

國立臺灣大學共同教育中心統計碩士學位學程



碩士論文

Master Program in Statistics
Center for General Education
National Taiwan University
Master Thesis

事件最大值數列與混合分布在降雨頻率分析之應用
Rainfall Frequency Analysis Using Mixture Distribution
of Event-Maximum Rainfalls Series

陳柏宇

Bo-Yu Chen

指導教授：鄭克聲 博士

Advisor: Ke-Sheng Cheng, Ph.D.

中華民國 108 年 6 月

June, 2019



國立臺灣大學碩士學位論文
口試委員會審定書

事件最大值數列與混合分布在降雨頻率分析之
應用

Rainfall Frequency Analysis Using Mixture
Distribution of Event-Maximum Rainfalls Series

本論文係 陳柏宇 君 (學號 R06h41003) 在國立臺灣大學統計碩士學位學程完成之碩士學位論文，於民國 108 年 6 月 24 日承下列考試委員審查通過及口試及格，特此證明。

口試委員： 鄭克聲 (簽名)
(指導教授)

黃文政 蔡政宏
周吉明 _____



誌謝

大學與研究所的時光轉瞬即逝，非常感謝家裡的支持與鼓勵讓我可以心無旁鶩的投入課業與研究之中。此外，十分感謝鄭老師的栽培與教誨，從大三成為老師的專題生以來受到老師許多的照顧與幫助，老師認真並且嚴謹做學問的態度一直是學生學習的目標。也謝謝黃文政老師、盧孟明老師與蔡政安老師擔任我的口試委員，老師們的建議讓我明白自己的不足之處，也讓這份論文更加完善。另外，謝謝 RSLAB 的各位成員——家宜學姊、家穎、松勳、阿勳、登偉、卉敏與和家，和你們的討論讓研究的過程多了許多想法與樂趣。謝謝國術社的大家以及 02 各位的鼓勵，看到每個人為了自己的目標而努力就能讓人再次鼓起幹勁。謝謝學程的各位同學，與大家一起修課、討論的過程讓我獲益良多。碩士畢業是結束，也是開始，期許自己在未來仍然可以繼續前進，更上層樓。



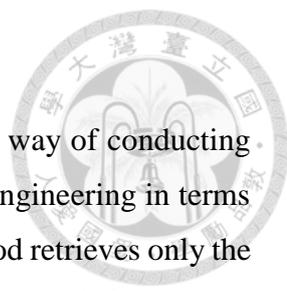
摘要

在水利工程與風險評估中降雨頻率分析是不可或缺的一步，其中年最大值數列法因為其簡單快速的步驟而最廣為使用。年最大值數列法以極端值理論 (extremal types theorem) 為基礎，使用廣義極端值分布 (Generalized Extreme Value Distribution, GEV) 配適年最大值數列並進行估計。然而年最大值數列法每年僅選取一筆資料的作法可能讓樣本數過少，不但增加估計的不確定性也使其容易受到極端值影響。除此之外，年最大值數列法在抽樣過程中沒有考慮降雨事件與降雨類型，因此實際上並不符合極端值理論的假設，不應以廣義極端值分布進行配適。考量以上因素，本研究提出基於降雨事件來進行估計的事件最大值數列法作為年最大值數列法的替代。事件最大值數列法將臺灣的降雨事件分為颱風、梅雨、對流雨和鋒面雨，各類型降雨的年最大值分布可以由事件數分布與事件降雨量分布混合而成，而考慮所有事件下的年最大值降雨分布則為各類型降雨年最大值分布的混合分布 (mixture distribution)。本研究首先以蒙地卡羅模擬比較年最大值數列法和事件最大值數列法所估計的設計降雨，結果顯示事件最大值數列法在偏誤 (bias) 和均方根誤差 (Root Mean Squared Error, RMSE) 的表現上均較佳。除此之外在五堵、頭汴坑與嘉義雨量站的頻率分析中，事件最大值數列法因為使用所有降雨事件進行估計而受極端值的影響較低，也不會選取跨事件樣本而高估設計降雨。尖峰流量分析結果顯示事件最大值數列法的年最大值降雨樣本之尖峰流量將大於等於年最大值數列法的樣本之尖峰流量，因此事件最大值數列法所估計的年最大值降雨更符合設計降雨的需求。兩種方法以拔靴法 (bootstrap) 所建立的信賴區間在最後進行模擬與實際資料比較，模擬結果顯示雖然均無法達到理論覆蓋率，事件最大值數列法的信賴區間覆蓋率仍然比年最大值數列法更接近理論值，並且隨著重現期 (return period) 而變化的程度較低。

