

## Copula 多元随机分析

### —方法简介

Copula是一种连接边缘分布和多变量联合分布的关联函数。以三变量 Copula为例，其一般表达为：

$$F(u, v, t) = C_{\theta}[F(u), F(v), F(t)]。$$

其中  $F(u)$ 、 $F(v)$  和  $F(t)$  是随机向量  $(U, V, T)$  的边缘分布， $\theta$  为未知参数。Pair-Copula是一种图解模型，能够基于一系列的二维Copula方法建立一个多维Copula函数。它可以提供多种弹性结构用于描述多元随机变量之间的互动关系。例如变量  $u, v, t$  的联合分布可以表示为：

$$f(u, v, t) = f(u)f(v)f(t)c_{u,t}(F(u), F(t)|\theta_{u,t})c_{v,t}(F(v), F(t)|\theta_{v,t})c_{u,v|t}(F(u|t), F(v|t)|\theta_{u,v|t})$$

其中， $f(u)$ 、 $f(v)$  和  $f(t)$  是变量  $u, v, t$  的边缘概率密度函数， $f(u, v, t)$  表示联合概率密度函数， $c_{u,v|t}(F(u|t), F(v|t)|\theta_{u,v|t})$  是变量  $t$  条件下变量  $u$  和  $v$  的二维 Copula 密度函数，其参数值为  $\theta_{u,v|t}$ 。

用 Copula 函数描述并模拟多变量之间的依赖性结构，可灵活反映多元随机变量间的互动关系，为多元随机风险分析及不确定性分析奠定基础。

### 案例 1：基于 Copula 多元随机分析的珠三角城市群能源管理技术

本研究开发了不确定性 Copula 随机分析技术，将 Copula 函数引入两阶段联合概率分析模型可以有效地解决约束中随机变量间的非线性互动性。其一般表达式为：

目标函数：

$$\text{Maximize: } f^{\pm} = \sum_{j=1}^n c_j^{\pm} x_j^{\pm} - \sum_{j=1}^n d_j^{\pm} \left( \sum_{i=1}^m p_i y_{ij}^{\pm} \right) - \sum_{t=1}^T e_t^{\pm} z_t$$

约束条件：

$$\sum_{j=1}^n a_{rj}^{\pm} x_j^{\pm} \leq b_r^{\pm}, \quad r = 1, 2, \dots, R$$

$$\sum_{j=1}^n (a_{kj}^{\pm} x_j^{\pm} + a'_{kij} y_{ij}^{\pm}) \geq \omega_{ki}^{\pm}, \quad i = 1, 2, \dots, m; \quad k = 1, 2, \dots, K$$

$$C \left( \sum_{j=1}^n (a_{sj}^{\pm} x_j^{\pm} + a'_{sij} y_{ij}^{\pm}) \leq b_s(t), \quad s = 1, 2, \dots, S' \right) \geq 1 - P, \quad i = 1, 2, \dots, m$$

$$z_t = 0 \quad \text{or} \quad 1, \quad t = 1, 2, \dots, T$$

$$x_j^{\pm} \geq 0, \quad j = 1, 2, \dots, n$$

$$y_{ij}^{\pm} \geq 0, \quad j = 1, 2, \dots, n; \quad i = 1, 2, \dots, m$$

其中，联合概率约束  $C \left( \sum_{j=1}^n (a_{sj}^{\pm} x_j^{\pm} + a'_{sij} y_{ij}^{\pm}) \leq b_s(t), \quad s = 1, 2, \dots, S' \right) \geq 1 - P$  可基

于 Copula 函数转化为以下线性形式

$$\sum_{j=1}^n (a_{sj}^{\pm} x_j^{\pm} + a'_{sij} y_{ij}^{\pm}) \leq b_s(t)^{(p_s^0)}, \quad s = 1, 2, \dots, S' \quad i = 1, 2, \dots, m$$

$$C(F_s(b_1(t)), \dots, F_s(b_s(t)), \dots, F_s(b_{S'}(t))) \geq 1 - P$$

式中， $b_s(t)$  的累计分布函数可以表示为  $F_s(b_s(t))$ ，对应的设计违反约束的概率为  $p_s^0$ ，满足  $F_s(b_s(t)) = 1 - p_s^0$ 。这种方法也是将联合概率约束转化为一列线性约束，但是要求 Copula 函数构造的联合概率分布在每个线性约束的违反概

