

DOI:10.20040/j.cnki.1000-7709.2023.20221633

基于条件重采样的多站洪水过程随机模拟方法研究

王权森^{1,2}, 李安强^{1,2}, 卢程伟^{1,2}

(1. 长江设计集团有限公司, 湖北 武汉 430010; 2. 流域水安全保障湖北省重点实验室, 湖北 武汉 430010)

摘要:以流域干支流多站径流过程自、互相关性为切入点,提出了基于条件重采样理论的多站洪水过程模拟方法,并对长江流域干流屏山、寸滩与宜昌站洪水过程进行模拟延拓,将模拟结果与多站季节性自回归模型模拟结果进行了对比验证分析。结果表明,相对于传统模型,所提方法能在满足径流序列基本统计特性及序列高阶自互相关结构的同时,能进一步考虑历史极端洪水特征,生成满足不同设计频率的多站点多情景模拟径流过程。

关键词:多站洪水过程模拟; 条件重采样; 高阶自互相关结构; 历史极端洪水特征

中图分类号: TV122⁺.1 文献标志码: A 文章编号: 1000-7709(2023)07-0098-04

1 引言

流域来水过程呈现的随机、高维、非线性、非平稳等特性给流域复杂水资源系统防洪减灾、径流适应性利用带来了一系列的难题^[1]。尤其是在现有数据资料背景下,大规模水利工程建设与运行管理均以长系列水文资料为基础进行水文分析计算^[2]。然而实测资料序列年限较短,无法准确反映流域水文特征。对此,现有研究大多采用随机过程理论和时间序列分析技术对水文时间序列进行模拟延拓,相关模型技术包括回归类模型^[3]、解集类模型及降雨洪水模拟模型^[4]等。然而,受模型复杂度的限制,现有多站随机模拟方法大多仅根据各站历史连续性实测洪水资料进行建模,建模过程较少考虑防洪区内各站点出现的历史超标水文极值^[5]。因此,模拟出极端条件下洪水序列的洪峰、洪量等模拟特征值与实际设计洪水特征值偏差较大,且模拟序列中各站点洪水特征值间的相关关系难以保持,模拟序列可靠性较低。为此,本文在传统洪水模拟技术的基础上,提出了一种基于重采样策略的多站径流过程随机模拟方法(Coupla-based Resampling Method, CRM),模拟序列在满足基本统计特征的同时,更能保证模拟序列的高阶自互相关结构及洪水特征极值出现的合理性。

2 研究方法

所提洪水模拟方法是由多站点洪水过程随机

模拟、洪水特征随机模拟及考虑极端洪水特征的多站径流过程匹配搜索三部分组成的混合模型。首先通过结合 Coupla 法对多站点日径流过程进行随机模拟,构造多站径流样本池,并选取主站统计模拟洪水特征序列;进一步模拟出考虑历史极端洪水的关键径流特征序列;最后,基于重采样方法,根据模拟的特征序列在第一步生成的径流样本池中选取匹配特征且具有时间因次的多站点洪水模拟过程。具体技术路线见图 1。

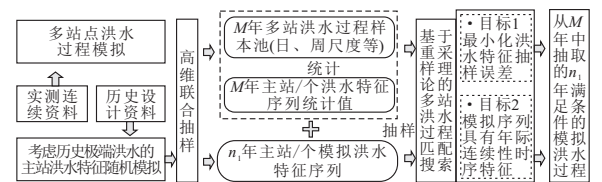


图 1 总体技术路线图

Fig. 1 The technical roadmap

2.1 多站点洪水过程随机模拟与主站模拟洪水特征统计

采用传统 Copula 法直接对高维相关变量进行联合抽样,具体步骤如下。

步骤 1 将 N_s 个站点的实测 m 年的洪水序列协整成洪水序列矩阵,若模拟时间尺度为天,则实测洪水序列矩阵 S^{obs} 为:

$$(S^{obs})_{m \times n} = \begin{bmatrix} S_{1,1}^{obs} & \dots & S_{1,165}^{obs} & S_{1,366}^{obs} & \dots & S_{1,730}^{obs} & \dots & S_{1,n}^{obs} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{m,1}^{obs} & \dots & S_{m,165}^{obs} & S_{m,366}^{obs} & \dots & S_{m,730}^{obs} & \dots & S_{m,n}^{obs} \end{bmatrix} \quad (1)$$

