

文章编号: 1000-7709(2018)01-0020-04

# 基于 R 软件的水文频率计算适线绘制

谢子波<sup>1,2</sup>, 鲁 帆<sup>2</sup>, 朱 奎<sup>1</sup>, 宋昕熠<sup>1,2</sup>

(1. 中国矿业大学 资源与地球科学学院, 江苏 徐州 221116; 2. 中国水利水电科学研究院 水资源研究所, 北京 100048)

**摘要:** 针对现有水文频率计算方法的缺点, 利用 R 软件编程完成了海森概率格纸绘制, 推求 P-Ⅲ 曲线的参数、计算水文变量序列的经验频率值与理论频率值及绘制理论频率曲线的工作, 最后结合目估适线法与离差平方和最小准则确定最佳参数组合。实例验证结果表明, R 软件中的统计函数适用于水文频率计算, 与其他软件相比, 节省了程序编制时间, 极大地提高了工作效率。利用编制的程序可快速匹配 P-Ⅲ 分布的不同参数值并计算曲线的离差平方和, 有效降低了计算量。研究成果可为水文频率计算工作提供参考。

**关键词:** R 软件; 水文频率计算; 目估适线; 离差平方和

**中图分类号:** TV11; P333 **文献标志码:** A

## 1 引言

水文频率曲线是推求工程水文设计值的重要工具, 以往的频率计算以人工为主, 调参和适线过程较复杂, 完成计算费时费力。随着计算机技术的发展, 许多软件都被应用到水文频率计算中。吴明官等<sup>[1]</sup>利用 Excel 完成了水文频率适线, 获得了符合工程要求的结果; 邢广军<sup>[2]</sup>基于 Visual Basic 开发的软件支持单一系列和综合频率曲线的绘制。但这些软件非专业统计软件, 可供调用函数较为有限, 程序编写过程难度较大。Matlab 等专业工具的应用进一步提高了水文计算效率<sup>[3,4]</sup>。R 软件是在 S 语言的基础上开发的一套集数据处理、计算和制图为一体的免费软件系统<sup>[5]</sup>。其应用范围已涵盖数据挖掘、机器学习、计量经济学、自然语言处理和空间统计学等诸多领域。因此, 本文研究了利用 R 软件计算水文频率的原理和实现频率曲线适线的过程, 并引入纵向离差平方和作为适线的辅助判断手段, 提高了适线的效率和准确度。研究成果可为水文频率计算工作提供参考。

## 2 R 软件计算水文频率的优势及原理

R 软件提供了有弹性的、互动的环境来分析、

可视及展示数据。提供了若干统计程序包, 以及一些集成的统计工具和各种数学计算、统计计算的函数, 用户只需根据统计模型, 指定相应的数据库及相关参数, 便可灵活机动地进行数据分析等工作, 甚至可以创造出符合需要的新的统计计算方法。与 Excel、Visual Basic、Matlab 等软件相比, 其在统计领域应用的优势具体见表 1。

表 1 不同软件在统计领域的应用比较  
Tab.1 Comparison of different software in statistics application

项目	Excel	Visual Basic	Matlab	R
是否开源	否	否	部分开源	完全开源
是否为商业软件	是	是	是	否
软件适用平台	Windows/Mac	Windows	Windows/Mac/Linux	Windows/Mac/Linux
统计函数丰富程度	可完成假设检验、相关分析、回归分析、方差分析等较多的统计计算工作	无自带统计函数	可完成参数估计、假设检验、回归分析、判别分析、聚类分析等绝大部分统计计算工作	可完成绝大部分统计计算工作且可以编制自己的函数来扩充现有 R 语言
绘图函数丰富程度	可直接绘制折线图、柱状图、散点图等简单统计图表, 只能展示原始数据, 难以表达统计意义	无法直接绘制统计图表。	可以直接绘制直方图、茎叶图、箱线图、散点图、交互效应图、QQ 图、三维图等绝大部分统计图形	可直接绘制绝大部分统计图形, 通过高、低水平绘图函数搭配, 可以更容易地绘制复杂图形

