河流输沙量复杂性分析上。论文将采用近似熵、样本熵和模

【作者简介】马志瑾(1983—), 女, 中国青海西宁人, 硕士, 高级工程师, 从事水文水资源研究。

糊熵算法, 尝试定量识别在变化环境下湟水干流控制站点输沙量序列的复杂性演化过程。

2 区域概况与数据来源

2.1 区域概况

湟水是黄河上游最大的一级支流, 由湟水干流和支流大通河组成。湟水干流是青海人的母亲河, 也是青海省主要的农业基地。湟水干流发源于青海省海晏县祁连山系大坂山南坡, 河道全长 373.9km, 流域面积 15342km², 平均比降 5.3‰。上游主源为麻皮寺河, 在海晏与哈利洞河汇合后称为西川河, 在西宁盆地纳北川河后蜿蜒东行, 穿过小峡、大峡、老鸦峡, 在民和县享堂与其最大一级支流大通河汇合后流入甘肃省, 于甘肃省永靖县上车村汇入黄河。

湟水干流东部的民和盆地与兰州盆地相连, 其边界均为山脉; 中部为西宁盆地, 有北川河、西川河和南川河 3 条较大河流汇入; 北部为拉脊山。整个流域由山地和盆地构成, 共有较大支流和支沟 117 条。长期的土壤侵蚀, 致使仅河源地带仍保留着高山草甸地貌, 下游的浅山区地貌由梁状逐渐演变为丘状, 每年都有大量的泥沙经湟水进入黄河, 是黄河上游地区主要产沙区之一^[7]。

2.2 数据来源

选取湟水干流中游和下游分界控制站西宁站和湟水干流入黄控制站民和站 1956 年至 2016 年的实测月平均输沙量数据, 数据 1956—1990 年、2005—2016 年摘自中华人民共和国水文年鉴《黄河流域水文资料》第一册黄河上游区上段(黑山峡以上)。资料符合整编要求, 精度满足分析要求。

3 计算方法

近似熵是一种用于量化时间序列的不规则性或者复杂性的度量方式, 它反映了时间序列中新子序列产生的概率, 也就是以一种条件概率的方式来衡量时间序列中新信息发生的可能性, 因此越复杂的时间序列对应的近似熵就越大。

样本熵同样也是用于衡量时间序列在单一尺度上的复杂度, 与近似熵相比, 具有更高的精度。

模糊熵是在样本熵的基础上改进得到的, 引入了模糊隶属度函数, 它是一种指数函数, 计算方法如下:

步骤 1~2 与近似熵相同。

步骤 3: 计算任意两个重构向量 $X(i)$ 和 $X(j)$ 之间的距离 $dm_{ij}^m[X(i), X(j)]$, 其中 dm_{ij}^m 表示两个重构向量 $X(j)$ ($1 \leq j \leq N-m+1$) 与 $X(i)$ 之间的距离。

步骤 4: 引入模糊隶属度函数:

$$A_{ij}^m = \exp\left[-\left(\frac{d_{ij}^m}{r}\right)^n\right]$$

步骤 5: 计算针对每个 i 的平均值:

$$C_i^m(r) = \frac{\sum_{j=1, j \neq i}^{N-m+1} A_{ij}^m}{N-m}$$

步骤 6: 定义在划分子序列个数为 m 时的平均相似率:

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

步骤 7: 按照上面 1~6 再计算当划分子序列个数为 $m+1$ 时的平均相似率。

步骤 8: 得模糊熵:

$$FsEn = \ln\phi^{m+1}(r) - \ln\phi^m(r)$$

注: 关于 m 和 r 的选取, m 通常选择 2 或者 3; r 通常选择 $r=0.1 * \text{std} \sim 0.2 * \text{std}$, std 是原时间序列的标准差。

4 结果分析与讨论

4.1 输沙量序列复杂性空间分布特性

以湟水干流西宁和民和站 1956-2016 年输沙量年序列标准差的 0.2 倍, 作为表示相似度比较的阈值 r ; 为了更多的保留原有系列的信息, 确定熵值计算维数为 1, 分别计算两站输沙量近似熵、样本熵和模糊熵值(见表 1)。

