基于相关系数的水文序列滑动周期识别方法

吴林倩^{1*}, 谢平^{1,2*}, 吴子怡¹, 桑燕芳³, 陈杰¹, 赵羽西⁴, 李雅晴¹, 王路¹

1. 武汉大学水资源与水电工程科学国家重点实验室, 武汉 430072;
2. 国家领土主权与海洋权益协同创新中心, 武汉 430072;
3. 中国科学院地理科学与资源研究所, 陆地水循环与地表过程重点实验室, 北京 100101;
4. 国家电网公司西南分部, 成都 610000

* 联系人, E-mail: wlq507@whu.edu.cn; pxie@whu.edu.cn

2018-11-30 收稿, 2019-04-30 修回, 2019-05-13 接受, 2019-08-21 网络版发表

国家自然科学基金(91547205, 91647110, 51579181, 51779176)和中国科学院青年创新促进会项目(2017074)资助

摘要 准确进行周期识别对于合理认识水文变异规律、开展中长期水文模拟预报、水安全评价等具有重要的意义. 本文通过滑动改变周期长度和初相位, 构造出多组模拟周期序列, 并以相关系数作为模拟周期序列与原始序列拟合精度的判别指标, 选择拟合程度最好的模拟周期作为最优周期识别结果, 从而提出水文序列滑动周期识别方法. 理论和公式推导表明: 相关系数可表征水文序列中的周期波动程度, 相关系数越大, 说明拟合精度越高, 对应的周期变异越显著. 统计试验结果显示: 相比于常用的周期识别方法, 文中方法的周期识别效率更高, 且在各因素影响下也最具稳定性和可靠性. 最后, 文中方法应用于澜沧江流域21个站的年降雨量序列进行周期识别, 结果显示该流域的降雨周期存在多时间尺度特征, 北部和中部的降雨周期长度较短, 多数在2~5 a左右, 而南部的降雨周期5~10 a居多, 个别站点的周期长度达到了18 a.

关键词 周期识别, 相关系数, 非一致性, 水文变异, 澜沧江

周期波动是水文过程演变的一种常见形式, 也是水文变异的一种重要表现形式^[1]. 准确识别和提取水文序列周期成分, 对于非一致性水文频率计算, 以及水资源、洪水、干旱等水安全评价具有重要的理论意义和实用价值. 目前常用的周期识别方法有简单分波法(SPWM)^[2]、谐波分析法(HAM)^[3]、功率谱分析(PSM)^[4]、最大熵谱分析(MESA)^[5]等. 这些方法在水文学领域的应用已较为成熟, 但受其适用条件的约束仍存在不足. 例如, 简单分波法计算简便, 但在判别周期是否显著时使用到 F 检验, 而 F 检验的应用前提是样本服从正态分布, 且各组数据总体的均方差相等^[6,7]; 此外, 假设检验中显著性水平 α 的选取也会对识别结果产生影响, 因此识别出的多个显著周期成分中可能含有

伪周期. 谐波分析法是基于傅里叶变换^[8]分析水文序列周期特性的方法, 计算速度快, 但往往不能兼顾高频和低频的需求^[9]. 功率谱分析通过功率谱图上的峰值确定对应周期, 但缺点同谐波分析, 也不能兼顾高频和低频的需求, 此外, 当样本长度与周期长度相当时识别效率也会受到影响^[10]. 最大熵谱分析基于最大熵原理, 摒弃了传统频谱方法低分辨率、主观判断等缺点, 但是, 该法的判别准则中涉及自回归模型(AR模型)阶数的确定^[11], 识别结果受阶数确定准则的影响并直接决定了周期识别的可靠性. 因此, 如何更加准确地识别出水文时间序列中存在的周期成分仍值得探索与创新.

本文以相关系数为判别依据, 通过构造与原水文序列拟合程度最好的周期成分来识别周期, 提出基于

引用格式: 吴林倩, 谢平, 吴子怡, 等. 基于相关系数的水文序列滑动周期识别方法. 科学通报, 2019, 64: 2549–2560

Wu L Q, Xie P, Wu Z Y, et al. Moving correlation coefficient-based method for the identification of periodicities in hydrologic time series (in Chinese). Chin Sci Bull, 2019, 64: 2549–2560, doi: [10.1360/N972018-01192](https://doi.org/10.1360/N972018-01192)

相关系数的滑动周期识别方法(简称SRPT方法),并辅以公式推导阐明识别原理以及具体识别步骤;最后,设计统计试验并结合实例分析来验证滑动周期识别方法的有效性.

1 滑动周期识别方法

为研究水文时间序列中含有的周期成分,需要先对序列中的趋势、跳跃等确定性成分进行识别和分离^[12].有关跳跃和趋势的识别目前已有多种方法,如识别跳跃点可以采用Pettitt法^[13]、滑动 F 检验法^[14]、Mann-Kendall法^[15]等.识别序列的趋势成分可以采用线性趋势相关系数检验法、Spearman秩次相关检验法^[16]、最小二乘法^[17].识别并剔除两种暂态成分后,可对剩余序列的周期成分进行检验识别.

周期成分一般包括简单周期、复合周期和近似周期3种形式^[1].本文仅针对含有简单周期成分的水文序列,验证SRPT方法对周期成分的识别和表达.对于复合周期,可以本方法为核心结合非平稳时间序列分解方法如EMD(Empirical Mode Decomposition)^[18]进行多个周期成分的准确识别,显著性判断与表达,以改进谱分析方法存在的问题^[19].水文序列在去除跳跃和趋势成分之后,假定剩余成分 $x_t(t=1,2,\cdots,n)$ 含有简单周期成分,可表达为

$$x_t = A + B \sin\left(\frac{2\pi}{T}t + t_0\right) + \eta_t, \quad (1)$$

式中, T 为周期长度, t 为时序, t_0 为初相位, η_t 为残余随机项; A , B 分别表示序列均值和周期的半振幅.

