

国内水文序列变异点分析方法及 在各流域应用研究进展

周园园^{1,2}, 师长兴¹, 范小黎^{1,2}, 杜俊^{1,2}

(1. 中国科学院陆地水循环及地表过程重点实验室, 中国科学院地理科学与资源研究所, 北京 100101;

2. 中国科学院研究生院, 北京 100049)

摘 要:变异点识别在区分气候变化和人类活动对水文过程的影响中具有十分重要的作用。近几十年来, 国内水文序列变异点的研究逐渐发展, 从最早的使用单一方法发展到由多种方法组成的水文序列变异点诊断系统和对部分方法的改进。同时, 各种方法在中国各主要流域得到广泛应用, 但是使用的方法不同, 水文序列时间年限不同, 得到的变异点的结果也不完全一致。本文将目前普遍采用的方法及在各流域的应用成果进行了系统总结, 为今后水文序列变异点的方法改进及其在流域中的应用发展提供参考。

关 键 词:时间序列分析; 变异点识别; 水文序列检验

1 引言

水文序列受气候变化和人类活动影响, 改变其原来的变化趋势, 为了区分人类活动和气候变化对水文序列的影响程度, 尤其是量化人类活动的影响, 研究中普遍使用的方法就是对序列进行趋势性检验, 找到使得水文序列前后变化不一致突变点。

黄河中下游径流量近几十年来发生了显著的变化, 其原因引起了众多学者的关注。黄土高原20世纪70年代大规模实施水土保持措施, 因此有研究者^[1-3]将60年代及以前的时期划分为基准期, 70年代及以后为措施期。事实上, 水土保持措施实施的初期并不能起到很大效果, 因此, 采用具物理机制的统计方法寻找突变点, 显得更具科学依据和说服力。目前国内对水文序列突变点的研究很多, 研究者采用不同的方法对各自的研究区进行分析, 并对方法不断改进, 比如将多种方法同时使用, 使得结果更准确更具说服力。本文将目前普遍采用的方法进行综合比较, 对以前研究者^[4-5]统计的方法进行补充, 另外将各流域应用情况进行总结, 目的是为今后水文序列变异点的方法发展及流域应用发展提供参考。

2 方法原理

水文序列变异诊断分为初步诊断、详细诊断两个阶段: 对水文序列进行初步分析时, 可以采用过程线法、累(差)积曲线和滑动平均^[6-8]等来检验序列的趋势性, 判断序列是否存在变异; 如果该序列可能存在变异, 则运用检验方法进行详细诊断, 结合流域水文过程可能变异点的成因分析, 从而得到最可能变异点。

检验方法可分为参数和非参数检验, 参数检验以总体分布和样本信息对总体参数作推断, 对总体有特殊的要求, 如t检验、F检验等; 非参数检验不需要利用总体的信息, 以样本信息对总体分布作出推断, 如秩和检验等。在水文序列检验时, 非参数检验, 对总体分布所加的条件较少, 较简单、实用, 因此应用也相对较广。

2.1 F检验、滑动F检验法、t检验、滑动t检验

这几种方法是统计学中常见和基本的方法, 都是通过分析时间序列可能变异点前后两个样本的方差构造T或F统计量, 判断计算的统计量值是否通过指定置信水平的检验, 从而判断时间序列是否存在变异点, 方法原理不再详述^[9-10]。

收稿日期: 2011-03; 修订日期: 2011-06.

基金项目: 国家重点基础研究发展计划项目(2010CB428404)。

作者简介: 周园园(1982-), 女, 河北石家庄人, 博士研究生, 主要从事河流地貌与河道演变研究。

E-mail: zhouyuanwinner@163.com

通讯作者: 师长兴(1963-), 男, 河北正定人, 研究员, 博士生导师, 主要从事流域物质迁移与环境变化。E-mail: shicx@igsnr.ac.cn

2.2 R/S 分析方法与 Hurst 指数

R/S 分析方法 (Rescaled Range Analysis), 即分形理论中的重新标度极差分析方法, 是 1965 年由英国学者 Hurst 提出的一种处理时间序列的方法, 在分形理论中有着重要的作用, 2002 年王孝礼等^[1]将该法应用在水文序列变点分析中, 给出了水文序列变点的定义。基本思想: 分析序列的统计特性变化规律, 以改变序列的时间尺度, 从而将小尺度的规律用于大尺度范围, 或者将大尺度的规律用于小尺度范围。其原理^[1]如下: 时间序列 $x(t)$ 的极差和标准差分别为:

$$\begin{aligned} R(\tau) &= \max_{1 \leq t \leq \tau} x(t, \tau) - \min_{1 \leq t \leq \tau} x(t, \tau) \\ S(\tau) &= \sqrt{(1/\tau) \sum [x(t) - (\bar{x})_\tau]^2} \end{aligned} \quad (1)$$

式中: 均值 $(\bar{x})_\tau = (1/\tau) \sum_{t=1}^{\tau} x(t)$, $1 \leq \tau \leq n$, n 为样本容量; 累积离差 $x(t, \tau) = \sum_{\mu=1}^t [x(\mu) - (\bar{x})_\tau]$, $1 \leq t \leq \tau$ 。

当 $x(t)$ 是相互独立、方差有限的随机序列, 即布朗运动时, Hurst 等人证明 $R(\tau)/S(\tau) = [(\pi/2)\tau]^H$, Hurst 指数 $H = 1/2$ 。

当 $x(t)$ 不是相互独立的分数布朗运动时可以证明 $R(\tau)/S(\tau) = (c\tau)^H$, c 为常数, 将等式两边取对数则得到: $\ln R(\tau)/S(\tau) = H \ln c + H \ln t$ 。在 $\ln R(\tau)/S(\tau)$ 和 $\ln t$ 的双对数坐标系下作散点图, 用最小二乘法得到散点的拟合直线, 则该直线的斜率即为 H 。