率条件下（即  $F_s(b_s(t))=1-p_s^0 (s=1,2,\dots,S')$ ）满足给定的联合概率。

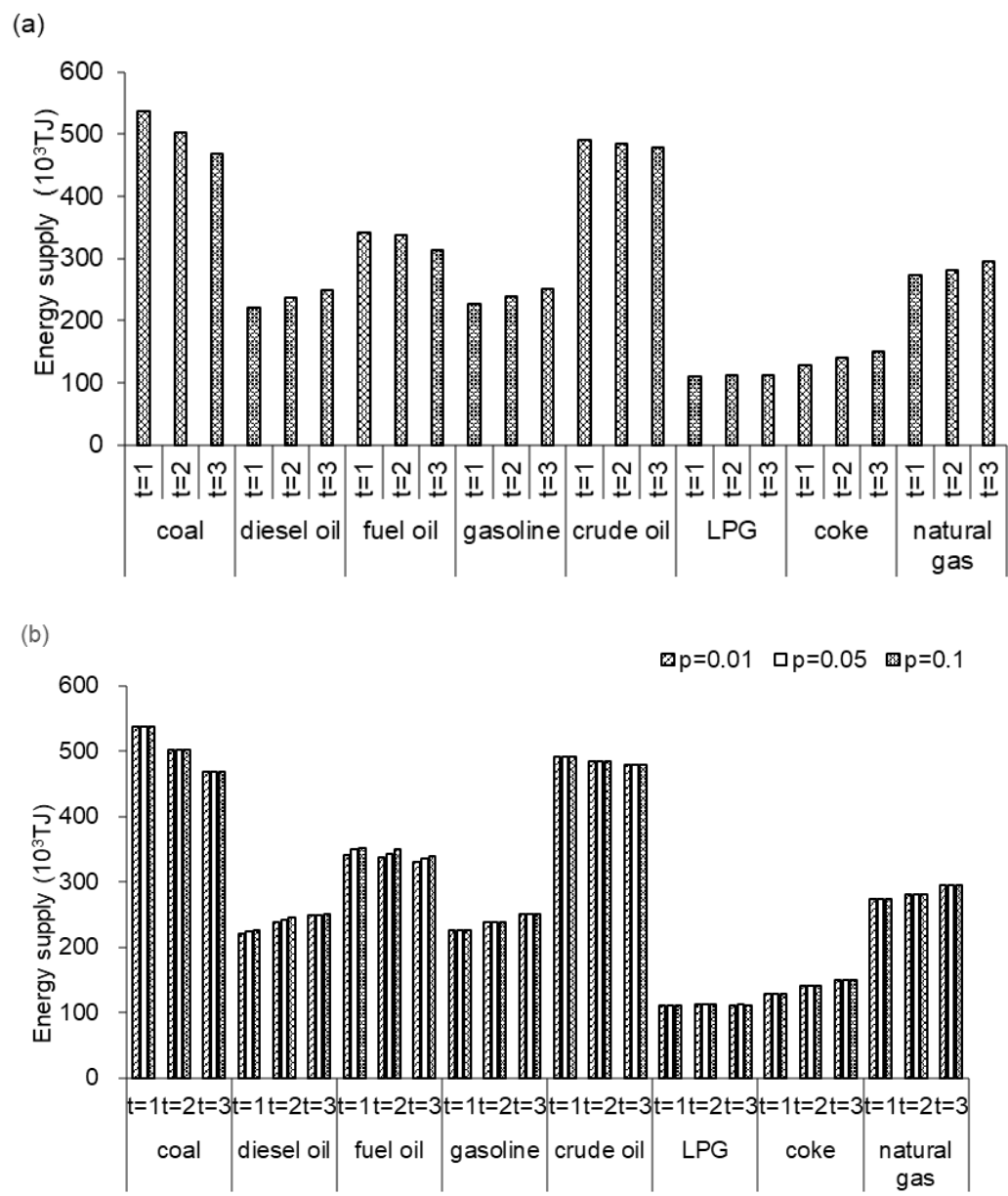


图 1 能源供应的解决方案（a）在不同的规划期间，（b）在不同的  $p$  水平下

图 1 给出了不同时期下的能源供应解决方案，其中包括 8 种能源类型（煤炭、柴油、燃料油、原油、液化石油气、焦炭和天然气）的能源供应结果。研究表明，除了煤炭和原油之外，其他能源的供应都随着时间的推移而增加，例如，煤炭供应将从第一时期的  $536.59 \times 10^3 \text{ TJ}$  减少为第三时期的  $469.03 \times 10^3 \text{ TJ}$ ，