将式(1)中的 A 和 η_t 合并为 u_t 作为水文纯随机成分,则序列 x_t 可表示为线性叠加的形式:

$$x_t = Bz_t + u_t, \quad (2)$$

其中, $z_t = \sin\left(\frac{2\pi}{T}t + t_0\right)$.

序列 x_t 与周期成分 $y_t = Bz_t$ 的相关程度可由相关系数 r 表示为

$$r = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2 \sum_{t=1}^n (y_t - \bar{y})^2}}. \quad (3)$$

对于发生了周期变化的水文序列,某一周期成分表达式中涉及的半振幅 B ,周期 T 和初相位 t_0 是确定存在的,因此式(3)中的相关系数可进一步表示为

$$r = \frac{\sum_{t=1}^n (x_t - \bar{x})(z_t - \bar{z})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2 \sum_{t=1}^n (z_t - \bar{z})^2}}. \quad (4)$$

相关系数是用以反映变量之间相关关系密切程度的统计指标.对于1个含有周期成分但周期未知的水文序列,可以先假设周期 T 和初相位 t_0 进行试算求解:当构造的正弦序列 $z_t = \sin\left(\frac{2\pi}{T}t + t_0\right)$ 与原始序列 x_t 拟合程度最好时,也即此正弦波动与序列中存在的周期波动最为接近时,对应的周期 T 和初相位 t_0 即为最优结果.然后再由最小二乘法求出 A , B ^[20].

$$\hat{B} = \frac{\sum_{t=1}^n (x_t - \bar{x})(z_t - \bar{z})}{\sum_{t=1}^n (z_t - \bar{z})^2}, \quad (5)$$

$$\hat{A} = \bar{x} - \hat{B}\bar{z}, \quad (6)$$

式中, $\bar{z} = \frac{1}{n} \sum_{t=1}^n z_t$, $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$, 可确定出含有简单周期成分的水文序列的表达式.

由式(3)和(5)可以建立相关系数 r 和半振幅 B 之间的关系^[21]:

$$r^2 = \frac{1}{1 + \frac{\sigma_u^2}{B^2 \sigma_z^2}}, \quad (7)$$

其中, 正弦序列 z_t 的方差 σ_z^2 与序列长度 n 、周期长度 T 和初相位 t_0 有关, 其公式为

$$\sigma_z^2 = \frac{1}{2} - \frac{\cos\left[\frac{2\pi}{T}(n+1) + 2t_0\right] \sin\frac{2\pi n}{T}}{2n \sin\frac{2\pi}{T}} - \frac{\left[1 - \cos\left(\frac{2\pi}{T}n + \frac{2\pi}{T} + 2t_0\right)\right] \left[1 - \cos\frac{2\pi n}{T}\right]}{2n^2 \left(1 - \cos\frac{2\pi}{T}\right)}. \quad (8)$$

纯随机成分的标准差 σ_u 与其均值 $\bar{u}$ 和变差系数 C_{vu} 有关, 其公式为

$$\sigma_u = \bar{u} C_{vu}. \quad (9)$$

分析式(7)可知, 对于周期 T 和初相位 t_0 的特定值, 正弦序列的方差 σ_z^2 和纯随机成分的方差 σ_u^2 是确定的, 此时 r 和 B 呈正相关关系. 相关系数值越大, 周期成分的振幅($2B$)越大, 说明序列的周期波动越显著. 因此, 在周期

未知的情况下, 提出构造并滑动周期成分来拟合原始序列的方法, 以相关系数为判别指标, 确定最可能的周期成分. 基于相关系数研究周期成分, 依据的是水文时间序列的线性叠加原理, 充分考虑了周期波动相对于叠加了纯随机成分的合成序列的相对变异程度. 因此, 相关系数可作为表征水文序列周期变异程度的有效指标, 基于相关系数的滑动周期识别方法有其合理性.

基于上述分析, 提出利用相关系数对水文序列周期成分进行滑动识别与显著性检验的方法, 此处称为“周期成分滑动相关系数识别方法(下文简称SRPT方法)”, 具体思路如下:

(1) 构造1个正弦函数拟合表达式 $z_t = \sin\left(\frac{2\pi}{T}t + t_0\right)$,

设定周期长度 T 由 $2 \sim n/2$ (n 为序列长度, 且 n 为奇数时, $n/2$ 取整数部分)以 l_1 为步长滑动, 初相位 t_0 由 $-\pi \sim \pi$ (由于 $t_0 = \pi$ 和 $t_0 = -\pi$ 表达的周期成分是相同的, 实际操作中二者取其—即可)以 l_2 为步长滑动, 从而得到 $\left(\frac{n/2-2}{l_1} + 1\right) \times \frac{2\pi}{l_2}$ 组 z_t (此处的步长 l_1 和 l_2 为可变的, 其中 l_1 的选取会影响周期结果的精度, 通常情况下可取 $l_1=1$, $l_2=0.001\pi$), 利用式(4)依次求取各 z_t 对应的序列与原始序列(线性叠加序列)的相关系数;

(2) 取所有结果中绝对值最大的相关系数 $|r_{\max}|$ 对其进行假设检验, 其中相关系数临界值可由服从 F 分布的 F 统计量临界值间接确定;

(3) 选择合适的显著水平 α , 若 $|r_{\max}| < r_\alpha$, 即未通过假设检验, 认为该序列无周期变异; 若 $|r_{\max}| > r_\alpha$, 即通过假设检验, 然后利用最小二乘法, 即式(5)确定相应 $z_{t_{\max}}$ 的半振幅 B , 得到 $y_{t_{\max}}$ 即为该水文序列的周期成分.

