

基于 Copula 函数的多变量水文不确定性处理器

刘章君^{1,2}, 郭生练¹, 何绍坤¹, 巴欢欢¹, 尹家波¹

(1. 武汉大学 水资源与水电工程科学国家重点实验室, 水资源安全保障湖北省协同创新中心, 湖北 武汉 430072;

2. 江西省水利科学研究院, 江西 南昌 330029)

摘要: 传统的水文不确定性处理器(HUP)属于单变量结构类型, 只能独立地给出各预见期实际流量的贝叶斯后验概率密度, 没有考虑它们之间的内在相关性。本文利用 Copula 函数推导了贝叶斯转移预报(BTF)方法中后验转移密度的解析表达式, 提出了基于 Copula 函数的贝叶斯转移预报(CBTF)方法和基于 Copula 函数的多变量水文不确定性处理器(CMHUP), 进而发展了基于 Copula 函数的贝叶斯极值预报(CBEF)方法, 并应用于三峡水库入库洪水预报中。结果表明: 所提方法实用有效, CBTF 方法和 CMHUP 可以定量地评估三峡水库入库流量转移预报的不确定性, 准确揭示了水文预报不确定性在时间上的演变特征, CBEF 方法则提供了预见期时段内最大入库流量预报的不确定性信息。所提方法不需要进行线性-正态假设, 能够很好地捕捉流量过程的非线性和非正态特征, 适用范围更加广泛, 对于支撑防洪减灾和水库运行调度具有重要的参考价值。

关键词: 水文预报; 贝叶斯理论; 水文不确定性处理器; 转移概率预报; 极值概率预报; Copula 函数

中图分类号: TV124

文献标识码: A

doi: 10.13243/j.cnki.slxb.20170675

1 研究背景

水文不确定性处理器(Hydrologic Uncertainty Processor, HUP)作为贝叶斯预报系统(Bayesian Forecasting System, BFS)^[1]的一个重要组件, 用来量化除定量降水预报不确定性以外的其他所有不确定性, 在实际中应用广泛^[2-5]。然而, HUP 属于单变量结构类型, 只能独立地给出各预见期实际流量的贝叶斯后验概率密度, 而没有考虑它们之间的内在相关性。Krzysztofowicz 和 Maranzano^[6]以单步马尔科夫转移密度函数为工具, 提出了基于贝叶斯转移预报(Bayesian Transition Forecast, BTF)方法的多变量水文不确定性处理器(Multivariate HUP, MHUP), 可以在给定确定性预报过程的条件下提供实际流量过程的后验联合概率密度函数, 考虑了不同预见期流量之间的相关性。在此基础上, 针对传统的贝叶斯概率洪水预报方法均只能提供某一时刻流量的概率预报, 无法满足决策者对某一时段内最大流量的预报不确定性信息的实际需求^[7-8], Krzysztofowicz 定义了极值概率预报的概念, 并利用 MHUP 输出产品给出了极值变量的超越概率分布函数, 提出了贝叶斯极值预报(Bayesian Extremum Forecast, BEF)方法^[9]。

BTF 方法是 MHUP 的基本组成构件, 通常采用基于正态分位数转换和线性-正态假设的亚高斯模型进行计算, 但这种正态分位数转换在外推极端事件时效果不稳健, 且逆转换时也可能使结果偏离最优值, 影响了该法的适用性^[10-11]。鉴于 Copula 函数能较好地模拟水文过程的非线性和非正态特征, 在构建多变量联合分布和条件分布中具有巨大优势^[12-15], 刘章君等^[11]提出了基于 Copula 函数的 HUP (CHUP), 不需要进行线性-正态假设, 适用范围更广, 应用更加灵活。事实上, BTF 方法中的先验转移密度和似然函数本质上也都是条件概率密度函数, 但目前没有文献将 Copula 函数引入 BTF 方

收稿日期: 2017-07-10; 网络出版日期: 2018-01-24

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1882.TV.20180124.1102.003.html>

基金项目: 国家自然科学基金重点项目(51539009); 国家重点研发计划项目(2016YFC0402206)

作者简介: 刘章君(1991-), 男, 江西吉安人, 博士生, 主要从事水文分析与计算研究。E-mail: liuzhangjun@whu.edu.cn

法、MHUP及BEF方法的研究中。

因此, 本文将采用 Copula 函数描述 BTF 方法中的先验转移密度和似然函数, 推导后验转移密度的解析表达式, 提出基于 Copula 函数的贝叶斯转移预报(CBTF)方法, 得到基于 Copula 函数的多变量水文不确定性处理器(CMHUP), 进而利用 CMHUP 输出的各预见期实际流量后验联合概率密度函数, 导出预见期时段内洪峰流量的超过概率分布函数, 发展基于 Copula 函数的贝叶斯极值预报(CBEF)方法。以三峡水库为例进行应用, 基于所提方法实现三峡水库入库流量的转移概率预报和极值概率预报。

2 多变量水文不确定性处理器

假设 H_0 表示预报时刻的当前实测流量, H_k 、 S_k ($k=1, 2, \dots, K$) 分别表示待预报的实际流量和确定性预报流量, K 为预见期长度; h_0 、 h_k 和 s_k 分别为随机变量 H_0 、 H_k 和 S_k 的实现值。多变量水文不确定性处理器的目的是在给定当前实测流量 $H_0 = h_0$ 、确定性预报过程随机向量 $\mathbf{S}_K = (S_1, \dots, S_K)$ 的实现值 $\mathbf{s}_K = (s_1, \dots, s_K)$ 的条件下, 推求得到实际流量过程随机向量 $\mathbf{H}_K = (H_1, \dots, H_K)$ 的后验联合概率密度函数。利用概率乘法定理, 后验联合概率密度函数 $\zeta(\mathbf{h}_K | h_0, \mathbf{s}_K)$ 可以因式分解为^[6]:

$$\zeta(\mathbf{h}_K | h_0, \mathbf{s}_K) = \phi_1(h_1 | h_0, s_1) \cdots \phi_k(h_k | h_0, \dots, h_{k-1}, s_k) \cdots \phi_K(h_K | h_0, \dots, h_{K-1}, s_K) \quad (1)$$

