

水文辗转相关插补延长研究

华家鹏,林 芸

(河海大学水资源环境学院,江苏 南京 210098)

摘要 利用 500 多种不同参数模拟计算结果对水文辗转相关插补延长有关问题进行了研究,结果表明:在资料插补延长中,如果参证站 X 资料较长而设计站 Z 与参证站 X 同期观测资料较短,那么就无法采用设计站 Z 与参证站 X 同期观测资料建立回归方程,这时可选用一个中间参证站 Y 进行辗转相关插补延长,但该站必须具备条件(a)参证站 Y 与设计站 Z 、参证站 Y 与参证站 X 均具有较长的同期观测资料,(b)参证站 Y 分别与设计站 Z 和参证站 X 均具有较强的相关性,即相关系数 $R_{XY} > 0.9, R_{YZ} > 0.9$,才能满足插补精度要求.同时还利用统计试验资料建立了辗转相关插补延长各相关系数之间的关系表达式.

关键词 水文计算 辗转相关 回归方程 相关系数

中图分类号 P333 **文献标识码** A **文章编号** 1000-1980(2003)05-0494-03

水文分析计算中常因设计站 Z 本身资料短缺而首先采用参证站 X 资料来插补延长设计站 Z 的资料,然后利用插补延长后的设计站 Z 的资料来推求水文变量的设计值 $Z_p^{[1\sim3]}$. 但在水文资料的插补延长中,又常常因参证站 X 与设计站 Z 同期观测资料较少,无法采用同期观测资料建立回归方程,不得不借用另一中间参证站 Y ,由参证站 X 的资料通过参证站 X 与参证站 Y 的同期观测资料建立的回归方程插补延长参证站 Y 的资料,再由插补延长后的参证站 Y 的资料通过参证站 Y 与设计站 Z 的同期观测资料建立回归方程来插补设计站 Z 的资料.这就是水文学上常说的辗转相关.水文界对采用辗转相关来展延水文系列基本上持否定态度^[1]. 当然,若参证站 X 与设计站 Z 资料本身相关关系较差,那么就不可能找到一个既与设计站 Z 有较好相关关系,又与参证站 X 有较好相关关系的中间参证站 Y ,这种情况下要通过辗转相关来达到插补延长的目的,显然是不可能的.但是,若设计站 Z 、参证站 X 和参证站 Y 之间的相关关系非常密切,因设计站 Z 与参证站 X 同期观测资料较短而无法通过建立回归方程来插补设计站 Z 资料,是否可在一定条件下通过辗转相关来达到插补延长水文系列的目的,这就是本文要研究的问题.由于三者之间的相关系数很难从理论上导出其解析式,因此本文采用统计试验方法模拟水文系列,并分析得出了三者之间相关系数的表达式.

1 偏态分布随机数的模拟方法

一般认为水文变量服从 P—Ⅲ 型分布,但模拟三维 P—Ⅲ 型分布水文系列较困难,而三参数对数正态分布同属偏态分布,参数个数相同,两分布比较相近,因此,本文采用三参数对数正态分布进行 3 个变量的模拟^[2]. 二维对数正态分布,其密度函数为

$$f(x,y) = \frac{1}{\sigma_1\sigma_22\pi\sqrt{1-r^2}(x-a_1)(y-a_2)} \exp\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{\ln(x-a_1)+b_1}{\sigma_1}\right)^2 - 2r\frac{\ln(x-a_1)+b_1}{\sigma_1}\frac{\ln(y-a_2)+b_2}{\sigma_2} + \left(\frac{\ln(y-a_2)+b_2}{\sigma_2}\right)^2\right]\right\} \\ (x > a_1, y > a_2, |r| < 1) \tag{1}$$

其中 $a_1, a_2, b_1, b_2, \sigma_1, \sigma_2, r$ 可分别由 2 组实测样本系列估计得出.

由式(1)通过积分,易知条件分布为

$$f(x/y) = \frac{1}{\sigma_2' \sqrt{2\pi}(y - a_2)} \exp\left\{-\frac{1}{2}\left[\frac{\ln(y - a_2) + b_2'}{\sigma_2'}\right]^2\right\} \tag{2}$$

其中

$$\sigma_2' = \sigma_2 \sqrt{1 - r^2} \tag{3}$$

$$b_2' = b_2 - r \frac{\sigma_2}{\sigma_1} [\ln(x - a_1) + b_1] \tag{4}$$

$$x = \exp(\sigma_1 \xi - b_1) + a_1 \tag{5}$$

$$y = \exp(\sigma_2 \eta - b_2) + a_2 \tag{6}$$

模拟步骤如下: (a)假定总体参数 $EX, EY, EZ, C_{VX}, C_{VY}, C_{VZ}, C_{SX}, C_{SY}, C_{SZ}, R_{XY}, R_{YZ}$. (b)由给定的设计站 Z 、参证站 X 和参证站 Y 总体参数,根据式(7)~(11)估计各自的 $a_1, a_2, b_1, b_2, \sigma_1, \sigma_2, r$.