当 $H = 1/2$ 时, 表明序列未来与过去无关, 或短程相关; 当 $H > 1/2$ 时, 表明未来的趋势与过去一致, 即该过程具有持续性, 且 H 越接近 1, 持续性越强; 当 $H < 1/2$ 时, 表明未来的趋势与过去相反, 即该过程具有反持续性, 且 H 越接近 0, 反持续性越强。R/S 法正是应用这个特点寻找序列变异点, 由于 H 随时间在变化, 因此将 H 发生最大变化时的点称为变异点。步骤为^[12]首先用 R/S 法分析前后两个序列 H 得到 H_1 、 H_2 , 再求出 $\Delta H = |H_1 - H_2|$, 最后逐一比较 ΔH , 最大的 ΔH 所对应的点就是该序列的变异点。R/S 分析法所需要的样本较大, 周寅康等^[13]研究发现样本个数大于 400 时 H 值趋于一个稳定值。

2.3 Brown-Forsythe 检验法

Brown-Forsythe 检验法是^[14-15]1974 年 Brown 和 Forsythe 提出的对单因子方差分析法 (ANOVA) 的改进方法, 克服了实践中较难满足的单因子方差分析法对样本要求正态分布、不同组数据的方差相等以及各组样本数量不能相差太大等条件, 区别就在

于统计量 F 的计算上, 方法如下:

$$F = \frac{\sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2 / \sum_{i=1}^m (1 - n_i/N) S_i^2}{\sum_{i=1}^m (1 - n_i/N) S_i^2} \quad (2)$$

式中: m 为分组数; n_i 为第 i 组中的样本数; N 为样本总数; \bar{x}_i 为第 i 组样本均值; \bar{x} 为样本总均值, S_i^2 为第 i 组的样本方差。统计量服从自由度为 $(m-1, f)$ 的 F 分布, 其中:

$$f = (n_i - 1) / \left(\frac{\sum_{i=1}^m (1 - n_i/N) S_i^2}{\sum_{i=1}^m [(1 - n_i/N) S_i^2]} \right)^2 \quad (3)$$

F 越大, 表明不同组间水平差异越大; 在给定的显著性水平 α 下, 若有 $F > F_\alpha$, 则认为不同组间水平差异明显。在实际应用中, 通常是首先采用趋势分析法分析序列的变化趋势, 然后用突变分析法分析序列的可能突变时间点, 最后结合流域的实际情况去掉非突变点, 得到合理的突变点。

张一驰等^[16]研究证明 Brown-Forsythe 检验法比单因子方差分析法在一定程度上放宽了对数据正态性、数据长度的要求, 特别是不同组数据方差需相同的要求, 因此当实测数据一定程度上违背一般统计方法的假设要求时, 仍可得到准确的识别结果, 使变异点识别更为准确迅速。

2.4 Bayesian 变点分析法

贝叶斯 (Bayesian) 变点分析法^[17-18]的原理如下: 假设序列 $x(t)$ 在时刻 t 前后的统计参数发生变化, t 时刻前后两部分所服从分布密度函数分别为^[19-20]:

$$\begin{aligned} x_{1i} &\rightarrow N(\mu_1, \sigma_1^2) \quad i = 1, 2, \dots, t \\ x_{2i} &\rightarrow N(\mu_2, \sigma_2^2) \quad i = t+1, t+2, \dots, n \end{aligned} \quad (4)$$

若只关注序列均值是否发生变化, 则可假设方差不变, 即 $\sigma_1^2 = \sigma_2^2 = \sigma^2$, 其中 σ^2 可以由序列的观测值来估计。由于不知道 μ_1 和 μ_2 的先验分布函数, 所以一般假设 μ_1 和 μ_2 服从相同的正态分布。

由贝叶斯定理和序列值 $R = \{R_t, R^{t+1}\}$ 可推导出分布参数 μ_1 和 μ_2 的后验分布。

根据观测信息 R_t 推导出 μ_1 的后验分布:

$$\mu_1 | x_{1i} \rightarrow N(\mu_1, \sigma_1^2), \quad \mu_1 = (n\mu_1 + \sum_{i=1}^t R_i) / (n+t) \quad (5)$$

$$\sigma_1^2 = \sigma^2 / (n+t), \quad n = \sigma^2 / \sigma_\mu^2$$

根据观测信息 R^{t+1} 推导出 μ_2 后验分布为:

$$\begin{aligned} \mu_2 &= (n\mu_2 + \sum_{i=t+1}^n R_i) / [n + (n-t)] \\ \sigma_2^2 &= \sigma^2 / [n + (n-t)] \end{aligned} \quad (6)$$

μ_1 和 μ_2 在给定情况下的水文序列值 $R = \{R_t, R^{t+1}\}$ 发生的联合分布密度函数:

$$x(R|t, \mu_1, \mu_2) =$$

$$\prod_{i=1}^t \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(r_i - \mu_1)^2}{2\sigma^2}\right) \times \prod_{i=t+1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(r_i - \mu_2)^2}{2\sigma^2}\right) \quad (7)$$

然后由贝叶斯法则推导出变点发生位置的后验分布密度函数:

$$x(t|R, \mu_1, \mu_2) = \frac{x(t|R, \mu_1, \mu_2)x(t)}{\sum_{i=1}^n x(R|i, \mu_1, \mu_2)x(t)} \quad (8)$$

式中: $x(t)$ 代表位置 t 的先验概率分布, 一般假定为均匀分布, 即 $x(t) = 1/n$, ($t = 1, 2, \dots, n$), 变点 t 的发生位置的后验概率密度函数可简化为:

$$x(t|R, \mu_1, \mu_2) = \frac{x(t|R, \mu_1, \mu_2)}{\sum_{i=1}^n x(R|i, \mu_1, \mu_2)} \quad (9)$$

变点发生位置 k 的后验发生概率的期望值为:

$$x(t|R) = \iint x(t|R, \mu_1, \mu_2) d\mu_1 d\mu_2 \quad (10)$$

最后由检验出的多个可能变异点, 进行成因分析, 确定时间序列 $x(t)$ 变异点。

2.5 Lee-Heghinian 法

里海哈林(Lee-Heghinian) 法^[2,21]是 1977 年 Lee 和 Heghinian 提出的基于贝叶斯理论的方法, 原理如下: 对于水文时间序列 $x(t)$, 假定总体为正态分布, 且可能变异点 τ 的先验分布为均匀分布的情况下, 推得 τ 的后验分布为:

$$f(\tau|x_1, x_2, \dots, x_n) = k[n/\tau(n-\tau)]^\varphi [R(\tau)]^{(n-2)/2} (1 \leq \tau \leq n-1) \quad (11)$$

$$R = \left[\sum_{i=1}^{\tau} (x_i - \bar{x}_{\tau})^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2 \right] / \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