式中: $\phi_k(h_k | h_0, \dots, h_{k-1}, s_k)$ 为后验条件概率密度函数。

由于 $\phi_k(h_k | h_0, \dots, h_{k-1}, s_k)$ 的计算非常复杂, 高维联合分布的参数估计在现有数据信息量情况下, 具有相当大的不确定性, 给求解带来很大困难。考虑到 s_k 作为 h_k 的确定性预报值, 通常与 h_k 的相关性最好, 是 h_k 最主要的影响因素, 自回归因子 $h_{k-1}, \dots, h_0$ 则是相对的次要因素。国内外学者的研究表明, 可以将实际日流量过程假设为一阶马尔科夫过程, 抓住 h_k 最主要的两个影响因子 s_k 和 h_{k-1} , 即 $\phi_k(h_k | h_0, \dots, h_{k-1}, s_k)$ 简化为 $\phi_k(h_k | h_{k-1}, s_k)$, 这样在很大程度上保证了结果的精度和可靠性, 是合理、实际、可行的^[16-17]。

相应地, 式(1)也可以简化为:

$$\zeta(\mathbf{h}_K | h_0, \mathbf{s}_K) = \phi_1(h_1 | h_0, s_1) \cdots \phi_k(h_k | h_{k-1}, s_k) \cdots \phi_K(h_K | h_{K-1}, s_K) \quad (2)$$

式中: $\phi_k(h_k | h_{k-1}, s_k)$ 称为转移概率密度函数。这样, 可以将计算 $\zeta(\mathbf{h}_K | h_0, \mathbf{s}_K)$ 这一复杂问题转化为计算式(2)的每一个因式 $\phi_k(h_k | h_{k-1}, s_k)$, 从而使问题变得易于处理。

3 基于 Copula 函数的贝叶斯转移预报方法

通过 CBTF 方法计算式(2)的每一个因式 $\phi_k(h_k | h_{k-1}, s_k)$ ($k=1, 2, \dots, K$), CBTF 方法的概念流程示意图如图 1 所示。实际流量 H_k 被认为同时与实际流量 H_{k-1} 和对应的预报流量 S_k 相关。根据贝叶斯公式, H_k 的后验转移密度函数为^[6]:

$$\phi_k(h_k | h_{k-1}, s_k) = \frac{f_k(s_k | h_{k-1}, h_k) r_k(h_k | h_{k-1})}{\int_{-\infty}^{+\infty} f_k(s_k | h_{k-1}, h_k) r_k(h_k | h_{k-1}) dh_k} \quad (3)$$

式中: $\phi_k(h_k | h_{k-1}, s_k)$ 为 H_k 的后验转移密度函数; $r_k(h_k | h_{k-1})$ 为先验转移密度函数, 代表了实际流量过程的先验不确定性; 对于确定的 $S_k = s_k$, 函数 $f_k(s_k | h_{k-1}, h_k)$ 为 H_k 的似然函数, 反映了确定性预报模型的预报能力。

3.1 Copula 函数 Copula 函数可以将多个随机变量的边缘分布连接起来构造联合分布。令 $G(x_1, x_2, \dots, x_n)$ 为一个 n -维分布函数, 其边缘分布分别为 $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ 。则存在一个

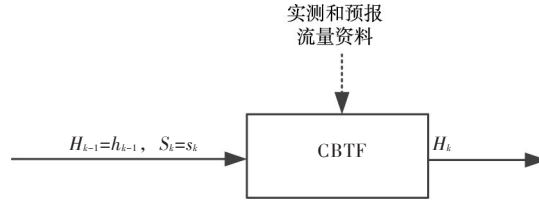


图1 CBTf方法概念示意图

n -Copula函数 C ，使得对任意 $x \in \mathbf{R}^n$ (x 为 n 维向量， $\mathbf{R}^n$ 为 n 维实数空间)^[13]：

$$G(x_1, x_2, \dots, x_n) = C_\theta(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (4)$$

式中： θ 为 Copula 函数的相关性参数。

Copula 函数可以将多个变量的边缘分布和它们之间的相关性结构分开来研究，且对边缘分布类型没有任何限制。采用 Copula 函数进行多变量建模时，通常包括两个步骤：①确定各个变量的边缘分布；②选择合适的 Copula 函数，描述变量之间的相关性结构。

本文选用 Archimedean Copula 函数族构造联合分布，采用 K-S 检验法对 Copula 函数进行拟合检验的基础上，基于 RMSE 准则评价联合分布的理论频率与经验频率拟合情况，选择 RMSE 值最小的 Copula 函数作为最优的 Copula 函数，其数学表达式、参数估计方法、K-S 检验法及 RMSE 值等，详见文献[12-15]。

3.2 先验分布 考虑到变量 H_{k-1} 和 H_k 具有相同的边缘分布，在表示概率分布函数和密度函数时统一为随机变量 H 。利用 Copula 函数， H_{k-1} 、 H_k 的联合分布可写为：

$$R_k(h_{k-1}, h_k) = C(F_H(h_{k-1}), F_H(h_k)) = C(u_1, u_2) \quad (5)$$

先验转移概率分布为给定 $H_{k-1} = h_{k-1}$ 时， H_k 的条件分布函数为^[18-19]：

$$R_k(h_k | h_{k-1}) = P(H_k \leq h_k | H_{k-1} = h_{k-1}) = P(U_2 \leq u_2 | U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1} \quad (6)$$

先验转移密度函数：

$$r_k(h_k | h_{k-1}) = \frac{dR_k(h_k | h_{k-1})}{dh_k} = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2} \frac{du_2}{dh_k} = c(u_1, u_2) f_H(h_k) \quad (7)$$

式中： $c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}$ 为二维 Copula 函数的密度函数。式(7)即为基于 Copula 函数推导的先验转移密度解析表达式。

3.3 似然函数 令 S_k 的边缘分布函数为 $u_3 = F_{S_k}(s_k)$ ，相应的概率密度函数为 $f_{S_k}(s_k)$ 。借助 Copula 函数， H_{k-1} 、 H_k 和 S_k 的联合分布可写为：

$$F_k(h_{k-1}, h_k, s_k) = C(F_H(h_{k-1}), F_H(h_k), F_{S_k}(s_k)) = C(u_1, u_2, u_3) \quad (8)$$