这有可能是由于煤炭高污染排放的特点和污染物减排的要求相矛盾引起的。相反，天然气的供应将从  $273.35 \times 10^3 \text{TJ}$  增加到  $294.59 \times 10^3 \text{TJ}$ ，这意味着天然气将成为未来的主要能源供应之一，这样的增长可能是与提倡新能源汽车的使用以及提高清洁能源消费比重有关。

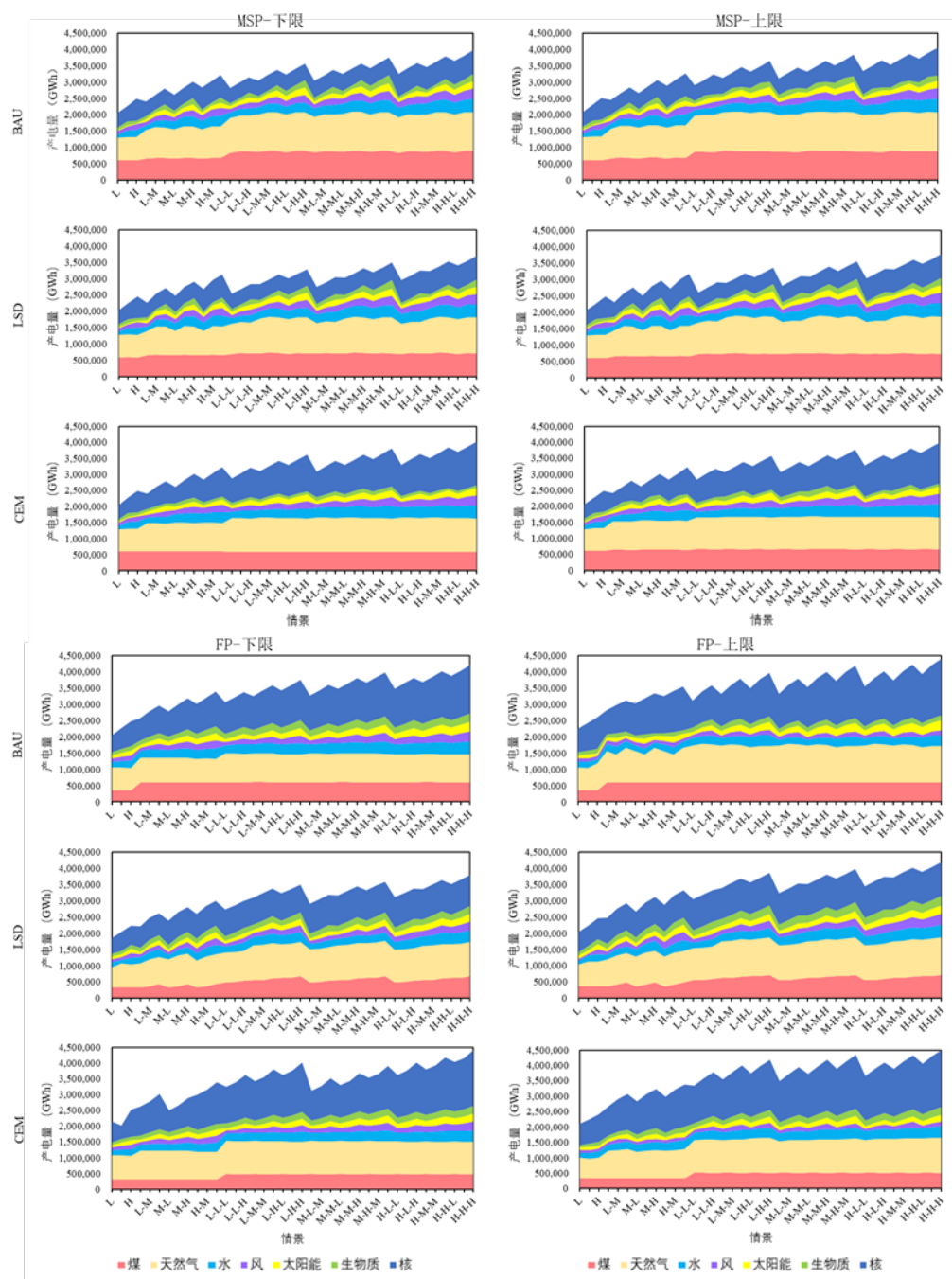


图 2 不同情景下发电计划的 FMSE 模型求解

珠三角作为全国经济发展的核心区域，该地区的电力需求快速增长，其电力供应面临着巨大挑战。图 2 展示了在规划时间跨度内，不同情景下珠三角地区不同能源发电技术的发电量。结果表明，由于人口增长和经济发展，在 FP 和 MSP 两种情景下，不同发电技术的发电量都将逐渐增加，以满足更高的电力需求。例如，FP-BAU 情景下，当第 3 期各时段需求水平较高（H-H-H）时，总发电量为 $[4206.94, 4406.69] \times 10^3 \text{GWh}$ ，而当第 1 期高（H）需求水平时，总发电量为 $[2483.63, 2596.30] \times 10^3 \text{GWh}$ 。同样，MSP-BAU 情景下，当第 3 期各时段需求水平较高（即 H-H-H）时，总发电量为 $[3970.53, 4053.35] \times 10^3 \text{GWh}$ ，而当第 1 期高（H）需求水平时，为 $[2483.63, 2508.47] \times 10^3 \text{GWh}$ 。根据结果可知，由于当地清洁能源大规模扩容，使得清洁发电单位成本最大化，FP 模型得到的两种方案的发电量均大于 MSP 模型得到的结果。因此，无论是 FP 模型还是 MSP 模型，清洁能源的发电量都随着需求水平的提高而增加，以满足未来较高的电力需求。与 FP 模型相比，MSP 中可再生能源和核电的总增量相对较低。