

# 统计相关条件下降水及洪水预报误差相关分析

董前进, 陈森林

(武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉 430072)

**摘要:**以三峡水库上游寸滩至万县区间降水预报误差和入库洪水预报误差相应数据为例,在探讨两者统计相关性的基础上,采用 Frank、Gumbel、Clayton 三种二元 Copula 连接函数分析了两预报误差的相关结构,以离差平方和最小为准则进行了 Copula 函数的选择,并与两预报误差独立情况下联合频率分布进行比较和分析。研究表明,降水预报误差和洪水预报误差的相关性对其二元联合分布有一定影响,同时,在两预报误差负相关条件下,其联合分布可做简化处理。本文研究结果可为水库预报调度风险管理提供决策参考。

**关键词:**降水预报误差;洪水预报误差;统计相关性;Copula;三峡水库

**中图分类号:**P333 **文献标识码:**A **文章编号:**1000-0852(2014)02-0014-05

洪水预报是水库实施预报调度的基础,降水预报可以较大程度提高洪水预报精度并延长洪水预见期,因此,两者的联合使用在当前的水库调度与管理中运用广泛,并发挥了重要作用。然而,一方面,由于气象、水文系统不确定性因素的存在,降水预报误差、洪水预报误差也客观存在且不可避免,一个客观事实是即使在相对发达国家,降水预报的准确率也仅 50%~60%,由此,对降水预报误差、洪水预报误差进行分析特别是分布规律进行探索是非常必要的;另一方面,虽然当前有一些文献对洪水预报误差和降水预报误差进行了研究<sup>[1-4]</sup>,但从数量上,这方面的探讨仍不够,对降水预报误差的分析相对更少,而对降水预报误差与洪水预报误差两者相关性的分析几乎空白,一种可能的解释是认为两者没有相关性可言,或者即使有,也是统计意义上的,即统计相关性。从降水预报与洪水预报受到相同气象条件影响的角度出发,两种误差也存在一定的物理联系,通常所认为的洪水预报误差和降水预报误差相互独立可能是为了计算的方便或者是弱相关条件下的一种简化。基于上述考虑,本文拟探讨统计相关条件下降水预报误差与洪水预报误差的相关问题,这能客观反映降水预报误差与洪水预报误差的真实情况,对于较为准确的计算水库调度水文风险有较大帮助。

## 1 理论与方法

### 1.1 二元变量联合经验分布频率计算

对于随机变量  $X, Y$  的样本  $x_i, y_i$  ( $i=1, 2, \dots, n$ ), 把  $(x_1, x_2, \dots, x_n)$  和  $(y_1, y_2, \dots, y_n)$  按照  $X$  的升序排列, 挑选排好次序数据中  $x_j \leq x_i$  并且  $y_j \leq y_i$  的数据对, 由 Gringorton 公式计算样本  $x, y$  联合分布的经验频率<sup>[5]</sup>:

$$P(X \leq x_i, Y \leq y_i) = \frac{\sum_{k=1}^i \sum_{l=1}^i n_{kl} - 0.44}{n + 0.12} \quad (1)$$

式中:  $P(X \leq x_i, Y \leq y_i)$  为二维联合分布概率;  $n_{kl}$  为  $X \leq x_i, Y \leq y_i$  的序号;  $n$  为挑选的总数据对。

### 1.2 Copula 连接函数

Copula 连接函数是定义在  $[0, 1]$  均匀分布的多维联合分布函数, 它因能独立于随机变量的边缘分布来构造联合分布, 并不要求具有相同的边缘函数, 对各变量的相关性也无强制要求, 具有形式灵活、构造简便的特点, 因而在金融、水利、电气等领域应用广泛<sup>[6]</sup>。对于二维随机变量而言, 它的表达式如下:

$$F(x, y) = C_\alpha(F_X(x), F_Y(y)) \quad (2)$$

式中:  $C_\alpha(\cdot)$  为 Copula 函数;  $\alpha$  为 Copula 参数;  $F_X(x)$ 、 $F_Y(y)$  分别为随机变量  $X$  和  $Y$  的边缘分布。在独立条件下, 有

