

何妹,赵华荣,姚越,等.基于 R 软件优化绘制皮尔逊Ⅲ型频率曲线[J].人民珠江,2021,42(2):7-11.

# 基于 R 软件优化绘制皮尔逊Ⅲ型频率曲线

何 妹,赵华荣,姚 越,金 鑫

(桂林理工大学环境科学与工程学院,广西 桂林 541006)

**摘 要:**皮尔逊Ⅲ型(P-Ⅲ型)频率曲线是水文分析过程常用的频率分析曲线。以某水文测站 47 年的年径流量实测资料,根据适线法和纵向离差平方和最小准则确定最优适线参数和频率分布曲线。利用 R 和 Notepad++ 软件实现了水文频率计算、拟合和制图过程,并且在拟合过程中增加自动查找最优适线参数;最终输出经验频率、统计参数初值、最小纵向离差平方和及最优频率曲线。应用结果表明:运用 R 软件进行水文频率分析,可以快速、高效地计算出不同分布参数和频率曲线的离差平方和,提高了拟合精度和工作效率,降低了水文频率计算的工作量,为水文频率曲线的优化绘制提供了一种新的实现方法。

**关键词:**R 软件;皮尔逊Ⅲ型曲线;水文频率计算;最小纵向离差平方和

**中图分类号:**TV11;P333   **文献标志码:**A   **文章编号:**1001-9235(2021)02-0007-05

## Optimization of Pearson Ⅲ Frequency Curve Based on R Software

HE Mei, ZHAO Huarong, YAO Yue, JIN Xin

(College of Environmental Science and Engineering, Guilin University of Technology, Guilin 541006, China)

**Abstract:** The Pearson Ⅲ frequency curve is a common frequency curve in hydrological analysis process. Based on the 47-year annual runoff data of a hydrological station, this paper defines the optimal fitting parameters and frequency distribution curve according to the curve fitting method and the minimum sum of squares of longitudinal deviation, carries out the hydrological frequency calculation, fitting and mapping through the R and Notepad++ software, and adds the automatic finding program of the optimal fitting parameters in the fitting process to output the empirical frequency, the initial value of statistical parameters, the minimum sum of squares of longitudinal deviation and the optimal frequency curve. The application results show that using R software for hydrological frequency analysis can quickly and efficiently calculate the sum of squares of deviation of different distribution parameters and frequency curves, improve the fitting accuracy and work efficiency, and reduce the workload of hydrological frequency calculation, so it is a new method for the optimal drawing of hydrological frequency curve.

**Keywords:** R software; Pearson Ⅲ curve; hydrological frequency calculation; minimum sum of squares of longitudinal deviation

P-Ⅲ型频率曲线适用于中国大部分地区的降雨径流序列,成为水文计算中常用的分布频率曲线。计算机技术在计算、绘图等方面都具有较强大的功能,在水文与水资源方面的应用也越来越广泛,由于

P-Ⅲ型频率曲线计算过程和图形绘制较复杂,可以充分发挥计算机的优势。对水文频率曲线的绘制和计算主要有:车国文<sup>[1]</sup>利用 Excel 绘制皮尔逊型曲线,其结果满足工程要求;李清富等<sup>[2]</sup>利用 MAT-

基金项目:国家自然科学基金项目(51369009);广西自然科学基金项目(2016GXNSFAA380116)

收稿日期:2020-09-05

作者简介:何妹,女,研究方向为水资源与水环境。E-mail:gluthm@163.com

通讯作者:赵华荣,男,研究方向为水资源与水环境。E-mail:gutzhr@163.com

LAB 编写 M 文件实现 P - III 型曲线的绘制;赵培颖等<sup>[3]</sup>研究了 Visual\_Basic 绘制 P - III 频率曲线的应用;许义和等<sup>[4]</sup>基于 Matlab 对 P - III 型曲线绘制软件的研发与应用;俞超锋等<sup>[5]</sup>基于 MATLAB 对 P - III 型频率曲线参数估算的优化;李传博等<sup>[6]</sup>实现了基于 C 语言的 P - III 型曲线绘制方法;谢子波等<sup>[7]</sup>利用基于 R 软件实现水文频率计算适线绘制。

