

DOI: 10.20040/j.cnki.1000-7709.2023.20220962

基于 GAMLSS 模型的滹沱河径流分析

李永建¹, 郝文龙^{1,2}, 朱长军¹, 曹 青^{2,3}

(1. 河北工程大学能源与环境工程学院, 河北 邯郸 056038; 2. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098; 3. 南京信息工程大学水文与水资源工程学院, 江苏 南京 210044)

摘要: 为探究当今环境下径流序列的非一致性特征, 更好地预判洪水事件的发生, 以海河流域的滹沱河为例, 采用 GAMLSS 模型, 以时间、水库指标和气候指标为协变量, 对滹沱河 1970~2012 年间的径流序列进行非一致性频率分析。结果表明, 滹沱河的径流序列存在明显的非一致性特征, 表现出趋势性和突变性; GAMLSS 模型对变化环境下径流模拟具有良好的效果, 可充分捕捉气候变化对洪水的影响, 优于传统的一致性模型; 大尺度气候指标 NINO1+2 和 NAO 的变化显著影响滹沱河的洪水事件, 该两项气候指标值增高时带来低洪水风险, 降低时带来高洪水风险。

关键词: GAMLSS; 非一致性; 水文频率分析; 滹沱河

中图分类号: TV123

文献标志码: A

文章编号: 1000-7709(2023)03-0014-05

1 概况

滹沱河区河流属海河流域的子牙河水系。受地理环境和气候条件所制约, 河流兼具山地型和夏雨型的双重特性。河流形态和河道特征为沟壑密度大、水系发育、坡陡流急、河流侵蚀切割严重。集水面积 11 936 km², 干流河长 260 km, 平均纵坡 2.17%^[1]。界河铺为滹沱河干流上的水文站, 控制流域面积 6 031 km²^[1]。在界河铺水文站上游的大中型水库共 4 座, 总库容 6 445 × 10⁴ m³, 分别为孤山水库、下茹越水库、神山水库、观上水库^[2]。针对滹沱河近年来流域环境变化显著所表现出的非一致性特征^[2], 以一致性假设为前提的传统水文频率分析方法受到挑战, 非一致性特征在水文序列的研究中日益被重视, 莫淑红等^[3,4]对玛纳斯河与小理河流域的径流进行模拟分析的过程中, 均得到 GAMLSS 模型的拟合度较高且优于传统的 P-III 型分布的结论。为此, 本文选择 GAMLSS 模型, 并在已有研究的基础上引用气候因子为协变量, 通过对协变量的相关性的筛选来优化模型, 对其径流量进行更精确的模拟和分析, 旨在明确非一致性特征对水文序列的影响。

2 数据来源与研究方法

2.1 数据来源

水文数据选取界河铺水文站 1970~2012 年间 43 年的实测径流数据。从日径流量中提取出逐年最大 1、5 d 径流量。另外, 选取 WP、NAO、NINO1+2、NINO3.4、NINO3、NINO4、SOI、PDO、AO 和 NP 10 个大尺度气候指标。单独月份可能会存在不正常的极值或波动, 故选取每年降水多发的 7 月~次年 6 月的连续 3 个月的均值作为预测因子。

2.2 研究方法

2.2.1 GAMLSS 模型

GAMLSS 模型^[5]相比广义线性模型和可加模型更具优势, 主要体现在其建模框架十分灵活, 且在描述解释变量与随机变量序列存在关系的方式可用线性函数、非线性或平滑函数来描述^[6]。本文从五种备选的两参数函数 (Weibull、Gamma、Lognormal、Logistic 和 Gumbel) 中通过 AIC 准则进行筛选 (表 1), 然后用 worm 图即残差图判断拟合质量。

收稿日期: 2022-05-08, **修回日期:** 2022-06-06

基金项目: 河北省自然科学基金项目 (E2020402057); 水文水资源与水利工程科学国家重点实验室“一带一路”水与可持续发展科技基金项目 (2020491111, 2020491211)