式中: k 为比例常数; n 为样本容量。由后验分布, 将满足 $\max_{1 \leq \tau \leq n-1} \{f(\tau|x_1, x_2, \dots, x_n)\}$ 条件的 τ 记为 τ_0 , 作为可能变异点, 最后结合变异点的成因分析, 得出最可能的变异点。

2.6 Pettitt 检验法

Pettitt 检验法最先由 Pettitt 用于检验突变点, 是一种非参数检验方法, 前提是序列存在趋势性变化, 其核心是通过检验时间序列要素均值变化的时间, 来确定序列跃变时间。该检验使用 Mann-Whitney 的统计量 $U_{t,N}$ 来检验同一个总体 $x(t)$ 的两个样本, 统计量 $U_{t,N}$ 的公式^[22-23]:

$$U_{t,N} = U_{t-1,N} + \sum_{i=1}^N \text{sgn}(x_t - x_i), t = 2, 3, \dots, N \quad (12)$$

式中: 若 $x_t - x_i > 0$, 则 $\text{sgn}(x_t - x_i) = 1$; 若 $x_t - x_i = 0$, 则 $\text{sgn}(x_t - x_i) = 0$; 若 $x_t - x_i < 0$, 则 $\text{sgn}(x_t - x_i) = -1$ 。

Pettitt 检验的零假设为序列无变异点, 其统计量

K_N 和相关概率的显著性检验公式如下:

$$K_{t,N} = \max |U_{t,N}|, (1 \leq t \leq N) \\ p \approx 2 \exp\{-6K_{t,N}/N^3 + N^2\} \quad (13)$$

式中: 若 $p < 0.05$, 则认为 t 点为显著变异点。由此检验出序列的一级变点; 然后以变点为界将原系列分为两个序列继续检测新的变点, 由此可检验出多级变点; 最后根据具体成因分析, 确定序列 $x(t)$ 变异点。

2.7 秩和检验、滑动秩和检验

秩和检验是一种非参数检验方法, 方法是将一个序列 $x(t)$ 分成两个样本序列 n_1 、 n_2 , 个数较小者为 n_1 , 较大者为 n_2 , 即 $n_1 < n_2$, 构造秩统计量 U :

$$U = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2)/12}} \quad (14)$$

式中: n 为样本容量; W 为 n_1 中各数值的秩之和。

U 服从标准正态分布, 若 $|U| > U_{0.05/2} = 1.96$, 则变异点显著; 否则, 变异点不显著。

滑动秩和检验法就是找出满足 $|U| > U_{0.05/2} = 1.96$ 的变异点中使得 U 统计量计算值达到最大值的点, 作为所求的最可能变异点。滑动秩和检验法对于不同变异特性(多参数组合变异)的水文系列具有较强的适应性, 在由渐变因素引起的变异诊断中具有优越性。

在秩和检验中^[24], 若时间序列越长、跳跃度越大, 检验能力越强; 若方差越大, 检验能力越弱, 若跳跃恰好发生在序列的中点, 检验能力最强。当数据服从不同的分布时, 检验能力也不尽相同, 以极值 III 型分布最强, 而对数正态分布最弱^[25]。

2.8 Spearman 秩次相关检验法

Spearman 秩次相关检验法^[26-28]是分析时间序列平稳性、检验趋势性的有效手段, 水文序列 $x(t)$

$$\text{秩次相关系数为: } r = 1 - \frac{6 \sum_{i=1}^n (R_i - i)^2}{n^3 - n} \quad (15)$$

式中: n 为样本容量; R_i 是把序列 $x(t)$ 从大到小排列时 x_i 所对应的序号。秩次相关系数较大, 趋势性越显著。在应用中, 通常是与其他检验方法相结合使用, 以判断序列的变异点。

2.9 Mann-Kendall 检验法

Mann-Kendall(M-K)检验法是目前应用较普遍的一种非参数检验方法, 对于时间序列 $x(t)$, 先确定其序列的对偶数 $x_i < x_j$ ($i < j$) 个数, 再计算序列的方差 $\text{var}(x_n)$ 及统计量 U ^[29-32]:

$$\sigma_{x_n}^2 = \text{var}(x_n) = n(n-1)(2n+5)/72$$

$$U_n = \frac{x_n - \bar{x}_n}{\sqrt{\text{var}(x_n)}} \quad (16)$$

式中: n 为样本容量。

$$x_n = \sum_{i=1}^n x_i; \bar{x}_n = E(x_n) = n(n-1)/4 \quad (17)$$

若 $U_n > 0$, 则有上升趋势; 反之, 则有下降趋势。若 $|U_n| > U_{0.05/2}$, 则序列趋势显著, 反之趋势不显著。周芬^[33]发现该法的检验能力与给定的显著性水平、样本容量、趋势度和变差系数有关, 随着显著性水平、样本容量和趋势度绝对值的增加, 检验法的识别能力增强, 随着变差系数的增加, 检验法的识别能力减小, 当趋势存在时, 检验法还与时间序列服从的分布类型和形状参数有关。

2.10 有序聚类分析法

有序聚类分析法^[34]是丁晶在 1986 年处理洪水时间序列干扰点时提出的, 用有序分类推估序列的可能显著干扰点 τ , 实质上就是推求最优分割点, 使同类之间的离差平方和较小, 而不同类之间的离差平方和较大, 其原理如下^[5]:

$$V_r = \sum (x_i - \bar{x}_r)^2, V_{n-r} = \sum (x_i - \bar{x}_{n-r})^2 \quad (18)$$

式中: \bar{x}_r 和 \bar{x}_{n-r} 为干扰点 τ 前和后序列的均值。离差平方和为: $S_n(\tau) = V_r + V_{n-r}$, 满足 $S_n^*(\tau) = \min |S_n(\tau)|$, $1 \leq \tau \leq n$ 的 τ 为最优分割点, 即为所求干扰点。该法适用于各种总体分布的序列, 但是较难处理干扰点在序列端点的情况。