例如，FP-BAU 情景下第 1 期需求水平由低向高变化时，生物质和核能发电量分别由 $[103.26, 103.61] \times 10^3$  和  $[528.23, 737.25] \times 10^3 \text{GWh}$  增加到  $[137.56, 138.61] \times 10^3$  和  $[756.11, 964.75] \times 10^3 \text{GWh}$ 。而在 MSP-BAU 情景下，当需求水平由低向高变化时，生物质和核能发电量分别由  $[86.11, 86.12] \times 10^3$  和  $[439.75, 439.80] \times 10^3 \text{GW}$  变化为  $[86.11, 86.11] \times 10^3$  和  $[661.71, 663.30] \times 10^3 \text{GWh}$ 。

## 案例 2：基于 MCMC-Copula 的多元水文系统风险分析

受水文系统复杂性、气候变化和人类活动的共同影响，水文水资源系统风险分析和管理中存在多重不确定性。传统的水文频率分析仅专注于极端水文事件的单一特征（如洪峰），忽略了多维风险（如洪峰、流量和持续时间）之间的互动关系。此外，随着我国水资源开发利用强度的增加和大量水利工程群的修建，传统的风险分析方法已无法反映多维风险的传递和累积。因此，必须在传统频率分析的基础上，同时考虑气候变化、人类活动以及其它多重不确定性要素对水文水资源系统的影响。

针对随机水文模拟中的多重不确定性，本研究基于多元 Copula 理论开发了多元水文耦合模拟及风险评估方法，并将其应用于极端水文事件的风险分析。该方法通过马尔可夫链蒙特卡洛（MCMC）算法量化模型边缘分布参数的不确定性，并基于后验分布评估联合重现期的不确定性。该方法还描述了双变量和多变量水文风险的概率分布特征。多元水文风险评估方法的开发包括以下步骤：（i）基于 Copula 方法建立风险分布；（ii）使用 MCMC 进行参数量化；（iii）确定条件分布与联合重现期；（iv）构建二元水文风险分析模型；（v）进行贝叶斯不确定性风险评估。