作者简介: 李永建 (1997-), 男, 硕士研究生, 研究方向为水文物理规律模拟及预报, E-mail: 996604138@qq.com

通讯作者: 郝文龙 (1988-), 男, 讲师, 研究方向为水文物理规律模拟及预报, E-mail: hwl5022202@163.com

表 1 五种函数的详细信息
Tab. 1 Details of five functions

| 分布函数 | 概率密度函数 | 连接函数 | |
|-----------|---|------------|------------|
| | | θ_1 | θ_2 |
| Weibull | $f_Y(y \theta_1, \theta_2) = \frac{\theta_2 y^{\theta_2 - 1}}{\theta_1} \exp\left[-\left(\frac{y}{\theta_1}\right)^{\theta_2}\right]$ $y > 0, \theta_1 > 0, \theta_2 > 0$ | log | log |
| Gamma | $f_Y(y \theta_1, \theta_2) = \frac{1}{(\theta_2^2)^{1/\theta_2}} \frac{y^{1/\theta_2 - 1} \exp(-y/\theta_2^2)}{\Gamma(1/\theta_2^2)}$ $y > 0, \theta_1 > 0, \theta_2 > 0$ | log | log |
| Lognormal | $f_Y(y \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2^2}} \frac{1}{y} \exp\left\{-\frac{[\log(y) - \theta_1]^2}{2\theta_2^2}\right\}$ $y > 0, \theta_1 > 0, \theta_2 > 0$ | identity | identity |
| Logistic | $f_Y(y \theta_1, \theta_2) = \frac{\exp(y - \theta_1/\theta_2)}{\sigma [1 + \exp(y - \theta_1/\theta_2)]^2}$ $-\infty < y < \infty, -\infty < \theta_1, \theta_2 > 0$ | identity | log |
| Gumbel | $f_Y(y \theta_1, \theta_2) = \frac{1}{\theta_2} \exp\left\{-\left[\frac{y - \theta_1}{\theta_2} - \exp\left(-\frac{y - \theta_1}{\theta_2}\right)\right]\right\}$ $-\infty < y < \infty, -\infty < \theta_1 < \infty, \theta_2 > 0$ | identity | log |

注:表中公式字母含义详见文献[7]。

2.2.2 水库指标

LÓPEZ J 等^[8]于 2013 年提出了一个无量纲的水库指标(R_{RI}),该指标能反映水库或大坝的建设对径流的影响,其值越大证明水库对径流影响越大。其计算公式为:

$$R_{RI} = \frac{\sum_{i=1}^N \left(\frac{A_i}{A_T}\right) \left(\frac{C_i}{C_T}\right)}{N} \quad (1)$$

式中, N 为水文站点上游水库或大坝的数量; A_i 为每个水库的控制流域面积; A_T 为水文站点的控制流域面积; C_i 为水库的总库容; C_T 为该水文站点的年均径流量。

2.2.3 非一致性判断

对径流时间序列,通过趋势性和突变性来判断其非一致性特征。采用非参数 Mann-Kendall (M-K)趋势性检验法和 Spearman 秩次检验法判断径流时间序列是否具有趋势性,采用 Mann-Kendall 突变性检验、Pettitt 检验和 Buishand 检验来判断径流序列是否存在突变点。

2.2.4 非一致性频率分析

首先拟合得出最适于该径流序列的分布函数,然后通过建立分布函数的参数与解释变量(时间、气候指标、水库指标)的关系对极端径流现象的非一致性特征进行归因分析和频率分析。

使用 Kendall 相关系数法选出对径流序列产生影响最为明显的气候指标,结合水库系数构造分布参数和解释因子的关系,最终建立两种以 GAMLSS 模型为基础的非一致性模型:①模型 1,变量为时间;②模型 2,变量为气候指标和水库指标(R_{RI})。