2.11 Yamamoto 法

Yamamoto 法^[35-36]是从信息与噪声两部分来讨论突变问题, 1986 年日本的 Yamamoto 最先采用此法处理地面气温、降水、日照时数等序列的突变问题。原理如下: 对于时间序列, 人为设定一个基准年, 将序列分为前后两个子序列, 则基准年的突变指数 S/N (即信噪比) 可表示为:

$$S/N = |x_1 - x_2| / (S_1 + S_2) \quad (19)$$

式中: x_1 、 x_2 、 S_1 、 S_2 分别为基准年前 m_1 年、基准年后 m_2 年时间段内的平均值和标准偏差。

通过在时间段内连续设置基准年, 得到突变指数 S/N 的时间序列。当 $S/N > 1.0$ 时定义为突变, $S/N > 2.0$ 时定义为强突变。定义统计量:

$$t_0 = (x_1 - x_2) / \sqrt{Sp^2 \left(\frac{1}{m_1} + \frac{1}{m_2} \right)}$$

$$Sp^2 = \frac{(m_1 - 1)S_1^2 + (m_2 - 1)S_2^2}{m_1 + m_2 - 2} \quad (20)$$

式中: m_1 和 m_2 分别为前后两段序列样本长度; Sp 是联合样本方差。

若 $S/N > 1.0$, 相当于 $t_0 > t_{a/2}$, 表示达到一定信度水平, 即从基准年后的 M 年前后时间段序列发生突变。在达到信度 S/N 可能连续出现的数年区间内, 取最大的 S/N 值作为突变年。 m_1 、 m_2 取值根据实际情况人为设定, 此方法能简单且直观地确定突变, 但是平均时段 M 的选择带有人为性。魏凤英等^[37]指出在具体计算时, 可以多次变动 M 值, 为避免突变点漂移, 通过反复比较, 确定突变点, 以提高计算结果的可靠性。

2.12 李氏指数法与小波分析

李氏指数 (Lipschitz exponent) 是数学上表征函数局部特征的一种度量, 方法^[38-39]如下: 设 $x(t)$ 在 t_0 附近具有以下特性 $|x(t_0 + h) - p_n(t_0 + h)| \leq A|h|^a$, $a \in (n, n+1)$; 式中 h 是一个充分小量, $p_n(t)$ 是过 $(t_0, x(t_0))$ 点的 n 次多项式, 称 $x(t)$ 在 t_0 处的李氏指数为 a , 则:

$$x(t) = x(t_0) + a_1 h + a_2 h^2 + \dots + a_n h^n + O(h^{n+1})$$

$$= p_n(t) + O(h^{n+1}) \quad (21)$$

一般而言, 函数在某一点的李氏指数表征了该点的奇异性的, a 越大该点的光滑度越高, 反之, a 越小该点的奇异性越大。

小波分析能反应时间序列的内在结构和局部变化特征, 分辨时间序列在不同尺度上的演变特征。对称性小波可以看成是一个平滑函数的二阶导数, 因此任意函数的对称小波变换是将该函数与平滑函数进行卷积运算, 然后对时间求二阶导数, 依据函数二阶导数的过零点在数学中是拐点, 则在此作为可能的变异点。连续小波变换定义为:

$$W_x(a, b) = |a|^2 \int_{-\infty}^{\infty} X(t) \bar{\varphi}\left(\frac{t-b}{a}\right) dt \quad (22)$$

式中: $\varphi(t)$ 为母小波函数, 连续可微; $\bar{\varphi}(t)$ 为 $\varphi(t)$ 的复共轭函数; $W_x(a, b)$ 为小波变换系数; a 为尺度因子; b 为时间平移因子。

若小波函数区间内满足连续可微, 则 $|W_x(a, b)| \leq ka^a$; 若将尺度因子 a 以二进制离散化, 即 $a = 2^j$, 则有 $|W_x(a, b)| \leq k2^{ja}$, 其中, j 为小波变换尺度特征, k 为常数。这样就将小波变换与李氏指数联系起来, 小波变换系数及模极大值将随尺度特征 j 或李氏指数 a 的变化而变化。

2.13 Kolmogorov-Smirnov 检验法

Kolmogorov-Smirnov(K-S) 检验法^[32]是基于观测值的累计频率曲线以及期望的理论频率曲线之间的差值而计算的, 目的是比较数据分布形式与已知的是否一致, 以及检验相似和不相似的显著性水平。原理如下: 计算累计频率曲线的理论分布形 $F_e(x)$, 再计算样本累计频率 $F_0(x): F_0(x) = k/n$ 。式中 k 为不大于 x 的次数, n 为总的观测次数。然后, 计算最大偏差 $D = \max |F_e(x) - F_0(x)|$; 给定假设检验: $H_0: F_0(x) = F_e(x), H_a: F_0(x) \neq F_e(x)$;

检验统计量: $D = \max |F_e(x) - F_0(x)|$; 则拒绝域为: $D \geq D_a(n)$ 。其中, $D_a(n)$ 为置信水平为 α 的 $K-S$ 临界值。若要检验正态分布, 则 $z_i = (x_i - \bar{x})/s$, $F_e(z_i)$ 可由正态分布表查得; s 为标准差。

2.14 最优信息二分割

最优信息二分割^[40-41]是 2001 年夏军等考虑到普遍应用的参数和非参数检验方法对样本分布假设多、资料要求条件高、不能给出变异量化的指标等, 提出了最优信息二分割模型。最优信息二分割模型是从灰色信息理论及差异信息原理出发, 量度水文时间、空间序列发生变异的信息理论。基本原理是采用不同分割方法, 计算分割点前后序列的差异信息测度, 对应分割点前后差异信息测度之差最小的点即为最可能变异点。步骤如下: ①从样本容量为 n 的 $x(t)$ 序列某起点 j 逐时段滑动, 组成新序列, 新序列的 j 点之前仍为原序列元素, 但 j 点之后的值是取序列前 j 点的均值, 即:

$$X_t = [x(1), x(2), \dots, x(j), \overline{x(j)}, \dots, \overline{x(j)}] \quad (23)$$

$$j = 1, 2, \dots, n$$

②计算每个 x_t 的差异信息熵 $I(x_t)$ 、差异信息测度 $I_d(x_t)$ 和相对差异信息测度 $I_a(x_0)$ 。计算 $I_d(x_t)$ 时需要确定 $I_{\min}(x_t)$ 值, 通常可以取 $I_{\min}(x_t)$ 的 0.5 ~ 0.95 作为参考点; ③点绘 $I_d(x_t)$ 或 $I_a(x_t)$ 的序列图, 分析其斜率的变化; ④将斜率相同的点视为同一类区段序列组, 计算 $I_a(x_t)$ 的增值差比 $dl_a(x_t) = [I_a(x_t) - I_a(x_t - 1)]/I_a(x_t - 1)$, 由增值差比的变化程度诊断序列 x 元素的变异分组和起始点。