给定 $H_{k-1} = h_{k-1}$ 、 $H_k = h_k$ 时的条件分布函数可表示为^[20-21]：

$$\begin{aligned} F_k(s_k | h_{k-1}, h_k) &= P(S_k \leq s_k | H_{k-1} = h_{k-1}, H_k = h_k) \\ &= P(U_3 \leq u_3 | U_1 = u_1, U_2 = u_2) = \frac{\partial^3 C(u_1, u_2, u_3) / \partial u_1 \partial u_2}{c(u_1, u_2)} \end{aligned} \quad (9)$$

相应的密度函数：

$$\begin{aligned} f_k(s_k | h_{k-1}, h_k) &= dF_k(s_k | h_{k-1}, h_k) / ds_k = \\ &= \frac{1}{c(u_1, u_2)} \frac{\partial^3 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3} \frac{du_3}{ds_k} = \frac{c(u_1, u_2, u_3)}{c(u_1, u_2)} f_{S_k}(s_k) \end{aligned} \quad (10)$$

式中： $c(u_1, u_2, u_3) = \frac{\partial^3 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3}$ 为三维 Copula 函数的密度函数。从另一个角度看，式(10)即为

似然函数的解析表达式, 给定 $H_{k-1}=h_{k-1}$ 、 $S_k=s_k$ 时可以计算 H_k 的似然函数值。

3.4 后验分布 将式(7)和式(10)代入式(3), 可以推导得到实际流量 H_k 后验转移密度函数的解析表达式:

$$\phi_k(h_k|h_{k-1}, s_k) = \frac{c(u_1, u_2, u_3)}{\partial^2 C(u_1, u_2, u_3)/\partial u_1 \partial u_3} f_H(h_k) \quad (11)$$

相应的后验转移分布函数为

$$\Phi_k(h_k|h_{k-1}, s_k) = \int_0^{h_k} \phi_k(h_k|h_{k-1}, s_k) dh_k \quad (12)$$

根据数理统计原理, 可以计算得到期望值作为确定性预报结果, 同时获取给定置信水平下的流量预报区间。实测流量的期望值 h_{ke} 通过下式求解:

$$h_{ke} = \int_0^\infty h_k \phi_k(h_k|h_{k-1}, s_k) dh_k \quad (13)$$

令实测流量 H_k 取值出现在分布两端的概率为 ζ , 就可以定义 H_k 的置信水平为 $(1-\zeta)$ 的区间估计。实测流量 H_k 置信区间的两个端点分别由以下两式给出^[22]:

$$\int_0^{h_{kl}} \phi_k(h_k|h_{k-1}, s_k) dh_k = \zeta_1 \quad (14)$$

$$\int_0^{h_{ku}} \phi_k(h_k|h_{k-1}, s_k) dh_k = 1 - \zeta_2 \quad (15)$$

式中: $\zeta_1 + \zeta_2 = \zeta$, 为显著性水平, 一般取 $\zeta_1 = \zeta_2 = \zeta/2$ 。

因此有

$$P(h_{kl} \leq H_k \leq h_{ku}) = 1 - \zeta \quad (16)$$

即 $[h_{kl}, h_{ku}]$ 为实测流量 H_k 置信水平 $(1-\zeta)$ 的区间估计, 根据置信区间可以对 H_k 的转移预报不确定性进行定量评价。

4 基于 Copula 函数的贝叶斯极值预报方法

4.1 极值概率预报的定义 令随机变量 Z_k 代表时段 $(0, k]$ 的最大实际流量, 这个新定义的一维连续随机变量可以看成是随机变量的极大值变量, 它由一个时间离散的流量随机过程 $\{H_1, \dots, H_k\}$ 衍生而得。

$$Z_k = \max\{H_1, \dots, H_k\} \quad (17)$$

对于任意给定的流量值 h , 设 $\overline{F}_k(h)$ 为最大实际流量 Z_k 的超过概率分布函数, 通过下式定义:

$$\overline{F}_k(h) = P(Z_k > h) = 1 - P(Z_k \leq h) \quad (18)$$

将式(17)代入, 并进行演算可得:

$$\overline{F}_k(h) = 1 - P(\max\{H_1, \dots, H_k\} \leq h) = 1 - P(H_1 \leq h, \dots, H_k \leq h) \quad (19)$$

实际上, $\overline{F}_k(h)$ 表示的是时段 $(0, k]$ 的最大实际流量 Z_k 超过给定值 h 的概率。换句话说, 也可以认为是 k 个时刻的实际流量 $\{H_1, \dots, H_k\}$ 中至少有一个超过给定值 h 的概率。

4.2 CBEF 方法的数学描述 CBEF 方法的概念流程示意图如图 2 所示。从式(19)可以看出, 对于任意给定的流量值 h , 求解超过概率分布函数 $\overline{F}_k(h)$ 的关键在于计算得到联合概率分布值 $P(H_1 \leq h, \dots, H_k \leq h)$ 。

CMHUP 输出的实际流量过程随机向量 $\mathbf{H}_k = (H_1, \dots, H_k)$ 的后验联合概率密度函数可以表示为下式:

$$\zeta(h_1, \dots, h_k|h_0, s_1, \dots, s_k) = \phi_1(h_1|h_0, s_1) \cdots \phi_k(h_k|h_{k-1}, s_k) \quad (20)$$

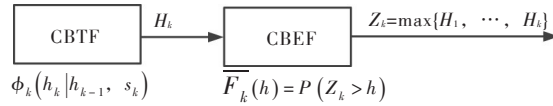


图2 CBEF方法概念示意图

相应的后验联合概率分布函数为 $\Xi(h_1, \dots, h_k|h_0, s_1, \dots, s_k)$ ，则不超过概率分布函数 $F_k(h)$ 可以表示为：

$$F_k(h) = P(H_1 \leq h, \dots, H_k \leq h) = \Xi(h, \dots, h|h_0, s_1, \dots, s_k) \quad (21)$$

则式(19)可以重新写为：

$$\overline{F}_k(h) = 1 - \int_{-\infty}^h \dots \int_{-\infty}^h \phi_1(h_1|h_0, s_1) \dots \phi_k(h_k|h_{k-1}, s_k) dh_1 \dots dh_k \quad (22)$$