3 结果与分析

3.1 非一致性诊断结果

3.1.1 趋势性诊断结果

M-K 检验法和 Spearman 检验法检验结果见表 2。由表 2 可看出,界河铺站的最大 1 d 径流量序列在经过检测后,两种检测方法的 p 值均小于 0.05,即认为在 1970~2012 年间界河铺站最大 1 d 径流量序列的趋势性在 95%置信水平下显著,具有趋势性,且呈显著下降趋势。

表 2 界河铺站两径流序列趋势性诊断结果

Tab. 2 Results of trend diagnosis of two runoff series at Jiehepu Station

| 径流序列 | Mann-Kendall 检验 | | Spearman 检验 | |
|------------|-----------------|-------|-------------|-------|
| | Z | p | r | p |
| 最大 1 d 径流量 | -1.989 | 0.047 | -0.330 | 0.031 |
| 最大 5 d 径流量 | -1.381 | 0.167 | -0.240 | 0.121 |

将线性回归函数和局部加权 loess 回归函数加入界河铺水文站的最大 1、5 d 径流量序列的散点图中,画出两种径流序列的拟合图,见图 1。由图 1 可看出,界河铺站的最大 1 d 径流量序列呈显著下降趋势,与表 2 所得结论一致。

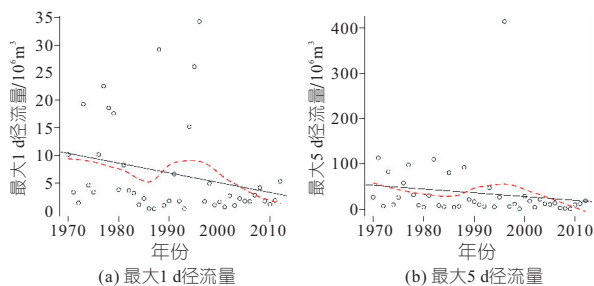


图 1 界河铺径流量线性回归和 Loess 回归拟合图
Fig. 1 Fitting diagram of linear regression and Loess regression of runoff at Jiehepu Station

3.1.2 突变点诊断结果

M-K 检验结果均通过 95%置信区间,界河铺

站最大 1 d 径流量突变点检验结果中(表 3), Pettitt 检验、M-K 检验的结果分别为 1984、1982 年,且均通过 95%置信水平。而 Buishand 检验的结果为 1997 年,且 p 值大于 0.05,故舍弃。综合来看,认为界河铺站的径流量序列在 1980 年代初发生了突变。界河铺站最大 5 d 径流量突变点检验结果中, Buishand 检验结果的 p 值过大,舍弃。Pettitt 检验、M-K 检验的结果分别为 1984、1982 年,可判断界河铺水文站的最大 5 d 径流量序列在 1980 年代初发生了突变。

表 3 界河铺站两径流序列突变点诊断结果

Tab.3 Diagnosis results of two runoff series mutation points at Jiehepu Station

| 径流序列 | M-K 检验 突变发生年份 | Pettitt 检验 | | Buishand 检验 | | | |
|------------|------------------|------------|--------|-------------|------------|-------|--------|
| | | 突变发 生年份 | U_k | p | 突变发 生年份 | U | p |
| | | 最大 1 d 径流量 | 1982 年 | 1984 年 | 224 | 0.049 | 1997 年 |
| 最大 5 d 径流量 | 1971、1983 年 | 1984 年 | 182 | 0.174 | 1997 年 | 0.152 | 0.394 |

下茹越水库于 1975 年建成蓄水,与突变年份较为接近,且其恰好位于界河铺上游的滹沱河干流上,结合其实际情况可判断下茹越水库的建设是众多导致突变原因中影响较大的因素,且突变影响存在着滞后性,分析其原因在于:①下茹越水库距界河铺水文站仍有一段距离,其河道曲折、山谷或沟壑等特殊地形、流域植被生长、农田取水灌溉等因素,均会使径流的突变产生滞后。②由于数据限制本文仅能对 1970 年开始的径流序列进行分析,而下茹越水库在 1975 年建成,这期间仅有 5 年的数据,导致分析不准确且突变节点滞后。