上述 P - III 型曲线计算和绘制过程中,大多采用商业软件,部分软件也不是针对统计计算设计的,操作过程比较复杂。如 EXCEL、Visual Basic、MATLAB 和 C 语言,在实现参数计算和图形绘制时,需要较强的计算及编程能力。文献[7]采用 R 软件绘制 P - III 型曲线过程中对最优参数的确定需要人工输入查找。为了弥补上述问题,本文采用 R 软件在文献[7]的基础上,增加了自动查找最优适线参数的功能。同时根据 P - III 型频率曲线在实际工程应用中可能会强调较大值或较小值拟合精度的情况,如对特大洪水频率计算需要考虑较大值拟合精度,用水保证率则需考虑较小值拟合精度。因此,将水文序列划分为较大值和较小值,增加了较大值和较小值拟合最优参数的查找并绘制对应水文频率曲线,可根据工程实际情况选择使用。

## 1 R 软件对水文频率计算的实现

### 1.1 R 软件运用原理及优势

R 软件是由奥克兰大学的 Ross Ihaka 和 Rontleman 两位统计学家基于 S 语言开发的一个面向对象的脚本语言,该语言免费开源,并以两者首字母“R”命名。它是一种比较适合统计计算的语言,可进行数据输入、输出、分支和循环,具备用户可自定义函数等功能<sup>[8]</sup>。

R 语言可以在命令窗口进行统计计算、预测分析和绘制精美图形,实现数据挖掘和可视化,完成较复杂的数据分析等工作。此外,R 语言有数千个的程序包,用户能通过选择相应的程序包快速将 R 应用到相关领域<sup>[9]</sup>。相比其他软件绘制 P - III 型水文频率曲线,R 语言更具有优势。R 语言与其他编程语言的对比见表 1。

表 1 P - III 型曲线绘制常用软件比较

项目	Excel	Visual Basic	Matlab	R
是否开源	否	否	部分开源	完全开源
是否商业化	是	是	是	否
软件使用平台	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、方差分析、判别分析与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、箱型图、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、中、低水平绘图函数的搭配,容易绘制复杂图形

P - III 型水文频率分析计算过程包括参数估计、理论频率、经验频率的计算及图形的绘制。以上内容在 R 软件中较容易实现。本文主要使用 openxlsx 和 e1071 程序包,其中 openxlsx 包是用于读取 excel 表格中水文序列值;e1071 程序包有 P - III 型曲线计算过程中所需要的\_gamma 函数。在水文频率计算分析中常用函数及对应 R 程序包见表 2。

表 2 水文频率计算分析常用函数类型

函数类型	R 函数	所属程序包
伽马分布(Gamma)	_gamma	states/e1071
正态分布(Normal)	_norm	states
指数分布(Exponential)	_exp	states/extRemes
泊松分布(Possion)	_posi	states/ extRemes/evdlyxes
逻辑斯蒂分布(Logistic)	_logis	states
β 分布(Beta)	_beta	base
广义极值函数(Generalized Value)	_gev	Ismev/evir/GEVcdn
威布尔分布(Weibull)	_weibull	States

由于 R 软件自身的图形编程界面不友好,一般都采用第三方图形界面。比较常用的编辑器有 Rstudio 和 Notepad + +。本次采用带 NPpToR 插件的 Notepad + + 编辑器编写代码,将编写好的程序代码通过 F8 功能键推送到 R 软件进行调试或运行。

P - III 型曲线在中国水文频率计算分析中应用

较广泛,该线性的不确定性影响是较低的<sup>[10]</sup>。P-Ⅲ型水文频率计算中常用的参数估计方法有矩法、适线法、图解法、权函数法、概率权重法和极大似然法等<sup>[11]</sup>。本文采用适线法推求水文参数,适线法可分为目估适线法和优化适线法,目估适线受到人为因素的影响,较难找到最佳拟合曲线。优化适线法是通过计算纵向离差平方和查找到最佳拟合曲线。水文频率分析计算过程见图 1。