式中：后验转移概率密度函数 $\phi_k(h_k|h_{k-1}, s_k)$ 由式(11)计算得到，采用文献[9]的数值方法求解式(22)。

5 实例应用

三峡工程是开发与治理长江的关键性骨干工程，具有防洪、发电、航运、补水等综合效益。三峡水库集水面积约为100万 km^2 ，坝址多年平均流量为14 300 m^3/s 。本文基于CBTF方法实现三峡水库汛期入库流量的转移概率预报，进而基于CBEF方法实现三峡水库汛期入库流量的极值概率预报。选用2003—2016年汛期(6月1日—9月30日)发生的22场实测入库洪水过程及长江水文气象预报中心发布的预见期1~3 d相应的确定性预报入库洪水过程资料。实测值和预报值均为每日8时入库流量，取样时以预见期对应的洪水发生的实际时间为基础，反推相应的预报发布时刻，同步样本数均为707次。

5.1 边缘分布的确定 边缘分布函数可以采用任意分布，选用的标准是使得假定理论分布与流量资料拟合较好。本文采用P-Ⅲ分布作为实际流量和预报流量的边缘分布，其密度函数表达式如下^[23]：

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - \gamma)^{\alpha-1} e^{-\beta(x-\gamma)} \quad (23)$$

式中： α 、 β 和 γ 分别为形状、尺度和位置参数。

考虑到变量 $H_k(k=1, 2, 3)$ 的边缘概率分布相同，将其统一表示为随机变量 H 。采用P-Ⅲ分布线型分别作为 H 、 $S_k(k=1, 2, 3)$ 的理论线型，利用L-矩法估计参数，结果见表1。采用K-S检验法分别对4个随机变量进行拟合检验，单变量K-S检验统计量 D 值也列于表1。结果表明，在5%的显著性水平下，K-S检验统计量 D 均小于相应的临界值(0.0511)，各变量均通过了检验。因此，估计的实际流量和预报流量边缘分布是合理可行的。

表1 边缘分布参数估计结果和K-S检验

变量	参数			K-S 检验 统计量 D
	α	β	γ	
H	3.08	0.00018	10083.58	0.0258
S_1	3.02	0.00018	10719.91	0.0330
S_2	3.13	0.00019	10835.52	0.0249
S_3	3.43	0.00021	10355.12	0.0293

5.2 联合分布的建立 变量 H_{k-1} 、 H_k 和 $S_k(k=1, 2, 3)$ 两两之间的Kendall秩相关系数矩阵见表2。可以发现，3个变量之间的相关性是不对称的， H_k 和 S_k 的相关性最高，且 H_{k-1} 和 H_k 、 H_{k-1} 和 S_k 的相关性基本相当。

考虑到 H_{k-1} 、 H_k 的联合分布与 H_0 、 H_1 的联合分布相同，因而只需要估计 H_0 、 H_1 的联合分布。分

别采用二维 Gumbel-Hougaard、Clayton 和 Frank Copula 函数分别构造 H_0 、 H_1 的联合分布，基于不同预见期的同步系列数据，分别得到相应的 Kendall 秩相关系数 τ ，根据 τ 与参数 θ 的关系分别计算 Copula 函数的参数值，结果见表 3。二维 K-S 检验统计量 D 和 RMSE 值也列于表 3。可知，在 5 % 的显著性水平下，3 种备选 Copula 函数的二维 K-S 检验统计量 D 均小于临界值 (0.0511)，并且二维 Frank Copula 函数的 RMSE 值均为最小，因而选择其构造 H_0 、 H_1 的联合分布。

表 2 变量 H_{k-1} 、 H_k 和 S_k 两两之间的 Kendall 秩相关系数矩阵

预见期	变量	τ
1d	H_0, H_1	0.725
	H_0, S_1	0.735
	H_1, S_1	0.904
2d	H_1, H_2	0.725
	H_1, S_2	0.752
	H_2, S_2	0.840
3d	H_2, H_3	0.725
	H_2, S_3	0.750
	H_3, S_3	0.752

表 3 H_0 和 H_1 联合分布参数估计、检验及优选结果

变量	Copula 类型	τ	θ	D	RMSE
H_0, H_1	Gumbel-Hougaard		3.64	0.0254	0.0113
	Clayton	0.725	5.29	0.0272	0.0120
	Frank		12.68	0.0201	0.0086

分别采用三维非对称 Gumbel-Hougaard、Clayton 和 Frank Copula 函数分别构造 H_{k-1} 、 H_k 和 S_k 的联合分布，基于不同预见期的同步系列数据，利用极大似然法估计的三维非对称 Copula 函数参数值、三维 K-S 检验统计量 D 和 RMSE 值列于表 4。可以发现，在 5 % 的显著性水平下，对于预见期 1 d 和 2 d 三维非对称 Gumbel-Hougaard Copula 函数和 Frank Copula 函数通过了 K-S 检验 (临界值 0.0511)，而预见期 3 d 只有三维非对称 Frank Copula 函数通过了检验。而且，3 个预见期三维非对称 Frank Copula 函数均给出了最小的 RMSE 值，因此选择其构造 H_{k-1} 、 H_k 和 S_k 的联合分布。

表 4 H_{k-1} 、 H_k 和 S_k 三维非对称 Copula 函数参数估计、检验及优选结果

变量	Copula 类型	$[\theta_1, \theta_2]$	D	RMSE
H_0, H_1, S_1	Gumbel-Hougaard	[3.26, 8.61]	0.0413	0.0175
	Clayton	[3.57, 10.54]	0.0644	0.0283
	Frank	[13.02, 30.96]	0.0261	0.0106
H_1, H_2, S_2	Gumbel-Hougaard	[3.30, 5.41]	0.0450	0.0190
	Clayton	[3.74, 6.33]	0.0661	0.0308
	Frank	[13.52, 23.30]	0.0232	0.0100
H_2, H_3, S_3	Gumbel-Hougaard	[3.17, 3.69]	0.0548	0.0217
	Clayton	[3.54, 4.19]	0.0750	0.0328
	Frank	[12.87, 14.95]	0.0262	0.0119

5.3 转移概率预报结果示例 任意给定实际流量 H_{k-1} 和确定性预报流量 S_k ($k=1, 2, 3$) 的取值，就可以求解实际流量 H_k 的后验转移概率密度和后验转移分布函数，实现入库流量的转移概率预报。在实际分析 H_k 的后验不确定性时，实际流量 H_{k-1} 的取值是未知的，其取值状态可以是可行域中的任何值。以预报时刻 2003 年 7 月 5 日 (20030705) 为例，预报时刻的实测流量 $h_0=33\ 700\ \text{m}^3/\text{s}$ 。图 3 和图 4 分别给出了 h_0 (为已知值)、 h_1 和 h_2 取各种不同状态时，预见期 1 ~ 3 d 实际入库流量 H_k 的后验转移密度函数曲线和后验转移分布函数曲线。可知，后验转移分布呈现正偏态特征，CBTF 方法可以准确地描述相邻两个预见期流量之间的后验转移概率密度和分布函数，从而定量评估转移预报的不确定性。

