

文章编号:0559-9350(2021)07-0793-14

基于ARMA模型的水文序列相依变异分级方法及验证

谢平¹, 霍竞群¹, 桑燕芳², 吴林倩¹, 李雅晴¹, 牛静怡¹

(1. 武汉大学 水资源与水电工程科学国家重点实验室, 湖北 武汉 430072;

2. 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)

摘要: 受自然和人为等因素的影响, 水文情势和地理环境不断发生显著变化, 不同水文要素形成的水文时间序列常呈现出一定的相依性。为定量研究水文序列中的这种相依现象, 本文以自回归滑动平均模型 ARMA 为例, 选取原始水文序列与其相依成分间的相关系数为衡量标准, 提出对相依变异强弱程度分级的一种方法。先用公式推导的方式从原理上阐明相关系数与序列的自回归系数和滑动平均系数存在的关系, 从而建立相关系数与序列自相关系数的联系, 再选择合理阈值作为分级界限, 把相关系数划分为 5 段区间, 对应描述 5 种不同强弱的相依变异程度。分别以较低阶数的 ARMA 模型为例, 通过统计试验验证了以相关系数作为分级指标的合理性。将所提方法分别应用于模拟时间序列和实测水文序列, 并结合物理成因从气候变化和人类活动两个方面对实测径流序列的相依变异分级结果进行了分析与验证, 结果表明该方法合理可靠。

关键词: 自回归滑动平均模型; 相关系数; 统计试验; 分级; 时间序列; 相依变异

中图分类号: P333

文献标识码: A

doi: 10.13243/j.cnki.slxb.20200762

1 研究背景

水文时间序列分析是深入了解水文现象特点与性质的重要途径, 以及探究水文过程复杂变化规律的关键方法^[1-2]。长久以来, 对水文时间序列的认识与研究都是基于一致性假设(即物理成因不变)的基本条件开展的^[1]。然而, 受自然因素与人为因素等的影响, 许多流域和地区的水文水资源状况发生了很大改变, 导致水文时间序列不再全部满足一致性假设^[3-4], 即发生了水文变异^[5], 表现为统计规律随时间发生明显改变, 且任意时刻数值均与前期时间内的历史值有关^[6]。

依据水文变异的具体表现形式与水文序列参数的变化特征, 又可分为跳跃、趋势、周期以及相依变异等^[7]。其中, 相依变异是水文变异的一种常见形式, 表现为水文序列后一数据与紧邻的前一(几)数据具有成因、数值等方面的继承性^[2]。当前, 对水文序列中相依变异成分的研究和分析主要是基于数理统计方法进行^[8-9], 如 Hurst 系数法^[10-11]、极差和轮次分析法^[12]、自相关分析法和谱分析法^[13]等。然而, 这些方法主要针对相依变异形式的识别和检验, 但对相依变异程度进行量化的研究则很少涉及。在实际的水文统计分析中, 不仅要明确水文序列是否发生了相依变异, 还要判定相依变异的强弱程度, 如果未对相依变异的显著性进行定量评估, 可能会对水资源安全评估以及水利工程建设等工作造成较大影响。相关系数可以描述序列间相关程度, 其绝对值越大, 则序列间的相关性越强, 故可将相关系数作为度量水文变异程度的指标^[14]。

常规的时间序列模型包括自回归模型、滑动平均模型和自回归滑动平均模型。因其具有时间相依的直观形式和较强的适用性, 故水文序列的相依性一般用它们来进行描述^[15]。其中, 自回归滑动

收稿日期: 2020-09-22; 网络首发时间: 2021-04-27

网络首发地址: <http://kns.cnki.net/kcms/detail/11.1882.TV.20210426.1758.002.html>

基金项目: 国家自然科学基金项目(91547205, 41971040, 51579181)

作者简介: 谢平(1963-), 博士生导师, 教授, 主要从事变化环境下水文水资源研究。E-mail: pxie@whu.edu.cn

通讯作者: 霍竞群(1996-), 硕士生, 主要从事变化环境下水文水资源研究。E-mail: huojingqun@whu.edu.cn

平均模型 ARMA (Auto-Regressive Moving Average) 是自回归模型 AR (Auto-Regressive) 和滑动平均模型 MA (Moving Average) 的混合, 且较 AR 和 MA 能更好地反映水文变量在时序变化上的统计特性, 即具有更大的弹性, 在水文学中应用更为广泛^[7]。国内外许多学者用其来描述水文现象在时间上的相依性。翟颖瑾等^[16]使用 ARMA 模型模拟水质污染的相依随机过程, 实现对长江未来水质情况的分析预测; 栗现文和杨佳等^[17-18]采用 ARMA 模型分析地下水位埋深时间序列中去除确定性组分之后的相依随机和纯随机组分, 从而达到对未来短期地下水位动态进行准确预测的目的; Dwivedi 等^[19]运用 ARMA 等时间序列模型对降雨和径流的随机相依水文过程建模, 从而实现降雨的预测和径流量的估计; Atan 等^[20]利用 ARMA 模型的相依性特征模拟洪水时间序列并对流域洪灾风险进行了评估与分析等。然而, 目前的研究主要是运用 ARMA 模型表征水文序列的相依特性, 对于水文序列中相依变异强弱的量化研究则鲜有提及。赵羽西等^[21]的研究虽涉及到对序列相依变异程度的分级量化, 却也仅适用于 AR 模型, 至于其它模型尚需进一步补充和完善。此外, 在运用 AIC、BIC 等定阶准则对含有 AR-MA 相依成分的水文序列进行定阶时, 只能确定模型的自回归阶数 p 与滑动平均阶数 q 的和 (即 $p+q$ 的值), 而在 AIC 和 BIC 函数最小值不唯一或阶数 p 和 q 有多种组合方式时, 并不能对其做出明确判断。

为此, 本文以适用性更为广泛的 ARMA 模型为例, 在已有研究^[21]的基础上, 以原序列与其相依成分间的相关系数为衡量标准, 提出了描述水文相依变异强弱的一种方法。通过相关系数表达式的推求说明选用该指标进行变异分级的基本原理, 并设计统计试验说明基本原理的合理性; 以 ARMA (1, 1) 和 ARMA (1, 2) 模型为例, 指出相关系数指标与自相关系数的联系; 最后通过模拟时间序列和实测水文序列对所提方法进行验证, 并结合水文序列的相依变异程度和相依成分与原序列的拟合效率系数为具有 ARMA (p, q) 相依成分的水文序列提供一种定阶的新思路。

2 分级原理与方法

2.1 相依变异分级原理 一般情况下, 在扣除确定性成分的影响之后, 剩余水文序列 x_t 中的相依成分可用自回归滑动平均模型 ARMA (p, q) 表示为:

$$x_t = u + \varphi_1(x_{t-1} - u) + \varphi_2(x_{t-2} - u) + \cdots + \varphi_p(x_{t-p} - u) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \quad (1)$$

式中: u 为序列 x_t 的均值; p 为模型的自回归阶数; q 为模型的滑动平均阶数; 参数 $\varphi_1, \varphi_2, \cdots, \varphi_p$ 是待定系数中的自回归系数; 参数 $\theta_1, \theta_2, \cdots, \theta_q$ 是待定系数中的滑动平均系数; $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ 为独立随机变量, 其标准差为 σ_ε 、均值为 0, 且与 $x_{t-1}, x_{t-2}, \cdots, x_{t-q}$ 独立无关。

式 (1) 中的 $\varphi_1(x_{t-1} - u) + \varphi_2(x_{t-2} - u) + \cdots + \varphi_p(x_{t-p} - u) - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$ 为相依成分, 用 η_t 表示, 故 x_t 能够表示为线性叠加的形式:

$$x_t = \eta_t + u_t \quad (2)$$

式中: $u_t = u + \varepsilon_t$, 其均值与标准差分别为 u 、 σ_ε , 为水文序列中的纯随机成分, 在我国水文统计分析与计算中通常认为其服从 P-III 型分布^[7]。

又因为 η_t 与 u_t 相互独立, 而 x_t 的期望 $E(x_t) = u$ 、 u_t 的期望 $E(u_t) = u$, 故可得相依成分序列的期望值 $E(\eta_t) = 0$ 。

x_t 与 η_t 间的相关系数表达式为:

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

把 $\bar{x}=u$ 、 $\bar{\eta}=0$ 、 $x_i=\eta_i+u_i$ 代入上式，可得：

$$r = \frac{\sum_{i=1}^n \eta_i^2 + \sum_{i=1}^n (u_i - u) \eta_i}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (\eta_i - \bar{\eta})^2}} \quad (4)$$

式中：

$$\sum_{i=1}^n (u_i - u) \eta_i = nE(u_i \eta_i - u \eta_i) = nE(u_i)E(\eta_i) - nuE(\eta_i) = 0 \quad (5)$$

因此，由式(4)可得：

$$r^2 = \frac{\sigma_\eta^2}{\sigma_x^2} \quad (6)$$

式中 σ_η^2 与 σ_x^2 分别为 η_i 和 x_i 的方差，它们之间具有以下关系：

$$\sigma_x^2 = \sigma_\eta^2 + \sigma_u^2 \quad (7)$$

式中 σ_u^2 为纯随机序列的方差，根据式(6)和式(7)可得：

$$r^2 = 1 - \frac{\sigma_u^2}{\sigma_x^2} \quad (8)$$

将式(1)进行移项并整理得：

$$x_i - u = \varphi_1(x_{i-1} - u) + \varphi_2(x_{i-2} - u) + \cdots + \varphi_p(x_{i-p} - u) + \varepsilon_i - \theta_1 \varepsilon_{i-1} - \theta_2 \varepsilon_{i-2} - \cdots - \theta_q \varepsilon_{i-q} \quad (9)$$

将式(9)两边同乘以 $x_i - u$ 并取数学期望得：

$$E(x_i - u)^2 = \varphi_1 E[(x_{i-1} - u)(x_i - u)] + \varphi_2 E[(x_{i-2} - u)(x_i - u)] + \cdots + \varphi_p E[(x_{i-p} - u)(x_i - u)] + E[(x_i - u)\varepsilon_i] - \theta_1 E[(x_i - u)\varepsilon_{i-1}] - \theta_2 E[(x_i - u)\varepsilon_{i-2}] - \cdots - \theta_q E[(x_i - u)\varepsilon_{i-q}] \quad (10)$$

上式两端除以 σ_x^2 后考虑到自相关系数的定义：

$$\frac{E[(x_{i-i} - u)(x_i - u)]}{\sigma_x^2} = \rho_i, \quad \rho_0 = 1 \quad (11)$$

可得：

$$1 - \rho_1 \varphi_1 - \rho_2 \varphi_2 - \cdots - \rho_p \varphi_p = \frac{E[(x_i - u)\varepsilon_i] - \theta_1 E[(x_i - u)\varepsilon_{i-1}] - \cdots - \theta_q E[(x_i - u)\varepsilon_{i-q}]}{\sigma_x^2} \quad (12)$$

式(12)两端再同乘以 σ_x^2 可得：

$$(1 - \rho_1 \varphi_1 - \rho_2 \varphi_2 - \cdots - \rho_p \varphi_p) \sigma_x^2 = E[(x_i - u)\varepsilon_i] - \theta_1 E[(x_i - u)\varepsilon_{i-1}] - \cdots - \theta_q E[(x_i - u)\varepsilon_{i-q}] \quad (13)$$

将式(9)两边同乘以 ε_i 并取数学期望得：

$$E[(x_i - u)\varepsilon_i] = \varphi_1 E[(x_{i-1} - u)\varepsilon_i] + \varphi_2 E[(x_{i-2} - u)\varepsilon_i] + \cdots + \varphi_p E[(x_{i-p} - u)\varepsilon_i] + E(\varepsilon_i^2) - \theta_1 E(\varepsilon_i \varepsilon_{i-1}) - \theta_2 E(\varepsilon_i \varepsilon_{i-2}) - \cdots - \theta_q E(\varepsilon_i \varepsilon_{i-q}) = E(\varepsilon_i^2) = \sigma_\varepsilon^2 \quad (14)$$

将式(9)两边同乘以 ε_{i-1} 并取数学期望得：

$$E[(x_i - u)\varepsilon_{i-1}] = \varphi_1 E[(x_{i-1} - u)\varepsilon_{i-1}] + \varphi_2 E[(x_{i-2} - u)\varepsilon_{i-1}] + \cdots + \varphi_p E[(x_{i-p} - u)\varepsilon_{i-1}] + E(\varepsilon_i \varepsilon_{i-1}) - \theta_1 E(\varepsilon_{i-1}^2) - \theta_2 E(\varepsilon_{i-1} \varepsilon_{i-2}) - \cdots - \theta_q E(\varepsilon_{i-1} \varepsilon_{i-q}) = -\theta_1 E(\varepsilon_{i-1}^2) = -\theta_1 \sigma_\varepsilon^2 \quad (15)$$

同理，依此类推，直至将式(9)两边同乘以 ε_{i-q} 再取数学期望得：

$$E[(x_i - u)\varepsilon_{i-q}] = \varphi_1 E[(x_{i-1} - u)\varepsilon_{i-q}] + \varphi_2 E[(x_{i-2} - u)\varepsilon_{i-q}] + \cdots + \varphi_p E[(x_{i-p} - u)\varepsilon_{i-q}] + E(\varepsilon_i \varepsilon_{i-q}) - \theta_1 E(\varepsilon_{i-1} \varepsilon_{i-q}) - \theta_2 E(\varepsilon_{i-2} \varepsilon_{i-q}) - \cdots - \theta_q E(\varepsilon_{i-q}^2) = -\theta_q E(\varepsilon_{i-q}^2) = -\theta_q \sigma_\varepsilon^2 \quad (16)$$

为此，式(13)可以化简为：

$$(1 - \rho_1 \varphi_1 - \rho_2 \varphi_2 - \dots - \rho_p \varphi_p) \sigma_x^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_\varepsilon^2 \quad (17)$$

将上式进行整理得：

$$\frac{\sigma_\varepsilon^2}{\sigma_x^2} = \frac{1 - \rho_1 \varphi_1 - \rho_2 \varphi_2 - \dots - \rho_p \varphi_p}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad (18)$$

因为 $\sigma_\varepsilon^2 = \sigma_u^2$ ，结合式(8)和(18)可得：

$$r^2 = 1 - \frac{1 - \rho_1 \varphi_1 - \rho_2 \varphi_2 - \dots - \rho_p \varphi_p}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \quad (19)$$

对式(19)进行初步验证：

当 $p=0$ 时，即为滑动平均模型 $MA(q)$ ，式(19)可化简为：

$$r^2 = 1 - \frac{1}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)} \quad (20)$$

当 $q=0$ 时，即为自回归模型 $AR(p)$ ，式(19)可化简为：

$$r^2 = \rho_1 \varphi_1 + \rho_2 \varphi_2 + \dots + \rho_p \varphi_p \quad (21)$$

式(20)和(21)与现有结论相符合^[22]，故初步认为式(19)合理可靠。

ARMA模型中的滑动平均系数 $\theta_1, \theta_2, \dots, \theta_q$ 和自回归系数 $\varphi_1, \varphi_2, \dots, \varphi_p$ 的参数解与样本自相关系数存在函数关系，而样本自相关系数能够描述水文序列的相依变异程度^[7]，故相关系数 r 能够表示样本序列的相依关系并可描述其相依程度。因此，可以用相关系数对水文序列的相依程度进行分级。

2.2 相依变异分级方法 基于以上推导证明的分级原理，借鉴水文变异分级方法^[21,23-25]，提出水文时间序列的相依变异分级方法。其中，选取 r_α 、 r_β 、0.6和0.8作为相依变异分级界限，相应地把相关系数 r 分为5段区间，对应描述不同强弱的相依变异程度，具体分级标准见表1。 r_α 和 r_β 分别是在已知模型阶数和序列长度的情况下，显著性水平为 α 和 β ($\alpha > \beta$) 下的分级阈值，通常取 $\alpha = 0.05$ ， $\beta = 0.01$ 。

以 $ARMA(p, q)$ 为例，利用所提方法对水文序列的相依变异程度量化分级的详细步骤如下：

(1) 去除水文序列中的确定性成分，得到含有独立随机或相依成分的剩余序列 x_t 。

(2) 判断 x_t 中是否存在相依成分。画出自相关与偏相关系数图，如果自相关系数均在容许限以内，认为剩余序列为纯随机成分；否则，认为剩余序列含有相依成分，且当自相关与偏相关系数均拖尾时，认为相依成分符合 $ARMA(p, q)$ 模型。