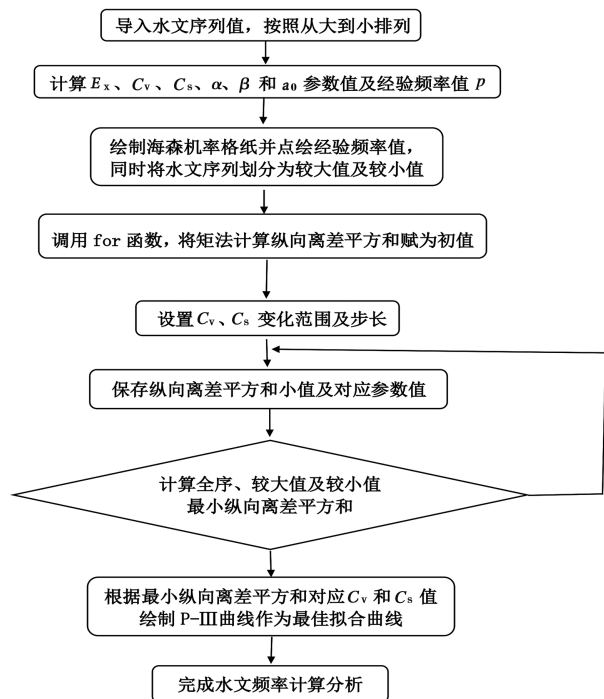


图 1 水文 P-Ⅲ曲线分析计算流程

## 1.2 水文频率计算原理

P-Ⅲ型曲线的概率密度函数为:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (1)$$

其中  $\Gamma(\alpha)$  为  $\alpha$  的伽马函数;  $\alpha, \beta, a_0$  分别代表 P-Ⅲ型的形状、尺度和位置参数, 并且  $\alpha, \beta$  均大于零<sup>[12]</sup>。 $\alpha, \beta, a_0$  通过均值 ( $E_x$ )、变差系数 ( $C_v$ ) 和偏态系数 ( $C_s$ ) 3 个统计参数关系可求出, 关系如下:

$$\alpha = \frac{4}{C_s^2} \quad (2)$$

$$\beta = \frac{2}{E_x C_v C_s} \quad (3)$$

$$a_0 = E_x \left(1 - \frac{2C_v}{C_s}\right) \quad (4)$$

其中根据矩法计算参数估计:

$$E_x = \frac{1}{n} \sum_{i=1}^n x_i \quad (5)$$

$$C_v = \frac{\sigma}{E_x} \quad (6)$$

$$C_s = \frac{nS_c}{n-3} \quad (7)$$

使用 R 软件 mean 函数计算均值  $E_x$ , sd 函数计算标准差  $\sigma$ , skewness 函数计算偏斜度  $S_c$ , 计算出以上参数后, 该概率密度函数也随之确定。由式 (8) 计算出对应的一组经验频率值:

$$p = \frac{m}{n+1} \quad (8)$$

式中  $m$ ——水文序列从大到小排列的序号;  $n$ ——实测序列年数;  $p$ ——经验频率。

在 P-Ⅲ型频率计算中, 需要求出某一个数值  $x$  所对应的概率  $P$ , 故将概率密度函数进行积分得出大于等于  $x_p$  频率  $P$ , 见式 (9):

$$P = P(x \geq x_p) = \int_{x_p}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} dx \quad (9)$$

通过以上参数可以变换得出式 (10):

$$x_p = E_x (1 + C_v \varphi) \quad (10)$$

$$\varphi = \frac{x - E_x}{E_x C_v} \quad (11)$$

根据上述参数之间的关系可求出离均势数  $\varphi$  和水文变量值  $x_p$ , 从而可以计算出理论频率  $P$ 。

## 2 R 软件在 P-Ⅲ型水文频率计算的应用

### 2.1 绘制海森概率格纸

绘制水文频率 P-Ⅲ型曲线之前, 先绘制机率格纸, 即海森概率格纸。海森概率格纸的横坐标数值是非均匀分布的数字坐标, 而纵坐标数值则是均匀排列的数字格式, 其中横坐标与频率值 (下侧概率) 的标准正态分布分位数  $P = 50\%$  有关,  $p = 0.01\%$  时横坐标值为 0。

海森概率格纸绘制方法如下: 首先, 将 (0.01, 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99.99, 99.99) 赋值到海森概率格纸的横坐标, 利用 qnorm 函数求出对应的分位数值, 并调用 axis 函数将所求的分位数替换横坐标刻度单位值; 调用 qgamma 函

数计算对应水文频率值作为纵坐标值,根据经验频率值  $P$  用 `qnorm` 函数计算对应  $x$  轴数值;最后调用 `abline` 函数完成海森概率格纸的绘制。

R 软件绘制机率格纸的相关函数<sup>[7]</sup>包括伽马函数分布函数(`pgamma`)、分位数函数(`qgamma`)、正态分布函数(`qnorm`)。函数调用方式为 `pgamma(q, hape, rate, scale = 1/rate)`; `qgamma(p, shape, rate, scale = 1/rate)`; `qnorm(p, mean, sd)`。R 软件中调用 `pgamma`、`qgamma` 函数的程序语句为:  $P = 1 - \text{pgamma}(x - a_0, a, \beta)$ ;  $x = \text{qgamma}(1 - P, a, \beta) + a_0$ 。