为了展示入库流量预报不确定性在时间上的演变特征，我们把实际发生的流量过程当作已知值 ($h_0=33\ 700\ \text{m}^3/\text{s}$, $h_1=36\ 100\ \text{m}^3/\text{s}$, $h_2=38\ 600\ \text{m}^3/\text{s}$ 和 $h_3=43\ 000\ \text{m}^3/\text{s}$)，对应的确定性预报结果过程为 ($s_1=35\ 500\ \text{m}^3/\text{s}$, $s_2=38\ 000\ \text{m}^3/\text{s}$ 和 $s_3=37\ 500\ \text{m}^3/\text{s}$)。三峡水库 20030705 入库流量预报先验和后验不确

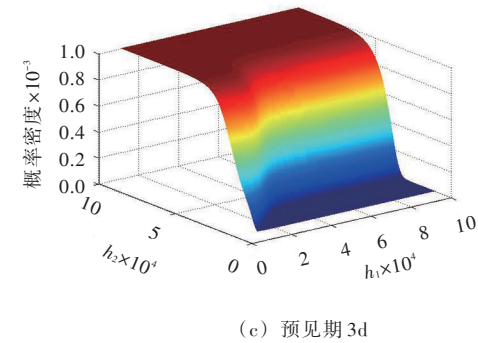
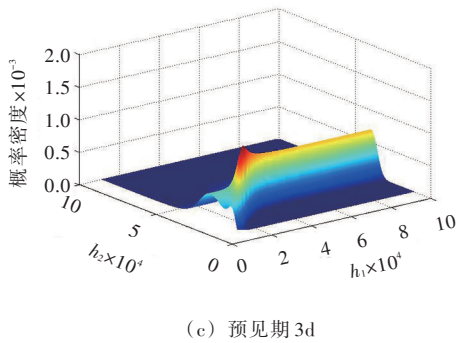
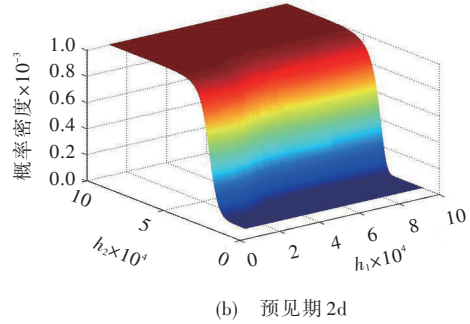
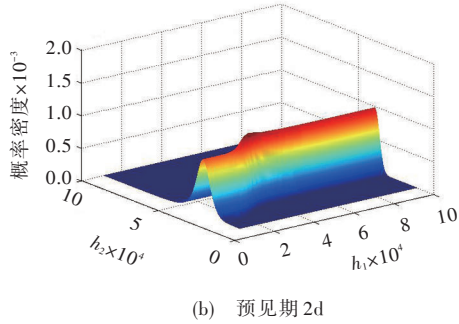
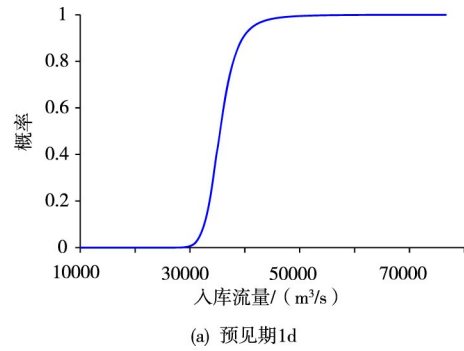
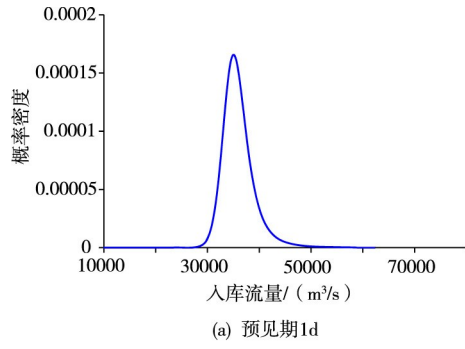


图3 三峡水库 20030705 入库流量后验转移密度曲线

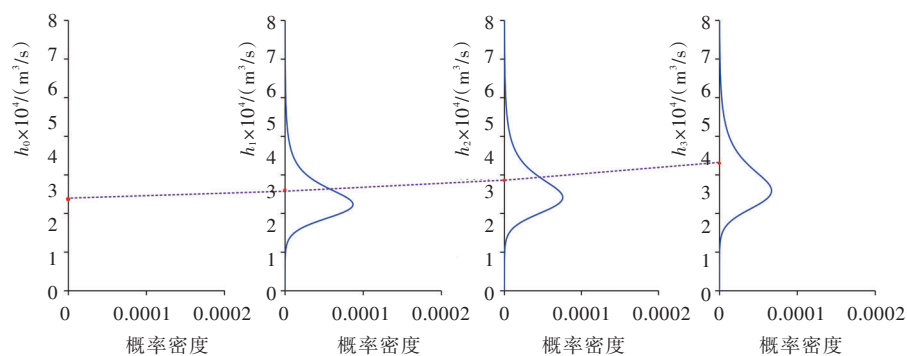
图4 三峡水库 20030705 入库流量后验转移分布曲线

定性随时间的演变过程见图 5，其中红色实点代表实际值，虚线为实际流量过程。可以发现，随着预见期的延长或者时间的推移，入库流量预报的先验和后验转移密度函数由陡变缓，表明预报不确定性增大，这与实际情况相符。但利用确定性预报信息经过贝叶斯修正之后的后验转移密度函数与相应的先验密度函数相比都更加尖瘦和集中，不确定性更小。CBTF 方法和 CMHUP 为分析水文预报不确定性在时间上的演变特征提供了有效工具。