式中, n 为模拟变量总数,若模拟站点总数为 N_s ,则 $n=365 N_s$; 每一行代表各站不同日截口的洪

收稿日期: 2022-08-08, 修回日期: 2022-09-08

基金项目: 湖北省博士后创新实践岗位资助项目(2022CXGW004)

作者简介: 王权森(1993-),男,博士后,研究方向为水文水资源、智慧水利, E-mail: quansen_wang@163.com

水序列。

步骤 2 选取合适的分布函数对每列洪水序列进行拟合,并计算矩阵 \mathbf{S}^{obs} 的列与列之间的皮尔逊相关系数 ρ^{obs} 。

步骤 3 选取合适的 Copula 函数,基于多站洪水序列变量间的相关性构建联合分布,联合分布函数参数可用皮尔逊相关系数 ρ^{obs} 进行估计:

$$F(Q_1, Q_2, \dots, Q_n) = C(F_1(x), F_2(x), \dots, F_n(x)) \quad (2)$$

式中, Q_n 为第 n 个变量的洪水向量; $C(\cdot)$ 为 Copula 函数; $F(\cdot)$ 为 n 个变量的联合分布函数; $F_n(\cdot)$ 为第 n 个变量的边缘分布。

步骤 4 根据式(2)推求洪水变量的条件分布,随机生成服从均匀分布的向量 $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$,令第 1 个洪水变量的抽样概率 $\mu_1 = \epsilon_1$,基于 u_1 和 ϵ_2 ,通过条件分布 $C(u_2 | u_1) = \epsilon_2$ 可计算得到 u_2 ,以此类推,通过条件分布 $C(u_i | u_1, \dots, u_{i-1}) = \epsilon_i (i = 1, 2, \dots, n)$,得到 n 个洪水变量的模拟抽样概率序列 (u_1, u_2, \dots, u_n) ,通过每个抽样变量边缘分布的反函数 $S_i^{\text{sim}} = F_i^{-1}(u_i), i = 1, 2, \dots, n$ 插值,将抽样概率值转换为具体洪水变量值。重复 M 次即可得到具有相关结构的 M 年的多站模拟洪水序列矩阵 $\mathbf{S}_{M \times n}^{\text{sim}}$,其结构与 \mathbf{S}^{obs} 类似,可作为多站洪水样本库。

步骤 5 引用主从站随机模拟的方法,综合考虑流域地理特征,选取下游某个关键控制站点作为主站。考虑主站历史大洪水设计资料,选取主站重要的洪水特征,洪水特征具体可包括年洪峰流量、年均流量、年最大 1 d 洪量等。统计 M 年日径流洪水样本库中 l 个主站的洪水特征矩阵 \mathbf{C}_s :

$$\mathbf{C}_s = \begin{bmatrix} C_{s_{1,1}} & \dots & C_{s_{1,l}} \\ \vdots & \vdots & \vdots \\ C_{s_{M,1}} & \dots & C_{s_{M,l}} \end{bmatrix} \quad (3)$$

式中, C_s 有 M 行 l 列, $C_{s_{i,j}}$ 为模拟序列中主站第 i 年第 j 个洪水特征,其中 $i = 1, 2, \dots, M, j = 1, 2, \dots, l$ 。