(3) 确定模型的阶数。不同于 $AR(p)$ 与 $MA(q)$ ，据自相关系数图和偏相关系数图并不能判断出 $ARMA(p, q)$ 模型的阶数，故选用两种定阶准则 (AIC 和 BIC 准则) 来判定其阶数 p 、 q 。

(4) 估计模型参数。用矩估计法^[26]对 $ARMA(p, q)$ 的参数 φ_i ($i=1, 2, \dots, p$) 和 θ_i ($i=1, 2, \dots, q$) 进行估计，由此得到相依成分序列 η_t 。

(5) 计算相关系数 r 。利用式(3)求得序列 x_t 与 η_t 间的 r 值。

(6) 判定变异等级。根据得出的相关系数 r 值对照表1找到对应的等级区间，由此判定出水文序列相依变异程度的强弱。

表1 相依变异程度分级

相依性变异程度	相关系数
无变异	$0 \leq r < r_\alpha$
弱变异	$r_\alpha \leq r < r_\beta$
中变异	$r_\beta \leq r < 0.6$
强变异	$0.6 \leq r \leq 0.8$
巨变异	$0.8 \leq r \leq 1.0$

3 分级方法验证

3.1 分级原理合理性验证 为验证水文相依变异分级方法对 $ARMA(p, q)$ 的适用性, 而实际中基于时间序列模型的水文序列预测与延展工作通常以低阶为主, 故以较低阶数 $(p+q)$ 的 $ARMA(1, 1)$ 和 $ARMA(1, 2)$ 模型为例, 分别设计统计试验, 以验证分级原理中式(19)推导过程的合理性。

3.1.1 $ARMA(1, 1)$ 模型 $ARMA(1, 1)$ 模型可以表示为 $x_t = \varphi_1(x_{t-1} - u) - \theta_1 \varepsilon_{t-1} + u_t$, 据式(19)可得:

$$r^2 = 1 - \frac{1 - \rho_1 \varphi_1}{1 + \theta_1^2} \quad (22)$$

用 Yule - Walker 方程估计自回归系数 $\varphi_i (i=1, 2, \dots, p)$ [27]:

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} \quad (23)$$

由 Yule - Walker 方程可得 $\rho_1 = \varphi_1$, 将其代入式(22)得到:

$$r^2 = 1 - \frac{1 - \varphi_1^2}{1 + \theta_1^2} \quad (24)$$

为满足平稳性和可逆性, 各系数需符合:

$$\begin{cases} -1 < \varphi_1 < 1 \\ -1 < \theta_1 < 1 \\ \varphi_1 > \theta_1 \end{cases} \quad (25)$$

统计试验详细步骤如下:

(1) 利用 P-III 型分布生成 $n=1000$ 、均值 $u=100$ 、变差系数 $C_{vu}=0.2$ 、偏态系数 $C_{su}=0.4$ 的序列 u_t , 然后随机生成 20 组自回归系数 φ_1 值和滑动平均系数 θ_1 值, 代入式(1)得到 20 个 $ARMA(1, 1)$ 序列 $x_i (i=1, 2, 3, \dots, 20)$;

(2) 相依成分 η_i 由所得 $ARMA(1, 1)$ 序列 x_i 扣除 u_t 求取;

(3) 每组系数数值内试验 1000 次, η_{ij} 为第 i 组第 j 次统计试验得到的相依成分序列 ($i=1, 2, \dots, 20; j=1, 2, \dots, 1000$);

(4) 依据式(3)计算得到每组原始序列与其相依成分序列的相关系数 $r_{ij} (i=1, 2, \dots, 20; j=1, 2, \dots, 1000)$, 再求出每组的平均相关系数 $r_i = \frac{1}{1000} \sum_{j=1}^{1000} r_{ij} (i=1, 2, 3, \dots, 20)$;

设定相对误差 $\delta = \frac{|r_i - r_a|}{r_a} \times 100\%$, 比较式(24) (为式(19)推导所得) 计算出的相关系数 r_a (理论值) 与统计试验得出平均相关系数 r_i (试验值), 结果见表 2。结果显示: 20 组不同 φ_1 与 θ_1 组合下的相对误差 δ 均在 5% 以内, 最大值也仅为 4.84% ($\varphi_1=0.7, \theta_1=-0.6$)。另外当 φ_1 值为 0 (即为 $MA(1)$ 模型) 时, 相对误差值较小, 这是由于相对误差随参数个数的增加而增大。说明在 $ARMA(1, 1)$ 模型中, 在相同 φ_1 值与 θ_1 值下, 试验值 r_i 与理论值 r_a 相近, 即分级原理对 $ARMA(1, 1)$ 模型是合理的。

3.1.2 $ARMA(1, 2)$ 模型 $ARMA(1, 2)$ 模型可以表示为 $x_t = \varphi_1(x_{t-1} - u) - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + u_t$, 将 $p=1, q=2$ 带入式(19)得:

$$r^2 = 1 - \frac{1 - \rho_1 \varphi_1}{1 + \theta_1^2 + \theta_2^2} \quad (26)$$

表2 ARMA(1, 1)模型中不同参数下的 r_a 、 r_i 及相对误差(δ)

φ_1	θ_1	r_a	r_i	$\delta/\%$	φ_1	θ_1	r_a	r_i	$\delta/\%$
0.9	-0.4	0.9144	0.8999	1.59	0.3	0.1	0.3147	0.3130	0.52
-0.1	-0.4	0.3828	0.3815	0.34	0.9	0.7	0.9341	0.9067	2.93
0.5	-0.2	0.5281	0.5190	1.72	0.7	-0.6	0.7906	0.7523	4.84
0.9	0.8	0.9403	0.9095	3.28	0.5	-0.3	0.5585	0.5431	2.76
0.5	0.1	0.5074	0.5046	0.54	0.3	-0.5	0.5215	0.5076	2.66
0.9	0.2	0.9041	0.8975	0.73	0.8	0.4	0.8305	0.8100	2.46
0.6	0.4	0.6695	0.6464	3.45	0.7	0.6	0.7906	0.7539	4.64
0.5	-0.8	0.7367	0.7029	4.59	0.2	-0.8	0.6439	0.6359	1.24
-0.3	-0.9	0.7052	0.6896	2.21	-0.1	-0.9	0.6731	0.6703	0.42
-0.4	-0.5	0.5727	0.5529	3.46	0	-0.9	0.6690	0.6687	0.04

因为由 Yule - Walker 方程得到 $\rho_1 = \varphi_1$ ，代入式(26)得：

$$r^2 = 1 - \frac{1 - \varphi_1^2}{1 + \theta_1^2 + \theta_2^2} \quad (27)$$

为满足平稳性和可逆性，各系数需符合如下条件：

$$\begin{cases} -1 < \varphi_1 < 1 \\ \theta_1 + \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1 \end{cases} \quad (28)$$

ARMA(1, 2)模型的统计试验步骤参照 ARMA(1, 1)模型，并保持参数不变。在符合式(28)要求的情况下，随机生成 20 个不同 φ_1 、 θ_1 、 θ_2 的系数值组合，依据模型表达式得到 ARMA(1, 2)序列，对比统计试验得到的平均相关系数 r_i 与由公式求出的相关系数 r_a ，结果见表 3。可以看出，20 组统计试验中只有两组结果的 δ 值超出 5%，但不足 6%，且参数个数越少，相对误差 δ 越小，由此验证了式(19)的合理性，说明本文所提分级原理对 ARMA(1, 2)模型也具有适用性。

表3 ARMA(1, 2)模型中不同参数下的 r_a 、 r_i 及相对误差(δ)

