

行政院國家科學委員會專題研究計畫 成果報告

以聯結函數探討雙變數水文事件之機率分佈

計畫類別：個別型計畫

計畫編號：NSC94-2211-E-032-009-

執行期間：94年08月01日至95年07月31日

執行單位：淡江大學水資源及環境工程學系

計畫主持人：蕭政宗

計畫參與人員：楊志傑、楊欣怡、黃亮芸、謝瑋珍

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 8 月 1 日

一、計畫中文摘要

水文事件之歷程通常極為複雜，但由於其對人類社會的影響甚鉅，因此許多水資源工程的規劃與設計均會利用水文事件的某些特性作為規劃與設計的依據，最常見的即為洪水頻率分析。然而，近年來許多的研究均指出使用單一變數並不足以完整描述一複雜的水文歷程，較佳的處理方式便是經由多變數聯合機率分佈(multivariate distribution)來研究水文事件，但此需要繁複的理論分析且適用的模式並不多，另外亦需要較多的數據來推估參數及驗證模式。因此，較常用的方法便是利用較易處理的雙變數分佈(bivariate distribution)來研究水文事件中較重要的兩個特性。常用的雙變數分佈模式有雙變數常態分佈、雙變數指數分佈、雙變數 gamma 分佈及雙變數極端值分佈等，但此類模式的邊際分佈(marginal distribution)皆為一樣，因此無法應用於具有關聯性但邊際分佈不一致的雙變數水文事件。

聯結函數(copula)的使用則可避免此一困難，聯結函數可聯結數個單變數分佈函數而構成多變數分佈函數。使用聯結函數建立多變數分佈函數之優點在於其能分離變數間之相關性與邊際分佈間的關係，因此，在已知邊際分佈的條件下，建立多變數分佈函數僅為探討變數間之相關性而已。聯結函數為 Sklar 於 1959 年所發展，雖然已歷經四十餘年的發展，但較少應用於實務研究，惟近十年來則廣被用於保險、金融、精算等領域，至於在水文方面的應用至近幾年才有學者加以研究。本計畫之目的即為利用聯結函數來建立具有關聯性之水文事件雙變數機率分佈，而聯結函數的參數推估將採用邊際推理函數法(the method of inference function for margins, 簡稱為 IFM)來推估，此法為二階段分離的推估方式，首先以最大概似法(the method of maximum likelihood)推估邊際分佈函數的參數，將所推估得之參數代入聯結函數內，而後再以最大概似法推估聯結函數的參數。建立水文事件之雙變數機率分佈後，即可進行雙變數頻率分析，此可獲得傳統單變數頻率分析所無法獲取之資訊，對複雜的水文歷程可有較深入且完整之瞭解，並對水資源工程之規劃與設計有所助益。

關鍵詞：聯結函數、邊際推理函數法、雙變數分佈、參數推估、最大概似法。

二、計畫英文摘要

Complicities inherent in hydrologic processes have made water resources engineering employing some hydrologic characteristics as planning and design criteria. One of the most common practices is the flood frequency analysis. Complex hydrologic events, such as droughts and floods, appear to be multivariate events that are characterized by a few correlated random variables. The multivariate analysis of hydrologic events because considerably more required data, sophisticated mathematical treatment, and the limited number of available models. A bivariate distribution is thus more common and easier to employ in describing the correlated hydrologic variables, e.g., bivariate normal distribution, bivariate exponential distribution, bivariate gamma distribution, bivariate extreme value distribution, etc.. However, such models cannot be applied to model correlated hydrologic variables with different marginal distributions.

Multivariate distribution construction using copulas can overcome such difficulties. Copulas are functions that join univariate distribution functions to form multivariate distribution functions. Copulas are able to model the dependence structure among random variables independently of the marginal distributions. Over four decades after Sklar (1959) introduced copulas, they have been used in insurance, finance, and hydrology recently. This study aims to construct a bivariate distribution of hydrologic events using two-dimensional copulas. To estimate the parameters in copulas, the method of inference function for margins or the IFM method is employed. Two separate estimation procedures are involved in this method. The maximum likelihood estimations of the univariate marginal distribution are performed first, then followed by an estimation of the copula dependence parameters. In addition to construct the bivariate distribution of hydrologic events, the bivariate frequency analysis is also performed to derive the needed information. Results obtained from this study can improve understanding of complex hydrologic processes and enhance design safety criterion of hydraulic structures.

Key words: copula, the method of inference function for margins (IFM), bivariate distribution, parameter estimation, the method of maximum likelihood.

三、報告內容

(一) 前言

水文事件如洪水或乾旱，其歷程通常極為複雜，但由於其對人類社會的影響甚鉅，例如，洪災常帶來居民生命的傷亡及巨額的財產損失，而乾旱所產生水源不足的問題更會影響到糧食的生產、生活品質的惡化及經濟發展的遲滯等。因此，水資源工程一般均會利用水文事件的某些特性來作為工程規劃與設計的依據。例如，最常見的洪水頻率分析，以歷年的年最大洪峰進行頻率分析，選定合適迴歸期的洪峰量作為決定堤防高度、溢洪道排水容量等工程的依據，其他如以降雨強度決定市區下水道的排水容量、以乾旱量作為水庫容量設計的依據等，皆是利用水文事件的特性作為水資源工程規劃設計基準的實例。

傳統的水資源規劃與設計以水文事件單一變數之單變數分析(univariate analysis)為主，如前述即是，因此在水文分析中單變數的機率分佈及相關的頻率分析有較完整的理論發展及應用實例(Kite, 1988; Stedinger 等, 1993)。但近年來有些學者認為使用單一變數並不足以完整描述一複雜的水文歷程，例如，Goel 等(1998)及 Yue 等(1999)即指出需利用包括洪峰(peak)、洪水體積(volume)、延時(duration)等變數方足以完整的描述一洪水歷程；Yevjevich (1967)也指出乾旱的研究必須考慮乾旱延時(duration)、乾旱量(magnitude)等。因此，對於複雜的水文歷程較佳的處理方式便是經由多變數聯合機率分佈(multivariate distribution)來研究水文事件，但此需要繁複的理論分析且適用的模式並不多，另外亦需要較多的數據來推估參數及驗證模式。因此實務應用的情況並不多見，較常採用的方法便是利用較易處理的雙變數分佈(bivariate distribution)來研究水文事件中較重要的兩個特性。例如，Gupta 等(1976)發展聯合機率分佈來描述洪峰及其發生時間，Singh 及 Singh(1991)利用雙變數指數機率分佈來研究降雨強度及降雨量，Bacchi 等(1994)利用雙變數指數機率分佈來研究降雨延時及降雨強度，Goel 等(1998)利用雙變數常態機率分佈來研究洪峰量及洪水體積，Yue 等(1999)利用雙變數極端值機率分佈來研究洪峰量、洪水體積及洪水延時之間的關係。

在探討洪水事件較常用的雙變數分佈模式有雙變數常態分佈(bivariate normal distribution)、雙變數指數分佈(bivariate exponential distribution)、雙變數 gamma 分佈(bivariate gamma distribution)、及雙變數極端值分佈(bivariate extreme value distribution)。前述常用雙變數分佈模式的特徵為其邊際分佈(marginal distribution)皆為一致，因此，此類雙變數分佈並無法應用於具有關聯性但邊際分佈不一致的雙變數水文事件，例如年最大洪水事件之洪峰量通常可以極端值分佈來代表，而相對之洪水體積可能以 gamma 或 lognormal 分佈有較佳的代表性，即洪水歷程由具極端值分佈之洪峰及具 gamma 或 lognormal 分佈之洪水體積來代表，因此前述水文分析中常用具有相同邊際分佈之雙變數模式便不適用。