$$\eta_i = \left(\frac{C_{SX_i} + \sqrt{C_{SX_i}^2 + 4}}{2} \right)^{1/3} - \left(\frac{-C_{SX_i} + \sqrt{C_{SX_i}^2 + 4}}{2} \right)^{1/3} \tag{7}$$

$$a_i = EX_i \left(1 - \frac{C_{VX_i}}{\eta_i} \right) \tag{8}$$

$$\sigma_j = \sqrt{\ln(1 + \eta_i^2)} \tag{9}$$

$$b_i = \frac{1}{2} \ln \left[\frac{1 + \eta_i^2}{(EX_i - a_i)^2} \right] \tag{10}$$

$$r_{i,j} = \frac{1}{\sigma_i \sigma_j} \ln [\sqrt{(e^{\sigma_i^2} - 1)(e^{\sigma_j^2} - 1)} \rho_{i,j} + 1] \tag{11}$$

(c)生成 4 个 $[0,1]$ 均匀分布的随机数 u_1, u_2, u_3, u_4 . (d)利用式(12)生成标准正态分布的随机数 n_1, n_2, n_3 .

$$n_1 = \sqrt{-2\ln u_1} \cos 2\pi u_2 \quad n_2 = \sqrt{-2\ln u_1} \sin 2\pi u_2 \quad n_3 = \sqrt{-2\ln u_3} \cos 2\pi u_4 \tag{12}$$

(e)以 n_1 代替式(5)中 ξ 可得到 x_1 . (f)以 x_1 代入式(4)得到 b_2' ,代入式(3)得到 σ_2' . (g)以 n_2, b_2', σ_2' 分别代替式(6)中的 η, b_2, σ_2 ,则得到在 $X = x_1$ 条件下的 y_1 . (h)同理可得到 $Y = y_1$ 条件下的 z_1 .

2 模拟结果分析

根据上述模拟方法,选用 500 种不同的总体参数,对每种总体参数模拟 100 组长度均为 30 的参证站 X 、参证站 Y 和设计站 Z 的样本,统计 X 和 Z 的相关系数的均值和均方差.模拟结果见表 1.

由 500 多种不同参数模拟计算结果可知: (a)设计站 Z 与参证站 X 的相关系数 R_{XZ} 仅与参证站 X 与参证站 Y 的相关系数 R_{XY} 和设计站 Z 与参证站 Y 的相关系数 R_{YZ} 有关,而与各自的参数无关; (b)设计站 Z 与参证站 X 的相关系数 R_{XZ} ,始终小于参证站 X 与参证站 Y 的相关系数 R_{XY} 和设计站 Z 与参证站 Y 的相关系数 R_{YZ} ; (c)设计站 Z 与参证站 X 的相关系数的均方差 $\sigma_{R_{XZ}}$ 随着设计站 Z 与参证站 X 的相关系数的增加而减小; (d)如果参证站 X 资料较长而设计站 Z 与参证站 X 同期观测资料较短,那么就无法采用设计站 Z 与参证站 X 同期观测资料建立回归方程,这时可选用一个中间参证站 Y 进行插补延长,但该站必须具备条件 (a)参证站 Y 与设计站 Z 、参证站 Y 与参证站 X 均具有较长的同期观测资料, (b)参证站 Y 与设计站 Z 、参证站 Y 与参证站 X 均具有较强的相关性,即相关系数 $R_{YZ} > 0.9, R_{XY} > 0.9$,才能满足插补精度要求.也就是说,只有满足以上条件,才能保证 X 和 Z 的相关系数 $R_{XZ} > 0.8$.

由 500 种不同总体参数模拟的结果可归纳出

$$R_{XZ} = 0.873 R_{小} + 0.460 R_{大} - 0.408 \tag{13}$$

复相关系数

$$R = 0.994$$

式中: $R_{小}$ —— R_{XY} 和 R_{YZ} 中的小值; $R_{大}$ —— R_{XY} 和 R_{YZ} 中的大值.

由式(13)可看出,设计站 Z 与参证站 X 的相关系数 R_{XZ} 主要取决于参证站 X 与参证站 Y 的相关系数 R_{XY} 和设计站 Z 与参证站 Y 的相关系数 R_{YZ} 2 个相关系数中的小值.