收稿日期: 2013-03-14

基金项目: 国家自然科学基金(51190094, 50909073); 湖北省自然科学基金(2010CDB08401); 中国长江三峡集团公司项目(CT-09-06-05)

作者简介: 董前进 (1979-), 男, 湖北安陆人, 副教授, 主要从事水资源系统工程研究。E-mail: dqjin@whu.edu.cn

$$C(F_X(x), F_Y(y)) = F_X(x)F_Y(y) \quad (3)$$

Archimedean Copula 函数是水文领域应用比较广泛的一类 Copula 函数,其中又包括 Gumbel-Hougaard Copula、Clayton Copula 和 Frank Copula 等函数形式,关于 Copula 函数形式及其选取与参数估计等,可参考文献[6-7]。

## 2 实例研究

### 2.1 数据准备

以三峡水库的预报调度为例,目前三峡水库的洪水预报及调度中已大量采用降水、洪水预报信息以达到充分发挥三峡水库综合效益的目的,尤其在三峡水库正常蓄水后,降水预报与洪水预报的优劣直接关系到水库蓄满兴利问题,而降水预报误差与洪水预报误差的存在在很大程度上限制了水库兴利效益的充分发挥,其风险分析问题成为三峡水库预报调度的关键制约因素。

当前,三峡水库的入库洪水预报采用了 MIKE 预报方法、水动力学方法及新安江三水源模型方法<sup>[8-10]</sup>,其中新安江水文模型方法得到的预报结果作为三峡水库入库流量预报的主要依据,其它模型预报结果作为预报参考在水库运用中灵活选择,洪水预见期分 12h、24h、48h 和 72h。三峡水库上游的数值降水预报主要结合 EC-20 时 500hpa 高度、EC-20 时风场、T213-20 时 500hpa 高度、JAP-20 时 6h 降水、GER-20 时 6h 降水等信息产生,降水预报作业范围包括混沱江、嘉陵江、乌江、宜宾-重庆(屏山至寸滩)、重庆-万县(寸滩至万县)、万县-宜昌等区域,其中,寸滩至万县降水预报资料年限最长,降水预报预见期分 12h、24h 和 48h,其预报结果均是区间范围预报,计算时以降水预报范围的均值作为降水预报的预报值,其与实测值(面雨量)的差作为降水预报误差。

根据长江干流洪水传播时间,本文洪水预报误差选择三峡水库 2003~2008 年汛期 12h 入库流量预报误差,降水预报误差选择 2003~2008 年汛期寸滩至万县区间 24h 降水预报误差。由于降水预报资料存在漏报、大雨以上量级降水预报次数较少的情况,降水预报误差主要探讨无雨预报、小雨预报情况下的误差,这样,洪水预报误差选择相同时间的误差数据与降水预报误差相对应。根据洪水预报误差和降水预报误差资料匹配情况,本文将分年对两种不同降水预报情况下的降水预报误差与洪水预报误差进行分析。2003~

2008 年历年汛期两种误差匹配情况如表 1 所示:

表1 2003~2008年历年汛期降水预报误差、洪水预报误差数据匹配情况

Table1 The matching of precipitation forecasting error and flood forecasting error for Three Gorges Reservoir during the flood seasons from 2003 to 2008

	2003	2004	2005	2006	2007	2008
无雨预报/对	42	62	64	83	80	67
小雨预报/对	43	35	32	20	17	27

### 2.2 相关系数

先对三峡水库汛期洪水预报误差和降水预报误差的相关性进行分析,得到两者误差的 Pearson 相关系数、Kendall 秩相关系数如表 2 所示。

表2 三峡水库汛期2003~2008年洪水、降水两预报误差相关系数

Table 2 The correlation coefficients of flood forecasting error and precipitation forecasting error for Three Gorges Reservoir during the flood seasons from 2003 to 2008

		2003	2004	2005	2006	2007	2008
无雨预报	$\rho$	0.0224	0.1503	-0.1180	0.2093	-0.0612	0.6682
	$\kappa$	0.1613	0.1782	-0.0647	0.1059	-0.0815	0.0884
小雨预报	$\rho$	-0.1515	0.1191	0.1202			
	$\kappa$	-0.0644	0.0993	0.1612			