2.2 考虑历史极端洪水的主站洪水特征随机模拟

考虑主站历史洪水特征设计资料,结合 Copula 法对主站洪水特征进行模拟延拓。具体模拟步骤包括洪水特征边缘分布拟合与洪水特征联合抽样模拟两部分。

针对洪水特征分布拟合,在中国往往采用 P-III 分布对此类洪水特征进行拟合,并结合站点实测序列与历史大洪水的洪水特征统计值,利用不连续矩法估计曲线的均值、 C_V 、 C_S 等参数^[6]。

进一步采用抽样方法对主站选取的洪水特征进行随机模拟,生成以下 n_1 年 l 个主站洪水特征序列 \mathbf{C}_M :

$$\mathbf{C}_M = \begin{bmatrix} C_{m_{1,1}} & \dots & C_{m_{1,l}} \\ \vdots & \vdots & \vdots \\ C_{m_{n_1,1}} & \dots & C_{m_{n_1,l}} \end{bmatrix} \quad (4)$$

式中, n_1 为模拟洪水特征序列的长度; l 为考虑洪水特征的个数; $C_{m_{i,j}}$ 为第 i 年第 j 个模拟洪水特征, $i = 1, 2, \dots, n_1, j = 1, 2, \dots, l$ 。

模拟的洪水特征序列 \mathbf{C}_M 中每个洪水特征的边缘分布考虑了历史大洪水资料,洪水特征极端值更加合理。与此同时,样本池内各洪水特征间的相关性结构也与实际历史洪水特征序列间的相关性结构保持一致。

2.3 基于重采样理论的多站洪水过程匹配搜索

基于生成的洪水特征矩阵 \mathbf{C}_M ,设定抽样目标,在多站径流过程样本库 \mathbf{S}^{sim} 中匹配搜索,得到考虑历史洪水特征且具有年际连续性时间因次的多站洪水序列。其中设定的抽样目标可分为两部分,第一部分为最小化洪水特征抽样误差,采用 \mathbf{C}_s 与 \mathbf{C}_M 之间各洪水特征的相对误差值之和来量化表征。第二部分则采用相关系数差值来代表各站点模拟洪水序列年与年之间的连续性特征与实测序列一致性,具体数学表达式如下:

$$\min O = \sum_{k=1}^l \frac{|C_{m_k} - C_{s_k}|}{C_{m_k}} + \sum_{j=1}^{N_s} \sum_{p=1}^L \sum_{q=n-l+p}^{N_a} (\rho_{p,q}^j - \zeta_{p,q}^j)^2 \quad (5)$$

式中, C_{m_k} 、 C_{s_k} 分别为特征矩阵 \mathbf{C}_M 与 \mathbf{C}_s 对应行数的第 k 个洪水特征值; N_s 为站点总个数; N_a 为一年内的总的截口数; $\rho_{p,q}^j$ 、 $\zeta_{p,q}^j$ 分别为模拟序列与实测序列第 j 个站点第 p 、 q 个截口洪水的相关系数,其中截口 p 与截口 q 的时滞为 L 。

在确立抽样目标 O 后,可通过 n_1 次不放回抽样模拟生成多站径流样本。对于第一次抽样,仅需满足模拟不同洪水特征相对误差值之和最小的目标,对于剩余 $n_1 - 1$ 次不放回抽样,则需严格按照目标 O 进行抽样。根据 n_1 个行号在模拟得到的洪水序列矩阵 \mathbf{S}^{obs} 中取出对应行的各站洪水序列,进而模拟出 n_1 年多站模拟洪水序列。

模拟序列的洪水特征误差可用模拟洪水特征值与实际洪水特征值之间的相对误差来表征,相关性结构误差可通过矩阵误差指标 E_r 来判断模拟序列的相关结构保持程度:

$$E_r = \frac{1}{n_1} \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (\boldsymbol{\rho}_{i,j}^{obs} - \boldsymbol{\rho}_{i,j}^{sim})^2} \quad (6)$$