考虑到原始序列包括纯随机成分与周期成分, 结合对式(7)的分析, 正弦序列 z_t 的方差 σ_z^2 、振幅和纯随机成分的方差 σ_u^2 均会影响 r 值进而影响周期识别结果. 其中, 周期成分的方差 σ_z^2 与序列长度 n 和初相位 t_0 有关, 纯随机成分的标准差 σ_u^2 与其均值 $\bar{u}$ 和变差系数 C_{vu} 有关, 这些因素均可进一步影响周期成分识别, 因此需要设计统计试验讨论此问题, 以说明本文方法的可靠性和稳定性.

2 统计试验

生成模拟序列时主要考虑两部分:

(1) 周期成分, 通过设定表达式 $y_t = B\sin\left(\frac{2\pi}{T}t + t_0\right)$

中的参数 B , T , t_0 得到;

(2) 纯随机成分, 通过设定均值 $\bar{u}$, 变差系数 C_{vu} 和偏态系数 C_{su} 生成服从P-III分布的序列. 为了对比分析, 选取应用较为广泛的3种周期识别方法, 包括谐波分析法(HAM)、功率谱分析法(PSM)和最大熵谱分析(MESA). 由于此处只讨论简单周期成分, 所以谐波分析法选择 F 检验值最大对应的周期成分为最可能的周期, 最大熵谱法是选择熵谱值最大值对应的周期为最可能的周期, 功率谱法求解周期依据公式 $T = \frac{2m}{f}$, 其中最大时滞 m 的取值直接影响了周期结果, 依据文献给出的经验范围^[22]统计试验中 m 的取值为 $n/10 \sim n/4$.

由于统计试验是在未知周期的前提下确定周期, 若从多个峰值点中选择相近值作为周期, 会使得结果偏向于真值. 为了减少人为因素的影响, 统计试验中取了多个最大时滞 m 下功率谱最大值对应的周期为最可能的周期; SRPT方法是在滑动过程中选择相关系数最大对应的周期成分. 各方法优劣的评价标准: 设 T 为模拟序列的理论周期值, T' 为各方法的计算周期值, 考虑到统计试验存在抽样误差, 所以设定允许误差 $\Delta T = |T - T'| = 1$, 对于某种识别方法, 设以周期 T 为理论真值的模拟水文序列共有 N 组, 其中计算周期 T' 在允许误差范围内的共有 M 组, 则定义识别效率 $\eta = \frac{M}{N} \times 100\%$, η 越大, 意味着该方法的识别效率越高、结果越可靠.

2.1 样本长度 n 的影响

为了分析不同的样本长度对周期识别效率的影响, 要固定周期长度然后设置不同的样本长度, 样本长度的选取不能过短也不能过长, 若周期长度为 T , 则样本长度至少应满足 $n \geq T$, 为了体现出周期波动, 此处分别取模拟序列的长度 n 为100, 150, 200, 250, $\dots$, 450, 500共9个数值, 将试验分为9组. 设纯随机成分均值 $\bar{u}=100$, 变差系数 $C_{vu}=0.2$, 偏态系数 $C_{su}=2C_{vu}$; 周期成分 $y_t = B\sin\left(\frac{2\pi}{T}t + t_0\right)$ 中半振幅 $B=30$, 初相位 $t_0=0$, 周期长度 T 取25. 在每组参数条件下生成100个序列, 并分别用已有的3种方法与本文所提方法对序列进行周期识别, 其中SRPT方法的步长取 $l_1=1$, $l_2=0.001\pi$ (后续统计实

验步长设定相同), 比较对周期 T 的识别效率, 统计试验结果整理见表1.

对于固定的周期长度, 样本长度越大, 序列中含有的周期个数越多, 周期波动在序列中越明显, 因此越容易识别. 由表1可以看出, MESA方法在样本长度小于200时, 识别效率不到80%, 说明较短的样本长度下, MESA方法有时会识别出伪周期. 随着样本长度的增加, MESA方法的识别效率增加, 在样本数大于250之后, 识别效率达到90%, 对于更长的样本序列, 识别效率可达到100%, 说明该方法的稳定性受样本长度影响. PSM的识别效率也受样本长度的影响, 但两者的线性规律不太明显, 这是由于该方法的识别结果还受到最大时滞 m 取值的影响; 最大时滞较小时, 不能识别出真实的周期成分; 增大最大时滞, 识别出的周期成分往往与最大时滞呈倍数关系, 又会造成伪周期的产生^[22], 因此在样本长度为100, 150和400时, 由功率谱值最大得到的最可能的周期长度分别为20, 30和26.67, 造成此时的识别效率为0, 而样本长度为250时恰好是10个周期长度, 所以对于长度为25的周期成分识别效率较高, 这些现象均说明PSM的周期识别很大程度上受到最大时滞的影响, 效果不稳定^[23]. HAM和SRPT方法在不同的样本长度下均有较高的识别效率, 当允许误差为 $\Delta T = |T - T'| = 1$ 时, 两种方法的识别效率基本都达到了100%, 说明了SRPT方法的可靠性和样本长度变化下的稳定性.