水文频率计算包括分布模型选择、参数估计、经验频率计算、理论频率计算和绘图适线等过程, 得益于 R 软件丰富的函数类型和程序包的支持,

收稿日期: 2017-04-09, 修回日期: 2017-05-25  
基金项目: 国家自然科学基金项目(51679252, 51409246); 国家重点基础研究发展计划课题(2013CB036406)  
作者简介: 谢子波(1994-), 男, 硕士研究生, 研究方向为水文与水资源, E-mail: 1642440041@qq.com  
通讯作者: 鲁帆(1981-), 男, 博士、教授级高级工程师, 研究方向为水文与水资源, E-mail: lf9805320@163.com

各种分布类型在 R 软件中均能较方便地实现。表 2 列举了不同类型的水文变量概率分布在 R 软件中的函数及其所属程序包。

表 2 R 软件中可应用于水文频率计算的分布函数

Tab.2 Distribution functions in R software which can be used in hydrological frequency calculation

分布类型	R 函数	所属程序包
指数分布(Exponential)	<code>_exp</code>	stats/ extRemes
伽马分布(Gamma)	<code>_gamma</code>	stats
对数正态分布(Lognormal)	<code>_lnorm</code>	stats
逻辑斯蒂分布(Logistic)	<code>_logis</code>	stats
正态分布(Normal)	<code>_norm</code>	stats
泊松分布(Poisson)	<code>_pois</code>	stats/ evdbayes/ extRemes
威布尔分布(Weibull)	<code>_weibull</code>	stats
帕累托分布(Pareto)	<code>_pareto</code>	actuar
布尔夫分布(Burr)	<code>_burr</code>	actuar
广义极值分布(Generalized Extreme Value)	<code>_gev</code>	ismev/ evir/ GEVcdn
广义帕累托分布(Generalized Pareto)	<code>_gpd</code>	evdbayes/ extRemes
耿贝尔分布(Gumbel)	<code>_gumbel</code>	ismev/ extRemes

注:各分布类型的密度函数、累计分布函数、分位函数和随机数函数,可通过在代表该分布的 R 函数前加上相应前缀 d、p、q、r 获得。

P-Ⅲ型曲线是水文频率计算中应用较广泛的频率曲线线型,也是《水利水电工程设计洪水计算规范》<sup>[6]</sup>中推荐使用的线型。参数估计方法包括矩法、极大似然法、概率权重法、权函数法、线型矩法等多种方法。本文选择 P-Ⅲ型曲线并采用常用的矩法进行参数估计。在 R 软件中实现水文频率计算的具体流程见图 1。

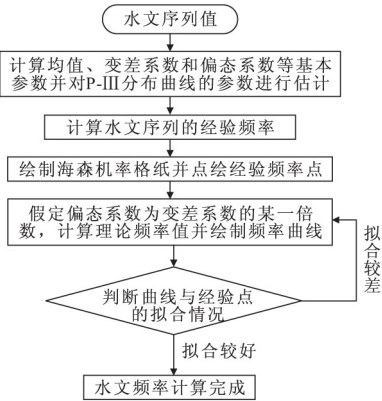


图 1 水文频率计算流程

Fig.1 The flow chart of hydrological frequency calculation

图 1 中水文频率计算的各个步骤在 R 软件的实现过程如下。

(1)参数估计。P-Ⅲ分布的参数包括形状、尺度和位置参数 $(\alpha, \beta, \alpha_0)$ ,可由样本均值 $(E_X)$ 、变差系数 $(C_V)$ 和偏态系数 $(C_S)$ 3 个参数相互推导出<sup>[7]</sup>。采用矩法进行参数估计,其计算公式为:

$$E_X = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

$$C_V = \sigma / E_X \tag{2}$$

$$C_S = \frac{1}{(n-3)\sigma^3} \sum_{i=1}^n (x_i - E_X)^3 = \frac{nS_c}{n-3} \tag{3}$$

式中, $x_i$  为序列中的某个值; $n$  为序列长度; $\sigma$  为序列标准差; $S_c$  为偏斜度。

在 R 软件中均值  $E_X$  可由 mean 函数计算得到,标准差  $\sigma$  可由 sd 函数计算得到,偏斜度  $S_c$  可由 skewness 函数计算得到。

两组参数间关系为:

$$\begin{cases} \alpha = 4 / C_S^2 \\ \beta = 2 / (E_X C_V C_S) \\ a_0 = E_X (1 - 2 C_V / C_S) \end{cases} \tag{4}$$