式中, E_r 为实测序列与模拟序列相关系数矩阵的 R_{RMSE} , 其值越小, 代表模拟序列的相关结构保持的越好; $\boldsymbol{\rho}_{i,j}^{sim}$ 、 $\boldsymbol{\rho}_{i,j}^{obs}$ 分别为第 i 行第 j 列模拟、实测洪水序列相关系数矩阵。

3 实例分析计算

以长江干流屏山站、寸滩站及宜昌站为例, 其中宜昌水文站位于长江干流上中游分界点, 年平均径流量为 $4\,510 \times 10^8 \text{ m}^3$, 是长江三峡水利枢纽工程的总出库控制站; 寸滩站位于宜昌站上游长江干流和嘉陵江汇合口处, 其多年平均径流量为 $3\,478 \times 10^8 \text{ m}^3$, 占宜昌站径流量的 80.50%; 屏山站位于寸滩站上游, 其多年平均汛期(5~10月)水量约占宜昌水量的 1/3, 是长江宜昌站洪水的基础来源。

输入数据采用屏山、寸滩与宜昌站 1965~2010 年经过还原后 46 年同期径流数据, 选取 P-III 分布对三个站点日径流边缘分布进行拟合。由于三站日尺度洪水变量高达 1 095 维, 因此采用 Gaussian Copula 函数拟合各站点不同截面洪水的联合分布, 并进行相关性抽样模拟得到有效保证高阶自互相关特征的 10 000 年三站日洪水过程。将各站点日径流的均值(M_{Mean})、变差系数(C_V)、偏态系数(C_S)与多站季节性自回归模型(MSAR)模拟结果进行对比, 两种方法的检验结果见图 2。由图 2 可知, 使用两种方法模拟的径流序列均能有效保持均值、 C_V 、 C_S 这些基本统计量的统计特征, 三者的相对误差几乎可忽略不计, 研究所提方法与传统模拟法均能合理保证实测径

流基本统计特征。

为进一步检验不同截面(日)之间相关性结构, 图 3 为各站点各截面一阶自相关系数、二阶自相关系数及宜昌站与上游屏山、寸滩站点的零阶互相关系数的检验结果。由图 3 可知, 与传统 MSAR 方法相比, 本文方法能够较好地贴近真实相关系数, 尤其从一阶与二阶相关系数检验结果来看, 所提方法优势更加明显。进一步结合式(6)方法检验涵盖全阶自互相关系数的误差值 E_r , 通过 MSAR 模型计算得到的 E_r 值为 0.149, 而基于所提方法计算得到的相关系数矩阵误差值仅为 0.018, 研究表明所提方法能够在现有算力条件下, 进一步充分保证模拟序列自身的低阶、高阶自相关特性和各站点之间的互相关特性, 可同时保证多站模拟径流时间序列所特有的相关性结构特征。

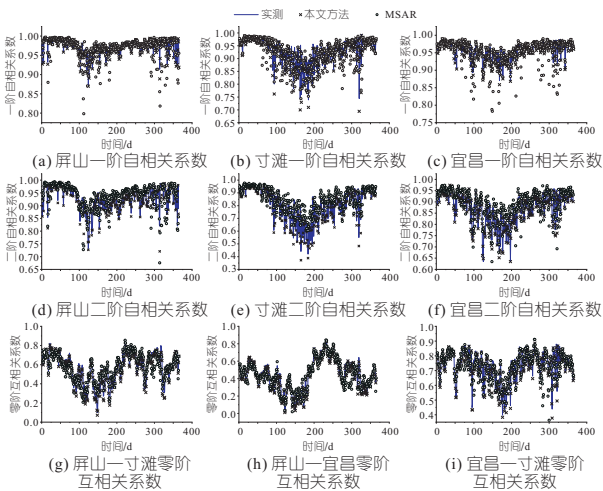


图 3 各站点日径流序列 1 阶、2 阶自相关系数及 0 阶互相关系数检验结果

Fig. 3 The auto-correlation coefficients with the lag-1 and lag-2 and the cross-correlation coefficients with lag-0 of simulated and observed daily streamflow