注:对于样本容量小于 30 的情况不予研究,下同。

由表 2 可知,在无雨预报条件下,三峡水库汛期 2003~2008 年洪水预报误差与降水预报误差线性相关系数  $\rho$  最大为 0.6682,最小为 0.0224,说明两种误差基本呈统计弱相关情况,其中正相关情形多于负相关情形,说明洪水预报误差与降水预报误差呈正向变化情况较多;秩相关系数  $\kappa$  最大为 0.1782,最小为 -0.0815,说明两种误差基本呈弱同向变化,也呈弱异向变化,同向变化年份较异向变化年份多,并且异向变化年份与负相关年份一致。小雨预报条件可同理分析,这里不再赘述。

通常认为洪水预报误差与降水预报误差是独立的,表 2 可给出两种误差独立的佐证,但是在个别年份两者的相关程度也比较大,如 2008 年, $\rho$  为 0.6682,两误差呈统计中相关情形,在这种情况下,不考虑两种误差的相关性势必会带来较大差错,从而可引起水库预报调度风险增加。

### 2.3 相关性结构

#### 2.3.1 两误差边缘分布

虽然洪水预报误差和降水预报误差通常认为服从正态分布<sup>[11]</sup>,但在对两种误差的正态性检验中所研究样本拒绝正态分布,而关于两种误差的边缘分布笔者

已另文研究,这里考虑在边缘分布未知的情况下,采用样条插值和核估计的方法拟合样本的边缘分布,限于

篇幅,图1仅给出了2004、2005年汛期两误差边缘分布拟合情况。

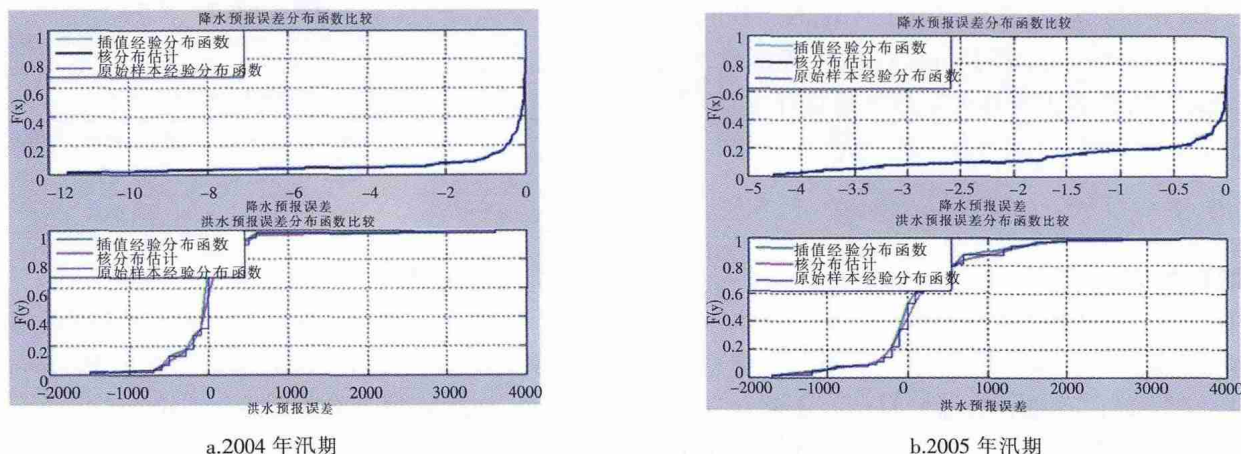


图1 两种误差边缘分布拟合

Fig.1 The fitting of marginal distribution for flood forecasting error and precipitation forecasting error during the flood seasons from 2004 to 2005

从图1可知,无论是降水预报误差还是洪水预报误差,采用样条插值和核估计方法与原始样本经验分布函数拟合较好,几乎完全重合。从2003~2008年整体汛期拟合情况看,两种拟合方法都与原始样本经验分布函数拟合较好,表明采用这两种拟合方法得到的误差分布函数是可行的,相对而言,样条插值较核估计方法拟合效果更好。