φ_1	θ_1	θ_2	r_a	r_i	$\delta/\%$	φ_1	θ_1	θ_2	r_a	r_i	$\delta/\%$
-0.3	-0.4	0.5	0.5955	0.5814	2.37	-0.1	-0.7	-0.3	0.6111	0.6076	0.57
0.4	-0.7	-0.4	0.7006	0.6738	3.83	0.9	-0.4	-0.6	0.9354	0.9067	3.06
0.5	-0.5	-0.8	0.7766	0.7407	4.63	0.5	-0.4	0.2	0.6124	0.5894	3.75
0.8	-0.5	-0.9	0.9084	0.8589	5.45	0.3	-0.4	-0.3	0.5215	0.5068	2.82
0	-0.5	0.2	0.4741	0.4742	0.01	0.8	-0.5	-0.3	0.8552	0.8217	3.92
0.1	0	0.4	0.3828	0.3814	0.38	0.4	-0.5	0.4	0.6358	0.6129	3.61
-0.2	-0.3	-0.4	0.4817	0.4725	1.90	-0.6	-0.8	-0.7	0.8364	0.7930	5.18
-0.1	-0.2	-0.9	0.6818	0.6790	0.41	0.6	-0.1	-0.6	0.7300	0.6941	4.92
0.3	-0.6	0.2	0.5916	0.5770	2.47	-0.1	-0.2	-0.8	0.6409	0.6380	0.45
-0.3	-0.5	0.2	0.5427	0.5297	2.40	0.9	-0.7	-0.7	0.9508	0.9145	3.82

此外，进行了针对 ARMA(2, 1)、ARMA(1, 3)、ARMA(2, 2)和 ARMA(3, 1)模型的统计试验，且根据统计试验结果，本文所提分级原理与方法对于其它阶数 $p+q$ 等于 4 阶以内的 ARMA 模型也是合理的，限于篇幅，详细分析结果不再展开赘述。因此，本文所提的分级原理与方法对较低阶数的 ARMA(p, q)模型具有普适性。

3.2 相关系数与自相关系数的关系 式(19)给出了 r 与自相关系数的关系，其成为本文所提分级方法的理论基础。为使该分级方法的合理性更具说服力，现继续探讨相关系数与自相关系数在 ARMA 模型中的具体联系，此处仅以 ARMA(1, 1)和 ARMA(1, 2)模型为例。

3.2.1 ARMA(1, 1)模型 利用统计试验法生成具有 ARMA(1, 1)相依成分的序列，随机序列 u_t 服从

P-III型分布，其长度 $n=1000$ 、均值 $u=100$ 、变差系数 $C_{vu}=0.2$ 、偏态系数 $C_{su}=0.4$ 。为单独研究两种待定系数的影响，选取两组 φ_1 、 θ_1 值。第一组滑动平均系数 θ_1 取固定值 0.2，自回归系数分别取 $\varphi_1=-0.9$ 、 -0.5 和 -0.2 ；第二组自回归系数 φ_1 取固定值 0.2，滑动平均系数分别取 $\theta_1=0.2$ 、 0.5 和 0.9 。图 1、图 2 为两组情况所得各阶自相关系数。

图 1 和图 2 显示， $|\rho_1|$ 不仅随着 $|\varphi_1|$ 的增大而增大，也随着 $|\theta_1|$ 增加而增加。而且，序列与相依成分的相关系数 r 也存在与 $|\rho_1|$ 一致的变化趋势，即随着 $|\varphi_1|$ 与 $|\theta_1|$ 的增大而增大。由此可知，对含有 $ARMA(1, 1)$ 相依成分的序列而言，相关系数可以刻画其相依变异程度。

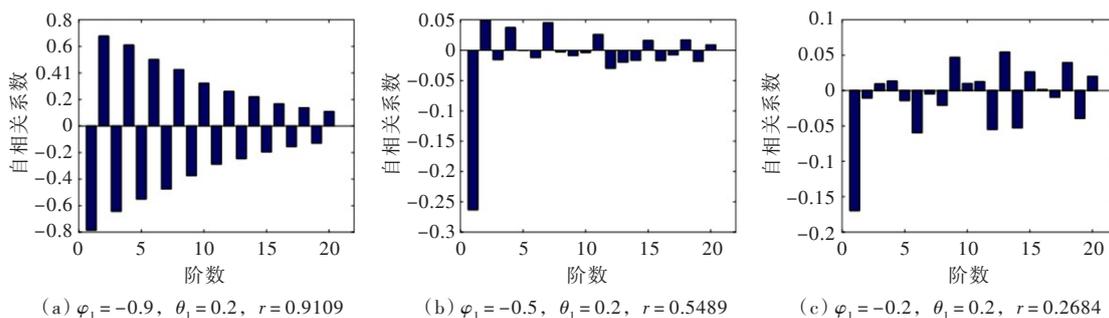


图 1 滑动平均系数 θ_1 固定时的各阶自相关系数

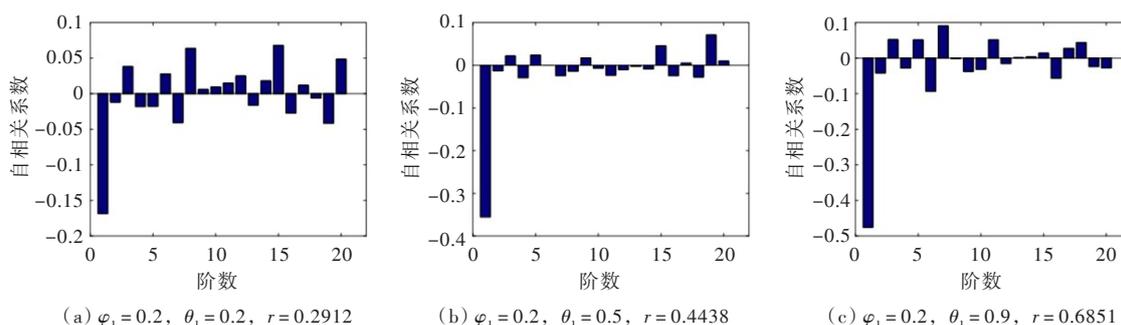


图 2 自回归系数 φ_1 固定时的各阶自相关系数

3.2.2 $ARMA(1, 2)$ 模型 自相关系数可以描述水文序列的相依程度^[7]，在 $ARMA(1, 2)$ 模型中，式 (27) 显示 r^2 与参数 φ_1 、 θ_1 、 θ_2 有关，而滑动平均系数和自回归系数的参数解与样本自相关系数存在函数关系，所以 r 与自相关系数具有函数关系。根据式 (28) 以及自相关系数、自回归系数和滑动平均系数的关系，选取多组符合条件的 ρ_1 、 ρ_2 的值，并由矩估计法^[26] 得到 $ARMA(1, 2)$ 模型的参数 φ_1 、 θ_1 、 θ_2 值；纯随机成分的生成同 3.2.1 节。 $ARMA(1, 2)$ 序列减去纯随机成分得到相依成分序列。求出每种 ρ_1 和 ρ_2 值下的相关系数，结果如表 4 所示。能够看出：一阶自回归系数 φ_1 的大小与 ρ_1 的值无关^[26, 28]，且当二阶自相关系数 ρ_2 固定不变时，相关系数 r 随着一阶自相关系数 ρ_1 的增加而增加；反之，当 ρ_1 固定时，相关系数 r 随着 ρ_2 的增加而减小。说明 r 与 $|\rho_1|$ 呈正相关关系、与 $|\rho_2|$ 呈负相关关系，故对于含有 $ARMA(1, 2)$ 相依成分的序列，其相依性强弱也能用相关系数进行描述。

3.3 模拟序列分析验证 将所提分级方法应用于具有相依成分的模拟序列进行验证。首先从常用的

表 4 不同 ρ_1 、 ρ_2 下序列的相关系数

	分组	ρ_1	ρ_2	φ_1	θ_1	θ_2	r
ρ_2 固定	1	-0.1	0.5	0.8000	0.0205	-0.0086	0.7529
	2	0.3	0.5	0.8000	0.0105	-0.0038	0.7663
	3	0.7	0.5	0.8000	0.0008	0.0009	0.7748
ρ_1 固定	1	0.5	-0.2	-0.5000	0.0230	0.0011	0.5230
	2	0.5	0.4	0.2500	-0.0089	-0.0133	0.2781
	3	0.5	0.8	0.1250	-0.0314	-0.0756	0.1406

显著性水平中选取第一显著性水平 $\alpha = 0.05$ 、第二显著性水平 $\beta = 0.01$ ；然后根据相依变异程度分级表(即表1)中各个变异等级对应的相关系数 r 的取值范围，并结合所提水文相依变异分级方法的理论

基础式(19)，求取各相依变异程度所对应的整体变量 $H = \frac{1 - \rho_1\varphi_1 - \rho_2\varphi_2 - \dots - \rho_p\varphi_p}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}$ 的取值区间；最后