聯結函數(copula)的使用則可避免此一困難，聯結函數可聯結數個單變數分佈函數(univariate distribution function)而構成多變數分佈函數(multivariate distribution function)，即在已知邊際分佈的條件下，利用聯結函數可建立一具有已知邊際分佈的多變數分佈。聯結函數為 Sklar 於 1959 年所發展，雖然已歷經四十餘年的發展，但較少應用於實務研究，惟近十餘年來則廣被用於保險、金融、精算等領域，至於在水文方面的應用則至近幾年才有學者加以研究。

(二) 研究目的

本計畫之目的為利用聯結函數(copula)建立水文事件中具關聯性二變數之雙變數機率分佈，並進行雙變數之頻率分析。由於甚多水文歷程需由具有相關聯之二變數才能有完整的描述，而此相關聯之變數通常不具有相同的分佈，因此，傳統的雙變數分佈無法模擬此

一現象。聯結函數則可聯結二個不同的邊際分佈而成為雙變數分佈，本計畫將選定一需要雙變數來描述之水文事件，先決定各變數之單變數分佈，即雙變數分佈之邊際分佈，再由聯結函數來建立該水文事件之雙變數機率分佈。由於聯結函數的形式有許多種，本計畫將另探討不同聯結函數之適用性。而聯結函數之參數推估，本計畫將採用邊際推理函數法(the method of inference function for margins, 簡稱為 IFM)來推估，此法為二階段分離的推估方式，首先以最大概似法(the method of maximum likelihood)推估邊際分佈函數的參數，將所推估得之參數代入聯結函數內，而後再以最大概似法推估聯結函數的參數。建立水文事件之雙變數機率分佈後，即可進行雙變數頻率分析，此可獲得傳統單變數頻率分析所無法獲取之資訊，可對複雜的水文歷程有較深入且完整之瞭解，對水資源工程之規劃與設計亦有所助益。

(三) 文獻探討

文獻上有關雙變數機率分佈的探討有雙變數常態分佈(bivariate normal distribution)(Goel 等, 1998; Yue, 1999)、雙變數指數分佈(bivariate exponential distribution)(Singh 及 Singh, 1991; Bacchi 等, 1994)、雙變數 gamma 分佈(bivariate gamma distribution)(Yue 等, 2001)、及雙變數極端值分佈(bivariate extreme value distribution)(Yue 等, 1999)

聯結函數(copula)近十餘年來則廣被用於保險、金融、精算等領域，例如 Frees 等(1996)利用聯結函數建立年金與死亡率間之關係；Frees 及 Valdez(1998)與 Klugman 及 Parsa(1999)則用於探討保險公司理賠與損失間之關聯性；Hürlimann(2004)應用聯結函數建立單一股票累積盈餘與股票指數間之聯合分佈；Mendes 及 Souza(2004)則應用於建立二不同地區股票指數間之關聯性。至於在水文方面的應用則至近幾年才有學者加以研究，De Michele 及 Salvadori(2003)及 Salvadori 及 De Michele (2004a)應用聯結函數於探討降雨強度與延時之關聯性；而 Favre 等(2004) 及 Salvadori 及 De Michele (2004b)則用於探討洪峰量與洪水體積之關聯性。

(四) 研究方法

I. 聯結函數(copula)

多變數分佈的發展從早期多變數常態分佈(multivariate normal distribution)開始，逐漸拓展到其他分佈，但多半是由單變數分佈(univariate distribution)直接延伸而成雙變數分佈(bivariate distribution)，例如雙變數指數分佈或雙變數 gamma 分佈等(Kotz 等, 2000)。Frees 及 Valdez(1998)指出此類的分佈應用上有以下的缺點：(1)其邊際分佈必須是一致的；(2)其延伸至多變數分佈(multivariate distribution)的方式不明確；(3)變數間之相關性會包括在邊際分佈內。由於有這些缺點，其應用上便受到很大的限制，但 1959 年 Sklar 所發展的聯結函數(copula)便可免除這些限制。

聯結函數能聯結數個單變數分佈函數而成為多變數分佈函數，使用聯結函數建立多變數分佈函數之優點在於其能分離變數間之相關性與邊際分佈間的關係，因此，在已知邊際分佈的條件下，建立多變數分佈函數僅為探討變數間之相關性而已。本計畫即為利用聯結函數來建立水文事件之雙變數分佈，因此，以下所述均為雙變數的情況。

Sklar 定理敘述當 $F_X(x)$ 及 $F_Y(y)$ 為雙變數分佈函數 $F_{X,Y}(x,y)$ 之邊際分佈時，存在有一聯結函數 C 使

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)) \quad (1)$$

相反的，對任何的單變數分佈函數 $F_X(x)$ 及 $F_Y(y)$ 與聯結函數 C 而言，前式所定義之 $F_{X, Y}(x, y)$ 即為一具邊際分佈函數 $F_X(x)$ 及 $F_Y(y)$ 之雙變數分佈函數，且若 $F_X(x)$ 及 $F_Y(y)$ 為連續時，則聯結函數 C 為惟一。

依據Sklar定理，在邊際分佈函數均為連續且其機率密度函數 (probability density function, pdf) 為 $f_X(x)$ 及 $f_Y(y)$ 的情況下，則雙變數的聯合機率密度函數可表為：

$$f_{X, Y}(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y) \quad (2)$$

其中 c 為聯結函數 C 的密度函數，定義為

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (3)$$

其中 u 及 v 代表具有相關性之二變數，即分別代表 $F_X(x)$ 及 $F_Y(y)$ ，因此其範圍在 0 與 1 之間。

但並非所有的函數均能作為聯結函數，由於聯結函數為雙變數分佈函數，因此其數值範圍必須在 0 與 1 之間。除此之外，聯結函數還必須具有以下特性 (Joe, 1997; Nelsen, 1999):

$$C(u, 1) = u \quad , \quad C(1, v) = v \quad (4)$$

$$C(u, 0) = 0 \quad , \quad C(0, v) = 0 \quad (5)$$

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) > 0 \quad \text{當 } 0 \leq u_1 \leq u_2 \leq 1 \text{ 且 } 0 \leq v_1 \leq v_2 \leq 1 \quad (6)$$

因聯結函數 $C(u, v)$ 對二變數 u 及 v 而言，必須是遞增且連續的，因此還需滿足下式：

$$0 \leq \frac{\partial C(u, v)}{\partial u} \leq 1 \quad , \quad 0 \leq \frac{\partial C(u, v)}{\partial v} \leq 1 \quad (7)$$

II. 常用之一參數二維聯結函數

雖然聯結函數必須符合上述的限制，仍有許多種形式，但較為常用之單一參數聯結函數 C 及其相對應之密度函數 c 有以下幾類 (Joe, 1997; Frees 及 Valdez, 1998; Nelsen, 1999; Cherubini 等, 2004):

1. Ali-Mikhail-Haq

$$C(u, v) = \frac{uv}{1 - \theta(1-u)(1-v)}, \quad -1 \leq \theta \leq 1 \quad (8a)$$

$$c(u, v) = \frac{[1 - \theta(1-u)(1-v)](1-\theta) + 2\theta uv}{[1 - \theta(1-u)(1-v)]^3} \quad (8b)$$

2. Clayton

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \quad \theta \geq 0 \quad (9a)$$

$$c(u, v) = (\theta + 1)(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta} - 2} (uv)^{-\theta - 1} \quad (9b)$$

3. Farlie-Gumbel-Morgenstern

$$C(u, v) = uv[1 + \theta(1-u)(1-v)], \quad -1 \leq \theta \leq 1 \quad (10a)$$

$$c(u, v) = 1 + \theta(1-2u)(1-2v) \quad (10b)$$

4. Frank

$$C(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right], \quad \theta \neq 0 \quad (11a)$$

$$c(u, v) = -\frac{\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{[e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta}]^2} \quad (11b)$$