3.2 非一致性频率分析结果

3.2.1 水库指标分析

图 2 为界河铺站径流序列实测值与水库指标的变化。由图 2 可看出,在 1975 年下茹越水库建成后径流量出现明显下降,表明该水库对其径流量产生较大影响,与表 3 突变点检验结果相吻合。所以,水库指标在一定程度上可反映水库对水文站径流序列的影响,可用来作为协变量对径流序列进行非一致性频率分析。

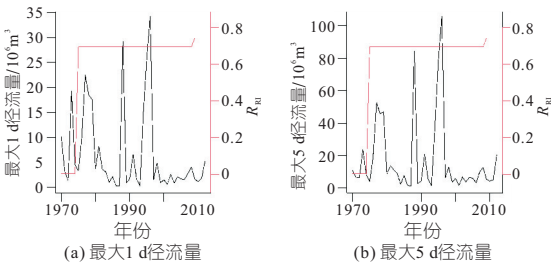


图 2 界河铺站径流序列实测值和水库指标 (R_{RI} 值) (R_{RI} values) at Jiehepu Station

Fig.2 Measured values of runoff series and reservoir index (R_{RI} values) at Jiehepu Station

3.2.2 模型 1 模拟结果

根据 AIC 准则,并以 GD 值为参考,从五种备选函数中选出 Lognormal 函数为最匹配界河铺站径流序列的函数(表 4),并经过模拟得出在位置参数 θ_1 随时间变化的情况下 AIC 值最小,模拟结果见表 5。

表 4 界河铺站径流序列拟合五种分布函数的 GD、AIC 值

Tab.4 GD and AIC values of five distribution functions fitted by runoff series of Jiehepu Station

| 径流序列 | 项目 | Lognormal | Logistic | Gamma | Gumbel | Weibull |
|--------|--------|-----------|----------|---------|---------|---------|
| 最大 1 d | AIC | 1 431.4 | 1 494.3 | 1 439.1 | 1 528.7 | 1 437.8 |
| | 径流量 GD | 1 427.4 | 1 490.3 | 1 435.1 | 1 524.7 | 1 433.8 |
| 最大 5 d | AIC | 1 511.3 | 1 576.6 | 1 522.4 | 2 368.2 | 1 521.1 |
| | 径流量 GD | 1 506.3 | 1 572.6 | 1 518.4 | 2 364.2 | 1 517.1 |

表 5 界河铺站径流序列的最优分布、位置参数和尺度参数模拟结果

Tab.5 Simulation results of optimal distribution, location parameters and scale parameters of runoff series at Jiehepu Station

| 站点 | 水文序列 | 最优分布 | 位置参数 θ_1 | 尺度参数 θ_2 |
|-----|------------|-----------|-------------------------------------|---------------------------|
| 界河铺 | 最大 1 d 径流量 | Lognormal | $\log(\theta_1) = -0.031t + 76.998$ | $\log(\theta_2) = -0.163$ |
| | 最大 5 d 径流量 | Lognormal | $\log(\theta_1) = -0.016t + 48.675$ | $\log(\theta_2) = -0.073$ |

得到最优模型后画出每个径流序列的分位数图见图 3。由图 3 可看出,各分位数曲线整体呈下降趋势,这与趋势性检验结果相一致。且随着时间的增加其波动性逐渐变小,与 5%分位数曲线和中位数曲线相比,95%分位数曲线下降程度最大,表明 95%分位数曲线的水文过程敏感性相对较大。

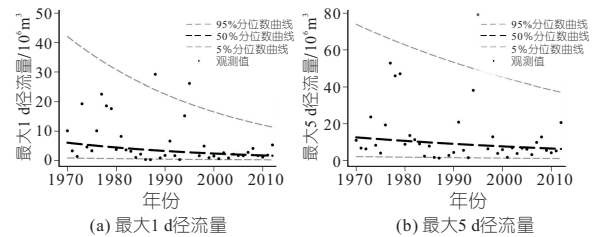


图 3 界河铺站径流序列在非一致性模型 1 条件下的分位数

Fig.3 Quantile of runoff series at Jiehepu Station in non-conformity model 1

表 6 为界河铺站径流序列在最优模型下的残差平均值、方差和 Filliben 系数。由表 6 可看出,模拟结果的 Filliben 系数均大于 0.983,意味着模