(2)经验频率计算。在对水文变量序列由大到小排序后,得到的新序列中每一项对应的经验频率  $p$  为:

$$p = m / (n + 1) \tag{5}$$

式中, $m$  为序数; $p$  为经验频率。

(3)理论频率计算。在计算出  $\alpha, \beta, \alpha_0$  值后,通过对 P-Ⅲ分布的密度函数进行积分,可计算出某一水文变量值  $x_p$  所对应的理论频率值  $P$ ,具体公式为:

$$P = P(x \geq x_p) = \int_{x_p}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} dx \tag{6}$$

式中, $\Gamma(\alpha)$  为  $\alpha$  的伽马函数。

在 R 语言中,伽马函数的分布函数 (pgamma) 的形式为:

$$p(0 \leq t \leq q) = \frac{1}{S^\alpha \Gamma(\alpha)} \int_0^q t^{\alpha-1} e^{-t/S} dt \tag{7}$$

式中, $q$  为分位数(即位置参数); $\alpha$  为形状参数; $S$  为尺度参数。

令  $P = 1 - p, \beta = 1/S, t = x - \alpha_0$ ,式(6)可化简为:

$$p(0 \leq t \leq x_p) = \frac{1}{S^\alpha \Gamma(\alpha)} \int_0^{x_p} t^{\alpha-1} e^{-t/S} dt \tag{8}$$

式(7)与式(8)的形式完全一致,因此在 R 软件中可直接调用 pgamma 函数与 qgamma 函数求理论频率值  $P$  与水文变量值  $x_p$ 。

R 软件中伽马函数的分布函数 (pgamma) 及其分位数函数 (qgamma) 的调用方式为 pgamma (q, shape, rate, scale = 1/rate); qgamma (p, shape, rate, scale = 1/rate)。R 软件中调用 pgamma、qgamma 函数的程序语句为:  $P = 1 - \text{pgamma}(x - \alpha_0, \alpha, S); x = \text{qgamma}(1 - p, \alpha, S) + \alpha_0$ 。需注意的是,由于 R 软件中伽马函数的尺度参数有两个且互为倒数关系 ( $\text{scale} = 1/\text{rate}$ ),故调用 pgamma、qgamma 函数语句中的尺度参数为  $S$ ,而非  $1/S$ 。

3 R 软件中的频率曲线适线过程

### 3.1 绘制概率格纸

海森概率格纸是水文频率计算中经常使用的一种特殊坐标系统,其纵坐标为正常的均匀数学坐标,横坐标(即频率值)来自于标准正态分布的分位数<sup>[1]</sup>。纵坐标的绘制过程较简单,利用 R 语言中的 seq 函数生成一个等差数列作为纵坐标的数值,再利用 abline 函数即可绘制出概率格纸的横向网格线。

在绘制横坐标的过程中,需调用 R 语言中的正态函数的分位数函数(qnorm),其调用方式为 qnorm(p,mean,sd)。首先将概率格纸的横坐标刻度(0.01,0.05,0.5,1,...,99,99.9,99.99)序列赋值给向量  $p$ ,然后利用 qnorm 函数计算出相应的分位数值,再调用 axis 函数将向量  $p$  中的概率值替换计算出来的分位数成为横坐标上的刻度,最后利用 abline 函数绘制出概率格纸的纵向网格线。

### 3.2 适线过程

用矩法进行参数估计时,由于  $C_s$  的抽样误差太大,故一般不计算  $C_s$ ,而是根据经验假定  $C_s$  为  $C_v$  的某一倍数。在绘制完海森概率格纸后,通过调用 pgamma、qgamma 函数计算出  $C_s$  与  $C_v$  在不同倍数时的理论频率值序列及其对应的水文变量序列值(即纵坐标值),通过调用 qnorm 函数将序列转化为正态函数的分位数(即横坐标值),再调用 lines 函数即可在概率格纸上绘制出理论频率曲线。通过观察频率曲线与经验频率点的吻合程度来选取最佳参数组合。

目估适线的结果主观性较强,难以获取最佳适线结果。为此本文引入了纵向离差平方和(即经验点和同频率的频率曲线纵坐标之差的平方和)来辅助判断适线情况,离差平方和的值越小则表明适线情况越好。将其与目估适线法相结合可获得更好的适线结果。