2.2 纯随机成分的影响

2.2.1 均值 $\bar{n}$

设模拟样本序列长度 n 为200, 均值 $\bar{n}$ 取100, 150, $\dots$,

450, 500共9个数值, 每个均值下对应产生100组样本, 纯随机成分的变差系数 $C_{vu} = 0.2$, 偏态系数 $C_s = 2C_v$;

周期成分 $y_t = B\sin\left(\frac{2\pi}{T}t + t_0\right)$ 的半振幅 $B=30$, 周期长度 $T=25$, 初相位 $t_0=0$; 用HAM, PSM, MESA, SRPT 4种方法分别进行周期成分识别, 统计试验结果见图1. 可以看出, 4种方法的识别效率皆随纯随机成分均值的增大而降低. 这是由于当变差系数 C_{vu} 固定时, 纯随机成分的均值越大, 方差 σ_u^2 越大, 对应的纯随机序列的离散程度越大, 纯随机成分的波动性越显著, 会使得周期成分在整个序列中的比重降低, 进而不易被识别. 依据式(7)计算各均值对应的相关系数值, 依据基于相关系数的周期变异分级理论^[21], 选用显著性水平 $\alpha=0.05$, $\beta=0.01$, 样本长度为200的序列对应的相关系数临界值下限 $r_\alpha=0.1388$, 上限 $r_\beta=0.1818$, 详见图1. 可以看出, 在均值大于250后, 相关系数已低于临界值下限, 说明此时的周期成分不显著, 因此SRPT方法的检验效率呈现下降趋势, 但识别效率在这几种方法中是最稳定的, 且保持着较高的准确率. HAM的识别效率基本与SRPT方法相当, 但在均值大于300后, 识别效率低于SRPT方法. 而MESA和PSM的识别效率受均值变化影响较大, 在均值大于150之后, 识别效率急剧下降, 在均值大于400后, 几乎不能识别出真实的周期成分, 会有很大的概率得到伪周期.

2.2.2 变差系数 C_{vu}

设模拟样本序列长度 n 为200, 均值 $\bar{n}$ 为200, 变差系数 $C_{vu}=0.1, 0.15, \dots, 0.45$, 偏态系数 $C_{su}=0.5$; 周期成分 $y_t = B\sin\left(\frac{2\pi}{T}t + t_0\right)$ 的参数半振幅 $B=30$, 周期长

表1 不同的样本长度下各方法的周期识别效率(%)

Table 1 Period identification efficiency of different methods under different sample lengths (%)

样本长度	PSM	MESA	HAM	SRPT
100	0	30	100	98
150	0	64	99	100
200	49	78	100	100
250	100	86	100	100
300	92	91	100	100
350	96	95	100	100
400	34	96	100	100
450	100	100	100	100
500	100	100	100	100

图1 (网络版彩色)不同的均值下各方法周期识别效率及相关系数

Figure 1 (Color online) Period identification efficiency and correlation coefficient of different methods under different mean values

度 $T=25$, 初相位 $t_0=0$. 用HAM, PSM, MESA和SRPT 4种方法分别进行周期成分识别, 结果见图2.

由试验结果可以看出, 各方法的周期识别效率均表现出随变差系数增大而减小的规律. 变差系数 C_v 反映序列的离散程度, C_v 越大, 序列的离散程度越大, 所以纯随机成分的随机波动性越强. 对比图1和图2可以看出两图呈现出了较一致的变化规律, 这与方差受均值和变差系数影响的规律是一致的. 相比而言, PSM与MESA方法随变差系数增大检验效率的变化规律相近, 均呈现出显著降低的趋势, 在 C_v 大于0.2后, PSM的检验效率已经低于50%. 而SRPT方法和HAM则表现出了较好的稳定性. 随着变差系数的增大, 相关系数在降低, 表示序列中周期成分的显著性在减弱, 在变差系数大于0.25后, 相关系数已低于临界值下限, 说明序列中的周期变异不显著, 因此导致了SRPT方法的识别效率降低. 在这4种方法中SRPT方法的稳定性最好.

2.3 周期成分参数的影响

为了分析叠加的不同周期成分对周期识别效率的影响, 固定纯随机成分, 设纯随机成分序列长度 n 为200, 均值 $\bar{x}$ 为200, 变差系数 $C_{vu}=0.2$, 偏态系数 $C_{su}=2C_{vu}$, 而周

期表达式 $y_t = B\sin\left(\frac{2\pi}{T}t + t_0\right)$ 中的参数不断变化, 分析其对各方法周期识别效率的影响.

2.3.1 振幅 B

设定纯随机序列的振动幅度为

$$A = |\text{quantile}(x_t, 3/4) - \text{quantile}(x_t, 1/4)|, \quad (10)$$

式中, $\text{quantile}(x_t, 3/4)$, $\text{quantile}(x_t, 1/4)$ 分别为纯随机项 x_t 的上下四分位数.

周期项振幅 B 在变化时应与随机项主要振动幅度 A 具有可比性, 所以设定 $B = A \times M$, 其中 M 为倍比系数, 取值0.5, 0.75, 1, 1.25, 1.5, 周期 $T=25$, 初相位 $t_0=0$, 统计试验结果整理见图3(a). 可以看出, 各方法的周期识别效率随半振幅的增大而逐渐提升. 变化半振幅, 改变的是周期波动在序列中的显著程度, 随着半振幅的增大, 周期成分在序列中占的比重增大, 更容易被识别出来. PSM在这4种方法中效果最差, MESA方法在半振幅较小时识别效率低, 随着半振幅增加到1.5 A , 识别结果与SRPT方法和HAM相近, 4种方法中, 表现最好的是SRPT方法, 且略优于HAM.