其中通过贝叶斯模型确定后验分布的过程如下：

步骤 1：使用最大似然法来估计边缘分布和 Copula 模型中参数的初始值。

步骤 2：计算边缘和联合概率的均方根误差（RMSE）和赤池信息量（AIC）值，以便选择最合适的边缘和联合分布。

步骤 3：根据在步骤 1 中获得的参数值设置先验分布。

步骤 4：将 MCMC 与 Metropolis-Hastings 算法结合使用，得出边缘分布和联合分布中参数的后验概率。

步骤 5：评估主重现期和联合重现期的不确定性。

步骤 6：揭示双变量水文风险的概率特征。

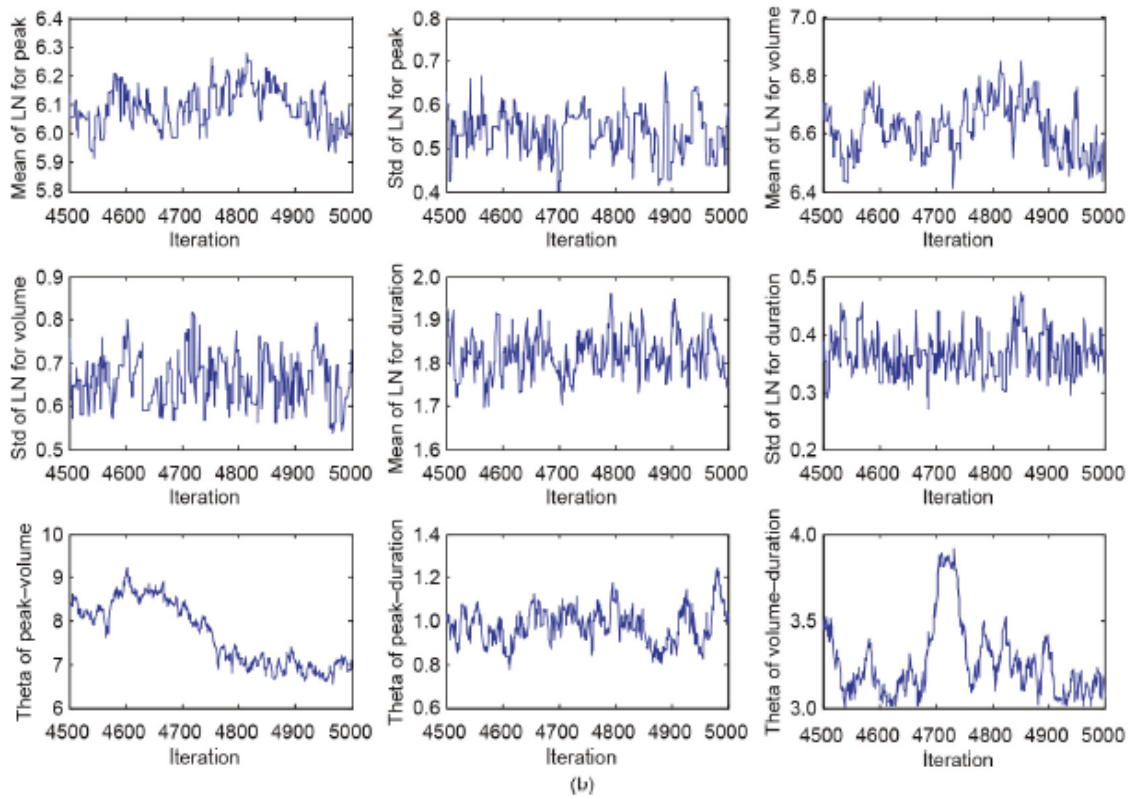
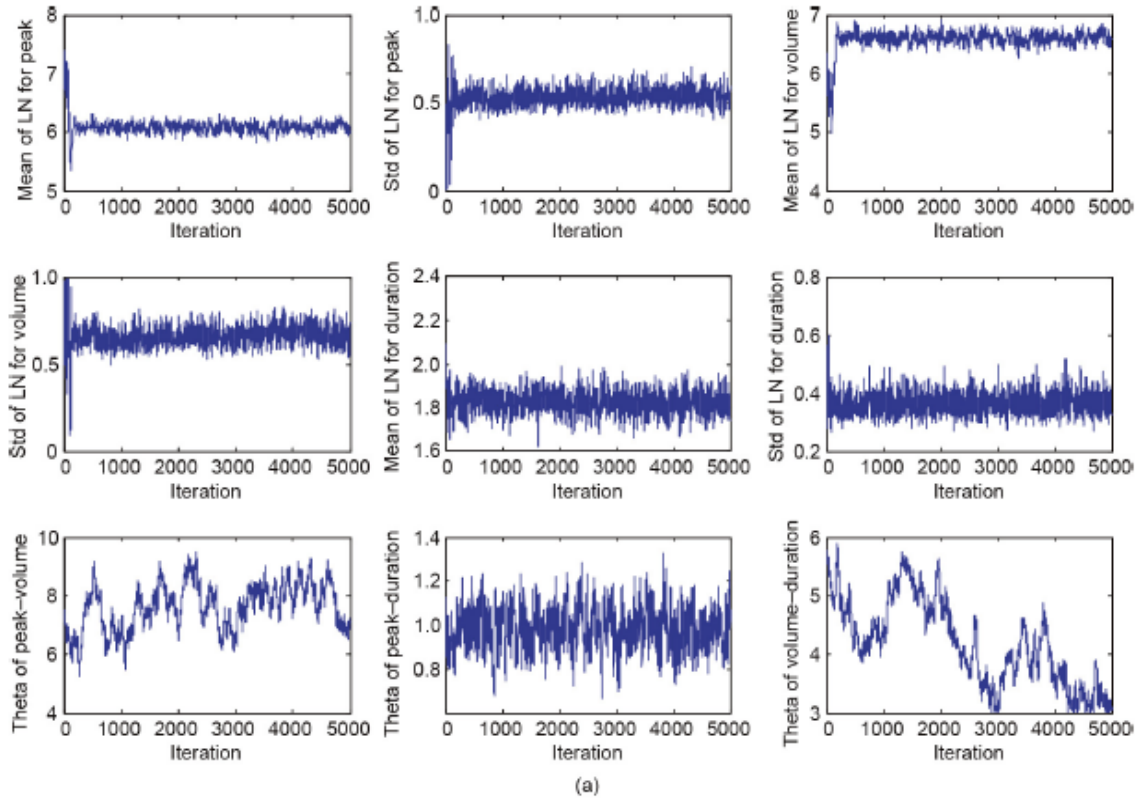