## 2.2 水文频率计算适线过程

由于目估适线法受到人为因素的影响。本文采用纵向离差平方和作为判断适线优劣的标准,通过计算机自动“捕捉”最佳适线参数,使适线更加准确。依照设计洪水频率曲线适线离差平方和最小准则,当纵向离差平方和最小,即  $\sum (x_i - x_p)^2 = \Delta_{\min}$  适线结果最佳,反之,较差。

采用矩法计算的参数估计值时,  $C_s$  偏差较大,通常将矩法所算出的参数值作为适线调整初值。适线过程不计算具体的  $C_s$  值,而是计算  $C_s$  为  $C_v$  某一倍数变化。在完成绘制海森概率格纸后,用 `qgamma` 函数和 `pgamma` 函数计算全序在  $C_s$  与  $C_v$  不同倍数下的理论频率和对应的水文变量值(即纵坐标)。通过 `qnorm` 将序列转变为正态函数的分位数(即横坐标)。最后,分别计算全序、较大值及较小值在  $C_v$  和  $C_s$  不同倍数下的纵向离差平方和,求出最小纵向离差平方和及对应参数作为最佳参数组合,再调用 `lines` 函数在概率格纸上绘制频率曲线。

## 3 实例计算

### 3.1 验证过程

本次实例采用文献<sup>[13]</sup>中某站 1949—2000 年的实测水文序列验证,具体步骤如下:①利用 R 软件读取 Excel 表格中的年径流量序列,使用相关函数计算序列长度( $n$ )、均值( $E_x$ )、变差系数( $C_v$ )和偏态系数( $C_s$ )等参数;②绘制海森概率格纸,根据式(8)计算经验频率值并点绘在概率格纸上;③将年径流量分成较大值序列与较小值序列;④调用 `for` 函数计算 P-Ⅲ型曲线参数,将矩法估算  $C_v$  和  $C_s$  值

作为初值,将  $C_v$  设定范围从  $C_v - 0.1$  到  $C_v + 0.1$ ;  $C_s$  设定范围在  $C_v \sim 4C_v$  之间,两者均按照 0.001 的步长增加;⑤确定最佳适线参数。根据经验点与频率曲线曲线的纵向离差平方和,查找全序、较大值和较小值纵向离差平方和的最小值,保存最小纵向离差平方和对应参数,退出循环。⑥图形绘制。根据最小纵向离差平方和对应参数调用 `lines` 函数绘制 P-Ⅲ曲线(图 2);⑦在 R 窗口显示部分水文频率所对应的水文变量值(表 3)。

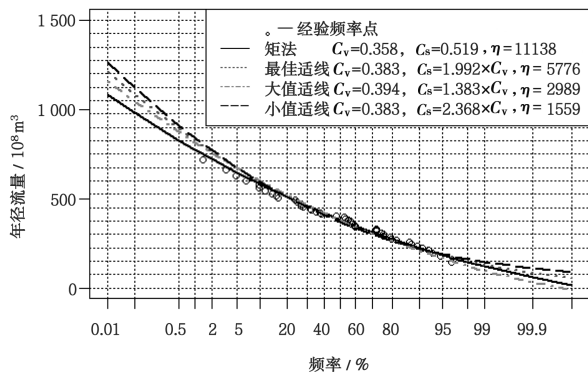


图 2 水文 P-Ⅲ曲线适线结果

表 3 部分水文频率下的水文变量

频率/%	矩法	最佳适线	较大值适线	较小值适线
0.01	1 083.69	1 215.25	1 161.85	1 264.88
1.00	777.23	829.16	818.40	843.65
20.00	509.44	513.95	520.58	500.61
50.00	383.12	376.25	381.24	372.74
80.00	274.10	265.67	261.88	266.03
98.00	145.44	149.00	122.55	161.47
99.00	120.81	128.93	96.14	144.80
99.90	59.59	83.11	31.02	109.09

注:除频率之外的其余数据单位均为  $10^8 \text{ m}^3$ 。

### 3.2 验证结果

本实例中经验频率、矩法和适线法结果见图 2,其中使用矩法计算的  $C_v$  和  $C_s$  值分别为 0.358 和 0.519,其纵向离差平方和为 11 138,偏差较大,精确度不高,一般不建议用该法拟合水文频率曲线。适线法对全序、大值及小值进行拟合,结果为:较大值最佳适线参数为  $C_v = 0.394$ 、 $C_s = 1.383 \times C_v$ ,其最小纵向离差平方和为 2 989;较小值最佳适线参数为  $C_v = 0.383$ 、 $C_s = 2.368 \times C_v$ ,最小纵向离差平方和为