2.3.2 周期 T

周期长度的取值和前面样本长度取值考虑的因素

图2 (网络版彩色)不同的变差系数下各方法周期识别效率及相关系数

Figure 2 (Color online) Period identification efficiency and correlation coefficient of different methods under different coefficients of variation

类似, 为了体现出周期波动, 在样本长度固定为200的情况下, 周期 T 取了5, 10, 15, 20, 25, 30共6个值, 半振幅 $B=30$, 初相位 $t_0=0$. 统计试验结果见图3(b). 当样本长度固定时, 随周期长度增加, 序列中的周期个数是减少的, 识别难度也在增加. 对于PSM, 最大滞后时间 m 的取值对识别结果影响很大, 对于样本长度为200的序列, 除在周期长度为 $T=200/10=20$ 时表现出了很高的识别效率外, 整体的规律是随周期长度的增加, 识别效率降低. MESA方法的识别效率也随周期长度的增加而降低, 但较PSM平缓. PSM和MESA法共同的缺点是不能同时兼顾高频和低频的需求, 因此在短周期情况下识别精度较高, 当周期长度较长时会得到与真实周期偏离较大的伪周期成分, 较难识别出真实周期. 对于HAM和SRPT两种方法, 识别效率在所取的周期长度中未受到周期长度的影响, 两者均为100%, 为了进一步分析两种方法在周期长度变化时的表现, 给出试验过程中100组数据的箱型图(图4). 可以看出, HAM在周期长度为15时, 整体识别结果偏大, 周期为30时, 整体识别结果偏小, 而SRPT方法不仅识别效率高, 且更精准, 识别结果的整体水平与设定的周期长度是一致的, 说明此方法更为精确.

2.3.3 初相位 t_0

初相位的取值范围是 $-\pi \leq t_0 \leq \pi$, 从中选取 $0, \pm\frac{\pi}{4}, \pm\frac{\pi}{2}, \pm\frac{3\pi}{4}, \pi$ 共8个值, 半振幅 $B=30$, 周期 $T=25$. 统计试验结果见图3(c). 由图3(c)可以看出初相位的变化对识别效率影响不大, SRPT方法和HAM的识别效率在不同的初相位下均达到了100%. 由于允许误差设定较小, MESA的识别效率在70%上下浮动, 以箱型图的形式给出了MESA和PSM的试验数据(图5), 当扩大允许误差范围至 $\Delta T = |T - T'| = 2$ 时, MESA的识别效率可提升至85%~90%, 说明MESA方法可以识别出周期, 但是识别结果不够精确; 对于PSM, 其识别效率在45%上下波动, 由图5可以看出, PSM识别出的周期总体较25偏小, 当扩大允许误差范围至 $\Delta T = |T - T'| = 2$, PSM的识别效率可达到60%左右.

综上, 通过对SRPT方法的检验效率的验证与比较, 结果表明, 在常用的周期识别方法中, SRPT方法有较高的检验效率, 且此方法的稳定性较高, 不受样本长度、周期长度和初相位的影响. 在均值和变差系数等参数发生变化且周期变异不显著的情况下, 检验效率有所降低, 但相较其他方法仍表现出优越性. 此外, 对

图3 (网络版彩色)不同周期成分参数下各方法的周期识别效率。(a) 半振幅; (b) 周期长度; (c) 初相位

Figure 3 (Color online) Period identification efficiency of different methods under different periodic component parameters. (a) Semi amplitude; (b) period length; (c) initial phase

于PSM, 由于最大时滞对周期识别的影响, 本文选定的识别准则得出的识别效率结果仅是为了在相同的标准下与其他方法形成参照, 在实际应用中周期的判定需依具体情况而定。

3 实例分析

本文选取澜沧江流域作为研究区域。澜沧江-湄公河在中国境内河段称为澜沧江, 位于中国的西南地区, 流经青海、西藏和云南三省, 自北向南呈条带状, 流域内地势北高南低。澜沧江上段径流以降雨补给为主, 兼

有融雪补给, 下段主要为降雨补给, 产水量稳定, 属西部多水地区。本文选取了分布在澜沧江流域的21个气象站1961~2014年共54 a的年降雨序列, 对其周期特性进行分析。

对于原始的年降雨序列, 首先应用滑动 F 检验法、Pettit方法、Spearman秩次相关检验法和Mann-Kendall法等方法识别并检验序列是否发生了趋势或跳跃变异, 其中显著性水平分别选取 $\alpha = 0.05$, $\beta = 0.01$ 。结果表明发生变异的序列, 其变异形式均为跳跃变异, 具体变异结果见表2(只列出发生变异的序列)。

将发生了跳跃变异的6个序列剔除跳跃成分后得到剩余序列^[24], 然后进行周期分析, 并选定显著性水平, 对相关系数进行显著性检验; 若检验结果显著, 说明此周期为显著周期, 否则认为该站点没有发生周期变异。同样应用SRPT, HAM, MESA和PSM四种方法对各序列进行周期分析。为了更精确地描述水文序列的周期特性, 其中SRPT方法的步长取 $l_1 = 0.1$, $l_2 = 0.0001\pi$ 。选用显著性水平 $\alpha = 0.05$, $\beta = 0.01$, 21个站点的年降雨序列长度为1961~2014年共54 a, 对应的相关系数临界值分别为 $r_\alpha = 0.2681$, $r_\beta = 0.3477$, 依据基于相关系数的周期变异分级方法^[21], 得到该站的周期变异程度, 分级标准见表3。若变异程度为中变异及以上, 则认为序列发生了周期变异。

下面以澜沧站为例, 应用4种周期识别方法得到的结果见表4。

分析表4可以看出, PSM在不同的最大时滞下给出的多个结果中, 依据其周期判定量-功率谱值最大, 得出的最可能的主周期为2 a, 此时的相关系数为0.3109, 为弱周期变异, 说明此周期成分与原始序列的波动性不太吻合; MESA方法得到的可能周期也为2 a; HAM的周期判定量是 F 统计量值, 取其最大对应的周期为18 a, 相关系数为0.3994, 为中等变异; 由SRPT方法也得到了18 a的周期, 在上述方法中最大, 说明此周期波动在原序列中最为显著, 因此澜沧站的周期定为18 a。给出澜沧站的周期成分与原始序列(图中数据均为模比系数)的匹配情况见图6, 可以直观看出拟合的周期成分与原序列周期波动较吻合, 峰值峰谷都能较好地对应, 也说明了周期识别结果的合理性。