表 6 界河铺站径流序列在最优模型下的残差平均值、方差和 Filliben 系数

Tab.6 Mean residual, variance and Filliben coefficient of runoff series at Jiehepu Station in the optimal model

| 站点 | 水文序列 | 平均值 | 方差 | Filliben 系数 |
|-----|------------|------------------------|-------|-------------|
| 界河铺 | 最大 1 d 径流量 | -0.1×10^{-15} | 1.024 | 0.992 |
| | 最大 5 d 径流量 | -0.1×10^{-15} | 1.024 | 0.983 |

型的残差很好地服从了正态分布。且残差的平均值和方差分别接近 0 和 1,说明残差偏离其正常值并不显著。

图 4 为最优模型下两径流序列拟合残差 worm 图。由图 4 可看出,模型的残差分布在 95% 置信区间内,进一步证明了建立的最优模型具有良好的模拟效果,以时间为协变量的非一致性模型可用于模拟滹沱河流域非一致性径流行为。

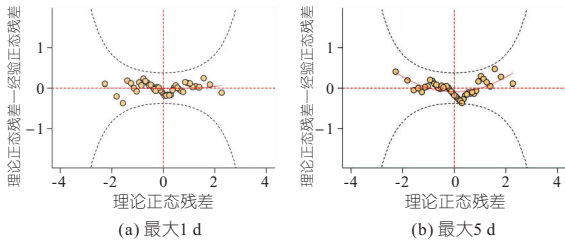


图 4 界河铺站径流序列在最优模型 1 条件下拟合残差的 worm 图

Fig. 4 Worm figure of fitting residuals of Jiehepu Station runoff series in optimal model 1

表 7 界河铺站径流在非一致性模型 2 条件下的 AIC 值、残差的平均值和方差位置参数 θ_1 、尺度参数 θ_2 和 Filliben 系数结果

Tab. 7 The results of AIC values, mean residual and variance location parameters θ_1 , scale parameters θ_2 and Filliben coefficients of event at Jiehepu Station in non-conformity model 2

| 径流序列 | AIC | 位置参数 θ_1 | 尺度参数 θ_2 | 平均值 | 方差 | Filliben 系数 |
|------------|--------|--|--|-------|-------|-------------|
| 最大 1 d 径流量 | 285.83 | $\log(\theta_1) = -0.105 N_{NINO1+2(2-4)} + 4.287$ | $\log(\theta_2) = -0.416 6 N_{NINO1+2(2-4)} + 10.918 9$ | 0.025 | 1.023 | 0.995 |
| 最大 5 d 径流量 | 240.23 | $\log(\theta_1) = -0.582 6 N_{NAO(3-5)} + 2.159 9$ | $\log(\theta_2) = 1.505 3 R_{RI} - 0.360 6 N_{NAO(3-5)} - 0.996$ | 0.024 | 1.023 | 0.988 |

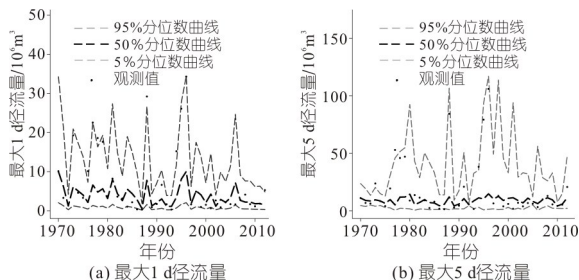


图 5 界河铺站各径流序列在非一致性模型 2 条件下的分位数曲线模拟

Fig. 5 Simulation of quantile curves of runoff series at the Jiehepu Station in non-conformity model 2