2.15 诊断系统

水文变异诊断系统是谢平等^[42]基于水文序列的确定性趋势成分与随机性变化成分理论, 并分析各种诊断方法的优缺点基础上提出的, 原理如下: 首先采用过程线法、滑动平均法、Hurst 系数法对水文

序列进行初步诊断, 从定性和定量上判断序列是否可能存在变异; 如果存在变异, 采用多种检验方法对水文序列进行详细诊断, 最后结合实际情况进行成因分析, 对各方法的计算结果进行确认, 从而得出最可能变异点。

2.16 小结

水文序列趋势检验的参数检验法对样本总体分布有相应要求, 若序列样本分布符合检验方法的假设条件才可采用, 相比来说非参数检验法较简单, 运用非参数检验法, 不必预先设定序列样本的总体分布。在上面的几种方法中 t 检验、 F 检验、Hurst 指数、Spearman 是对时间序列趋势性检验的方法, 在突变检验方法中 R/S 法对具有“Hurst 效应”的序列具有独特识别能力, 但是要求序列长度较长; 秩和检验的检验能力随着时间序列的总体分布不同而不同, 若时间序列越长、跳跃度越大, 检验能力越强, 若跳跃恰好发生在序列的中点, 检验能力最强; 滑动秩和检验法对于不同变异特性(多参数组合变异)的水文系列具有较强的适应性, 在由渐变因素引起的变异诊断中具有优越性; Mann-Kendall 法对显著性水平高、样本容量大和趋势度绝对值大的序列识别能力强, 但是序列变差系数越大, 检验法的识别能力越小减小; 小波分析从时间序列的内部结构辨析出发, 识别时间序列多时间尺度的变化、分布及突变; Yamamoto 法简单易操作, 其参数可由人为设定, 但是也增加了结果的不可靠性; Brown-Forsythe 检验对非正态分布的水文序列具有较好识别能力; 有序聚类法方法使用简单, 适用于各种序列总体分布情况, 但是较难处理突变点出现在序列端点的情况; 最优信息二分割法从灰色系统理论出发, 可以判断差异信息序列多变点或单变点的变点的位置和估计变化的跃度, 拓宽了水文序列变点分析途径; 雷红富等^[2]用实例分析了上述几种方法的识别能力, 指出秩和检验效果最优, Brown-Forsythe、贝叶斯法、有序聚类法、滑动 T 法次之, Mann-Kendall 法、滑动 F 再次之、信息二分割法和 R/S 检验最差。不同方法各具优缺点, 实际应用中可多种方法综合运用, 使结论更准确。

3 水文序列变异点分析法在各流域的应用

已有许多学者对中国各流域水文序列变异点

进行了广泛的分析研究,而且检验方式不仅限于使用一种方法,更有研究者将几种方法同时使用,组成水文诊断系统,使得各方法相互补充和验证,取得更准确和更具说服力的结果。

3.1 黄河流域

黄河流域水文与泥沙系列突变点的研究成果较其他流域要多。综观这些成果,同一流域的上、下游和干、支流发生突变的时间并不完全相同,甚至对于同一个水文站采用的方法、水文资料的项目和时间不同,得到的结论也不完全一致。

其中,使用资料时间年限最长的是从1470年(500多年的径流资料)开始的三门峡站(1470-1998年)和陕县站(1470-2000年)。陕县站可能存在的4个可能突变点1958、1965、1976和1985年,1985年是张少文等^[39]采用小波变换、李氏指数分析得出的;沈楠等^[43]采用Hurst等5种方法得到三门峡站的转折点在1932年,与刘攀等^[25]分析1919-1989年序列得到的结果仅差1年,但是刘攀采用的是线性回归等8种检验法,与李春晖等^[36]的Yamamoto法得到的1969-1972年的结果相差很大。

始于20世纪30、40年代的渭河张家山、状头和林家村站的资料也相对较长,成静清等^[44]采用13种检验方法的诊断系统,与拜有存^[3]采用滑动秩和检验分析的结果相差很大,分别为1985和1994年。就黑河流域而言,张学真等^[45]和杨明金等^[46]采用的方法不同,但是都发现了两个突变点,其中一个是在1979年,但另外一个突变点时间不同。

时间年限最短的是成静清等^[44]分析的延河安塞、延川站和与渭河吴旗站20世纪80年代初开始二十几年的序列,都没有发现突变点,但是发现枣园站从1971年开始的序列突变点在1994年,而朱恒峰等^[47]采用Spearman、有序聚类法分析的1959-1989年输沙量突变点在1970年,与径流序列突变点不同。

时间年限比较集中的是始于20世纪50、60年代的资料。突变时间结论相差不大的是河-龙区间、无定河流域和利津站。穆兴民等^[48]采用历时曲线、双累积曲线等5种方法分析径流、泥沙序列,得出70年代初、末两个突变点,与杨志峰等^[49]采用M-K法分析的径流序列结果类似。在无定河流域除了白家川站^[50]的降水量和绥德站^[44]的径流量无突变点以外,其他站点和项目的突变点都集中在1971年或1978年左右^[19,51-54]。樊辉等^[54]采用M-K、Pettitt、

Bayesian法分析利津站径流和泥沙通量,发现都存在3个变点,而且都在1968年、1985年存在突变。径流序列一致性较好的站点是志丹、刘家河、张村驿和交口河,成静清等^[44]采用诊断系统方法在其中没有发现突变点。