同理得到其余20站的年降水序列的周期, 结果见表5。

由表5可以看出, 除囊谦、巴塘和维西站的周期变

图4 不同周期长度下的周期识别结果. (a) SRPT方法; (b) HAM

Figure 4 Period identification result under different period lengths. (a) SRPT method; (b) HAM

图5 不同的初相位下的周期识别结果. (a) MESA方法; (b) PSM

Figure 5 Period identification result under different initial phases. (a) MESA method; (b) PSM

表2 年降水量序列跳跃变异识别结果^{a)}

Table 2 Identification of jump variation of annual average precipitation series

站点	变异情况
囊谦	1997↑
贡山	1971↑
丽江	2004↓和1997↑
大理	2004↓
耿马	2001↓
景洪	1973↓

a) 1997↑表示在1997年发生了向上的跳跃变异, 其余同理. 其中丽江检验出两个跳跃点

异为弱变异外, 其余站点均为中变异. 对这些变异程度为中变异的站点剔除周期成分后, 再利用Hurst系数和Bartels统计量联合检验^[25]判断序列是否为纯随机, 若为纯随机说明序列只含有这1个周期成分; 若不为纯随机成分, 则继续使用本法进行周期循环识别直至剩余序列为纯随机成分为止. 检验结果显示剔除周期成分后的剩余序列均为纯随机序列. 对澜沧江流域的周期

表3 年降水序列周期变异分级标准

Table 3 The gradation criterion of period variation of annual precipitation series

相关系数 r 下限~上限	周期变异程度
[0.0000, 0.2681]	无变异
[0.2681, 0.3477]	弱变异
[0.3477, 0.6000]	中变异
[0.6000, 0.8000]	强变异
[0.8000, 1.0000]	巨变异

变异情况整理如图7所示.

根据结果, 澜沧江流域北部8个站点(杂多、玉树、石渠、囊谦、丁青、德格、昌都和巴塘)中, 囊谦和巴塘不存在明显的周期成分, 其余站点中除德格和丁青两站存在12 a左右的周期震荡外, 多数站点的降雨周期在2~5 a之间, 说明这些站点的降雨短期变化较为剧烈; 中部区域的6个站点(德钦、贡山、维西、丽江、保山和大理)中, 维西站不存在明显的周期成分, 其余站点的降雨周期均在2~3 a左右, 说明中部的降雨也存

表 4 各方法对澜沧站年降雨序列的周期识别结果^{a)}

Table 4 Period identification results of annual average precipitation series at Lancang station by various methods

周期分析方法	周期	周期判定量	相关系数	变异程度
PSM	2	0.2348	0.3109	弱
	2	0.1940	0.3109	弱
	14	0.1731	0.1069	无
	16	0.1581	0.3149	弱
	18	0.1455	0.3994	中
	20	0.1346	0.3347	弱
	2	0.1279	0.3109	弱
	2	0.1252	0.3109	弱
	2	0.1229	0.3109	弱
MESA	2	0.9955	0.3109	弱
HAM	18	4.8407	0.3994	中
	2.8	3.2323	0.3354	弱
SRPT	18	0.3994	0.3994	中

a) PSM的周期判定量为功率谱值, MESA的周期判定量为最大熵谱值, HAM的周期判定量为 F 检验值, SRPT方法的周期判定量为相关系数值

在短期变化剧烈的特点, 或与中部地形复杂有关^[26], 剩余分布在南部的7个站点(景东、耿马、双江、澜沧、

思茅、景洪、勐腊)中, 双江站的降雨周期最短为2.3 a, 其次是景洪站为3.4 a, 澜沧站的降雨周期最长为18 a, 其余站点周期长度在5~10 a间, 在整个流域的周期长度中处于中长尺度水平. 从整个流域来看, 2~5 a尺度的周期变化在整个流域均有分布, 尤其是北部和中部较为集中, 说明整个流域的降雨量均表现出短时间尺度上的变化规律; 5~10 a的周期主要集中在南部, 10~20 a左右的长时间尺度的周期在全流域也有分布, 但数量较少, 故澜沧江的降雨量存在多时间尺度特征.

4 结论

本文针对水文序列中含有的周期成分, 选取相关系数作为评判指标, 提出了一种水文序列周期变异识别方法. 该方法通过计算水文序列与周期成分之间的相关系数来判断滑动过程中的周期是否为序列中的真实周期. 由公式推导得出可能影响周期识别的多个因素, 包括样本长度、纯随机成分的均值、变差系数和周期成分中的振幅、周期长度以及初相位. 经分析得到以下结论和认识:

(1) 统计试验验证了本文所提方法的可靠性和在多种因素影响下的稳定性. 在此基础上, 对比了各变量

图 6 (网络版彩色)澜沧江流域澜沧站周期成分拟合图

Figure 6 (Color online) The curve fitting of periodic components of Lancang station in Lancang River Basin

表5 澜沧江流域各站点滑动周期识别结果

Table 5 Identification results of slip period of stations in Lancang River Basin

站码	站名	相关系数	周期变异程度	主周期成分
56018	杂多	0.3816	中	4.9
56029	玉树	0.4531	中	2.2
56038	石渠	0.3864	中	2.2
56116	丁青	0.4003	中	12.1
56125	囊谦	0.3374	弱	2.6
56137	昌都	0.4047	中	2.8
56144	德格	0.3529	中	12.1
56247	巴塘	0.3265	弱	2.1
56444	德钦	0.3558	中	2.0
56533	贡山	0.3949	中	2.5
56548	维西	0.3295	弱	3.1
56651	丽江	0.4206	中	3.0
56748	保山	0.3643	中	2.8
56751	大理	0.4358	中	5.9
56856	景东	0.4073	中	8.5
56946	耿马	0.3747	中	5.7
56950	双江	0.3916	中	2.3
56954	澜沧	0.3994	中	18.0
56959	景洪	0.4700	中	3.4
56964	思茅	0.3690	中	8.1
56969	勐腊	0.4837	中	6.0