表 1 辗转相关统计试验结果

Table 1 Statistical results of multivariate correlation

序号	R_{XY}	R_{YZ}	C_{VX}	C_{VY}	C_{VZ}	C_{SX}	C_{SY}	C_{SZ}	ER_{XZ}	$\sigma_{R_{XZ}}$
1	0.95	0.95	1.0	1.0	1.0	1.5	1.5	1.5	0.897 4	0.038 6
2	0.95	0.95	1.0	1.0	1.0	1.5	1.5	1.5	0.906 2	0.038 0
3	0.90	0.90	0.5	0.5	0.5	0.5	0.5	0.5	0.801 2	0.074 2
4	0.90	0.90	0.5	0.5	0.5	0.5	1.0	0.5	0.807 8	0.077 4
5	0.90	0.90	0.5	0.5	0.5	0.5	0.5	1.0	0.816 7	0.069 3
6	0.90	0.90	1.0	0.5	0.5	1.5	1.5	1.5	0.812 5	0.065 0
7	0.80	0.80	0.5	0.5	0.5	0.5	1.0	0.5	0.614 2	0.142 7
8	0.80	0.80	1.0	0.5	0.5	1.5	1.5	1.5	0.612 1	0.128 6
9	0.80	0.80	0.5	0.5	0.5	0.5	0.5	0.5	0.625 1	0.122 6
10	0.80	0.80	0.5	0.5	0.5	0.5	0.5	1.0	0.634 7	0.114 9
11	0.60	0.60	0.5	0.5	0.5	0.5	0.5	0.5	0.393 0	0.164 6
12	0.60	0.60	0.5	0.5	0.5	0.5	1.0	0.5	0.349 5	0.183 6
13	0.60	0.60	0.5	0.5	0.5	0.5	0.5	1.0	0.350 2	0.158 0
14	0.60	0.60	1.0	1.0	1.0	1.5	1.5	1.5	0.390 2	0.163 7
15	0.60	0.60	1.0	0.5	0.5	1.5	1.5	1.5	0.355 8	0.180 6
16	0.40	0.40	0.5	0.5	0.5	0.5	0.5	1.0	0.131 0	0.187 4
17	0.40	0.40	0.5	0.5	0.5	0.5	1.0	0.5	0.161 9	0.188 4
18	0.40	0.40	0.5	0.5	0.5	0.5	0.5	0.5	0.154 3	0.165 9
19	0.40	0.40	1.0	0.5	0.5	1.5	1.5	1.5	0.132 7	0.174 1
20	0.40	0.40	0.5	0.5	0.5	0.5	0.5	0.5	0.154 3	0.165 9
21	0.40	0.80	1.0	1.0	1.0	1.5	1.5	1.5	0.317 5	0.167 9
22	0.80	0.40	1.0	1.0	1.0	1.5	1.5	1.5	0.320 2	0.199 2
23	0.30	0.90	1.0	1.0	1.0	1.5	1.5	1.5	0.270 3	0.176 5
24	0.90	0.30	1.0	1.0	1.0	1.5	1.5	1.5	0.270 4	0.179 0
25	0.95	0.25	1.0	1.0	1.0	1.5	1.5	1.5	0.232 9	0.183 1

注： ER_{XZ} 为 X 和 Z 相关系数的均值。

3 结 论

一般情况下应尽量避免采用辗转相关插补水文资料. 如要采用辗转相关插补水文资料, 则必须满足本文所述条件.

参考文献：

[1] 刘光文. 水文分析与计算[M]. 北京 水利电力出版社, 1989. 16.
[2] 华东水利学院. 水文学的概率统计基础[M]. 北京 水利出版社, 1981. 295—299.
[3] 华家鹏. 偏态分布水文变量间的曲线回归问题[J]. 水文, 1988, 8(2) : 14—19.

Study on multivariate correlation-based interpolation
and extension in hydrological computation

HUA Jia-peng, LIN Yun
(College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China)

Abstract Some problems in interpolation and extension of hydrological data by multivariate correlation are studied based on the simulated result of more than five hundred different parameters. The following conclusions are drawn: in hydrological data interpolation and extension, if the data series of reference station X is long, while the series of synchronous observation data of station Z to be designed and reference station X are short, it is difficult to set up a regression equation based on the data series, so the multivariate correlation method should be used for data interpolation and extension by introduction of an intermediate reference station Y . The station introduced should meet the following requirements: (a) long-term synchronous observation data should be available of stations Y and Z , and stations Y and X ; (b) reference station Y is well correlated with station Z and X , the correlation coefficients R_{XY} and R_{YZ} being larger than 0.9, so as to ensure the accuracy of interpolation. Furthermore, expressions reflecting the relationships between correlation parameters are deduced based on statistical results.

Key words hydrological computation; multivariate correlation; regression equation; correlation coefficient