5. Galambos

$$C(u, v) = uv \exp\{[(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{\frac{1}{\theta}}\}, \quad \theta \geq 0 \quad (12a)$$

$$c(u, v) = \frac{C(u, v)}{uv} \{1 - [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{\frac{1}{\theta}-1} [(-\ln u)^{-\theta-1} + (-\ln v)^{-\theta-1}] \\ + [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{\frac{1}{\theta}-2} [(-\ln u)(-\ln v)]^{-\theta-1} [1 + \theta + [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{\frac{1}{\theta}}]\} \quad (12b)$$

6. Gumbel-Hougaard

$$C(u, v) = \exp\{-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{\frac{1}{\theta}}\}, \quad \theta \geq 1 \quad (13a)$$

$$c(u, v) = C(u, v) \frac{[(-\ln u)(-\ln v)]^{\theta-1}}{uv} [(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{\frac{2}{\theta}-2} \{(\theta-1)[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{\frac{1}{\theta}} + 1\} \quad (13b)$$

7. Plackett

$$C(u, v) = \frac{1}{2} \frac{1}{\theta-1} \{1 + (\theta-1)(u+v) - [(1 + (\theta-1)(u+v))^2 - 4\theta(\theta-1)uv]^{\frac{1}{2}}\}, \quad \theta \geq 0 \quad (14a)$$

$$c(u, v) = [(1 + (\theta-1)(u+v))^2 - 4\theta(\theta-1)uv]^{\frac{3}{2}} \theta [1 + (\theta-1)(u+v-2uv)] \quad (14b)$$

其中 θ 為代表變數 u 及 v 間相關性之參數，其範圍隨聯結函數而異，亦列於上式中。

上述類型聯結函數相關的研究有 Frees 等(1996)利用 Frank 聯結函數建立年金與死亡率間之關係；Frees 及 Valdez(1998)曾應用 Gumbel-Hougaard、Frank、Clayton 等聯結函數於探討保險公司理賠與損失間之關聯性；Klugman 及 Parsa(1999)則應用 Clayton 聯結函數建立保險公司理賠與損失之之雙變數函數；Hürlimann(2004)曾比較 Gumbel-Hougaard、Galambos、Frank、Clayton 等聯結函數於建立單一股票累積盈餘與股票指數間聯合分佈之適用性；Mendes 及 Souza(2004)則應用 Farlie-Gumbel-Morgenstern、Galambos 等聯結函數建立二不同地區股票指數間之關聯性；De Michele 及 Salvadori(2003)及 Salvadori 及 De Michele (2004)應用 Frank 聯結函數探討降雨強度與延時之關聯性；而 Favre 等(2004)則應用 Farlie-Gumbel-Morgenstern、Frank、Clayton 等聯結函數於探討洪峰量與洪水量之關聯性。

上述類型的聯結函數並非均能應用於所有的情況，例如 Farlie-Gumbel-Morgenstern 聯結函數僅能模擬二變數間相關係數(correlation coefficient)介於 $-1/3$ 與 $+1/3$ 之間的情況(Joe, 1997)。因此本計畫除以聯結函數建立水文事件之雙變數分佈外，亦將探討各聯結函數之適用性。

III. 聯結函數參數之推估

一般用於代表二變數間相關性之 Spearman 相關係數 ρ (Spearman's correlation coefficient)與聯結函數間之關係可表為(Joe, 1997)：

$$\rho = 12 \iint F_X(x)F_Y(y)dF_{X,Y}(x, y) - 3 = 12 \iint uv dC(u, v) - 3 \quad (15)$$

緣由此式可建立相關係數 ρ 與聯結函數參數 θ 間之關係，亦即可利用較易計算之相關係數 ρ 來推估參數 θ ，但並非所有類型聯結函數之 ρ 均可得到解析式，若無法得到解析式，並不容易利用此法來計算，因此本計畫不採用此法來推估參數。

聯結函數中用於代表變數間相關性的參數 θ 將以 Joe(1997)所建議的邊際推理函數法(the method of inference function for margins, 簡稱為 IFM)來推估。IFM 法為二階段分離的推估

方式，首先以最大概似法(the method of maximum likelihood)分別推估邊際分佈函數的參數，而後將所推估得之邊際分佈函數的參數代入聯結函數內，再以最大概似法推估聯結函數的參數，詳細過程說明如下。

假設 $f_X(x)$ 及 $f_Y(y)$ 分別為變數 X 及 Y 的機率密度函數，且均為二參數之機率密度函數，以 α 及 β 來代表 $f_X(x)$ 之參數，寫為 $f_X(x; \alpha, \beta)$ ；以 γ 及 δ 來代表 $f_Y(y)$ 之參數，寫為 $f_Y(y; \gamma, \delta)$ 。若有 n 組成對的觀測資料： (x_1, y_1) 、 (x_2, y_2) 、 \dots 、 (x_n, y_n) ，則變數 X 及 Y 的概似函數(likelihood function)可分別表為：

$$L_X(x; \alpha, \beta) = \prod_{i=1}^n f_X(x_i; \alpha, \beta) \quad (16a)$$

$$L_Y(y; \gamma, \delta) = \prod_{i=1}^n f_Y(y_i; \gamma, \delta) \quad (16b)$$

為便於求解，將上二式分別取對數，即成為對數概似函數(log-likelihood function)，可分別寫為：

$$\ln L_X(x; \alpha, \beta) = \sum_{i=1}^n \ln f_X(x_i; \alpha, \beta) \quad (17a)$$

$$\ln L_Y(y; \gamma, \delta) = \sum_{i=1}^n \ln f_Y(y_i; \gamma, \delta) \quad (17b)$$

將(17a)式最大化求取參數，即將(17a)式分別對參數 α 及 β 微分一次並令為0，聯立求解即為所需之參數，推估所得之參數以 $\hat{\alpha}$ 及 $\hat{\beta}$ 來代表；相同的步驟用於(17b)式所求得之參數以 $\hat{\gamma}$ 及 $\hat{\delta}$ 來代表。將所求得之邊際分佈函數參數 $\hat{\alpha}$ 、 $\hat{\beta}$ 、 $\hat{\gamma}$ 、 $\hat{\delta}$ 代入雙變數分佈函數之密度函數(2)式內，則雙變數分佈函數之概似函數如下：

$$L(x, y; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \theta) = \prod_{i=1}^n c(F_X(x_i; \hat{\alpha}, \hat{\beta}), F_Y(y_i; \hat{\gamma}, \hat{\delta})) f_X(x_i; \hat{\alpha}, \hat{\beta}) f_Y(y_i; \hat{\gamma}, \hat{\delta}) \quad (18)$$

將上式取對數，則雙變數分佈函數之對數概似函數可寫為：

$$\ln L(x, y; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \theta) = \sum_{i=1}^n \ln c(F_X(x_i; \hat{\alpha}, \hat{\beta}), F_Y(y_i; \hat{\gamma}, \hat{\delta})) + \sum_{i=1}^n \ln f_X(x_i; \hat{\alpha}, \hat{\beta}) + \sum_{i=1}^n \ln f_Y(y_i; \hat{\gamma}, \hat{\delta}) \quad (19a)$$

若以下式來代表聯結函數密度函數之對數概似函數，

$$\ln L_C(x, y; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \theta) = \sum_{i=1}^n \ln c(F_X(x_i; \hat{\alpha}, \hat{\beta}), F_Y(y_i; \hat{\gamma}, \hat{\delta})) \quad (20)$$

並將(20)式、(17a)式及(17b)式代入(19a)式內，則雙變數分佈函數之對數概似函數可改寫為：

$$\ln L(x, y; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \theta) = \ln L_C(x, y; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \theta) + \ln L_X(x; \hat{\alpha}, \hat{\beta}) + \ln L_Y(y; \hat{\gamma}, \hat{\delta}) \quad (19b)$$