## 4 实例验证

### 4.1 验证过程

为验证程序的适用性,利用文献[8]中某站 1949~2000 年的年径流量数据完成一次水文频率计算,具体过程如下:①在 R 软件主窗口中输入水文变量序列,计算出序列长度、均值、变差系数( $C_v$ )等参数。②利用相关函数绘制出海森概率格纸,根据式(5)计算出经验频率值并在概率格纸上绘出经验点。③根据相应公式进行参数估计,再计算出经验频率和对应的水文变量值,利用相关函数绘制出频率曲线。④结合目估适线与纵

向离差平方和来判断经验点与频率曲线的拟合程度,调整  $C_v$  及  $C_s$  与  $C_v$  的比值以获得最好的适线结果(图 2)。⑤程序运行结束后在主窗口内会显示部分常用频率所对应的水文变量值,方便后续研究使用(图 3)。

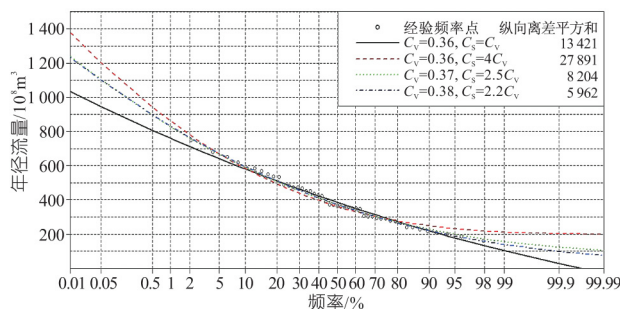


图 2 水文频率适线结果

Fig.2 The curve fitting result in hydrological frequency calculation

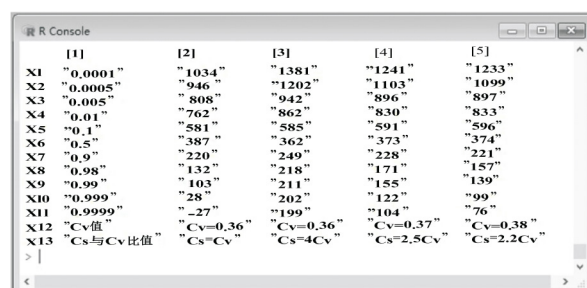


图 3 部分频率下的水文变量值

Fig.3 Part of the hydrological variable values in specific frequency

### 4.2 适线效果评价

文献[8]中实例原有的适线结果为  $C_v = 0.37, C_s = 2.5C_v$ ,纵向离差平方和为 8 204(图 2)。在利用编写的程序进行调整后,最终确定  $C_v = 0.38, C_s = 2.2C_v$ ,在保持较好适线结果的同时,将纵向离差总平方和的值降低到了 5 962,提高了适线的准确性。

利用 Visual Basic 等语言编制的水文频率计算程序由于无法直接调用各种相关函数,需编写大量程序语句,工作量较大,程序设计和控制较为复杂<sup>[2]</sup>。利用 Excel 软件进行水文频率分析的结果能满足工程要求,但实现过程特别是海森概率格纸的绘制依然较复杂<sup>[1,3]</sup>。Matlab 与 R 软件均具有丰富的统计函数和较好的适线准确度,但 Matlab 在循环计算时运行效率较低,执行速度较慢<sup>[4]</sup>。综上所述,R 软件在水文频率计算中具有较好的适应性,具有一定的推广价值。

## 5 结论

a. R 软件相对于其他方法具有简单易学、内嵌函数多、编程速度快、绘图方便、适线准确度高

等诸多优势,在水文统计研究和实践工作中具有良好的应用前景。

b. 下一步应继续探索 R 语言在水文不确定分析和水文非一致性分析中的应用。

#### 参考文献:

- [1] 吴明官,任中海,周庆欣,等. Excel 在水文频率计算中的应用[J]. 水文, 2001, 21(5): 45-47.
- [2] 邢广军. P-Ⅲ曲线拟合软件的研发与应用[J]. 水文, 2009, 29(2): 74-77.
- [3] 李永坤,张少焱,张书函,等. 基于 MATLAB GUI 的