2.3.2 两误差二元经验与理论联合频率分布

根据公式(1)计算得到两种误差在不同降雨条件下的经验联合频率分布,根据Copula函数参数不同统计方法<sup>[12]</sup>,采用极大似然估计法对Gumbel Copula、Clayton Copula和Frank Copula参数进行估计,得到三种Copula函数在无雨、小雨预报两种条件下的参数 $\alpha$ 估计值如下表3所示,表3同时给出了由三种不同Copula函数理论联合分布与经验联合分布频率的离差平方和 $d$ ,限于篇幅,图2仅给出了2003年汛期三

种Copula理论联合分布与经验联合分布拟合图,从图中可以看出,三种Copula理论联合分布与经验联合分布拟合较好,经计算三种情况理论与经验联合频率分布相关系数均在0.99以上,其余年份的理论与经验联合分布拟合效果也好,可以满足计算要求。

由表3可知,在无雨预报条件下,降水预报误差与洪水预报误差在正相关情形选用Clayton Copula函数进行两误差的理论联合频率分布拟合较好,在负相关情形选用Frank Copula连接函数;在小雨预报条件下,选用Gumbel Copula或者Frank Copula连接函数拟合较好,其中2004年汛期选用Gumbel Copula,2005年汛期选用Frank Copula,2003年汛期呈负相关情形选用Frank Copula。总体来说,在正相关情形下,三种Copula理论分布与经验分布的拟合略有差异,但差别并不太大,说明三种Copula在降水预报误差与洪水预报误差相关性分析中适用性较好。

表3 三峡水库汛期2003~2008年洪水、降水两预报误差Copula参数 $\alpha$ 与离差平方和 $d$

Table 3 The copula coefficient  $\alpha$  and sum of deviation square  $d$  for flood forecasting error and precipitation forecasting error for Three Gorges Reservoir during the flood seasons from 2003 to 2008

		2003		2004		2005		2006		2007		2008	
		$\alpha$	$d$	$\alpha$	$d$	$\alpha$	$d$	$\alpha$	$d$	$\alpha$	$d$	$\alpha$	$d$
无雨 预报	Frank	1.4828	0.0120	1.6466	0.0266	-0.5847	0.0141*	0.9614	0.0125	-0.7373	0.0116*	0.8008	0.0179
	Gumbel	1.1923	0.0123	1.2169	0.0282	-	-	1.1184	0.0134	-	-	1.0970	0.0197
	Clayton	0.3845	0.0115*	0.4337	0.0221*	-	-	0.2368	0.0119*	-	-	0.1940	0.0160*
小雨 预报	Frank	-0.5819	0.0199*	0.9008	0.0198	1.4821	0.0203*	-	-	-	-	-	-
	Gumbel	-	-	1.1102	0.0190*	1.1922	0.0211	-	-	-	-	-	-
	Clayton	-	-	0.2205	0.0220	0.3843	0.0215	-	-	-	-	-	-