1 559。全序列最佳适线参数为  $C_v = 0.383$ 、 $C_s = 1.992 \times C_v$ ,其最小纵向离差平方和为 5 776,全序列适线比文献[7]的纵向离差平方和少 186 表明拟合结果更佳。

为了检验 R 优化成果,对部分文献的实例数据用 R 软件再次进行参数拟合,并与原文献结果进行比较。由于原文献中没有给出纵向离差平方和,因此采用原参数通过本文的方法进行计算,结果见表 4。采用本文适线方法所得的纵向离差平方和比原参数计算结果都要小一个数量级,说明本文所采用方法可以大大提高拟合精度。

表 4 不同软件适线结果与本文适线结果比较

原文献适线结果				本文适线结果			文献来源
$C_v$	$C_s$	纵向离差平方和	采用软件	$C_v$	$C_s$	纵向离差平方和	
0.360 0	0.900 0	1 611 579	C 语言	0.468	1.287	817 874	[6]
0.251 8	0.667 5	12 741	Fortran	0.221	0.391	5 865	[14]
0.800 0	2.800 0	3 570 205	MATLAB	0.903	2.662	343 546	[15]

## 4 结论

a) R 软件结合 Notepad + + 具有操作简单,适用于小规模软件开发等功能。其软件包统计函数多,编程及绘图方便,计算精度较高,实用性强,易于操作,便于实际应用推广。

b) 通过实例验证结果可知 R 软件不仅可以精准地计算出水文序列的特征参数,绘制海森概率格纸,而且能快速获取最小纵向离差平方和及对应参数,并根据最佳适线参数组合 ( $C_v = 0.383$ 、 $C_s = 1.992 \times C_v$ )绘制 P - III 曲线的最优曲线,该法相比同软件绘制所得的纵向离差平方和少了 186,达到优化效果。

c) 通过不同软件适线结果比较可知 R 软件拟合精度最大提高一倍,若应用于工程实际,可以为工程计算提高精度和节约大量时间。

## 参考文献:

[1] 车国文. 皮尔逊Ⅲ型曲线的 Excel 绘制法[J]. 辽宁师专学报(自然科学版), 2018,20(4):9-11.

[2] 李清富, 闫鹏飞, 孙静涛, 等. 水文频率曲线绘制的 Matlab 程序设计[J]. 河南科学, 2013,31(8):1250-1254.

[3] 赵培颖, 金冶, 张忠孝. Visual Basic 在绘制 P - III 频率曲线中的应用[J]. 水利规划与设计, 2008,33(2):55-57.

[4] 许义和, 魏晓妹. 基于 Matlab 的 P - III 型曲线绘制软件的研发与应用[J]. 水电能源科学, 2010,28(7):15-17.

[5] 俞超锋, 刘新成. 基于 MATLAB 的 P - III 型频率曲线参数估算的优化[J]. 人民黄河, 2015,37(8):24-26.

[6] 李传博, 郑亮. 基于开源的 C 语言科学计算库(GNU Scientific Library)的水文频率计算软件开发[J]. 信息通信, 2016,34(9):53-54.

[7] 谢子波, 鲁帆, 朱奎, 等. 基于 R 软件的水文频率计算适线绘制[J]. 水电能源科学, 2018,36(1):20-23.

[8] 李素云. 数据分析与 R 软件[M]. 2 版. 北京: 科学出版社, 2017.

[9] 汤银才. R 语言与统计分析[M]. 北京: 高等教育出版社, 2008.

[10] 张立杰, 蒙潇然. 线型选择对西江流域水文频率计算不确定性的影响[J]. 人民珠江, 2017,38(2):37-41.

[11] 姜尧志. 水文频率参数估计的生物地理学优化算法[J]. 人民珠江, 2015,36(6):55-57.

[12] 王德智, 张桂菊, 吴畏, 等. 皮尔逊Ⅲ型频率曲线图解适线软件开发[J]. 水利科技与经济, 2009,15(12):1111-1113.

[13] 詹道江, 徐向阳, 陈元芳. 工程水文学[M]. 4 版. 北京: 中国水利水电出版社,2010.

[14] 赵晓慎, 王卫东. 皮尔逊 - III 型分布曲线最佳配线及自动化绘图[J]. 人民黄河, 2007,29(4):25-26.

[15] 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.

(责任编辑:程 茜)