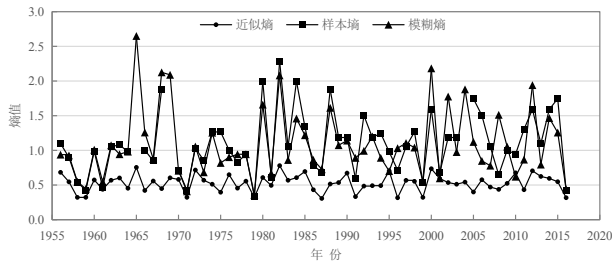
表 1 水文站年输沙量序列年均值与熵值信息表

站点	1956-2016 年均输沙量 (万 t)	样本熵	模糊熵	近似熵
西宁站	243	1.09	1.09	0.53
民和站	1286	0.79	0.74	0.45

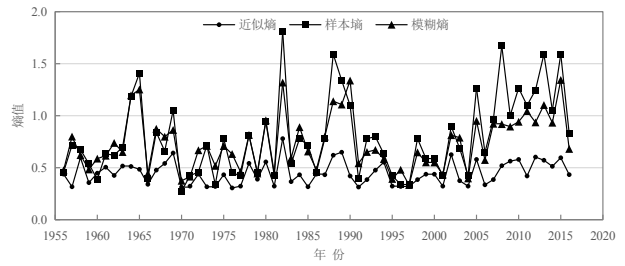
中游和下游分界站西宁站与入黄河控制站民和站两站熵值呈递减趋势。由熵的性质可知, 熵值越大, 反映的复杂程度越高, 可预测性也就越低。由此可见湟水干流复杂性存在明显的空间差异, 整体表现为西宁比民和站更复杂, 说明影响西宁站输沙量的原因较多, 影响民和站输沙量的不确定性成分相对较少。造成复杂性差异的原因, 可能有集水面积、水沙组成、下垫面状况以及人口活动强度等。但最主要是西宁站以上的输沙量仅占全干流的 19%, 即湟水干流输沙量的 81% 均来自于西宁站到民和站区段。属于相对比较天然阶段, 变幅大是由于产沙地区的地貌特性对特殊降水分布产生相应, 不同的水沙组合, 随机性成分增多, 从而导致熵值增大。

4.2 输沙量序列复杂性年际变化特性

以湟水干流西宁和民和站 1956-2016 年月输沙量逐年序列标准差的 0.2 倍, 作为改序列表示相似度比较的阈值 r ; 为了更多的保留原有系列的信息, 确定熵值计算维数为 1, 分别计算两站输沙量近似熵、样本熵和模糊熵逐年熵值, 计算结果见图 1。由西宁和民和站三种熵值的对比可以看出: 近似熵对数据敏感度低, 年际变化幅度较小, 两站熵值均在 0.31~0.78 之间浮动, 基本无趋势变化, 不适合进行输沙量复杂性分析; 样本熵与模糊熵对数据敏感性较高, 年际变化幅度较大, 且两者熵值曲线呈现出较为相似的演化过程。两者虽然样本熵对数据更为敏感, 变化幅度更大, 但西宁站样本熵计算时在 1965 年、1969 年和 2004 年三年均出现奇异值, 使得样本熵在计算时出现了不稳定情况。