注:标\*号为选定的Copula函数。

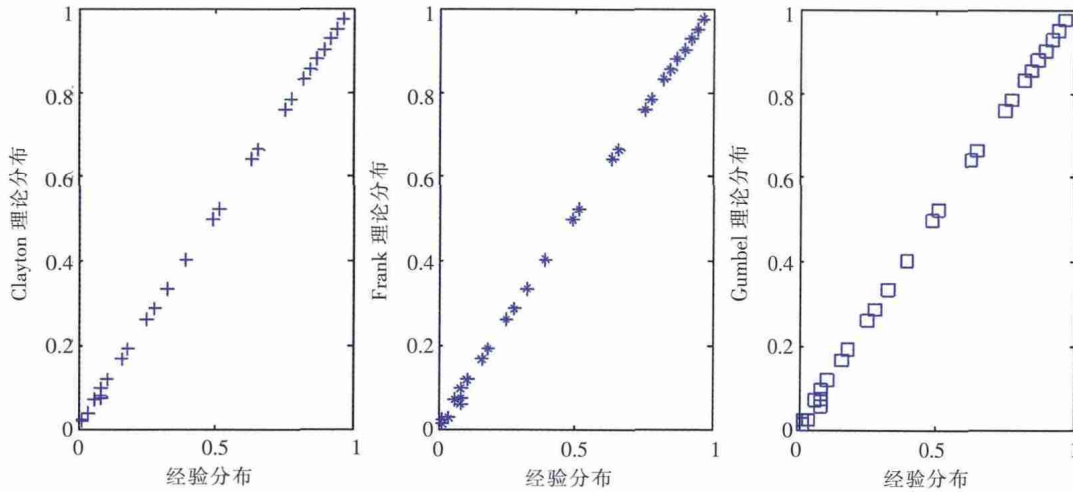


图2 2003年汛期降水、洪水两预报误差三种理论与经验联合分布拟合  
 Fig.2 The fitting of three theoretical distributions and empirical joint distribution for flood forecasting error and precipitation forecasting error during the flood season in 2003

进行经验与理论联合频率分布 K-S 统计检验得知,所选的 Copula 理论频率分布与经验频率分布在显著性水平  $\alpha=0.05$  条件下可认为服从同一分布,表明所选 Copula 拟合优度较好。

2.3.3 两误差相关性影响分析

根据表 3 选定的 Copula 连接函数,与两预报误差在独立情况下联合分布相比较,表 4 给出了理论联合分布与独立联合分布频率差值的最大与最小值。

表4 三峡水库汛期2003~2008年洪水、降水两预报误差理论与独立联合分布差值特征值

Table 4 The characteristics of differentials between flood forecasting error and precipitation forecasting error for Three Gorges Reservoir during the flood seasons from 2003 to 2008

		2003	2004	2005	2006	2007	2008
无雨 预报	最大	0.0339	0.0478	0	0.0286	0	0.0162
	最小	0	0	-0.0180	0	-0.0227	0
小雨 预报	最大	0	0.0241	0.0436	—		
	最小	-0.0181	0	0	—		

由表 4 分析,在洪水、降水两预报误差正相关情形下,由所选 Copula 连接函数所得理论联合分布比独立条件下联合分布略大,由差值最大值统计,最大偏大 0.0478,最小偏差比较小,基本可认为两者无差异,说明在正相关情形下,考虑洪水、降水两预报误差的相关性是有益处的,否则会导致两预报误差联合分

布偏小,从而影响水库预报调度风险决策;而在两预报误差负相关情形下,由所选 Copula 连接函数所得理论联合分布比独立条件下联合分布略小,最小偏小达 0.0227,最大偏小比较小,可认为两者无差异,说明在负相关情形下,考虑洪水、降水两预报误差的相关性会导致减小水库预报调度风险,为水库调度安全起见,在两预报误差弱相关条件下,可不考虑两预报误差的相关性。

3 结语

降水预报误差与洪水预报误差通常认为是不相关的两个变量,其相关性对两者的联合分布无影响,由本文研究可知:

(1)在两预报误差正相关条件下,考虑了两者相关的二元联合分布比独立条件下联合分布频率略有增加,此时计算两预报误差的相关性对水库调度安全来说比较稳妥。

(2)在两预报误差负相关条件下,为水库调度安全起见,可不考虑两者的相关性。

值得一提的是,由于本文的研究是在两预报误差基本呈弱统计相关情形得到的结果,这种结果不能随意外推,其它情形需进一步研究后加以应用。

参考文献:

[1] 刁艳芳,王本德,刘冀. 基于最大熵原理方法的洪水预报误差分布研究[J]. 水利学报, 2007,38(5):591-595. (DIAO Yanfang, WANG Bende, LIU Ji. Study on distribution of flood forecasting errors by the method based on maximum entropy[J]. Journal of Hydraulic