对常用的3种周期识别方法识别效率的影响。结果表明,PSM和MESA方法受样本长度的影响较大,而HAM和SRPT方法在样本长度变化时表现较稳定;4种方法的识别效率均随均值和变差系数的增大导致的周期变异不显著而降低;周期成分参数中,随振幅的增大4种方法的识别效率均有所增加,PSM和MESA方法的识别效率随周期长度增加而降低,HAM和SRPT方法较稳定,初相位基本不影响各种方法的检验效率。总体而言,在统计试验讨论的多种影响因素中,SRPT方法的识别效

参考文献

- Wang W S, Ding J, Jin J L. Statistical Hydrology (in Chinese). 2nd Ed. Beijing: China Water & Power Press, 2008. 38–40 [王文圣, 丁晶, 金菊良. 随机水文学. 第2版. 北京: 中国水利水电出版社, 2008. 38–40]
- Zhang S Q, Ding S S. A Simplified Method of Time Series Forecasting (in Chinese). Proceedings of the 3rd National Climate Conference. Beijing: China Meteorological Press, 1964. 36–39 [章少卿, 丁士晟. 一种简化了的时间序列预报方法的讨论. 第三届全国气候学术会议论文集. 北京: 气象出版社, 1964. 36–39]

图7 (网络版彩色)澜沧江流域站点分布图及周期变异情况

Figure 7 (Color online) Distribution and periodic variation of stations in Lancang River Basin

率稳定且高,体现了基于相关系数建立的SRPT方法的可行性。

(2) 将该周期识别方法应用于澜沧江流域多站的降雨量分析,结果表明北部和中部的降雨周期较小,降雨量短期变化较为剧烈,南部的周期长度多在5~10 a左右,个别站点周期长度较长,说明澜沧江流域的降雨量存在多时间尺度波动特征。

(3) 传统周期识别方法在识别水文时间序列的复合周期和近似周期时还存在很多问题。为克服相关方法的不足,可尝试以此方法为核心并结合非平稳时间序列的分析方法(如EMD等)识别和描述其他周期成分,但相关问题仍需进一步研究。

- 3 Thomson D J. Spectrum estimation and harmonic analysis. *Proc IEEE*, 2005, 70: 1055–1096
- 4 Zhao Z C, Wang S W. Periodicity of drought and flood evolution in China in recent 100 years (in Chinese). *Meteorol*, 1979, 5: 19–22 [赵宗慈, 王绍武. 近百年我国旱涝演变的周期性. *气象*, 1979, 5: 19–22]
- 5 Burg J P. A new analysis technique for time series data. NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics. 1968, 12–13
- 6 Xiao Z G. A comparative study of several periodic analysis methods for hydrological time series (in Chinese). Master Thesis. Nanjing: Hohai University, 2006. 15–38 [肖志国. 几种水文时间序列周期分析方法的比较研究. 硕士学位论文. 南京: 河海大学, 2006. 15–38]
- 7 Zhuang C Q, He C X. Basis of Applied Mathematical Statistics (in Chinese). 3rd ed. Guangzhou: South China University Technology Press, 2006 [庄楚强, 何春雄. 应用数理统计基础. 第3版. 广州: 华南理工大学出版社, 2006]
- 8 Pan W J. Fourier Analysis and Application (in Chinese). Beijing: Peking University Press, 2000. 84–94 [潘文杰. 傅里叶分析及其应用. 北京大学出版社, 2000. 84–94]
- 9 Yang H. Research of hydrological time series cycle analysis method (in Chinese). *Chin Water Power Electr*, 2015, (5): 63–66 [杨华. 水文时间序列周期分析方法的研究. *中国水能及电气化*, 2015, (5): 63–66]
- 10 Cao H X, Luo Q L. Maximum entropy spectrum analysis of historical meteorological series (in Chinese). *Chin Sci Bull*, 1979, 24: 351–355 [曹鸿兴, 罗乔林. 气象历史序列的最大熵谱分析. *科学通报*, 1979, 24: 351–355]
- 11 Padmanabhan G, Rao A R. Maximum entropy spectral analysis of hydrologic data. *Water Resour Res*, 1988, 24: 1519–1533
- 12 Sang Y F, Xie P, Gu H T, et al. Discussion on several major issues in the studies of hydrological nonstationarity (in Chinese). *Chin Sci Bull*, 2017, 62: 254–261 [桑燕芳, 谢平, 顾海挺, 等. 水文过程非平稳性研究若干问题探讨. *科学通报*, 2017, 62: 254–261]
- 13 Pettitt A N. A non-parametric approach to the change-point problem. *J R Stat Soc*, 1979, 28: 126–135
- 14 Chen G C, Xie P. Slide F test of change-point analysis (in Chinese). *J Chin Hydrol*, 2006, 26: 57–60 [陈广才, 谢平. 水文变异的滑动F识别与检验方法. *水文*, 2006, 26: 57–60]
- 15 Mann H B. Nonparametric tests against trend. *Econometrica*, 1945, 13: 245–259
- 16 Wu X Z, Wang Z J. Nonparametric Statistical Methods (in Chinese). Beijing: Higher Education Press, 1996. 106–108 [吴喜之, 王兆军. 非参数统计方法. 北京: 高等教育出版社, 1996. 106–108]
- 17 Hirsch R M, Helsel D R, Cohn T A, et al. Statistical analysis of hydrologic data. In: *Handbook of Hydrology*. New York: McGraw-Hill, 1993. 17.1–17.55
- 18 Huang N E, Zheng S, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Mathematical, Physical and Engineering Sciences*, 1998, 454: 903–995
- 19 Tang J. The application of power spectrum analysis in periodic analysis (in Chinese). *J Shaanxi Univ Technol (Nat Sci Ed)*, 2013, (5): 71–74 [唐洁. 功率谱分析方法在周期分析中的应用. *陕西理工学院学报(自然科学版)*, 2013, (5): 71–74]
- 20 Liu J J, Xie P, Xu B, et al. Hydrological cycle variation classification method based on maximum entropy spectrum and coefficient of correlation (in Chinese). *Water and Regional Sustainable Development - Proceedings of the Ninth Chin Water Forum*. 2011. 273–276 [刘静君, 谢平, 许斌, 等. 基于最大熵谱和相关系数的水文周期变异分级方法. *水与区域可持续发展——第九届中国水论坛论文集*. 2011. 273–276]
- 21 Xie P, Zhao Y X, Sang Y F, et al. Correlation coefficient-based method for gradation of significance level of periodicities in hydrologic time series and its verification (in Chinese). *J Hydroelectr Eng*, 2018, 37: 33–43 [谢平, 赵羽西, 桑燕芳, 等. 基于相关系数的水文周期变异分级方法及验证. *水力发电学报*, 2018, 37: 33–43]
- 22 Wang Q, Mo Z S. Discussions on some key problems in power spectra analysis (in Chinese). *J Geol*, 1990, 57: 363–372 [王强, 么枕生. 功率谱分析中几个重要问题的讨论. *地理学报*, 1990, 57: 363–372]
- 23 Wu Y X, Dong L H, Wang D, et al. Analysis of precipitation cycles in Zhalong wetland based on power spectral (in Chinese). *J Qiqihar Univ (Nat Sci Ed)*, 2014, (3): 79–82 [吴月祥, 董立红, 王丹, 等. 基于功率谱的扎龙湿地降水周期分析. *齐齐哈尔大学学报(自然科学版)*, 2014, (3): 79–82]
- 24 Xie P, Xu B, Zhang S A, et al. Study on Regional Water Resources Variability under Changing Environment (in Chinese). Beijing: Science Press, 2012. 62–65 [谢平, 许斌, 章树安, 等. 变化环境下区域水资源变异问题研究. 北京: 科学出版社, 2012. 62–65]
- 25 Li B B, Xie P, Li X N, et al. Joint analysis method of hydrological variation based on Hurst coefficient and Bartels test (in Chinese). *J Bas Sci Eng*, 2014, (3): 481–491 [李彬彬, 谢平, 李析男, 等. 基于Hurst系数与Bartels检验的水文变异联合分析方法. *应用基础与工程科学学报*, 2014, (3): 481–491]
- 26 Chen S J, Li L J, Li J Y, et al. Spatial and temporal variations of precipitation over the Lancang River Basin in recent 55 years (in Chinese). *J Geo-Info Sci*, 2017, 19: 365–373 [陈素景, 李丽娟, 李九一, 等. 近55年来澜沧江流域降水时空变化特征分析. *地球信息科学学报*, 2017, 19: 365–373]