图 3. 模型中未知参数的演变：（a）MCMC 过程、（b）训练后。

本研究首先假设模型所有参数的先验分布为正态分布，其中正态分布的期

望值设置为通过最大似然估计（MLE）和矩估计（MOM）方法获得的初始值。然后在 MCMC 过程中进行 5000 次迭代，最后 500 次迭代被设置为 Copula 模型参数的后验参数。图 3 展示了 MCMC 过程中边缘分布和 Copula 模型中未知参数的演变。可以看出，使用贝叶斯方法估计的参数逐步收敛到稳定值。模型边缘分布和 Copula 函数中未知参数的后验分布如图 4 所示。

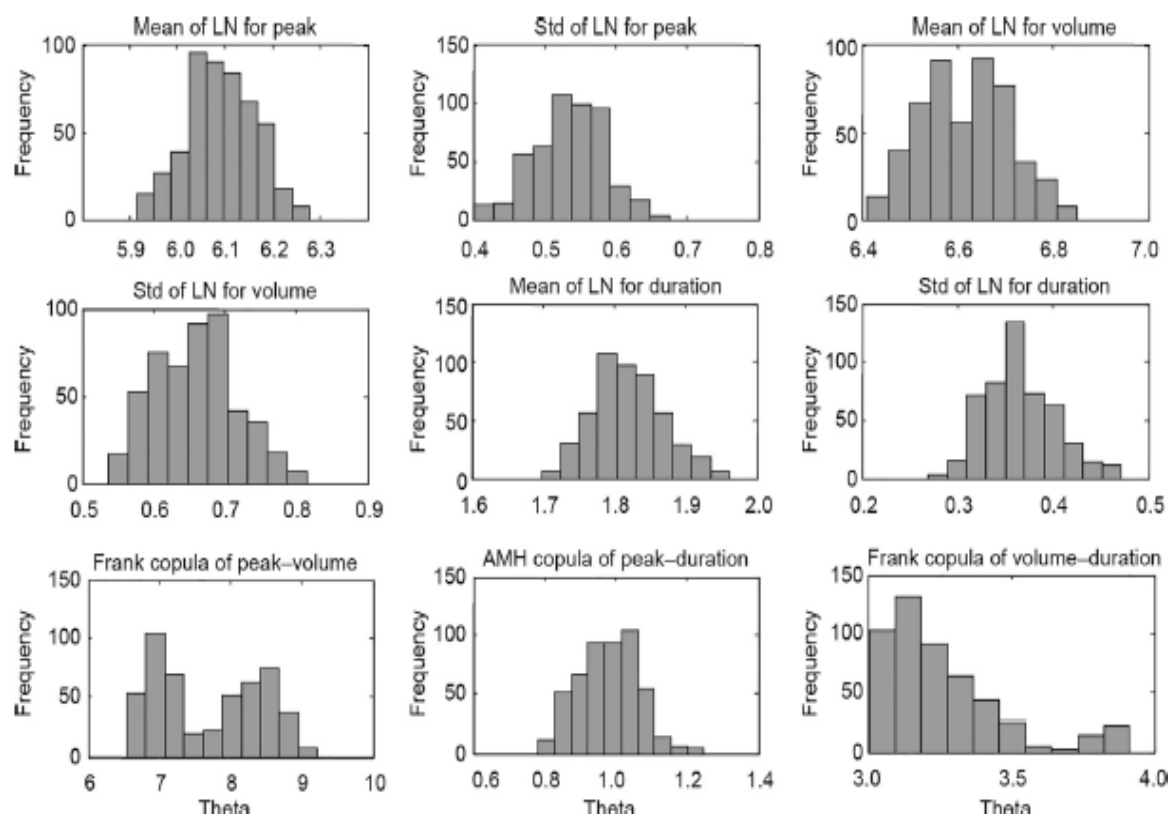


图 4. 边缘分布和 Copula 中未知参数的后验分布。

基于 MCMC 获得的参数抽样，可以获得设计值的样本分布，该方法不仅可以获得点估值，还可以提供其预测区间。通过贝叶斯方法获得的边缘分布和 Copula 函数中的参数以概率分布的方式呈现，可用来表征洪水变量的边缘和联合分布存在的不确定性。图 5 比较了洪水变量的拟合概率和观测概率，包含了 95% 的预测区间。该图表明，模型的预测区间总体包含了观测值。