同时以宜昌站为主站, 结合其历史设计资料及实测洪水资料, 模拟出满足不同洪水特征边缘分布特征与特征间相关性结构的宜昌站 1 000 年洪水特征序列集。在此基础上, 以满足主站(宜昌站)洪水特征误差最小与洪水序列连续性特征为目标, 在 10 000 年三站模拟日径流序列中进行采样, 得到考虑历史大洪水资料的 1 000 年三站同期日径流模拟洪水序列, 统计模拟洪水序列中宜昌站 1 000 年逐年模拟洪水特征值, 绘制洪水特征边缘分布见图 4。由图 4 可知, 基于重采样得到的日径流序列推求得到的洪水特征值整体分布与设计特征分布一致, 所提方法能够进一步综合利用历史设计资料, 有效模拟出合理、可靠的极端

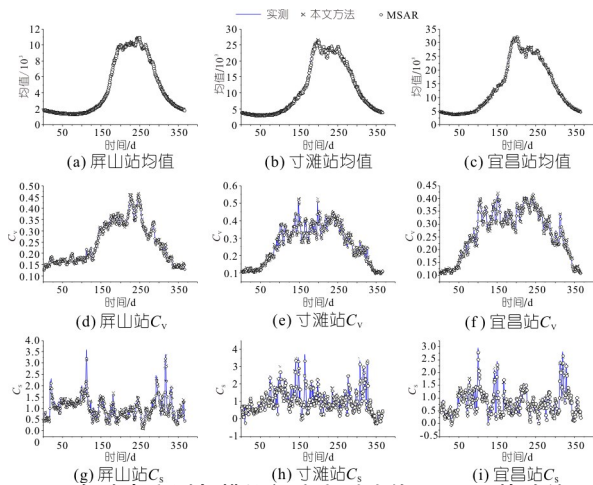


图 2 各站点实测与模拟径流序列均值、 C_V 、 C_S 检验结果
Fig. 2 Statistical results including the mean, C_V and C_S of simulated and observed daily flood series

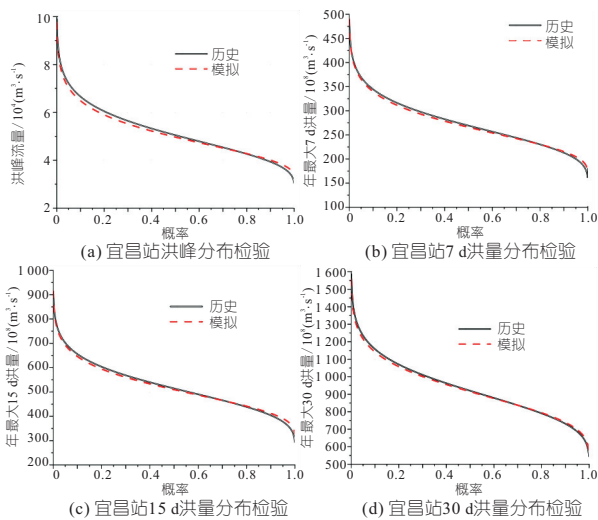


图 4 宜昌站实测与模拟洪水特征分布检验结果

Fig. 4 The distributions of the flood characteristics for the simulated and historical flood series at the target station Yichang station

洪水情景。但该方法对基础样本池(10 000 年序列)的样本大小要求较高,该方法距离实时洪水场景快速模拟生成的业务需求还有一定差距,有待进一步研究。

此外,统计各站点实测序列与模拟序列每年第 1、2 d 两个截面与前一年最后 2 d 的相关系数矩阵,计算实测相关系数矩阵与模拟相关系数矩阵的绝对误差值,见表 1。由表 1 可知,各站模拟序列年际相关系数绝对误差值最大仅为 0.028 9,几乎可忽略不计,总体而言,经过重采样生成的模拟洪水序列能较好保持实测洪水的年际相关特征。