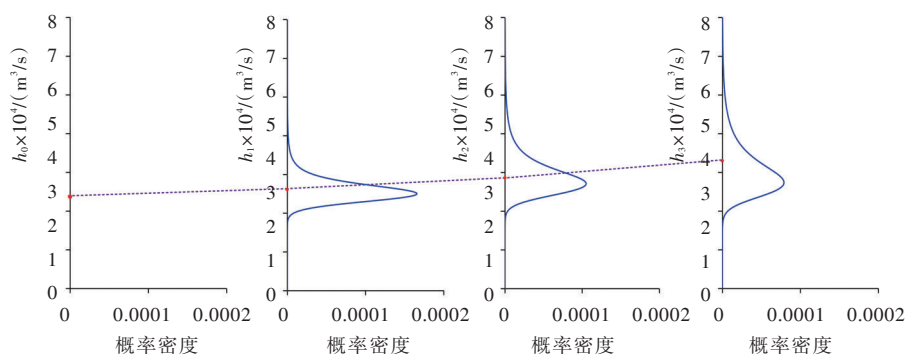
图 6 给出了三峡水库“20160701”号洪水过程预见期 1 ~ 3 d 的后验期望值和 90 % 置信区间及实际值。可以看出，洪水流量的量级越大，相应的不确定性一般也越大。后验期望值预报与实际流量序列拟合效果总体较好，但拟合效果随预见期的延长而降低。此外，90 % 预报区间也随着预见期的延长均变宽，预报不确定性增大，但基本上可以包住实际流量，表明概率区间预报是可靠的，可以为水库防洪调度决策提供更多的风险信息。

5.4 转移概率预报结果评价 基于整个研究数据集（样本数为 707），从后验期望值预报的精度和概率预报的整体性能对 CMHUP 的转移概率预报结果进行评价，并与 CHUP 的结果进行对比分析。从理论定义上讲，CHUP 描述的是 k 步流量状态转移 $H_0 \rightarrow H_k$ ($k=1, 2, 3$) 的不确定性，而 CMHUP 描述的则是单步流量状态转移 $H_{k-1} \rightarrow H_k$ 的不确定性。需要指出的是，预见期 1 d 时两者结果是相同的，描述的都是状态转移 $H_0 \rightarrow H_1$ 的不确定性，以下重点分析预见期 2 d、3 d 的情形。

(1) 后验期望值预报。三峡水库预见期 1 ~ 3 d 的确定性预报与 CHUP、CMHUP 后验期望值预报结果的精度评价指标 R^2 和 RE 分别见表 5。可以看出，相比确定性预报，CHUP、CMHUP 后验期望值预



(a) 先验



(b) 后验

图5 三峡水库 20030705 入库流量预报不确定性随时间的演变过程

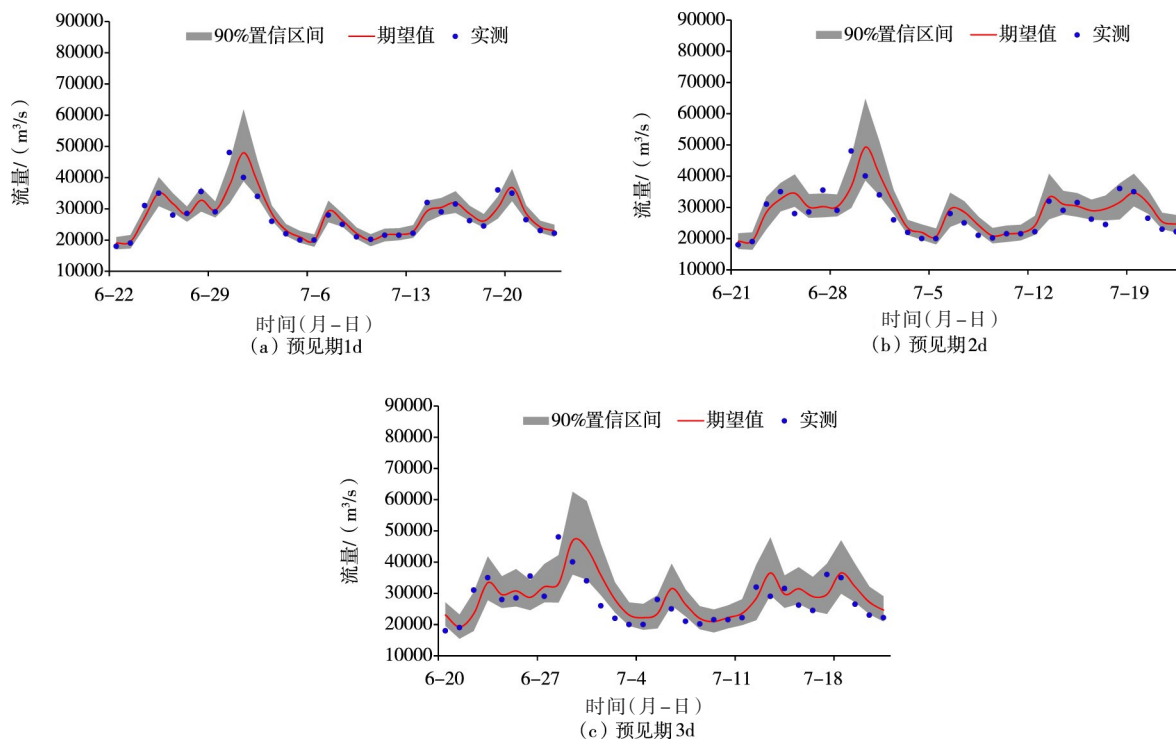


图6 三峡水库“20160701”号入库洪水流量实测值、期望值与90%置信区间

报结果 R^2 均有一定程度的提高, RE 显著地减小, 而且 CMHUP 的后验期望值预报相比 CHUP 略有改善。由于三峡水库的防洪库容较大, 水库运行调度主要受洪水总量控制, 因而准确地预报径流总量具有重要意义, 相比于确定性预报而言, CHUP、CMHUP 后验期望值预报在这方面显示出了很大优势。还可以发现随着预见期的延长, 由于确定性预报精度的下降, 导致 CHUP、CMHUP 后验均值预

表5 不同预见期的确定性预报与CHUP、CMHUP后验期望值预报结果比较（单位：%）

预见期/d	确定性预报		CHUP		CMHUP	
	R^2	RE	R^2	RE	R^2	RE
1	96.15	0.09	96.19	0.01	96.19	0.01
2	90.75	-0.59	91.63	0.04	91.85	0.04
3	80.90	-2.71	83.29	0.11	84.56	0.09