由上式可知，雙變數分佈函數之對數概似函數由聯結函數密度函數之對數概似函數及二邊際分佈之對數概似函數組成，由於邊際分佈之參數均已求出，因此上式僅剩聯結函數之參數 θ 未知。將(19b)式最大化以求取參數，即對參數 θ 微分一次並令為0，即可求得聯結函數之參數 $\hat{\theta}$ ，將所求得之參數 $\hat{\theta}$ 代回聯結函數內即可得到雙變數機率分佈函數。

至於應選取何聯結函數則有賴觀測資料來判定，一般可以相同一組觀測資料就不同的聯結函數分別以IFM法求取其參數，Klugman及Parsa(1999)建議可以具有最大對數概似函

數值的聯結函數來代表該組觀測資料，本計畫將以此法來判斷何聯結函數最適合所選取之水文事件。

(五) 結果與討論

本計畫主要研究目的為利用聯結函數建立水文事件中具關聯性二變數之雙變數機率分佈並進行雙變數之頻率分析，因此將以水資源工程較常用於分析的洪水及乾旱事件為例說明之。

I. 個案研究 I：乾旱頻率分析

本計畫以位於台灣南部之烏山頭雨量站月雨量資料定義乾旱，並建立乾旱事件乾旱延時-乾旱量雙變數分佈以進行雙變數乾旱頻率分析。

I.1 乾旱資料

本計畫所使用的雨量資料為烏山頭雨量站 1932 至 2001 年共計 70 年的月雨量資料，首先將之轉換為標準化降雨指數(standardized precipitation index, SPI) (McKee 等, 1993)，而後定義 SPI 小於 0 為乾旱的開始，連續負值 SPI 為一乾旱事件，乾旱延時(drought duration)及乾旱量(drought magnitude)為本計畫探討乾旱事件的二項重要特性。在本計畫所使用 70 年的資料中總計有 194 個乾旱事件，其乾旱量與乾旱延時之關係如圖 1 所示。

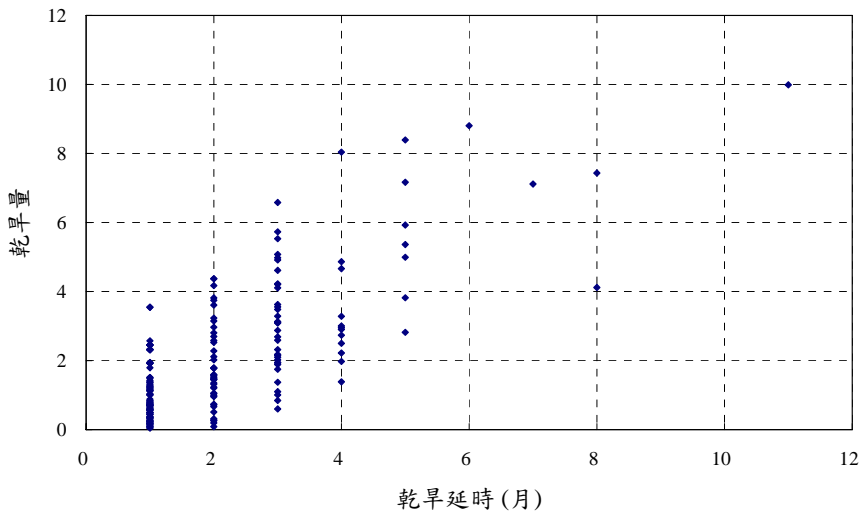


圖 1. 烏山頭雨量站乾旱事件之乾旱量-乾旱延時關係圖

I.2 乾旱量及乾旱延時之單變數機率分佈

許多學者都以指數分佈(exponential distribution)來代表乾旱延時(Zelenhastic 及 Salvai, 1987)，指數分佈的機率密度函數如下所示：

$$f_D(d) = \frac{1}{\lambda} e^{-d/\lambda}, \quad d > 0 \quad (21)$$

經以最大概似法推估參數得 $\hat{\lambda} = 1.529$ 。乾旱延時觀測值與理論指數分佈之關係如圖 2 所示，二者非常接近，亦即指數分佈可充分的描述乾旱延時的特性。

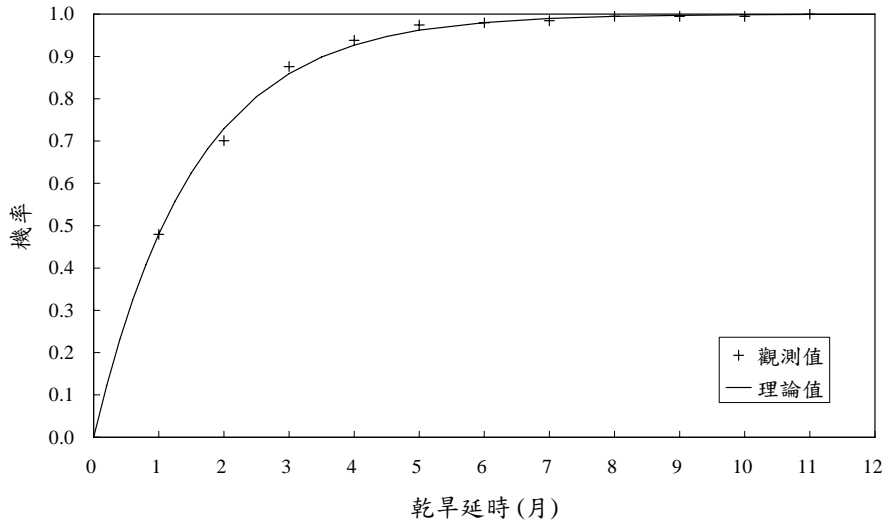


圖 2. 理論指數分佈與乾早延時觀測值關係圖

乾早量通常以 gamma 分佈來代表(Zelenhastic 及 Salvai, 1987; Mathier 等, 1992; Shiau 及 Shen, 2001), gamma 分佈之機率密度函數如下所示:

$$f_s(s) = \frac{s^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-s/\beta}, \quad s > 0 \quad (22)$$

其參數以最大概似法推估得 $\hat{\alpha} = 1.245$ 及 $\hat{\beta} = 1.640$ 。理論 gamma 分佈與乾早量觀測值之比較如圖 3 所示, 亦相當接近, 代表 gamma 分佈可用於模擬本計畫所使用之乾早量資料。

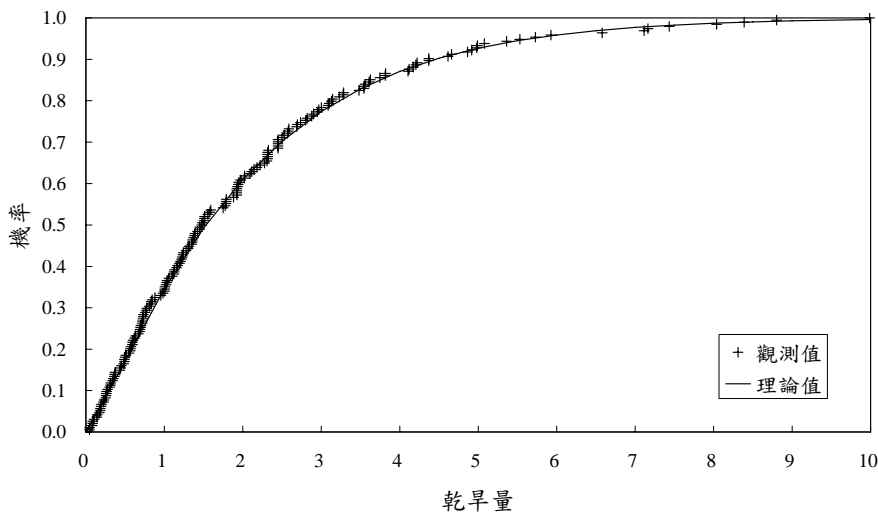


圖 3. 理論 gamma 分佈與乾早量觀測值關係圖

I.3 以聯結函數為基礎之乾早量-乾早延時雙變數機率分佈

由於乾早量及乾早延時不具有相同的邊際分佈, 常用之雙變數指數與雙變數 gamma 分佈便無法使用於上述的情況, 因此本計畫以聯結函數來建立雙變數乾早機率分佈。因 Farlie-Gumbel-Morgenstern 聯結函數僅能模擬相關係數介於 $-1/3$ 與 $+1/3$ 之間的情況, 而本計畫所使用之乾早量及乾早延時之相關係數為 0.756 , 因此無法用於代表乾早事件。