选取满足整体变量 H 在取值区间内的系数值组合，生成具有相依变异成分的模拟序列。下面仍以 $ARMA(1, 1)$ 和 $ARMA(1, 2)$ 模型为例进行验证。纯随机成分在模拟序列中的长度 $n = 100$ 、均值 $u = 100$ 、变差系数 $C_{uu} = 0.2$ 、偏态系数 $C_{su} = 0.4$ ，且服从 P-III 型分布。

3.3.1 $ARMA(1, 1)$ 模型 在 $ARMA(1, 1)$ 模型中，它的自回归部分和滑动平均部分的阶数分别为 $p = 1$ 、 $q = 1$ ，根据式(22) — (24)可知此时整体变量 H 可表示为 $H = \frac{1 - \varphi_1^2}{1 + \theta_1^2}$ 。各相依变异程度对应的整体变量 H 的取值区间见表5。

表5 $ARMA(1, 1)$ 模型各相依变异程度区间所对应的整体变量 H 的取值范围

$H = \frac{1 - \varphi_1^2}{1 + \theta_1^2}$ 取值范围	相关系数 r 取值范围	相依变异程度
(0.9401, 1.0000]	[0.0000, 0.2447)	无变异
(0.9094, 0.9401]	[0.2447, 0.3010)	弱变异
(0.6400, 0.9094]	[0.3010, 0.6000)	中变异
(0.3600, 0.6400]	[0.6000, 0.8000)	强变异
(0.0000, 0.3600]	[0.8000, 1.0000)	巨变异

在上述条件下，分别选取一阶自回归与一阶滑动平均 $\varphi_1 = 0.12$ 、 $\theta_1 = 0.10$ ， $\varphi_1 = 0.20$ 、 $\theta_1 = 0.16$ ， $\varphi_1 = 0.25$ 、 $\theta_1 = 0.24$ ， $\varphi_1 = 0.50$ 、 $\theta_1 = 0.70$ 和 $\varphi_1 = 0.80$ 、 $\theta_1 = 0.95$ 共5组系数值，由这5组系数值求得的整体变量 H 分别对应相依变异程度的5个等级。模拟生成这5种情形下的序列，绘于图3，并计算出所生成序列与其相依成分之间的相关系数。

图3为相依无、弱、中、强、巨变异下模拟序列与其相依成分示意图。从图3可以直观看出，各相依成分与原序列的拟合程度依次加强，表明模拟序列的相依变异程度依次加强；求得各图的模拟序列与相依成分间的相关系数值分别为0.1556、0.2530、0.3370、0.7047、0.9004，满足表5所列出的各级相依变异程度对应的相关系数的取值区间，也从定量角度很好地验证了所提方法的合理性。

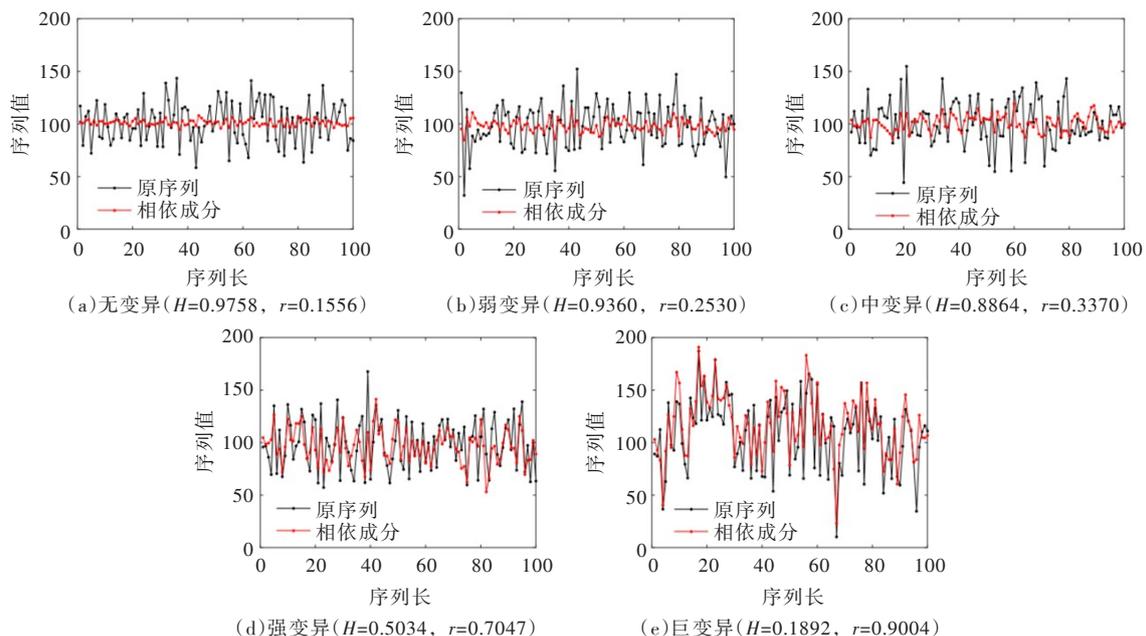


图3 $ARMA(1, 1)$ 模型不同 H 值下相依变异分级序列

3.3.2 ARMA(1, 2)模型 ARMA(1, 2)模型的阶数为 $p=1$ 、 $q=2$ ，由式(26)、(27)可知整体变量 H 可表示为 $H = \frac{1 - \varphi_1^2}{1 + \theta_1^2 + \theta_2^2}$ ，各相依变异程度所对应的整体变量 H 的取值范围见表6。

表6 ARMA(1, 2)模型各相依变异程度区间所对应的整体变量 H 的取值范围

$H = \frac{1 - \varphi_1^2}{1 + \theta_1^2 + \theta_2^2}$ 取值范围	相关系数 r 取值范围	相依变异程度
(0.9222, 1.0000]	[0.0000, 0.2789)	无变异
(0.8891, 0.9222]	[0.2789, 0.3331)	弱变异
(0.6400, 0.8891]	[0.3331, 0.6000)	中变异
(0.3600, 0.6400]	[0.6000, 0.8000)	强变异
(0.0000, 0.3600]	[0.8000, 1.0000)	巨变异

此种条件下，分别在相依变异的5个等级对应的整体变量 H 区间内各取一组满足条件的 φ_1 、 θ_1 、 θ_2 组合值，同时一阶自回归系数 φ_1 、一阶滑动平均系数 θ_1 以及二阶滑动平均系数 θ_2 还需满足式(28)的限制条件。选取的5组系数值分别为 $\varphi_1=0.02$ 、 $\theta_1=0.07$ 、 $\theta_2=0.08$ ， $\varphi_1=0.15$ 、 $\theta_1=0.18$ 、 $\theta_2=0.20$ ， $\varphi_1=0.30$ 、 $\theta_1=0.28$ 、 $\theta_2=0.30$ ， $\varphi_1=0.50$ 、 $\theta_1=-0.50$ 、 $\theta_2=-0.40$ ， $\varphi_1=0.80$ 、 $\theta_1=-0.80$ 、 $\theta_2=-0.85$ 。所生成序列与其相依成分的示意图如图4所示，图4依次为相依变异从弱到强的5个等级，所对应的相关系数 r 值分别为0.1077、0.2975、0.4703、0.6842、0.9207，均满足表6中各变异等级对应的 r 值区间。图4直观地显示了模拟序列相依变异的程度依次增强，而相关系数 r 的值也依次增加，故所提分级方法的合理性得到了验证。