报的精度都相应地降低。

(2) 概率预报。采用可靠性(α -index)、分辨率(π -index)和连续概率排位分数(CRPS, 确定性预报退化为平均绝对误差MAE)等指标来评定概率预报的整体性能^[24-25]。CHUP、CMHUP不同预见期的 α -index、 π -index和CRPS等指标值列于表6。可知, 不同预见期的 α -index值均大于0.9, 表明概率预报结果可靠性高。此外, 随着预见期的延长, π -index值逐渐减小, 概率预报的分辨率和精度降低, 说明入库流量的预报不确定性增加。

表6 不同预见期的CHUP、CMHUP概率预报评价指标值比较

预见期/d	确定性预报	CHUP			CMHUP		
	CRPS/MAE/(m ³ /s)	α -index	π -index	CRPS/(m ³ /s)	α -index	π -index	CRPS/(m ³ /s)
1	1097	0.9147	14.09	970	0.9147	14.09	970
2	1847	0.9269	10.55	1526	0.9210	10.99	1505
3	2782	0.9306	6.92	2188	0.9247	7.74	2133

随着预见期的延长, CHUP、CMHUP计算的綜合指标CRPS值也呈现不断增加的趋势, 意味着概率预报的性能和总体效果降低。然而, 基于CHUP、CMHUP得到的概率预报CRPS值, 始终小于相应的确定性预报, 彰显了概率预报的有效性。而且, CMHUP相比于CHUP也有不同程度的改善, 表明入库流量单步状态转移的不确定性要低于多步转移。CMHUP计算的预见期1~3 d的CRPS值相比于确定性预报分别降低11.58%、18.52%和23.33%。

5.5 最大入库流量超过概率分布函数 根据样本空间 $\{H_k: H_k > \gamma_{H_k}\}$, 取 $h_{k\min}=10\ 100\ \text{m}^3/\text{s}$; 同时考虑入库流量概率分布的特点, 取累积概率超过0.9999的流量分位数值作为 h_k 取值的上限 $h_{k\max}=100\ 000\ \text{m}^3/\text{s}$ 。因此, 本文 $h_k(k=1, 2, 3)$ 的取值区间为 $10\ 100 \sim 100\ 000\ \text{m}^3/\text{s}$, 离散间距为 $100\ \text{m}^3/\text{s}$, 相应的离散点数据为900。以预报时刻2003年7月5日(20030705)为例, 对于给定的流量值 h (离散取值和 h_k 相同), 根据计算的转移概率预报结果, 可以计算得到不同预见期内最大入库流量的超过概率分布函数 $\overline{F_k}(h)$, 超过概率分布曲线见图7。

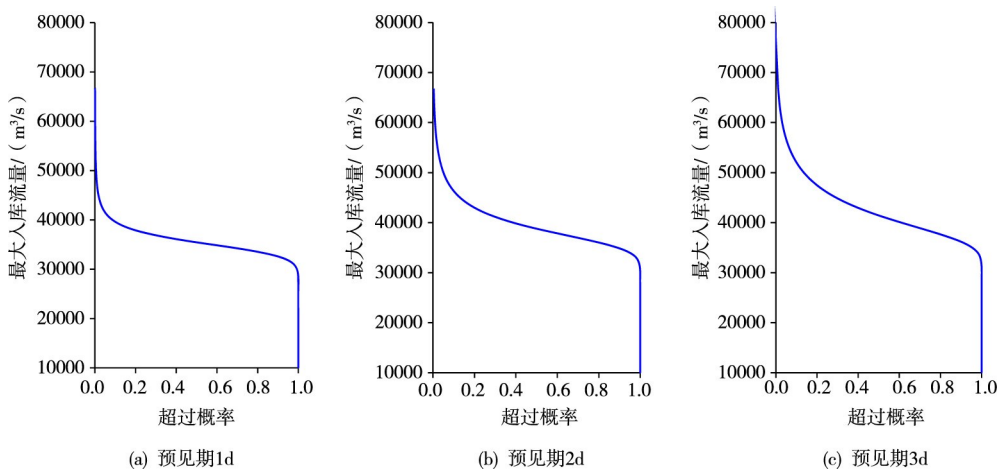


图7 三峡水库20030705不同预见期内最大入库流量超过概率分布曲线

从图中可以读取不同预见期和给定流量值 h 对应的最大入库流量超过概率,例如1 d内最大入库流量超过35 000 m³/s的概率 $\overline{F}_1(35000)=0.58$,2 d内最大入库流量超过40 000 m³/s的概率 $\overline{F}_2(40000)=0.39$,3 d内最大入库流量超过50 000 m³/s的概率 $\overline{F}_3(50000)=0.14$ 。另外,还可以计算得到3 d内最大入库流量对应的90%置信区间为[35100, 56700]m³/s,提供了预见期时段内最大入库流量预报的不确定性和风险信息。

6 结论

本文提出了基于Copula函数的贝叶斯转移预报(CBTF)方法和基于Copula函数的多变量水文不确定性处理器(CMHUP),进而发展了基于Copula函数的贝叶斯极值预报(CBEF)方法。以三峡水库为例进行应用,主要研究结论如下:(1)CBTF方法不需要进行线性-正态假设,可以在原始流量数据空间给出贝叶斯后验转移密度函数的解析表达式,能够很好地捕捉流量过程的非线性和非正态特征,适用范围更加广泛。(2)CMHUP可以在给定当前实测流量和确定性预报过程的条件下,更加灵活地给出考虑实际流量过程内在相关性结构的后验联合概率密度函数,为分析水文预报不确定性在时间上的演变特征提供了有效工具。(3)CBEF方法给决策者提供了预见期时段内最大入库流量预报的不确定性和风险信息,对防洪预报预警和减灾具有重要的参考价值。

本文采用Copula函数取代原有基于正态分位数转换和线性-正态假设的亚高斯模型,不仅在理论上存在优势,而且实用有效。然而,本文方法在假定实际流量过程服从一阶马尔科夫过程的基础上推导得到,对于流量过程服从两阶及以上马尔科夫过程的情形本文方法的思路仍然适用,但需要采用更高维的Copula函数,计算更加复杂,有待进一步研究。另外,如何将定量降水预报不确定性与本文所提方法进行耦合,估计转移预报和极值预报的总不确定性,也需要进一步探讨。