- Engineering, 2007,38(5):591-595.(in Chinese))
- [2] 周惠成,李丽琴,王本德. 洪水预报误差分布的极大熵法[J]. 大连理工大学学报, 2007,47(3):408-413. (ZHOU Huicheng, LI Liqin, WANG Bende. Maximum entropy method used in error distribution of flood forecast[J]. Journal of Dalian University of Technology, 2007,47(3):408-413. (in Chinese))
- [3] 王启光, 封国林, 郑志海, 等. 长江中下游汛期降水优化多因子组合客观量化预测研究[J]. 大气科学, 2011,35(2):287-297. (WANG Qiguang, FENG Guolin, ZHENG Zhihai, et al. A study of the objective and quantifiable forecasting based on optimal factors combinations in precipitation in the middle and lower reaches of the Yangtze river in summer[J]. Chinese Journal of Atmospheric Sciences, 2011,35(2):287-297. (in Chinese))
- [4] 罗雨, 张立凤. 一次梅雨暴雨预报中的误差演变及可预报性分析[J]. 气象学报, 2010,68(3):411-420. (LUO Yu, ZHANG Lifeng. A case study of the error growth evolution in a meiyu front heavy precipitation forecast and an analysis of the predictability[J]. Acta Meteorologica Sinica, 2010,68(3):411-420. (in Chinese))
- [5] Zhang L. Multivariate Hydrological Frequency Analysis and Risk Mapping [D]. Louisiana: Department of Agricultural and Mechanical College, Louisiana State University, USA, 2005.
- [6] 韦艳华,张世英. Copula 理论及其在金融分析上的应用[M]. 北京:清华大学出版社,2008. (WEI Yanhua, ZHANG Shiyong. Copula Theory and Its Application to Analysis of Finance[M]. Beijing: Tsinghua University Press, 2008. (in Chinese))
- [7] Nelson R. B. An introduction to Copulas [M]. Springer, New York, 1999.
- [8] 王玉华,鲍正风,杨旭. 三峡水库入库流量短期洪水预报[J]. 水利水电技术, 2011,42(2):62-65. (WANG Yuhua, BAO Zhengfeng, YANG Xu. Short-term flood forecast on inflow to Three Gorges Reservoir [J]. Water Resources and Hydropower Engineering, 2011, 42(2):62-65. (in Chinese))
- [9] 李春红,王建平,王玉华,等. 三峡特大流域洪水预报方案运用与分析 [J]. 水利学报, 2011,42 (5):624-630. (LI Chunhong, WANG Jianping, WANG Yuhua, et al. Operation and analysis of flood forecasting scheme for the Three Gorges Project[J]. Journal of Hydraulic Engineering, 2011,42(5):624-630. (in Chinese))
- [10] 王船海,郭丽君,芮孝芳,等. 三峡区间入库洪水实时预报系统研究[J]. 水科学进展, 2003,14 (6):677-681. (WANG Chuanhai, GUO Lijun, RUI Xiaofang, et al. Study on real-time flood forecasting system for the Three Gorges Reservoir [J]. Advances in Water Science, 2003,14(6):677-681. (in Chinese))
- [11] 长江水利委员会.水文预报方法[M]. 北京:水利电力出版社, 1993. (Changjiang Water Resources Commission. Hydrology Forecasting Method [M]. Beijing: Water Resources and Electric Power Press, 1993. (in Chinese))
- [12] 肖义. 基于 Copula 函数的多变量水文分析研究[D]. 武汉: 武汉大学博士论文, 2007. (XIAO Yi. Multivariate Hydrological Analysis and Estimation Methods Based on Copula Function [D]. Wuhan: Doctoral Dissertation of Wuhan University, 2007. (in Chinese))

## Correlation Analysis of Precipitation Forecasting Errors and Flood Forecasting Errors Based on Statistical Correlations

DONG Qianjin, CHEN Senlin

(State Key Laboratory of Water Resources and Hydropower Engineering Sciences, Wuhan University, Wuhan 430072, China)

**Abstract:** Taking the precipitation forecasting errors for the reach from Cuntan to Wanxian and inflow flood forecasting errors for Three Gorges Reservoir as an example, the Frank copula, Gumbel copula and Clayton copula were used to analyze the relevance formation of two forecasting errors based on discussing the statistical correlations for the two forecasting errors, and the selection of copula function was performed on the rule of minimum of sum square variation, then the theoretical bivariate joint distribution frequencies were compared to the results of dependent condition. The results demonstrate that considering the correlations between the two forecast errors has some impact on frequency of joint distribution, and on the negative correlation condition, the calculation of the joint distribution between two forecast errors can be simplified. These results may give guidance to the forecasting operation risk management of reservoirs.

**Key words:** precipitation forecasting error; flood forecasting error; statistical correlation; copula; Three Gorges Reservoir