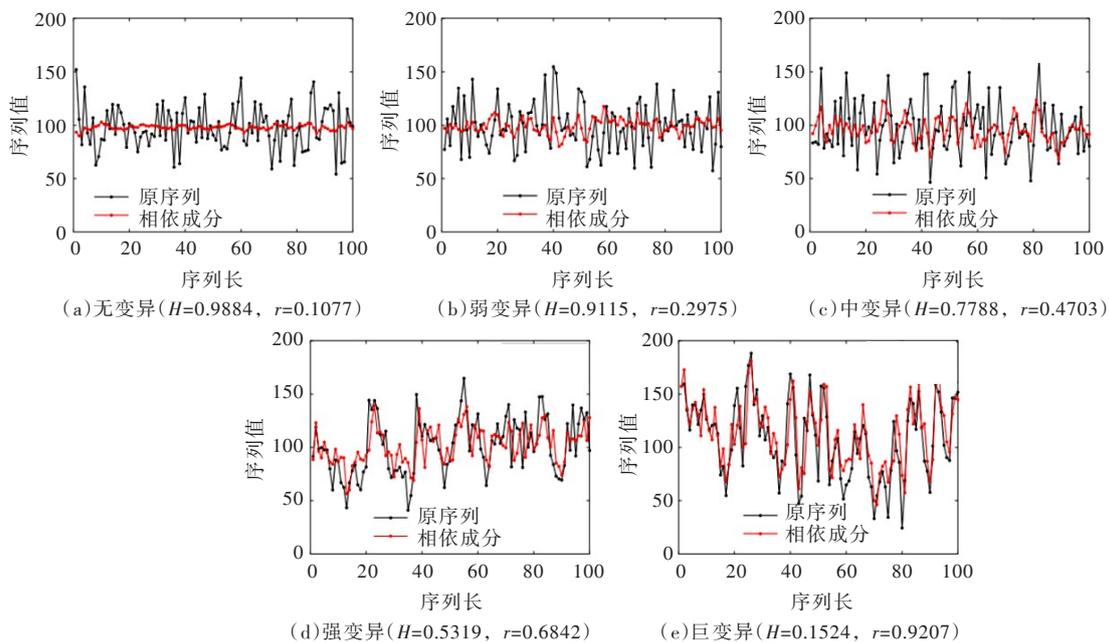


图4 ARMA(1, 2)模型不同 H 值下相依变异分级序列

4 实测序列分析

为进一步说明所提分级方法的合理性与适用性，选取实测水文序列对其进行验证，包括黄河花园口站的年径流序列(1919—2000年)、长江汉口(武汉关)站的月径流序列(2010—2015年)及长江沙市(二郎矾)站的月径流序列(1991—1998年)。按照相依变异分级方法的步骤顺序，要先扣除水文序列中的确定性成分，得到含有独立随机或相依成分的剩余序列，可以采用综合加权法^[1]、滑动相关系

数法^[29]、Brown-Forsythe法^[30]等检测序列的跳跃点；采用分段趋势相关系数识别法^[31]、过程线法、Kendall检验法^[32]等鉴别趋势成分；采用滑动周期相关系数识别法^[33]、周期图法、最大熵谱分析法^[34-35]等检验序列的周期成分。

显著性水平分别选取 $\alpha = 0.05$ 、 $\beta = 0.01$ ，对以上序列进行变异诊断分析^[36]，其中结果如表7所示。结果表明，花园口站的年径流序列在1932年发生了明显的跳跃变异，这主要是气候变异直接导致的^[37]；汉口站的月径流序列和沙市站的月径流序列都存在12 a的周期变异，这种周期波动现象主要和区域的气候条件及水文地理特征密切相关^[24]。

表7 实测径流序列变异诊断结果

站点	变异诊断结果
花园口站	跳跃(1932)
汉口站	周期(12a)
沙市站	周期(12a)

图5为各站点实测径流去除确定性成分之后的自相关系数和偏相关系数图，蓝线为容许度为95%时的界限。由图可知，花园口站的自相关系数均在容许限以内，说明花园口站年径流序列不含有相依成分；而汉口站和沙市站序列的自相关系数和偏相关系数均呈现出拖尾的特性，表明汉口站和沙市站的月径流序列都存在相依成分，且为 $ARMA(p, q)$ 相依成分。故采用 $ARMA(p, q)$ 模型对汉口站和沙市站的月径流的剩余序列进行建模， $ARMA(p, q)$ 中的阶数 p, q 运用AIC与BIC准则进行判定：

$$AIC(p, q) = n \ln \sigma_{\varepsilon}^2 + 2(p + q) \quad (29)$$

$$BIC(p, q) = n \ln \sigma_{\varepsilon}^2 + \ln n(p + q) \quad (30)$$

式中： n 与 σ_{ε}^2 分别为序列长、残差方差。AIC准则和BIC准则类似，都是确定在有最小函数值时所对应的 p 和 q 为最佳阶数。

可得在 $p + q$ 等于3和2时，汉口站和沙市站有最小的AIC与BIC值，由此可确定沙市站月径流的

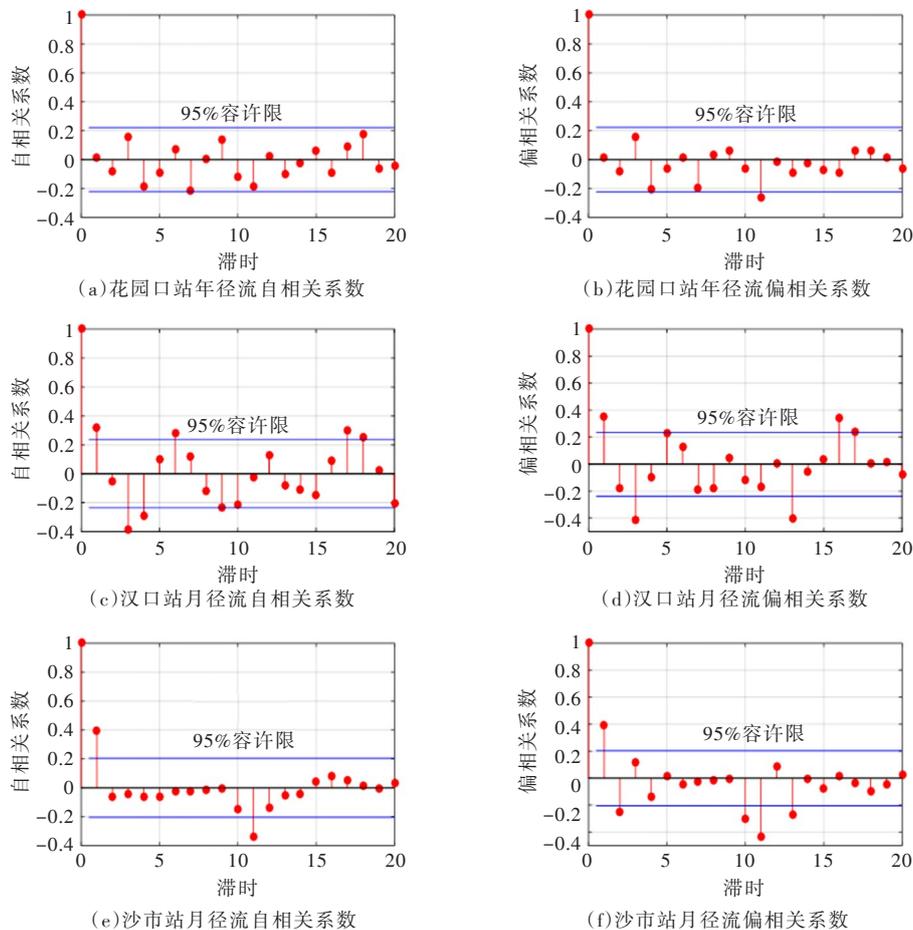


图5 实测水文序列的自相关系数及偏相关系数

剩余序列符合 $ARMA(1, 1)$ 模型, 因在 $ARMA(p, q)$ 模型中 q 的取值一般小于 p ^[28], 故初步判断汉口站月径流的剩余序列符合 $ARMA(2, 1)$ 模型, 还需要根据相依成分序列与其原始水文序列的拟合效果进一步进行验证。采用 $ARMA(2, 1)$ 和 $ARMA(1, 1)$ 模型分别对汉口站和沙市站的剩余序列建模, 由矩估计法^[26]估计模型未知参数, 并求出相关系数, 具体结果如表 8 所示。可以看出: 花园口站年径流剩余序列由 AIC 和 BIC 准则判定的阶数为 1 阶, 但与其相依成分之间的相关系数为 0.0240, 比 $r_\alpha = 0.2172$ 小, 因此该序列无相依变异, 阶数也最终确定为 0; 汉口站序列的相关系数 $r = 0.3418$, 在区间 $[r_\alpha, r_\beta]$ 内, 故其相依变异程度为弱变异; 沙市站序列的相关系数 $r = 0.4022$, 在区间 $[r_\beta, 0.6]$ 内, 因此其相依变异程度为中变异。

表 8 实测序列相依变异强弱分级

站点	序列长度	AIC 准则	BIC 准则	$p+q$	p	q	r_α	r_β	相关系数	相依变异程度
花园口站	82	1(0)	1(0)	1(0)	0	0	0.2172	0.2830	0.0240	无变异
汉口站	72	3	3	3	2	1	0.3284	0.3907	0.3418	弱变异
沙市站	96	2	2	2	1	1	0.2498	0.3071	0.4022	中变异