利用前述之 IFM 推估 Ali-Mikhail-Haq、Clayton、Frank、Galambos、Gumbel-Hougaard、Plackett 等六種聯結函數的參數, 其值列於表 1, 其相對之對數概似函數值亦列於表 1。圖 4 顯示此六種聯結函數的等值線, 亦即乾早量-乾早延時的聯合機率分佈。由表 1 可知

Galambos 聯結函數的對數概似函數值最大，因此本計畫以 Galambos 聯結函數來代表烏山頭雨量站乾旱量-乾旱延時的聯合機率分佈，並以此分佈進行雙變數乾旱頻率分析，其累積分佈函數如下所示：

$$F_{D,S}(d,s) = C(F_D(d), F_S(s))$$

$$= F_D(d)F_S(s)\exp\left[(-\ln F_D(d))^{-0.934} + (-\ln F_S(s))^{-0.934}\right]^{-1/0.934}$$
(23)

其中 $F_D(d)$ 及 $F_S(s)$ 為乾旱延時及乾旱量之單變數分佈函數，分別為指數及 gamma 分佈：

$$F_D(d) = 1 - e^{-d/1.529}$$
(24a)

$$F_S(s) = \int_0^s \frac{t^{0.245}}{1.640^{1.245} \Gamma(1.245)} e^{-t/1.640} dt$$
(24b)

表 1. 代表乾旱事件聯結函數之參數及對數概似函數值

聯結函數	參數	對數概似函數值
Ali-Mikhail-Haq	0.731	-663.5
Clayton	0.515	-669.2
Frank	3.830	-649.3
Galambos	0.934	-632.1
Gumbel-Hougaard	1.629	-633.3
Plackett	5.253	-649.5

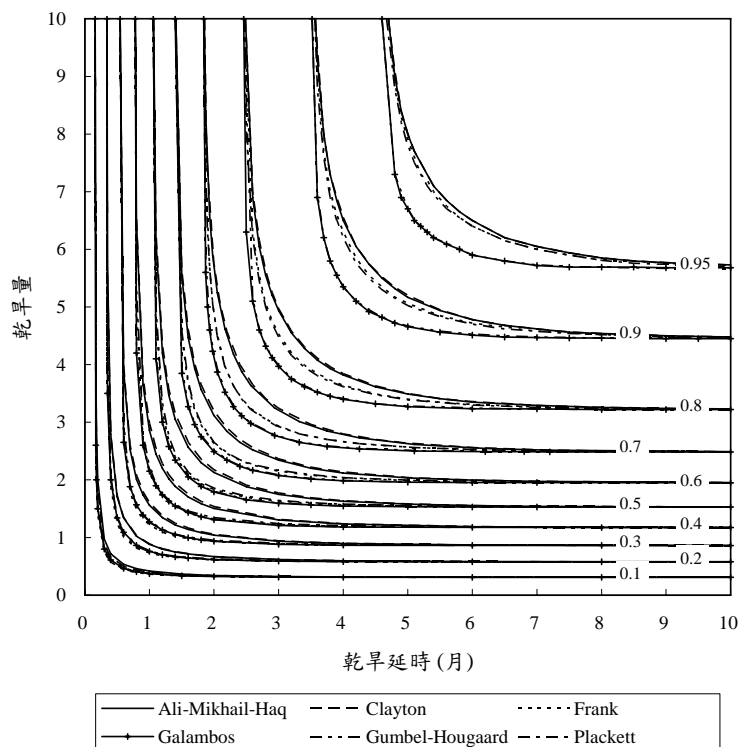


圖 4. 乾旱事件聯結函數等值線圖

I.4 乾旱量-乾旱延時雙變數乾旱頻率分析

頻率分析以迴歸期(return period)來代表，乾旱量及乾旱延時之單變數迴歸期與其分佈函數間之關係可表為：

$$T_s = \frac{E(L)}{1 - F_s(s)} \quad (20)$$

$$T_D = \frac{E(L)}{1 - F_D(d)} \quad (21)$$

其中 $E(L)$ 為平均乾旱間距，此值由觀測資料推估得 4.337 月，因此迴歸期為 2、5、10、20、50、100 年的乾旱量及乾旱延時分別列於表 2。

表 2. 不同迴歸期之乾旱量及乾旱延時

迴歸期 (年)	乾旱延時(月)	乾旱量
2	2.62	3.40
5	4.02	5.02
10	5.08	6.23
20	6.14	7.42
50	7.54	8.99
100	8.60	10.17

至於雙變數之迴歸期則可以下列二種方式定義 (Shiau, 2003)，(1) $D \geq d$ 且 $S \geq s$ 之迴歸期，及(2) $D \geq d$ 或 $S \geq s$ 之迴歸期，分別如下二式所示：

$$T_{DS} = \frac{E(L)}{P(D \geq d, S \geq s)} = \frac{E(L)}{1 - F_D(d) - F_S(s) + F_{DS}(d, s)} \quad (22)$$

$$= \frac{E(L)}{1 - F_D(d) - F_S(s) + C(F_D(d), F_S(s))}$$

$$T'_{DS} = \frac{E(L)}{P(D \geq d \text{ or } S \geq s)} = \frac{E(L)}{1 - F_{DS}(d, s)} = \frac{E(L)}{1 - C(F_D(d), F_S(s))} \quad (23)$$