4 结论

本文提出了一种基于条件重采样技术的多站洪水过程随机模拟方法,对长江流域屏山、寸滩、宜昌站日尺度洪水进行模拟,模拟的洪水过程在

表 1 各站点实测序列与模拟序列年际相关系数矩阵绝对误差

Tab. 1 Historical and simulated correlation coefficients for transitional (year to year) days at each site

站点	项目	第 1 d	第 2 d	第 364 d	第 365 d
屏山	第 1 d	0	0.001 2	0.011 8	0.025 2
	第 2 d	0.001 2	0	0.012 5	0.027 0
	第 364 d	0.011 8	0.012 5	0	0.008 7
	第 365 d	0.025 2	0.027 0	0.008 7	0
寸滩	第 1 d	0	0.002 6	0.010 9	0.014 1
	第 2 d	0.002 6	0	0.011 3	0.000 2
	第 364 d	0.010 9	0.011 3	0	0.000 1
	第 365 d	0.014 1	0.000 2	0.000 1	0
宜昌	第 1 d	0	0.006 6	0.013 2	0.018 7
	第 2 d	0.006 6	0	0.002 9	0.018 9
	第 364 d	0.013 2	0.002 9	0	0.000 9
	第 365 d	0.018 7	0.018 9	0.000 9	0

满足统计特征与多站年际、年内相关结构的同时,进一步保持了下游关键控制站点关键洪水特征的边缘分布及其相关特性,使模拟的极端洪水情景更加可靠。研究成果可为流域水资源系统规划管理和风险评估与决策提供数据支撑与决策支持。

参考文献:

- [1] 邢子强, 严登华, 鲁帆, 等. 人类活动对流域旱涝事件影响研究进展 [J]. 自然资源学报, 2013, 28 (6): 1070-1082.
- [2] 杨洪平. 水文水资源随机模拟现状及发展方向 [J]. 水科学与工程技术, 2009(3): 6-8.
- [3] YUREKLI K, KURUNC A, SIMSEK H. Prediction of daily maximum streamflow based on stochastic approaches [J]. Journal of spatial hydrology, 2004, 4(2): 1-12.
- [4] 王文圣, 金菊良, 李跃清. 水文随机模拟进展 [J]. 水科学进展, 2007(5): 768-775.
- [5] 马细霞, 朱慧青. 基于信息熵和统计参数的径流随机模拟序列检验方法 [J]. 水电能源科学, 2019, 37 (4): 5-8, 51.
- [6] 王燕生. 工程水文学 [M]. 北京: 中国水利电力出版社, 1992.

Research on Stochastic Simulation Method of Multi-station Flood Series Based on Conditional Resampling

WANG Quan-sen^{1,2}, LI An-qiang^{1,2}, LU Cheng-wei^{1,2}

(1. CISPDR Corporation, Wuhan 430010, China;

2. Hubei Key Laboratory of Basin Water Security, Wuhan 430010, China)

Abstract: Taking the self- and cross-correlation of the multi-station streamflow series of the main and tributary of the basin as the starting point, a multi-site flood series simulation method based on the conditional resampling theory was proposed, and simulate the flood series of Pingshan, Cuntan and Yichang station on the main stream of Changjiang River. The simulation results were compared with the results of the multi-station seasonal autoregressive model (MSAR) for verification and analysis. The results show that compared with the traditional model, the proposed method can not only satisfy the basic statistical characteristics of the runoff series and the higher-order self- and cross-correlation structures of the series, but also further consider the historical extreme flood characteristic, and could generate multi-station simulated flood scenarios with different design frequency.

Key words: multi-station flood series simulation; conditional resampling; higher-order self-and cross-correlation structures; historical extreme flood characteristics