图 6 为各站点实测径流序列与其相依成分的示意图。图 6(b) 与图 6(c) 为分别使用 $ARMA(2, 1)$ 模型和 $ARMA(1, 2)$ 模型对汉口站月径流序列进行拟合得到的原始水文序列及其相依成分示意图, 其中 $ARMA(2, 1)$ 模型的相关系数为 0.3418、 $ARMA(1, 2)$ 模型的相关系数为 0.3343, 从相关系数的大小初步判断 $ARMA(2, 1)$ 模型优于 $ARMA(1, 2)$ 。进一步引入一个能够验证水文模型模拟结果优劣的评价指标即纳什效率系数 (NSE)^[38], 其表达式为:

$$E = 1 - \frac{\sum_{t=1}^T (Q_0^t - Q_m^t)^2}{\sum_{t=1}^T (Q_0^t - \bar{Q}_0)^2} \quad (31)$$

式中: T 为序列长度; Q_0^t 为实测序列的第 t 个值; Q_m^t 为模拟序列的第 t 个值; \bar{Q}_0 表示实测序列的均值。 E 的取值范围为 $(-\infty, 1]$, E 越接近 1 表示拟合效果越好, 模型可信度越高; E 接近 0, 表示模拟结果总体可信, 但拟合过程存在一定的误差; E 远小于 0, 则模型是不可信的。经计算, 使用 $ARMA(2, 1)$ 模型和 $ARMA(1, 2)$ 模型对汉口站月径流序列进行拟合所得的纳什效率系数分别为 $E_1 = -0.6104$ 和 $E_2 = -41.5755$, 考虑到汉口站序列的相依变异程度为弱变异, 而相依变异程度越高,

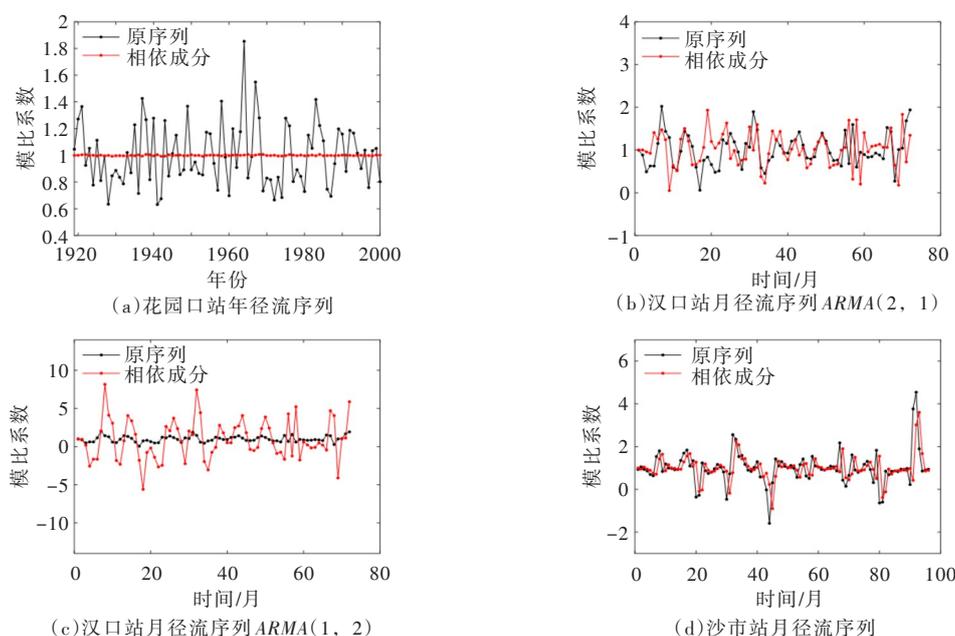


图 6 各站点原始水文序列及其相依成分示意

拟合效果越好,即拟合误差越小,所以汉口站的相依成分序列与原序列间的拟合效果相比于相依中、强、巨变异要差,此为拟合过程存在误差的原因,从而使用 $ARMA(2, 1)$ 模型对汉口站月径流序列进行拟合所得的纳什效率系数为 $E_1 = -0.6104$ 得到了很好的解释。故根据拟合效果与汉口站月径流序列的相依变异程度可以判定汉口站月径流的剩余序列符合 $ARMA(2, 1)$ 模型,这为AIC和BIC函数最小值不唯一或阶数 p 和 q 有多种组合方式时,提供了一种对 $ARMA(p, q)$ 相依成分进行定阶的新思路。图6直观地显示,实测径流序列与其相依成分的拟合程度逐渐加强,说明实测序列的相依变异等级越来越高,故该分级方法的适用性在实测水文序列的应用中得到了验证。

受自然和人为等因素的影响,水文要素常常呈现出一定的相依特性,故从气候和人为两方面对各站径流序列的相依变异做出物理成因分析^[21]。花园口站位于黄河下游,它的年径流序列无相依变异主要是其总体受气候变化与人类活动的影响较小^[23];相较于气候变化,人类活动对长江中游径流变化的影响更加显著^[39],沙市水文站地处湖北省荆州市,作为长江中游荆江河段的关键控制站,它上游葛洲坝和三峡水利枢纽的运行和建设是导致此站月径流中变异的重要原因,其次还受到沿程护岸工程以及洲滩守护工程的影响^[40-41];汉口站地处湖北武汉,是对汉江入汇长江后的水情进行控制和监测的重要站点,虽然受到长江上游水利工程的影响,但汉江的水利工程设施相对较少,汇入长江后,水文相依变异程度减轻为弱变异^[13]。

5 结论

为定量研究水文序列中相依性的强弱,本文提出了一种基于自回归滑动平均模型($ARMA$)的水文序列相依变异分级方法,分析验证后得出结论如下:

(1)由分级原理可知, $ARMA$ 模型中的自回归系数和滑动平均系数的大小影响着相依成分与水文时间序列间相关系数 r 的取值,而滑动平均系数和自回归系数的参数解与样本自相关系数存在函数关系,基于此建立了相关系数 r 与能够表征水文序列相依程度的样本自相关系数之间的联系,即相关系数 r 可以表示样本序列的相依关系并描述其相依程度。此外,本文所推导的相关系数 r 与 $ARMA(p, q)$ 模型参数间的关系(即式(19))在以往的文献中并未提及,该关系可为后续对自回归滑动平均模型与相关系数的研究提供参考。

(2)运用统计试验证明了分级原理与方法对较低阶数的 $ARMA$ 模型的合理性。①公式计算和统计试验所得 r 值差别很小,证明了式(19)的合理性;②在 $ARMA(1, 1)$ 模型中,相关系数 r 和自相关系数 $|\rho_1|$ 均随自回归系数的绝对值 $|\varphi_1|$ 和滑动平均系数的绝对值 $|\theta_1|$ 的增大而增大;在 $ARMA(1, 2)$ 模型中, r 与 $|\rho_1|$ 呈正相关关系、与 $|\rho_2|$ 呈负相关关系,说明以相关系数作为指标衡量序列的相依性强弱是合理的;③根据整体变量 H 的取值范围,生成模拟水文序列,所得相关系数处于对应变异程度的取值区间,进一步说明此方法合理可靠;④而本文所提的分级原理与方法对于更高阶数的 $ARMA(p, q)$ 模型是否适用尚需进一步研究与探讨。

(3)实测序列分析表明:黄河花园口站年径流序列无相依变异、长江汉口(武汉关)站月径流序列为相依弱变异、长江沙市(二郎矾)站月径流序列为中等程度的相依变异。并结合物理成因从气候变化和人类活动两个方面很好地解释了各站点实测水文序列相依性依次加强的原因,验证了所提分级方法的合理性及对实测水文资料的适用性。

(4)提供了一种具有 $ARMA(p, q)$ 相依成分的水文序列定阶的新思路:在AIC和BIC函数最小值不唯一或阶数 p 和 q 有多种组合方式时,可以先根据原始水文序列的相依变异程度初步判定模型阶数是否合适,再根据其相依成分的拟合效率系数最终确定模型阶数。

参 考 文 献:

[1] 顾海挺,谢平,桑燕芳,等.水文跳跃变异综合加权识别方法——以澜沧江下游戛旧水文站径流序列为例

- [J]. 应用生态学报, 2018, 29(4): 1061–1070.
- [2] 李雅晴, 谢平, 桑燕芳, 等. 水文序列相依变异识别的 RIC 定阶准则——以自回归模型为例[J]. 水利学报, 2019, 50(6): 721–731.
- [3] PIAO S, CIAIS P, HUANG Y, et al. The impacts of climate change on water resources and agriculture in China [J]. *Nature*, 2010, 467(7311): 43–51.
- [4] LENG G, TANG Q, RAYBURG S. Climate change impacts on meteorological, agricultural and hydrological droughts in China[J]. *Global and Planetary Change*, 2015, 126: 23–34.
- [5] 谢平, 陈广才, 雷红富, 等. 论变化环境下的地表水资源评价方法[J]. 水资源研究, 2007, 28(3): 1–3.
- [6] MILLY P C D, BETANCOURT J, FALKENMARK M, et al. Climate change – Stationarity is dead: Whither water management?[J]. *Science*, 2008, 319(5863): 573–574.
- [7] 王文圣, 金菊良, 丁晶. 随机水文学[M]. 北京: 中国水利水电出版社, 2008.
- [8] SCHWANKL L J, RAGHUWANSHI N S, WALLENDER W W. Time series modeling for predicting spatially variable infiltration[J]. *Journal of Irrigation and Drainage Engineering*, 2000, 126(5): 283–287.
- [9] MIRZA M Q, WARRICK R A, ERICKSEN N J, et al. Trends and persistence in precipitation in the Ganges, Brahmaputra and Meghna river basins[J]. *International Association of Scientific Hydrology Bulletin*, 1998, 43(6): 845–858.
- [10] CARBONE A, CASTELLI G, STANLEY H E. Time-dependent Hurst exponent in financial time series[J]. *Physica A—statistical Mechanics Its Applications*, 2004, 344(1/2): 267–271.
- [11] CAPODAGLIO A G, MOISELLO U. Simple stochastic model for annual flows[J]. *Journal of Water Resources Planning & Management*, 1990, 116(2): 220–232.
- [12] YEVEJEVICH, VUJICAM. Stochastic Processes in Hydrology[M]. Water Resources Publications, 1972.
- [13] 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[J]. 应用基础与工程科学学报, 2014, 22(3): 481–491.
- [14] 赵羽西, 谢平, 桑燕芳, 等. 基于相关分析的水文趋势变异分级原理及验证[J]. 科学通报, 2017, 62(26): 3089–3097.
- [15] 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[J]. 科学通报, 2017, 62(4): 254–261.
- [16] 翟颢瑾, 高晶. 长江未来水质污染的时间序列分析[J]. 沈阳师范大学学报(自然科学版), 2006(1): 22–24.
- [17] 栗现文, 向东进, 周金龙, 等. 时间序列分析方法在干旱区地下水位动态预测中的应用[J]. 工程勘察, 2011, 39(12): 28–32.
- [18] 杨佳, 钱会. 时间序列分析在地下水位动态预测中的应用[J]. 水资源与水工程学报, 2015, 26(1): 58–62.
- [19] DWIVEDI D K, SHRIVASTAVA P K. Rainfall and runoff estimation of micro watersheds of coastal Navsari[J]. *Journal of Soil and Water Conservation*, 2019, 18(1): 43–51.
- [20] ATAN I B, METCALFE A V. Estimation of seasonal flood risk using a two-stage transformation[J]. *Water Resources Research*, 1994, 30(7): 2197–2206.
- [21] 赵羽西, 谢平, 桑燕芳, 等. 基于相关系数的水文相依性变异分级方法——以自回归模型为例[J]. 应用生态学报, 2018, 29(4): 1089–1097.
- [22] MAITY R. Statistical Methods in Hydrology and Hydroclimatology[M]. Singapore: Springer Nature Singapore Pte Ltd, 2018.
- [23] 谢平, 唐亚松, 李彬彬, 等. 基于相关系数的水文趋势变异分级方法[J]. 应用基础与工程科学学报, 2014, 22(6): 1089–1097.
- [24] 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[J]. 水力发电学报, 2018, 37(12): 33–43.
- [25] 吴子怡, 谢平, 桑燕芳, 等. 基于相关系数的水文序列跳跃变异分级原理与方法[J]. 应用生态学报, 2018, 29(4): 1042–1050.
- [26] 宋安超. 自回归滑动平均模型参数估计方法的仿真比较[J]. 现代经济信息, 2011(17): 278–279.
- [27] 安鸿志, 陈兆国, 杜金观, 等. 时间序列的分析与应用[M]. 北京: 科学出版社, 1983.
- [28] 何书元. 应用时间序列分析[M]. 北京: 北京大学出版社, 2003.

- [29] 吴子怡, 谢平, 桑燕芳, 等. 水文序列跳跃变异点的滑动相关系数识别方法[J]. 水利学报, 2017, 48(12): 1473-1481, 1489.
- [30] LI J, TAN S, WEI Z, et al. A new method of change point detection using variable fuzzy sets under environmental change[J]. Water Resources Management, 2014, 28(14): 5125-5138.
- [31] 吴子怡, 谢平, 桑燕芳, 等. 基于相关系数的水文序列分段趋势识别方法[J]. 水力发电学报, 2019, 38(7): 77-86.
- [32] KISI O, AY M. Comparison of Mann-Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey[J]. Journal of Hydrology, 2014, 513: 362-375.
- [33] 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[J]. 科学通报, 2019, 64(24): 2549-2560.
- [34] SANG Y-F, WANG Z, LIU C. Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis[J]. Journal of Hydrology, 2012, 424: 154-164.
- [35] 曹鸿兴, 罗乔林. 气象历史序列的最大熵谱分析[J]. 科学通报, 1979(8): 351-355.
- [36] 谢平, 陈广才, 雷红富, 等. 水文变异诊断系统[J]. 水力发电学报, 2010, 29(1): 85-91.
- [37] 李春晖, 郑小康, 庞爱萍, 等. 黄河流域 1919~1998 年径流量突变分析[J]. 人民黄河, 2008(6): 33-34, 96.
- [38] 徐驰, 曾文治, 黄介生, 等. 基于高光谱与协同克里金的土壤耕作层含水率反演[J]. 农业工程学报, 2014, 30(13): 94-103.
- [39] 代稳, 吕殿青, 李景保, 等. 气候变化和人类活动对长江中游径流量变化影响分析[J]. 冰川冻土, 2016, 38(2): 488-497.
- [40] 陈吉琴, 王艳, 肖博文. 长江干流沙市水文站汛期流量变化趋势分析[J]. 黄河水利职业技术学院学报, 2020, 32(1): 1-4.
- [41] 黄勇, 袁晶, 高宇, 等. 长沙沙市河段近期河道演变分析[J]. 人民长江, 2019, 50(1): 18-23.

ARMA model-based classification method of hydrological series dependence variability and its verification

XIE Ping¹, HUO Jingqun¹, SANG Yanfang², WU Linqian¹, LI Yaqing¹, NIU Jingyi¹

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

2. 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)

Abstract: Hydrological processes in many basins and regions worldwide are changing significantly because of the natural and human factors such as continuous global climate change and frequent human activities in recent years. The hydrological time series formed by different hydrological elements often shows certain dependence. This paper takes ARMA model as an example, and selects the correlation coefficient between the original hydrological series and its dependence component as an index. A new classification method for significance evaluation of hydrological dependence variability was proposed to quantitatively study this dependence phenomenon in hydrological series. By deriving the expression of correlation coefficient between the original series and its dependence component, the relationship between correlation coefficient and autocorrelation coefficient is constructed. And then choosing reasonable thresholds of correlation coefficient, this method divides significance degree of dependence into five levels: no, weak, mid, strong, and drastic. The lower order ARMA models were taken as examples, the reasonability of the index used in this method was verified through Monte-Carlo experiments. The proposed method is applied to the simulated time series and the observed hydrological series, and the results of the dependence variability classification for the observed runoff series are analyzed and verified from the aspects of climate change and human activities combined with the physical causes. The results show that the method is reasonable and reliable. Therefore, it is helpful to understand the complex evolution law of hydrological process and quantitatively study the impact of environmental change on hydrological variability.

Keywords: auto-regressive moving average model; correlation coefficient; statistical test; classification; time series; dependence variability

(责任编辑: 耿庆斋)