风险具有一定的可信度。总体优于模型 1 的模拟结果。

图 6 为界河铺站径流序列在最优模型 2 条件下拟合残差的 worm 图。由图 6 可看出,模型的残差分布在 95% 置信区间内,进一步证明了建立的最优模型具有良好的模拟效果。

为更直观地体现气候指标对径流序列的影响,以 $N_{NAO(3-5)}$ 、 $N_{NINO1+2(2-4)}$ 两项指标为自变量,做出径流量随气候指标变化的曲线见图 7。由图 7 可看出:①随着两项指标的增大径流量序列呈波动性减小趋势,表明该气候指标升高对应低洪水风险,降低对应高洪水风险。②滹沱河的径流序列分位数曲线存在一定的波动性,其原因

3.2.3 模型 2 模拟结果

低频气候指数会对径流变化产生影响^[9]。本文根据 Kendall 相关系数法在 10 个大尺度气候指标中筛选出在 1970~2012 年间与界河铺站径流量相关性最强的气候指标,即 3~5 月的 NAO (简称 $N_{NAO(3-5)}$)、2~4 月的 NINO1+2 (简称 $N_{NINO1+2(2-4)}$)。通过建立对界河铺站径流序列产生显著影响的潜在气候指标和水库指标与分布参数的关系,对界河铺站两径流序列进行非一致性频率分析,其结果见表 7。

在非一致性条件下由显著影响因子为协变量对径流序列进行频率模拟的结果见图 5。由图 5 可看出,在加入了显著影响因子为协变量后模拟结果由曲线变为波动性的折线,直观上与实测值的拟合度更高。同时波动性将气候指标对径流序列产生的影响更直观地体现出来,部分峰值与实测值非常接近,这表明根据气候指标来预测洪水

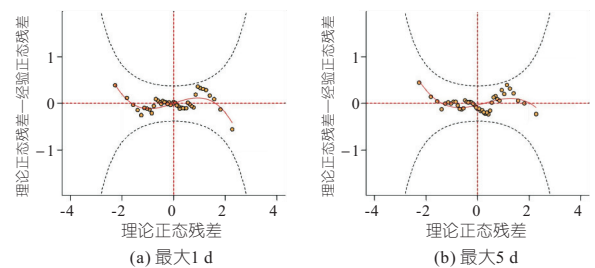


图 6 界河铺站径流序列在最优模型 2 条件下拟合残差的 worm 图

Fig. 6 Worm figure of fitting residuals of Jiehepu Station runoff series in optimal model 2

在于水库指标 R_{RI} 对径流序列存在着干扰作用。界河铺径流序列选中了水库指标 R_{RI} 值作为协变量进行分析,而水库指标只有在水库建成时才会发生变化,产生的 4 次波动分别对应 4 座水库的建成。界河铺最大 5 d 径流量序列的分位数图波动很小,这是因为最大 5 d 径流量对上游水库带来的径流变化具有缓释作用,不像最大 1 d 径流量变化的非常剧烈。

4 结论

a. 滹沱河的径流序列在 1970~2012 年间呈显著下降趋势,且存在显著突变点,众多因素中下

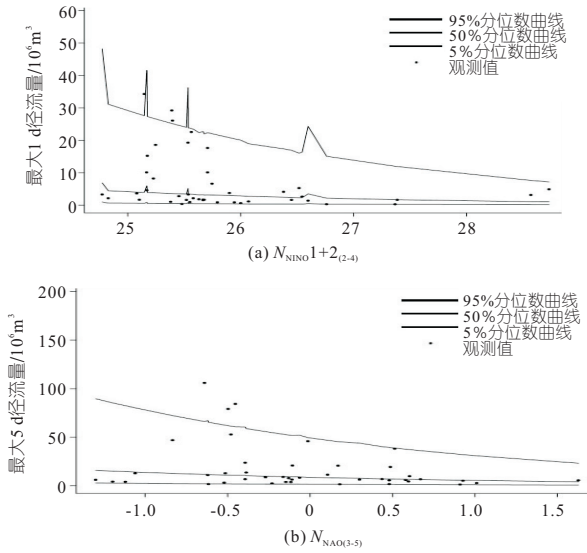


图 7 界河铺站径流序列分位数在非一致性模型 2 条件下随气候指标的变化

Fig. 7 Variation of quantile of runoff series at Jiehepu Station with climate index in non-conformity model 2