Summary for “基于相关系数的水文序列滑动周期识别方法”

Moving correlation coefficient-based method for the identification of periodicities in hydrologic time series

Linqian Wu^{1*}, Ping Xie^{1,2*}, Ziyi Wu¹, Yanfang Sang³, Jie Chen¹, Yuxi Zhao⁴, Yaqing Li¹ & Lu Wang¹

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China;

² Collaborative Innovation Center for Territorial Sovereignty and Maritime Rights, Wuhan 430072, China;

³ Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;

⁴ State Grid Corporation Southwest Branch, Chengdu 610000, China

* Corresponding authors, E-mail: wq507@whu.edu.cn; pxie@whu.edu.cn

Accurate identification of periodicities is an important issue for understanding the hydrological variability, medium- to long-term hydrologic simulation and forecasting, as well as water security assessment. Many methods have been developed presently for handling the issue, including the simple partial wave method (SPWM), harmonic analysis method (HAM), power spectrum method (PSM) and maximum entropy spectral analysis (MESA). However, the results of periodic identification are not good as expected due to the complexity of hydrological variability and defects of methods. In this article, a new moving correlation coefficient-based method is proposed for the identification of periodicities in hydrological time series. The specific steps of periodicity identification by the proposed methods is described as follows: (1) Generate a simulated periodic series by using the sinusoidal function; (2) use the correlation coefficient as an index to evaluate the fitting accuracy between the simulated periodic series and the original series, which generally includes the periodic components and stochastic components; (3) change the periodic value and the initial phase of the sinusoidal function, and repeat the above steps to calculate the correlation coefficient values; and (4) find the maximum magnitude of the correlation coefficient, and its corresponding simulated periodic series is regarded as the expected periodic component, and its significance is evaluated by doing hypothesis test. Theoretical analysis and formula derivation demonstrate that the index of correlation coefficient can reflect the significant degree of periodic variability of the original hydrological time series. Bigger correlation coefficient value corresponds to more significant periodic variability. Results of Monte-Carlo experiments confirm the advantages of the proposed method compared with other conventional methods such as HAM, PSM and MESA. The former has high identification efficiency and can keep its stability, even if encountering the influences of some advance factors, including the sample length, mean value, coefficient of variation, the semi amplitude, period length and initial phase. By applying the proposed method to analyze the annual precipitation series at 21 meteorological stations in the Lancang River Basin, results indicate that the annual precipitation presents periodic variations at multi-time scales; comparatively, annual precipitation series in the northern and central parts of the basin mainly presents a 2- to 5-year periodic variability, versus the 5- to 10-year periodic variability in the south part.

periodicity identification, correlation coefficient, inconsistency, hydrologic variability, Lancang River

doi: [10.1360/N972018-01192](https://doi.org/10.1360/N972018-01192)