3.2 珠江流域

在珠江流域的分析中,涉及到的水文序列项目更丰富,不仅涉及径流、输沙量序列,还有年均潮差、最高水位、分流比、分沙比序列等。

方法上,谢平等^[55]提出了水文序列变异点诊断系统,将各方法的结果相比较,以此得到更合理的变异点。在研究中谢平采用了14种检验方法,涉及了输沙、分流比、分沙比序列,其中分析发现1960-2002年北江三水站年径流、输沙、分流比和分沙比突变点都在1992年,而西江马口站年径流无突变点,年输沙突变点在1994年。

另外,珠江流域的研究中还涉及到了最高水位和年均潮差序列。肖宜等^[41]采用差异信息理论分析大洲站1900-1997年最高水位序列,得出的突变点为1993年。张蔚等^[56]采用M-K、Pettitt法,分析了东江、西江、北江主要水文站的年均潮差序列,发现突变时间集中在20世纪80年代到90年代初。

在径流序列研究方面,时间年限最长的是闽江竹岐站1936-2004年资料,于延胜等^[57]采用差积曲线、秩检验法与R/S、M-K法得出的结果完全不同,分别是1954年、1990年。谭莹莹等^[58]采用滑动平均、过程线和Hurst指数法,分析1954-2005年东江流域汛期径流序列,发现不存在变异点,非汛期径流都在1973年突变,而年径流多在1972年突变。

另外武江和元江流域的径流序列在20世纪90年代都存在突变点^[59-60]。

3.3 长江流域

长江流域分析使用的水文资料长短不同,最早始于1870年,最新的资料持续到2008年,主要是对径流序列的分析,采用的水文站点也相对集中。始于1870年的汉口站径流序列在1926年发生突变,与始于1882年的宜昌站径流序列的突变时间相同,这是秦年秀等^[61]采用M-K、滑动t检验分析得出结论。但是,利用宜昌站同样的径流序列,覃爱基等^[62]、熊立华等^[19]分别采用Kendall、Bayesian法,得出的结论却是无变异。

研究发现,同一站点相同时段的年径流、汛期和枯水期径流序列变异点也不完全相同。胡安森

等^[63]采用 Lee-Heghinian、有序聚类和秩和检验法, 分析了汉江流域 3 个水文站 1961-2000 年的年径流序列, 发现变点在 1990 年, 汛期径流变点分别在 1990、1991、1991 年, 枯水期径流变点分别在 1996、1963、1963 年; 顾世祥等^[64]采用 M-K、R/S 法, 不仅分析了部分气象因子, 还分析了水稻灌溉定额和农业综合灌溉定额序列, 发现序列的转折点集中在 20 世纪 70 年代末至 80 年代初和 80 年代末至 90 年代初两个时期^[44]。

另外, 黄川等^[65]采用滑动平均、Spearman、Kendall、线性回归法, 分析了屏山站 1954-1992 年径流序列, 发现资料一致性较好, 不存在变异点。

3.4 海河流域

海河流域采用的水文资料集中在 1956-2000 年, 谢平等^[66]采用 17 种检验法的水文诊断系统, 对流域 5 个水库进行分析, 发现密云、西大洋、王快水库年径流序列均在 1959 年发生变异, 而黄壁庄水库则在 1967 年发生变异, 最晚的是官厅水库在 1969 年发生变异; 另外, 谢平等^[67]用滑动 F 检验法分析潮白河水资源分区的年径流时发现其突变点在 1974 年, 与陈广才等^[9]的结果相同。另外, 刘克岩等^[68]和胡庆芳等^[69]发现白洋淀降雨、径流序列和北三河(北运河、潮白河、蓟运河)洪水量序列的突变点都在 1979 年。

3.5 内陆河流域

张一驰等^[16]运用自己建立的基于 Brown-Forsythe 检验识别方法, 分析了 1956-2000 年开都河大山口站径流序列, 发现存在 1973 年、1986 年两个变异点; 另外, 张一驰等^[70]采用 M-K、Mann-Whitney 法分析了开都河流域巴音布鲁克气象站和大山口水文站 1958-2002 年的年降雨序列, 发现流域上游降雨在 1973 年发生跳跃, 中游在 1987 年发生突变。蔺学东等^[71]用 M-K、Pettitt 法, 分析了拉萨、唐加水文站的 1956-2003 年年径流序列, 发现序列在 1970 年发生突变。

4 总结

近几十年来, 中国大部分流域的水文情势发生了剧烈变化, 确定水文序列变异点是分析其驱动因素的关键步骤之一。早期研究者多采用定性的方法判断, 其后采用单一检验方法进行统计分析, 发展到提出采用水文序列变异点诊断系统进行对比

判断, 以及将现有的检验方法进行改进使用, 这个过程显示了水文序列变异点研究在逐步的成熟和完善。从各流域的研究成果可以看出, 应用的方法不同, 水文序列时间年限不同, 得到的变异点的结果也不完全一致, 这表明方法和时间年限都对水文序列变异点的识别产生影响。虽然目前有不少的研究成果, 但也不排除检验方法和流域应用的局限性, 因此对检验方法和应用还有待进一步的研究和发展。