(a) 西宁站输沙量熵值年际变化图



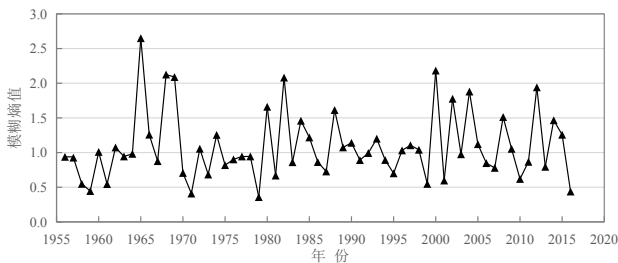
(b) 民和站输沙量熵值年际变化图

图1 输沙量熵值年际变化图

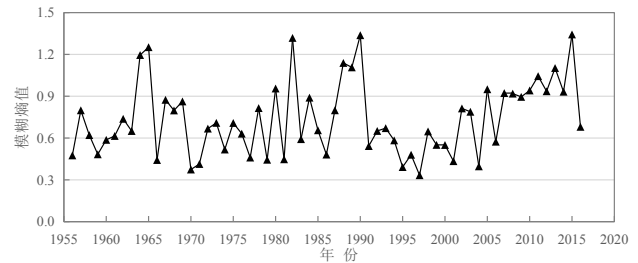
分析其原因,主要是因为样本熵是以 $-\ln(\text{CP})$ 为模型,在西宁站计算时出现了重构向量与模版之间的匹配概率为零的情况,即出现 $\ln(0)$ 的情况,导致了奇异点。为了避免出现 $\ln(0)$ 的情况,在计算重构向量距离的时候,需要剔除与自身的比较。综上所述,三种熵值计算方法,模糊熵具有敏感度高,计算简便以及数据稳定等优势,更适合进行输沙量复杂性分析。

以模糊熵为例进行西宁和民和站输沙量复杂性分析见图2,由图中我们可以看到两站输沙量熵值变化幅度较大的地方,主要集中在三个时间段,分别为1964—1969年段、1980—1990年段以及2000年—2016年段。

1964—1969年段,属于相对比较天然阶段,变幅大是由于产沙地区的地貌特性对较特殊降水分布产生相应,不同的水沙组合,随机性成分增多,从而导致熵值增大。



(a) 西宁站输沙量模糊熵年际变化图



(b) 民和站输沙量模糊熵年际变化图

图2 输沙量模糊熵年际变化图

1980—1990年段波动主要有两个方面造成:第一,中国于20世纪80年代初实行农村改革,进行了大规模的开垦荒山荒坡和焚林造田造成;第二,与修建水库、水电站以及小型的私人河沙开采有关^[12]。

2000—2016年段变幅大,主要是2000年起青海省开始实施退耕还林还草,湟水流域也开始实施退耕种草种树、封山育林、治理水土流失,截至2010年湟水流域完成退耕还林还草面积达14.7万 hm^2 ,各种淤地坝、水土保持、生态修复工程取得了显著的成效。

5 结论

①湟水干流输沙量的时空变化。在空间上,西宁站的熵值明显高于民和站,说明影响西宁站输沙量不确定性成分比民和站要多;在年际变化上,两站的变化主要集中在1964—1969年、1980—1990年和2000—2016年三个时间段,造成因素分别主要为降水分布,大规模的开垦、焚林造田和水电建设以及2000年以后的“退耕还林还草”。

②湟水干流输沙量序列的复杂性不具有上、下游延续

性,而与输入来源占绝对优势部分的复杂程度相关。

③在用三种熵进行输沙量复杂性分析计算时,模糊熵更适用。近似熵敏感度低,没有办法很好的表示输沙量在年际上的复杂性变化;样本熵虽敏感性强,复杂性表现度强,但是以 $-\ln(\text{CP})$ 为模型,会出现 $\ln(0)$ 的情况;模糊熵敏感性强,又因引入了模糊隶属度函数,其参数的改变对复杂性度量的结果影响很小,算法简单,且结果相对稳定,为输沙量序列的复杂性研究提供了一种新的方法。

参考文献

- [1] 冯国章,宋松柏,李佩成.水文系统复杂性的统计测度[J].水利学报,1998(11):76-81.
- [2] 王文均,叶敏,陈显维.长江径流时间序列混沌特性的定量分析[J].水科学进展,1994,5(2):87-94.
- [3] 王远坤,王栋.基于样本熵理论的长江干流径流序列复杂性分析[J].河海大学学报(自然科学版),2015,43(3):203-207.
- [4] PINCUS S M. Approximate entropy as a measure of system complexity[J]. Proceeding of the National Academy Sciences, 1991,88(6):2297-2301.