上述二迴歸期分別如圖 5 及圖 6 所示。

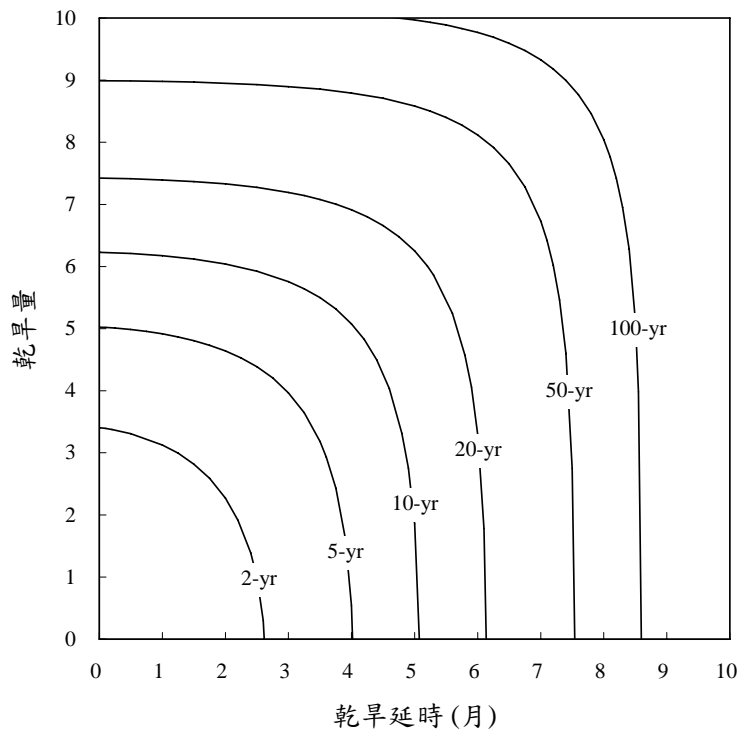


圖 5. 乾旱量-乾旱延時聯合迴歸期 T_{DS}

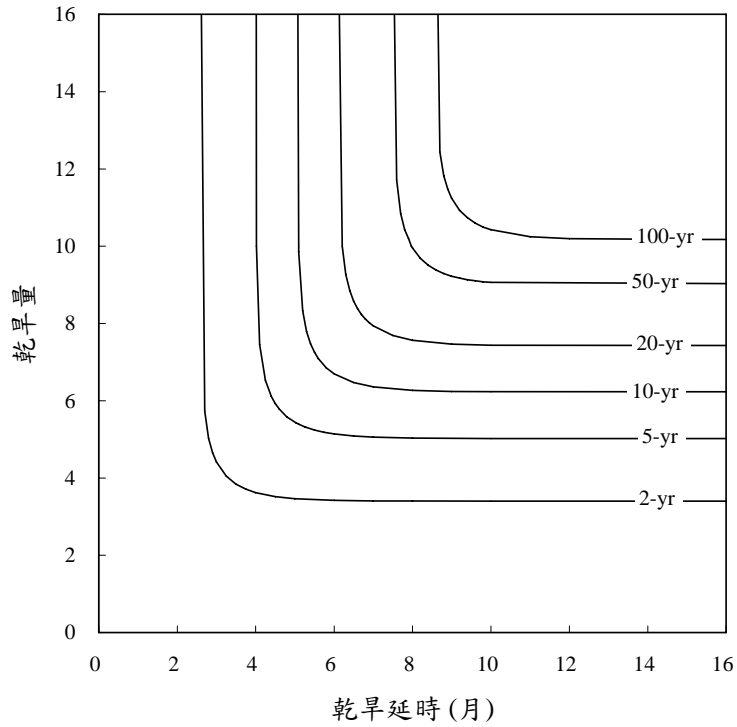


圖 6. 乾旱量-乾早延時聯合迴歸期 T'_{DS}

II. 個案研究 II：洪水頻率分析

本計畫以位於台灣中部濁水溪流域之桶頭(2)流量站之年最大洪水資料建立洪水事件之洪峰-洪水量雙變數分佈，並進行雙變數洪水頻率分析。

II.1 洪水資料

本計畫所使用的洪水資料為濁水溪流域之桶頭(2)流量站 1959 至 2002 年之年最大洪水資料，共計有 41 場洪水紀錄，其洪峰與洪水量之關係如圖 7 所示。

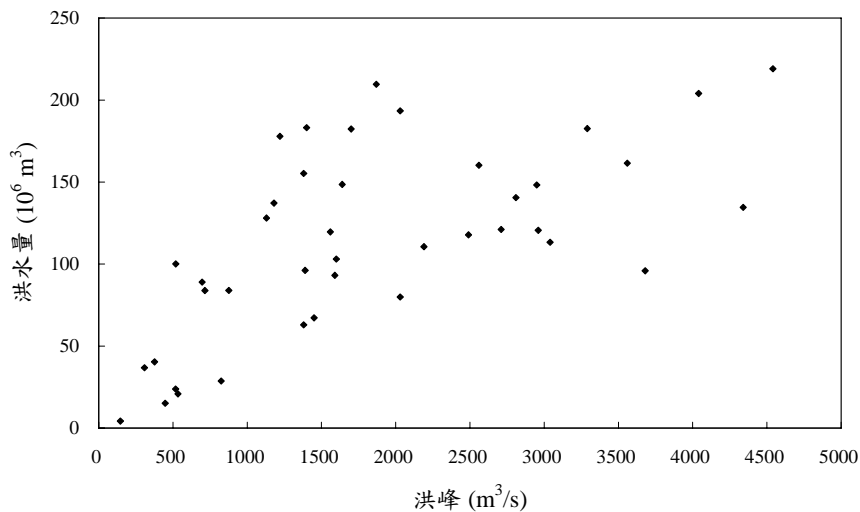


圖 7. 洪水事件之洪峰-洪水量關係圖

II.2 洪峰及洪水量之單變數機率分佈

洪峰可以 Gumbel 分佈來代表，其機率密度函數如下所示：

$$f_Q(q) = \frac{1}{\gamma} \exp\left(-\frac{q-u}{\gamma} - \exp\left(-\frac{q-u}{\gamma}\right)\right) \quad (24)$$

經以最大概似法推估參數得 $\hat{\gamma} = 919.981$ 及 $\hat{u} = 1303.334$ ，洪峰觀測值與理論 Gumbel 分佈之關係如圖 8 所示，由於甚為接近，因此 Gumbel 分佈可用於代表洪峰。

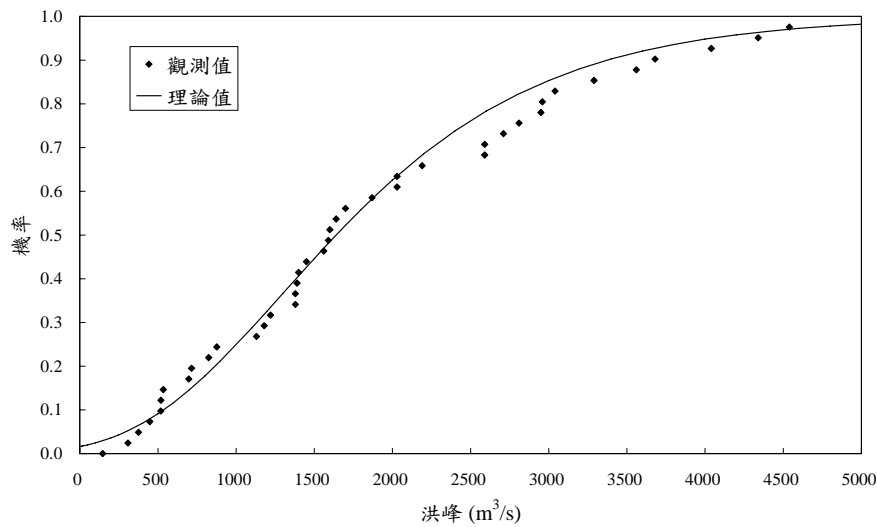


圖 8. 理論 Gumbel 分佈與洪峰觀測值關係圖

洪水量則以混合對數常態分佈(mixture lognormal distribution)來代表，其機率密度函數如下所示：

$$f_V(v) = p \left[\frac{1}{v\sigma_1\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln v - \mu_1}{\sigma_1}\right)^2\right) \right] + (1-p) \left[\frac{1}{v\sigma_2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln v - \mu_2}{\sigma_2}\right)^2\right) \right] \quad (25)$$

參數以最大概似法推估得 $\hat{p} = 0.778$ 、 $\hat{\mu}_1 = 4.861$ 、 $\hat{\sigma}_1 = 0.314$ 、 $\hat{\mu}_2 = 3.350$ 、 $\hat{\sigma}_2 = 0.897$ ，洪水量觀測值與理論混合對數常態分佈之關係如圖 9 所示，亦甚為接近。

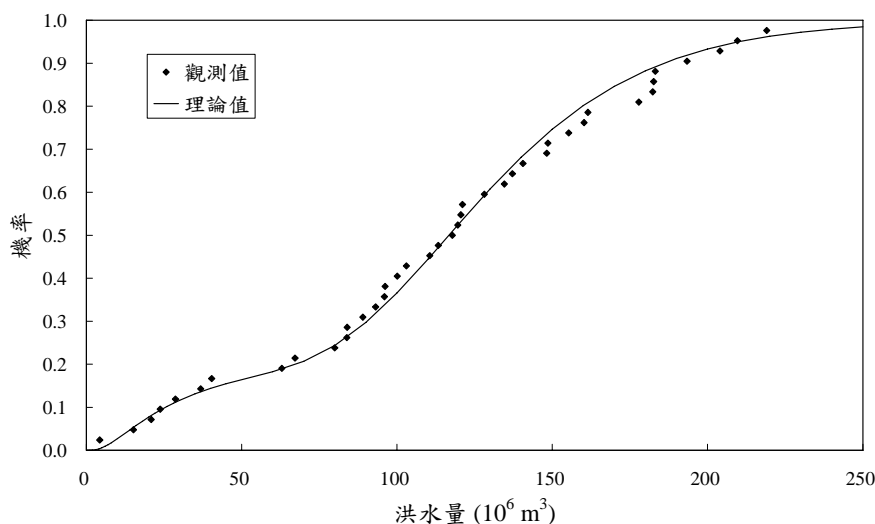


圖 9. 理論混合對數常態分佈與洪水量觀測值關係圖

II.3 以聯結函數為基礎之洪峰-洪水量雙變數機率分佈

由於洪峰及洪水量亦不具有相同的邊際分佈，常用之雙變數分佈便無法使用於上述的情況，因此本計畫以聯結函數來建立雙變數洪水機率分佈。Farlie-Gumbel-Morgenstern 聯