参考文献

- [1] 陈浩, 陆中臣. 黄河中游流域系统水沙过程变异研究. 水土保持研究, 2001, 8(4): 46-52.
- [2] 雷红富, 谢平, 陈广才. 水文序列变异点检验方法的性能比较分析. 水电能源科学, 2007, 25(4): 36-40.
- [3] 张建业, 王国庆, 贺瑞敏. 黄河中游水文变化趋势及其对气候变化的响应. 水科学进展, 2009, 20(2): 153-158.
- [4] 张文林. 巢湖流域水文时间序列的变点分析[D]. 合肥: 合肥工业大学, 2006.
- [5] 拜存有, 苏莹, 郭旭新. 流域水文过程变点分析研究综述. 水资源与水工程学报, 2010, 21(1): 83-86.
- [6] 郭鹏, 陈晓玲, 刘影. 鄱阳湖湖口、外洲、梅港三站水沙变化及趋势分析(1955-2001). 湖泊科学, 2006, 18(5): 458-463.
- [7] 莫莉, 穆兴民, 王勇. 近 50 多年来北洛河水沙变化特征及原因分析. 泥沙研究, 2009(6): 30-36.
- [8] 高振英, 田庆明, 刘晓云, 等. 近 58 年河西走廊地区气温变化及突变分析. 干旱区研究, 2010, 27(2): 194-203.
- [9] 陈广才, 谢平. 水文变异的滑动 F 识别与检验方法. 水文, 2006, 26(2): 57-60.
- [10] 丁晶, 邓育仁. 随机水文学. 成都: 成都科技大学出版社, 1988.
- [11] 王孝礼, 胡宝清, 夏军. 水文时序趋势与变异点的 R/S 分析法. 武汉大学学报: 工学版, 2002, 35(2): 10-12.
- [12] 徐宗学, 张楠. 黄河流域近 50 年降水变化趋势分析. 地理研究, 2006, 25(1): 27-34.
- [13] 周寅康, 张捷, 王腊春, 等. 长江下游地区近五百年洪涝序列的 R/S 分析. 自然灾害学报, 1997, 6(2): 78-84.
- [14] Brown M B, Forsythe A B. The small sample behavior of some statistics which test the equality of several means. Technometrics, 1974, 16(1): 129-132.
- [15] Brown M B, Forsythe A B. The ANOVA and multiple comparisons for data with heterogeneous variances. Biometrics, 1974, 30(4): 719-724.
- [16] 张一驰, 周成虎, 李宝林. 基于 Brown-Forsythe 检验的水文序列变异点识别. 地理研究, 2005, 24(5): 741-748.

- [17] Chib S. Estimation and Comparison of Multiple Change-point Models. *Journal of Econometrics*, 1998, 86 (2): 221-241.
- [18] Perreault L, Bernier J, Bobbe B, et al. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology*, 2000, 235(3-4): 221-241.
- [19] 熊立华, 周芬, 肖义. 水文时间序列变点分析的贝叶斯方法. *水电能源科学*, 2003, 21(4): 39-41.
- [20] 邹传仁, 张俊芝. 基于 Bayes 方法和模糊检验的年最高水位的概率模型研究. *浙江工业大学学报*, 2006, 34(5): 571-575.
- [21] Lee A F S, Heghinian S M. A shift of the mean level in a sequence of independent normal random variable: A Bayesian Approach. *Technometrics*, 1977, 19(4): 503-506.
- [22] Pettitt A N. A non-parametric approach to the change-point problem. *Applied Statistics*, 1979, 28(2): 126-135.
- [23] 张晓萍, 张櫓, 王勇, 等. 黄河中游地区年径流对土地利用变化时空响应分析. *中国水土保持科学*, 2009, 7(1): 19-26.
- [24] Yue S, Wang C Y. Power of the Mann-Whitney test for detecting a shift in median or mean of hydro-meteorological data. *Stochastic Environmental Research and Risk Assessment*, 2002, 16(4): 307-323.
- [25] 刘攀, 郭生练, 肖义, 等. 水文时间序列趋势和跳跃分析的再抽样方法研究. *水文*, 2007, 27(2): 49-53.
- [26] 许全喜, 石国钰, 陈泽方. 长江上游近期水沙变化特点及其趋势分析. *水科学进展*, 2004, 15(4): 420-426.
- [27] 戴会超, 王玲玲, 蒋定国. 三峡水库蓄水前后长江上游近期水沙变化趋势. *水力学报*, 2007, 10(增刊): 226-231.
- [28] 王开军, 黄添强. 基于趋势秩的 Spearman 相关方法. *福建师范大学学报: 自然科学版*, 2010, 26(1): 38-41.
- [29] Mann H B. Non-parametric Test Against Trend. *Econometric*, 1945, 13(3): 245-259.
- [30] Kendall M G. Rank Correlation Methods. London: Charles Griffin, 1975.
- [31] Li L J, Zhang Lu, Wang Hao, et al. Assessing the impact of climate variability and human activities on streamflow from the Wuding River basin in China. *Hydrological Processes*, 2007, 21(25): 3485-3491.
- [32] 褚健婷, 夏军, 许崇育. 海河流域气象和水文降水资料对比分析及时空变异. *地理学报*, 2009, 64(9): 1083-1092.
- [33] 周芬. Kendall 检验在水文序列趋势分析中的比较研究. *人民珠江*, 2005(2): 35-37.
- [34] 丁晶. 洪水时间序列干扰点的统计推估. *武汉水利电力学院学报*, 1986(5): 36-40.
- [35] Yamamoto R, Iwashima T, Sanga N K. An analysis of climate jump. *Meteorological Society of Japan*, 1986, 64 (2): 273-281.
- [36] 李春晖, 郑小康, 庞爱萍, 等. 黄河流域 1919-1998 年径流量突变分析. *人民黄河*, 2008, 30(6): 33-34.
- [37] 魏凤英, 曹洪兴. 中国、北半球和全球的气温突变分析及其趋势预测研究. *大气科学*, 1992, 19(2): 140-148.
- [38] 郑泽权, 谢平, 蔡伟. 小波变换在非平稳水文时间序列分析中的初步应用. *水电能源科学*, 2001, 19(3): 49-51.
- [39] 张少文, 张学成, 王玲, 等. 黄河天然年径流长期突变特征的小波与李氏指数分析. *水文*, 2005, 25(5): 16-18.
- [40] 夏军, 穆宏强, 邱训平. 水文序列的时间变异性分析. *长江职工大学学报*, 2001, 18(3): 1-4.
- [41] 肖宜, 夏军, 申明亮, 等. 差异信息理论在水文时间序列变异点诊断中的应用. *中国农村水利水电*, 2001(11): 28-30.
- [42] 谢平, 陈广才, 李德. 水文变异综合诊断方法及其应用研究. *水电能源学*, 2003, 23(2): 11-16.
- [43] 沈楠, 李春晖. 黄河流域近 500 多年来径流量演变特征. *水资源与水工程学报*, 2009, 20(5): 37-40.
- [44] 成静清, 宋松柏. 基于混合分布非一致性年径流序列频率参数的计算. *西北农林科技大学学报: 自然科学版*, 2010, 38(2): 229-234.
- [45] 张学真, 梁俊峰, 胡安焱. 人类活动对黑河水文过程的影响分析. *干旱区资源与环境*, 2007, 21(10): 98-103.
- [46] 杨明金, 张勃, 王海青, 等. 黑河流域 1950 年-2004 年出山径流变化规律分析. *资源科学*, 2009, 31(3): 413-419.
- [47] 朱恒峰, 赵文武, 康慕谊. 延河流域输沙时变过程的显著干扰点分析. *人民黄河*, 2008, 30(8): 80-88.
- [48] 穆兴民, 巴桑赤烈, Zhang L, 等. 黄河河口镇至龙门区间来水来沙变化及其对水利水保措施的响应. *泥沙研究*, 2007(2): 36-41.
- [49] 杨志峰, 李春晖. 黄河流域天然径流量突变性与周期性特征. *山地学报*, 2004, 22(2): 140-146.
- [50] 高鹏, 穆兴民, 李锐, 等. 黄河支流无定河水沙变化趋势及其驱动因素. *泥沙研究*, 2009(5): 22-28.
- [51] 杨新, 延军平, 刘宝元. 无定河年径流量变化特征及人为驱动力分析. *地理科学进展*, 2005, 20(6): 637-641.
- [52] 王国庆, 贾西安, 陈江南, 等. 人类活动对水文序列的显著影响干扰点分析: 以黄河中游无定河流域为例. *西北水资源与水工程*, 2001, 12(3): 13-15.
- [53] 王蕊, 夏军. 近 40 年黄河中游径流情势变化分析. *水文*, 2007, 27(5): 74-77.
- [54] 樊辉, 刘艳霞, 黄海军. 1950-2007 年黄河入海水沙通量变化趋势及突变特征. *泥沙研究*, 2009(5): 9-16.
- [55] 谢平, 陈广才, 雷红富. 西北江三角洲马口站和三水站