茹越水库对其径流量影响较大,是导致径流发生突变的主要原因。

b. 在引入气候因子为解释变量后两种径流序列模拟结果的 Filliben 系数分别提升了 0.03 和 0.05,模型的拟合度提高。表明 GAMLSS 模型对变化条件下的径流模拟具有良好的环境适应性,相比传统模型可捕捉径流量受气候变化而产生的变动。

c. 大尺度气候因子 NINO1 + 2 和 NAO 影响滹沱河的径流变化,NINO1 + 2 值低于 24, NAO 值低于 -1,预示着很高的洪水风险,对水文工作具有参考意义。

参考文献:

[1] 于宪民. 滹沱河界河铺站水文特性分析[J]. 科技情报开发与经济, 2011, 21(26): 182-184.

[2] 崔炳玉. 气候变化和人类活动对滹沱河区水资源变化的影响[D]. 南京:河海大学, 2004.

[3] 莫淑红, 李晨星, 邢华, 等. 基于 GAMLSS 模型的小理河流域年径流研究[J]. 应用基础与工程科学学报, 2022, 30(1): 40-49.

[4] 郑锦涛, 陈伏龙, 张鑫厚, 等. 基于 GAMLSS 模型的玛纳斯河设计年径流分析[J]. 气候变化研究进展, 2018, 14(3): 257-265.

[5] RIGBY R A, STASINOPOULOS D M. Generalized additive models for location, scale and shape (with discussion)[J]. Applied statistics, 2005, 54: 507-554.

[6] 江聪, 熊立华. 基于 GAMLSS 模型的宜昌站年径流序列趋势分析[J]. 地理学报, 2012, 67(11): 1505-1514.

[7] 郝文龙. 变化环境下汉江流域极端水文气象事件频率分析与研究[D]. 南京:河海大学, 2019.

[8] LÓPEZ J, FRANCÉS F. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates [J]. Hydrology & earth system sciences, 2013, 17: 3189-3203.

[9] HAO W L, SHAO Q X, WEI P, et al. Evaluation and projection of the annual maximum streamflow in response to anthropogenic and climatic effects under nonstationary conditions in the Hanjiang River Basin, China [J]. Journal of water and climate change, 2022, 13 (4): 1855-1877.

Analysis of Runoff of Hutuo River Based on GAMLSS Model

LI Yong-jian¹, HAO Wen-long^{1,2}, ZHU Chang-jun¹, CAO Qing^{2,3}

(1. College of Energy and Environmental Engineering, Hebei University of Engineering, Handan 056038, China; 2. State Key Laboratory of Hydrological Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China; 3. School of Hydrology and Water Resources, Nanjing University of Information Engineering, Nanjing 210044, China)

Abstract: In order to investigate the non-stationary characteristics of runoff series in changing environments and to better predict the occurrence of flood events, taking Hutuo River in the Haihe River Basin as the research object, GAMLSS model was used to conduct a non-coherent frequency analysis of the runoff series during 1970 and 2012 by taking time, reservoir indicators and climate indicators as covariates. The results show that the runoff series of Hutuo River has obvious non-stationary characteristics with trend and abrupt changes. The GAMLSS model can better simulate the runoff series in changing environments and describe the flood regimes under the influence of climate change, thus outperforming the traditional model. Changes of the large-scale climate indicators NINO1 + 2 and NAO significantly affect the flood events in the Hutuo River, with increasing values of these two climate indicators indicating low flood risk and decreasing values indicating high flood risk.

Key words: GAMLSS; non-conformity; hydrological frequency analysis; Hutuo River