結函數亦無法用於代表洪水事件，因本計畫所使用洪水事件之洪峰及洪水量之相關係數為 0.640。

利用前述之 IFM 推估 Ali-Mikhail-Haq、Clayton、Frank、Galambos、Gumbel-Hougaard、Plackett 等六種聯結函數的參數，列於表 3，其對數概似函數值亦列於表 3。圖 10 顯示此六種聯結函數的等值線，即洪峰-洪水量的聯合機率分佈。由表 3 可知 Clayton 聯結函數的對數概似函數值最大，因此本計畫以 Clayton 聯結函數來代表桶頭(2)流量站洪峰-洪水量的聯合機率分佈，並以此分佈進行雙變數洪水頻率分析，其累積分佈函數如下所示：

$$F_{Q,v}(q, v) = \left[(F_Q(q))^{-1.579} + (F_V(v))^{-1.579} - 1 \right]^{-1/1.579} \quad (26)$$

其中 $F_Q(q)$ 及 $F_V(v)$ 為洪峰及洪水量之單變數分佈函數，分別為 Gumbel 及理論混合對數常態分佈：

$$F_Q(q) = \exp\left(-\exp\left(-\frac{q-1303.334}{919.981}\right)\right) \quad (27a)$$

$$F_V(v) = \int_0^v 0.778 \left[\frac{1}{0.787t} \exp\left(-\frac{1}{2} \left(\frac{\ln t - 4.861}{0.314} \right)^2 \right) \right] + 0.222 \left[\frac{1}{2.248t} \exp\left(-\frac{1}{2} \left(\frac{\ln t - 3.35}{0.897} \right)^2 \right) \right] dt$$

(27b)

表 3. 代表洪水事件聯結函數之參數及對數概似函數值

聯結函數	參數	對數概似函數值
Ali-Mikhail-Haq	0.996	-552.8
Clayton	1.579	-551.4
Frank	4.516	-555.8
Galambos	0.977	-556.2
Gumbel-Hougaard	1.700	-556.3
Plackett	6.818	-556.0

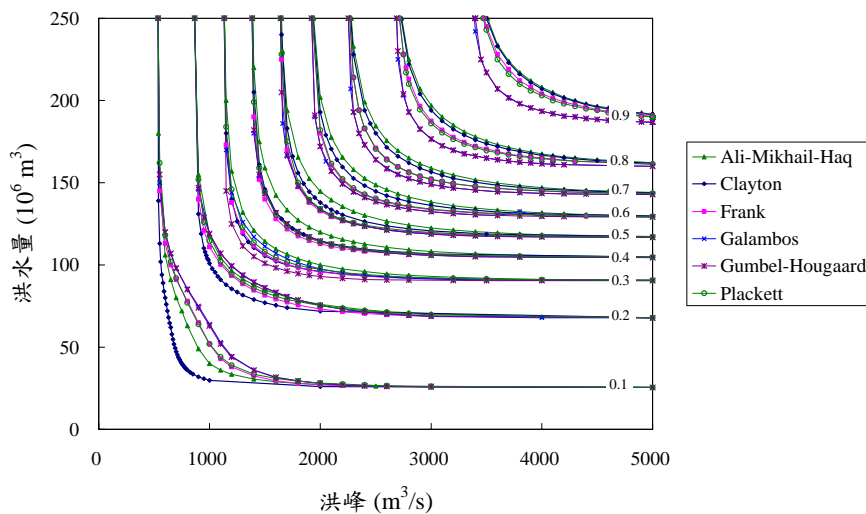


圖 10. 洪峰-洪水量聯結函數的等值線圖

II.4 洪峰-洪水量雙變數洪水頻率分析

洪峰及洪水量之單變數迴歸期與其分佈函數間之關係可表為：

$$T_Q = \frac{1}{1 - F_Q(q)} \quad (28)$$

$$T_V = \frac{1}{1 - F_V(v)} \quad (29)$$

迴歸期為 2、5、10、20、50、100 年的洪峰及洪水量分別列於表 4。

表 4. 不同迴歸期之洪峰及洪水量

迴歸期(年)	洪峰(m ³ /s)	洪水量(10 ⁶ m ³)
2	1640	116.6
5	2864	159.7
10	3374	185.8
20	4036	210.1
50	4893	241.3
100	5535	265.1

洪峰-洪水量雙變數迴歸期定義如下：

$$T_{QV} = \frac{1}{1 - F_Q(q) - F_V(v) + [(F_Q(q))^{-1.579} + (F_V(v))^{-1.579} - 1]^{-1/1.579}} \quad (30)$$

$$T'_{QV} = \frac{1}{1 - [(F_Q(q))^{-1.579} + (F_V(v))^{-1.579} - 1]^{-1/1.579}} \quad (31)$$