- 水文泥沙序列变异分析//何根寿. 中国水力发电工程学会水文泥沙专业委员会第七届学术讨论会论文集(上册). 北京: 中国水利水电出版社, 2007: 10-14.
- [56] 张蔚, 严以新, 郑金海, 等. 珠江三角洲年际潮差长期变化趋势. 水科学进展, 2010, 21(1): 77-83.
- [57] 于延胜, 陈兴伟. 水文序列变异的差积曲线-秩检验联合识别法在闽江流域的应用: 以竹岐站年径流序列为例. 资源科学, 2009, 31(10): 1717-1721.
- [58] 谭莹莹, 谢平, 陈丽, 等. 东江流域径流序列变异分析. 变化环境下的水资源响应与可持续利用//中国水利学会水资源专业委员会, 中国水利水电科学研究院, 大连理工大学. 中国水利学会水资源专业委员会 2009 学术年会论文集. 辽宁: 大连理工大学出版社, 2009: 98-104.
- [59] 汪丽娜, 李粤安, 陈晓宏. 武江流域降雨、径流突变性研究. 广东水利水电, 2009(8): 4-5.
- [60] 叶长青, 甘淑, 王文玲, 等. 元江-红河干流径流时序特性及突变分析. 长江流域资源与环境, 2008, 17(6): 886-891.
- [61] 秦年秀, 姜彤, 许崇育. 长江流域径流趋势变化及突变分析. 长江流域资源与环境, 2005, 14(5): 589-594.
- [62] 覃爱基, 陈雪英, 郑艳霞. 宜昌径流时间序列的统计分析. 水文, 1993, 13(5): 15-21.
- [63] 胡安焱, 刘秀花, 姜凌. 汉江流域水文序列变点分析. 灌溉排水学报, 2009, 28(6): 128-135.
- [64] 顾世祥, 何大明, 李远华. 金沙江河谷灌溉需水转折趋势分析. 水科学进展, 2008, 19(3): 352-360.
- [65] 黄川, 娄霄鹏, 刘元元. 金沙江流域泥沙演变过程及趋势分析. 重庆大学学报: 自然科学版, 2002, 25(1): 21-23.
- [66] 谢平, 陈广才, 夏军, 等. 基于降雨径流关系分析的非一致性年径流序列频率计算方法//刘恒, 贾金生. 中国水利学会第二届青年科技论坛论文集. 郑州: 黄河水利出版社, 2005: 406-412.
- [67] 谢平, 陈广才, 雷红富, 等. 水文变异诊断系统. 水力发电学报, 2010, 29(1): 85-91.
- [68] 刘克岩, 张橹, 张光辉, 等. 人类活动对华北白洋淀流域径流影响的识别研究. 水文, 2007, 27(6): 6-10.
- [69] 胡庆芳, 王银堂. 北三河流域洪水资源量的演变特性及驱动因素评估. 水利水运工程学报, 2010(1): 30-36.
- [70] 张一驰, 李宝林, 程维明. 开都河流域径流对气候变化的响应研究. 资源科学, 2004, 26(6): 69-76.
- [71] 蔺学东, 张德铨, 姚治君. 拉萨河流域近 50 年来径流变化趋势分析. 地理科学进展, 2007, 26(3): 58-67.

Advances in the Research Methods of Abrupt Changes of Hydrologic Sequences and Their Applications in Drainage Basins in China

ZHOU Yuanyuan^{1,2}, SHI Changxing¹, FAN Xiaoli^{1,2}, DU Jun^{1,2}

(1. Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Detecting abrupt changes in hydrologic sequences plays an important role in discriminating the impacts of climate change and human activities. In recent decades, researches on abrupt changes of hydrologic sequences have advanced greatly in China from using only one method in the earliest stage to applying diagnosis system of multiple methods and improved methods in the study in the late stage. Different results were given by studies using different methods and/or different time series of hydrologic sequences in the same drainage basins. In this paper, the theories of methods determining abrupt changes in hydrologic sequences and their application in the main watersheds in China are reviewed in order to provide reference for further development of these methods and their applications.

Key words: time series analysis; detecting abrupt changes; hydrologic sequences test

本文引用格式:

周园园, 师长兴, 范小黎, 等. 国内水文序列变异点分析方法及在各流域应用研究进展. 地理科学进展, 2011, 30(11): 1361-1369.