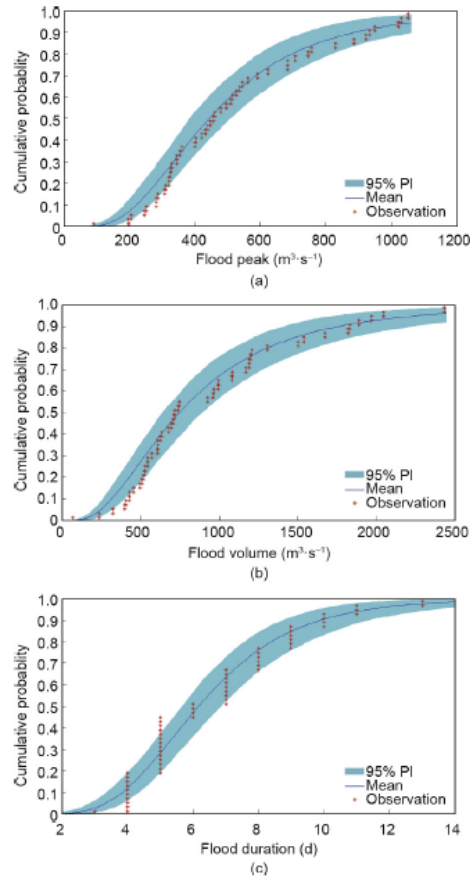


图 5. 比较洪水变量的拟合概率和观测概率：（a）洪峰、（b）洪水量、（c）洪水持续时间。

本研究还对洪水变量之间的相互作用及其对水文风险的影响进行了评估。对于洪水事件，水利工程所面临的风险主要是高峰值流量，所以洪水峰值流量是分析水文风险时必须考虑的基本因素。本研究分析了洪峰流量和洪峰持续时长的双变量洪水风险。图 6 展示了在不同设计流量和服务年限下，洪水风险相对于洪水量的变化趋势。对于所有设计流量和服务年限，洪峰流量和体积的双变量风险在一段时间内保持恒定，之后便随着洪灾持续时间的增加而降低。此外，由于边缘和联合分布的不确定性，水文风险也因此存在不确定性。根据这些参数的后验分布，可以得出双变量风险的平均值和 95% 的预测区间。图 6 表明，较低的设计流量和较长的服务年限会导致更大的水文风险不确定性。

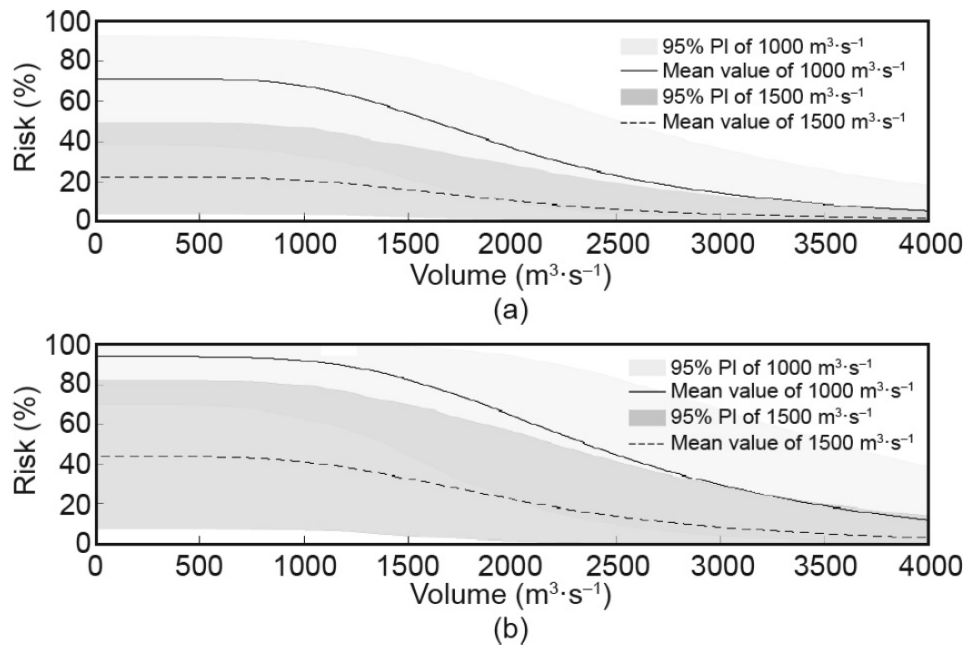


图 6. 不同洪峰情况下的双变量洪灾风险：（a）20 年服务年限、（b）50 年服务年限。

本研究将 MCMC 与 Copula 理论相结合，开发了多元水文风险评估方法，并对香溪河流域进行了耦合模拟和系统风险评估。结果表明，所开发的方法能有效地表征随机水文特征并量化随机水文过程中的不确定性，在洪水风险评估、径流频率分析以及工程水文设计等方面均表现良好。本研究的主要创新之处在于将贝叶斯推理引入到 Copula 框架中，以量化边缘分布和 Copula 结构的不确定性，并表征边缘和联合分布的不确定性之间的相互作用。研究发现香溪河洪水变量的边缘分布可用对数正态分布表示，所获得的预测区间可涵盖观测值。研究还发现，较低的洪水设计流量和较长的工程使用年限可导致双变量水文风险。本研究实现了对多重水文风险的辨识和表征，尤其是未来水文极端事件的多重随机特征。所提出的方法不仅能够反映多种水文变量自身的随机不确定性，而且涵盖了多变量间关系的多层次交互作用，可为复杂条件下水文水资源系统管理提供科学依据。