- P-Ⅲ型频率曲线的实现[J]. 人民黄河, 2016, 38(5): 19-21, 26.
- [4] 王剑峰,李宏伟,宋松柏,等. 基于 MATLAB GUI 技术的水文频率计算[J]. 人民黄河, 2010, 32(10): 42-44.
- [5] 李素兰. 数据分析与 R 软件[M]. 北京: 科学出版社, 2013.
- [6] SL44-2006, 水利水电工程设计洪水计算规范[S]. 北京: 中国水利水电出版社, 2006.
- [7] 程银才,范世香,李明华. 一种新的水文频率计算方法[J]. 水文, 2008, 28(1): 59-60, 69.
- [8] 詹道江,徐向阳,陈元芳. 工程水文学(第四版)[M]. 北京: 中国水利水电出版社, 2010.

## Hydrological Frequency Calculation and Curve Fitting Based on R Software

XIE Zi-bo<sup>1,2</sup>, LU Fan<sup>2</sup>, ZHU Kui<sup>1</sup>, SONG Xin-yi<sup>1,2</sup>

(1. School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China;  
2. Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100048, China)

**Abstract:** Aiming at the weakness of existing methods, the hydrological frequency calculation and curve fitting were realized based on R software programming. The main contents included: drawing the Heyson probability paper; parameter estimation of P-III distribution; calculation of empirical and theoretic frequencies of hydrological series; drawing the theoretic frequency curve. The empirical fitting method and ordinary least square were used to determine the optimal parameters. The test results prove that the statistical functions in R software are suitably used in hydrological frequency calculation, and it can decrease the programming time and greatly improve work efficiency compared to other software. Different parameters in P-III distribution could be matched efficiently and calculated out the ordinary least square when the program was running, which effectively reduces the computational load and greatly improve efficiency. The research results can provide reference for hydrological frequency calculation.

**Key words:** R software; hydrological frequency calculation; empirical fitting method; ordinary least square

\*\*\*\*\*  
(上接第 4 页)

## Long-term Precipitation Series in the Middle and Lower Reaches of Yangtze River and Its Relationship with PDO

WAN Zhi-wei<sup>1</sup>, JIA Yu-lian<sup>1</sup>, JIANG Mei-xin<sup>1</sup>, HONG Yi-jun<sup>2</sup>

(1. Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China; 2. Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** To analyze the relationship between long-term precipitation series in the middle and lower reaches of the Yangtze River and PDO, based on the precipitation data of Jiujiang Customs Weather Station from 1885 to 1937 and modern weather record, this study reconstructs the long-term precipitation series of Jiujiang in the middle and lower reaches of the Yangtze river. The regression model of Nanchang, Ji'an and Jiujiang modern precipitation data was used to reconstruct the 1938-1950 precipitation data. The results show that the annual average precipitation in Jiujiang is 1 435.4 mm, and the smallest precipitation is 764.8 mm in 1898; the largest precipitation is 2 165.7 mm in 1954. The results of the linear propensity estimation show that there is no significant increase trend of precipitation in Jiujiang in recent 132 years and the rate is 32.5 mm/100a. The results of the cumulative anomaly show that the change in precipitation is presented as a cycle of 30 years, which is being increase at first stage and then decrease in the second stage. Precipitation cycle reached its peak in 1885, 1920, 1955, 1975 and 2000. The process is currently at the top of the sixth peak, the next period of precipitation will show an upward trend. The results of power spectrum analysis and multi-taper method show that the precipitation series of Jiujiang has 5 a, 2-4 a in high frequency oscillation, and 35 a, 21-22 a and 11-16 a cycle in low frequency oscillation. Wavelet transform shows that there is a complicated relationship between precipitation and PDO at inter-decadal scale, precipitation changes during 1885-1900 were behind the PDO of about 1/4 cycle; during 1900-1980, precipitation was negatively correlated with PDO; since 1980, PDO was behind the precipitation of about 1/4 cycle. Cross wavelet analysis shows that the spectrum value of precipitation and PDO was higher in the scale of 8-16 years, and passed the 95% confidence test. This study can provide basis for the precipitation change research in the middle and lower reaches of the Yangtze river, as well as the flood control and drought relief, water conservancy construction and efficient use of water resources.

**Key words:** reconstruction of precipitation series; PDO; Jiujiang; the middle and lower reaches of the Yangtze river