参 考 文 献:

- [1] KRZYSZTOFOWICZ R . Bayesian theory of probabilistic forecasting via deterministic hydrologic model[J] . Water Resources Research, 1999, 35(9): 2739-2750 .
- [2] KRZYSZTOFOWICZ R, KELLY K S . Hydrologic uncertainty processor for probabilistic river stage forecasting [J] . Water Resources Research, 2000, 36(11): 3265-3277 .
- [3] 李向阳,程春田,林剑艺 . 基于BP神经网络的贝叶斯概率水文预报模型[J] . 水利学报, 2006, 37(3): 354-359 .
- [4] 鄢来标,康玲 . 河道洪水概率预报方法研究[J] . 水力发电, 2008, 34(12): 40-41 .
- [5] 蒋晓蕾,梁忠民,王春青,等 . BFS-HUP模型在潼关站洪水概率预报中的应用[J] . 人民黄河, 2015, 37(7): 13-15 .
- [6] KRZYSZTOFOWICZ R, MARANZANO C J . Hydrologic uncertainty processor for probabilistic stage transition forecasting[J] . Journal of Hydrology, 2004, 293(1): 57-73 .
- [7] KRZYSZTOFOWICZ R . A theory of flood warning systems[J] . Water Resources Research, 1993, 29(12): 3981-3994 .
- [8] 程晓陶,李娜,王艳艳,等 . 防汛预警指标与等级划分的比较研究[J] . 中国防汛抗旱, 2010, 20(3): 26-31 .
- [9] KRZYSZTOFOWICZ R . Probabilistic flood forecast: Exact and approximate predictive distributions[J] . Journal of Hydrology, 2014, 517(1): 643-651 .
- [10] MADADGAR S, MORADKHANI H . Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging[J] . Water Resources Research, 2014, 50(12): 9586-9603 .
- [11] 刘章君,郭生练,李天元,等 . 贝叶斯概率洪水预报模型及其比较应用研究[J] . 水利学报, 2014, 45(9): 1019-1028 .
- [12] SALVADORI G, de MICHELE C . On the use of copulas in hydrology: theory and practice[J] . Journal of Hydrologic Engineering, 2007, 12(4): 369-380 .

- [13] 郭生练, 闫宝伟, 肖义, 等. Copula 函数在多变量水文分析计算中的应用及研究进展[J]. 水文, 2008, 28(3): 1-7.
- [14] 孙鹏, 张强, 陈晓宏. 基于 Copula 函数的鄱阳湖流域极值流量遭遇频率及灾害风险[J]. 湖泊科学, 2011, 23(2): 183-190.
- [15] 冯平, 李新. 基于 Copula 函数的非一致性洪水峰量联合分析[J]. 水利学报, 2013, 44(10): 1137-1147.
- [16] MARANZANO C J, KRZYSZTOFOWICZ R. Identification of likelihood and prior dependence structures for hydrologic uncertainty processor[J]. Journal of Hydrology, 2004, 290(1-2): 1-21.
- [17] ENGELAND K, STEINSLAND I. Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times[J]. Water Resources Research, 2014, 50(1): 182-197.
- [18] 刘曾美, 覃光华, 陈子桑, 等. 感潮河段水位与上游洪水和河口潮位的关联性研究[J]. 水利学报, 2013, 44(11): 1278-1285.
- [19] 陈璐, 卢伟伟, 周建中, 等. 水文预报不确定性对水库防洪调度的影响分析[J]. 水利学报, 2016, 47(1): 77-84.
- [20] 侯芸芸, 宋松柏, 赵丽娜, 等. 基于 Copula 函数的三变量洪水频率研究[J]. 西北农林科技大学学报(自然科学版), 2010, 38(2): 219-228.
- [21] 梁忠民, 郭彦, 胡义明, 等. 基于 copula 函数的三峡水库预泄对鄱阳湖防洪影响分析[J]. 水科学进展, 2012, 23(4): 485-492.
- [22] 刘章君, 郭生练, 李天元, 等. 设计洪水地区组成的区间估计方法研究[J]. 水利学报, 2015, 46(5): 543-550.
- [23] 中华人民共和国水利部. 水利水电工程设计洪水计算规范(SL44-2006) [S]. 北京: 中国水利水电出版社, 2006.
- [24] THYER M, RENARD B, KAVETSKI D, et al. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis[J]. Water Resources Research, 2009, 45(12): 1211-1236.
- [25] 李明亮, 杨大文, 陈劲松. 基于采样贝叶斯方法的洪水概率预报研究[J]. 水力发电学报, 2011, 30(3): 27-33.

Multivariate hydrologic uncertainty processor based on Copula function

LIU Zhangjun^{1, 2}, GUO Shenglian¹, HE Shaokun¹, BA Huanhuan¹, YIN Jiabo¹

(1. State Key Laboratory of Water Resources and Hydropower Engineering Science,

Hubei provincial Collaborative Innovation Center for Water Resource Security, Wuhan University, Wuhan 430072, China;

2. Jiangxi Provincial Institute of Water Sciences, Nanchang 330029, China)

Abstract: The traditional hydrologic uncertainty processor (HUP) belongs to the univariate structure type, which only independently provides a marginal Bayesian posterior probability density function of observed discharge for each lead time and does not consider and characterize the inherent dependence among these variables. In this paper, the analytical expression of Bayesian posterior transition density was derived by using Copula function, and therefore the Copula-based BTF (CBTF) method and Copula-based multivariate HUP (CMHUP) was proposed. Subsequently, the Copula-based BEF (CBEF) was developed. Application results of Three Gorges Reservoir (TGR) indicate that the proposed methods are practical and effective, of which the CBTF method and CMHUP not only can quantitatively evaluate the uncertainty of transition forecast for inflows of the TGR, but also reveal the evolution characteristic with time of uncertainty in hydrological forecasting. Moreover, the uncertain information about the maximum inflow forecast within specified lead time is provided by the CBEF method. The proposed methods relax the linear-normal assumption and capture the nonlinear and non-Gaussian characteristics of discharge process adequately, which lead to more extensive application scope and support the flood control and disaster mitigation, and reservoir operation better.

Keywords: hydrological forecasting; Bayesian theory; hydrologic uncertainty processor; probabilistic transition forecast; probabilistic extremum forecast; Copula function

(责任编辑: 李 琳)