上述二迴歸期分別如圖 11 及圖 12 所示。

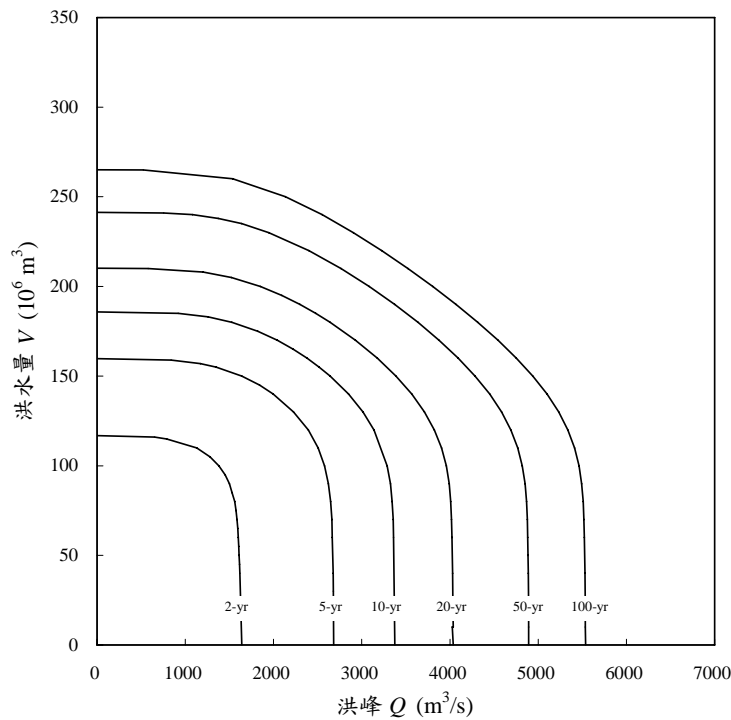


圖 11. 洪峰-洪水量聯合迴歸期 T_{DS}

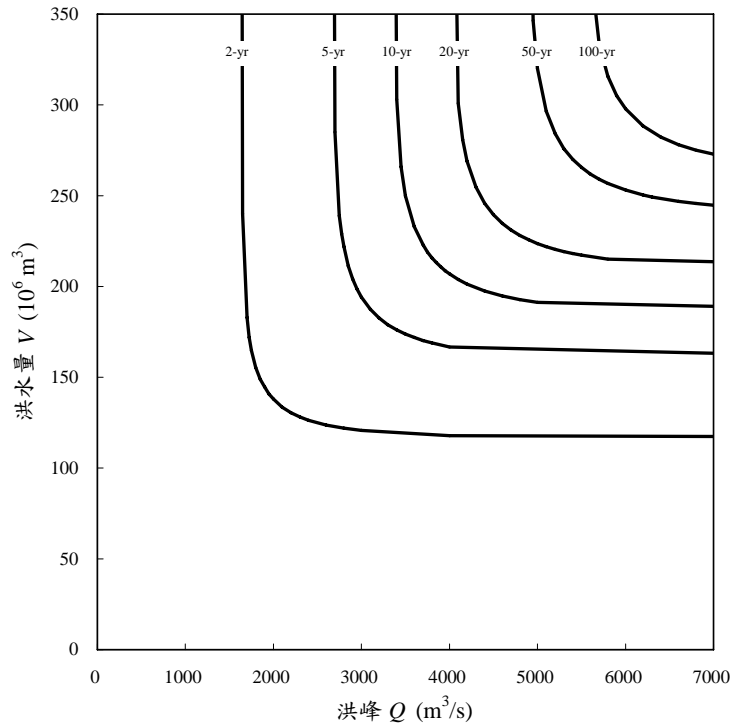


圖 12. 洪峰-洪水量聯合迴歸期 T'_{DS}

(六) 結論與建議

1. 本計畫以聯結函數(copula)建立水文事件雙變數機率分佈，不僅可適用於邊際分佈不一致的情況，更可以簡化建立雙變數機率分佈的計算過程。
2. 本計畫將聯結函數應用於水資源工程規劃與設計常碰到的極端水文事件，由於極端水文事件較為複雜，傳統之單變數分析不足以了解其過程，而且其邊際分佈通常不一致，利用聯結函數可避免此一困擾。
3. 本計畫以烏山頭雨量站之乾旱事件及桶頭(2)流量站之洪水事件為例，說明建立極端水文事件雙變數機率分佈之過程，並將之應用於雙變數頻率分析，演算結果顯示本計畫建議之方法可增加對極端水文事件的了解程度。

四、參考文獻

- Bacchi, B., Becciu, G. and Kottegoda, N. T.: 1994, 'Bivariate exponential model applied to intensities and durations of extreme rainfall', *Journal of Hydrology* **155**, 225-236.
- Cherubini, U., Luciano, E. and Vecchiato W.: 2004, *Copula Methods in Finance*, John Wiley & Sons, Chichester.
- De Michele, C. and Salvadori, G.: 2003, 'A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas', *Journal of Geophysical Research* **108**(D2), 4067, doi:10.1029/2002JD002534.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. and Rosso, R.: 2004, 'Bivariate statistical approach to check adequacy of dam spillway', *Journal of Hydrologic Engineering ASCE* **10**(1), 50-57.
- Favre, A. -C., El Adlouni, S., Perreault, L., Thiémonge, N. and Bobée, B.: 2004, 'Multivariate hydrological frequency analysis using copulas', *Water Resources*

Research **40**, W01101, doi: 10.1029/2003WR002456.

- Frees, E. W., Carriere, J. and Valdez, E.: 1996, 'Annuity valuation with dependent mortality', *Journal of Risk and Insurance* **63**(2), 229-261.
- Frees, E. W. and Valdez, E. A.: 1998, 'Understanding relationships using copulas', *North American Actuarial Journal* **2**(1), 1-25.
- Goel, N. K., Seth, S. M. and Chandra, S.: 1998, 'Multivariate modeling of flood flows', *Journal of Hydraulic Engineering ASCE* **124**(2), 146-155.
- Gupta, V. K. and Duckstein, L.: 1976, 'A stochastic analysis of extreme droughts', *Water Resources Research* **11**(2), 221-228.
- Hürlimann, W.: 2004, 'Fitting bivariate cumulative returns with copulas', *Computational Statistics and Data Analysis* **45**(2), 355-372.
- Joe, H.: 1997, *Multivariate Models and Dependence Concepts*, Chapman and Hall, New York.
- Kite, G.W., 1988. *Frequency and Risk Analysis in Hydrology*. Water Resources Publications, Littleton.
- Klugman, S. A. and Parsa, R.: 1999, 'Fitting bivariate loss distributions with copulas', *Insurance: Mathematics and Economics* **24**(1-2), 139-148.
- Kotz, S., Balakrishnan, N., and Johnson, N. L., 2000, *Continuous Multivariate Distributions*, John Wiley and Sons, Inc., New York.
- Mathier, L., Perreault, L., Bobe, B. and Ashkar, F.: 1992, 'The use of geometric and gamma-related distributions for frequency analysis of water deficit', *Stochastic Hydrology and Hydraulics* **6**(4), 239-254.
- McKee, T. B., Doesken, N. J. and Kleist, J.: 1993, 'The relationship of drought frequency and duration to time scales', *Proceedings of the 8th Conference on Applied Climatology*, 179-184.
- Mendes, B. V. M. and Souza, R. M.: 2004, 'Measuring financial risks with copulas', *International Review of Financial Analysis* **13**(1), 27-45.
- Nelsen, R. B.: 1999, *An Introduction to Copulas*, Springer-Verlag, New York.
- Salvadori, G. and De Michele, C.: 2004a, 'Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals', *Geophysical Resources Letters* **31**, L04502, doi:10.1029/2003GL018767.
- Salvadori, G. and De Michele, C.: 2004b, 'Frequency analysis via copulas: Theoretical aspects and applications to hydrological events', *Water Resources Research* **40**, W12511, doi:10.1029/2004WR003133.
- Shiau, J. T. and Shen, H. W.: 2001, 'Recurrence analysis of hydrologic droughts of differing severity', *Journal of Water Resources Planning and Management ASCE* **127**(1), 30-40.
- Shiau, J. T.: 2003, 'Return period of bivariate distributed hydrological events', *Stochastic Environmental Research and Risk Assessment* **17**(1-2), 42-57.
- Singh, K. and Singh, V. P.: 1991, 'Derivation of bivariate probability density functions with exponential marginals', *Stochastic Hydrology and Hydraulics* **5**(1), 55-68.
- Sklar, K.: 1959, *Fonctions de repartition à n dimensions et leurs marges*, Publ. Inst. Stat. Univ. Paris, 8, 229-231.
- Stedinger, J.R., R.M. Vogel, and E. Foufoula-Georgiou, 1993. *Frequency Analysis of Extreme Events*, in *Handbook of Hydrology*, Maidment, D. R. edited, McGraw-Hill Inc., New York.
- Yevjevich, V.: 1967, *An objective approach to definitions and investigations of continental*

hydrologic droughts, Hydrologic Paper No.23, Colorado State University, Fort Collins.

Yue, S.: 1999, 'Applying bivariate normal distribution to flood frequency analysis', *Water International* **24**(3), 248-254.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P. and Bruneau, P.: 1999, 'The Gumbel mixed model for flood frequency analysis', *Journal of Hydrology* **226**, 88-100.

Yue, S., Ouarda, T. B. M. J. and Bobée, B.: 2001, 'A review of bivariate gamma distributions for hydrological application', *Journal of Hydrology* **246**, 1-18.

Zelenhastic, E. and Salvai, A.: 1987, 'A method of streamflow drought analysis', *Water Resources Research* **23**(1), 156-168.

五、計畫成果自評

本計畫是國內首度以聯結函數建立水文事件雙變數機率分佈之研究，並將應用雙變數頻率分析，研究成果證實本計畫所建議方法之可行性，研究所得成果不僅達到預期目標，並於研究期間投稿國際期刊二篇論文，且已獲得接受，研究成果豐碩。