關鍵字：年最大值數列、頻率分析、事件最大值數列、混合分布



ABSTRACT

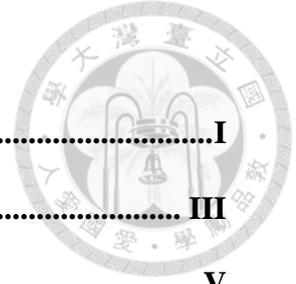


The Annual Maximum Series (AMS) method is a conventional way of conducting rainfall frequency analysis, which plays a crucial role in hydrology engineering in terms of hydrological risk assessment. Given any design duration, the method retrieves only the maximum rainfall within a year and approximate the Annual Maximum Rainfall (AMR) distribution by the Generalized Extreme Value (GEV) distribution according to the Extremal Types Theorem. However, the GEV approximation is inappropriate since AMS is prone to have insufficient sample size and does not take storm events and storm types into account. To overcome the above problems, the Event Maximum Series (EMS) method is proposed. The EMS method classifies storm events in Taiwan into Typhoon, Meiyu, frontal rain and convective storm. The AMR distribution of a given storm type can be derived from the corresponded event occurrence distribution and event rainfall distribution, and the AMR distribution of all events is a mixture distribution of different types of AMR distribution. As a result, the EMS method provides a more suitable and effective design-rainfall than the traditional approach. The EMS method outperforms the AMS approach in many ways. In Monte Carlo simulation, the EMS method is superior to AMS method in terms of the bias and Root Mean Squared Error (RMSE). Three stations in Taiwan are selected for frequency analysis and peak flow analysis, the results show that EMS method can avoid overestimation, capture larger peak flow events and is less affected by outliers. Finally, simulation and real data analyses of confidence interval (CI) through bootstrap method are performed. Although CI of both method does not achieve the theoretical coverage rate, the coverage rate of EMS method is more stable in different return period.

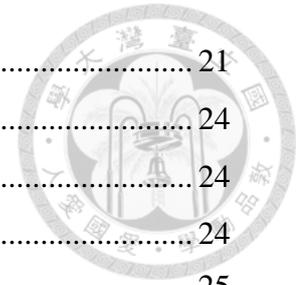
Key words: Annual Maximum Rainfall, Event-based, frequency analysis, mixture distribution



目錄

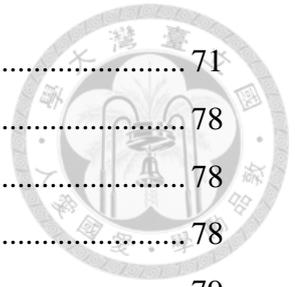


誌謝	I
摘要	III
Abstract	V
目錄	VII
圖目錄	XI
表目錄	XIII
第一章 前言	1
1.1 研究動機與目的	1
1.2 研究架構與流程	3
第二章 文獻回顧	5
2.1 極端值理論與頻率分析	5
2.1.1 年最大值數列法	5
2.1.2 部分延時數列法	6
2.1.3 事件數列法	8
2.2 降雨事件切割與過度離散	10
2.2.1 降雨事件切割	10
2.2.2 過度離散	11
2.3 不確定性分析與拔靴法	13
第三章 年最大值數列法	15
3.1 降雨頻率分析	15
3.2 年最大值降雨數列法	15
3.3 機率分布	16
3.3.1 常見分布	16
3.3.2 極端值理論與廣義極端值分布	19
3.4 參數推估	20
3.4.1 動差法	20
3.4.2 最大概似法	21



3.4.3 線性動差法	21
3.5 機率適合度檢定	24
3.5.1 機率圖	24
3.5.2 Kolmogorov-Smirnov Test	24
3.5.3 線性動差比圖	25
3.6 降雨強度—延時—頻率曲線	27
第四章 事件最大值數列法	31
4.1 年最大值與混合分布	31
4.2 事件發生次數分布	32
4.2.1 卜瓦松分布	32
4.2.2 負二項分布	32
4.3 事件最大值數列法	33
4.3.1 降雨事件切割	33
4.3.2 事件最大值數列	34
第五章 蒙地卡羅模擬	39
5.1 降雨資料生成與頻率分析模擬	39
5.2 頻率分析結果討論	40
5.3 極端值理論模擬	40
5.4 極端值模擬結果討論	41
第六章 真實資料分析	49
6.1 研究測站	49
6.2 事件切割結果	49
6.3 降雨頻率分析	49
第七章 結果與討論	65
7.1 設計降雨討論	65
7.2 尖峰流量分析	68
7.2.1 SCS 三角形單位歷線	68
7.2.2 研究測站	70
7.2.3 尖峰流量比較	70
7.3 不確定性分析	71

7.3.1 拔靴法與信賴區間.....	71
7.3.2 年最大值數列抽樣.....	78
7.3.3 事件最大值數列抽樣.....	78
7.3.4 信賴區間模擬.....	78
7.3.5 真實資料設計降雨之信賴區間.....	79
第八章 結論.....	87
參考文獻.....	89



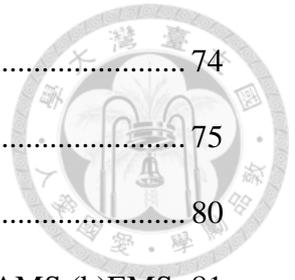


圖目錄



圖 3-1 年最大值數列法示意圖	17
圖 3-2 嘉義雨量站 1969 年 24 小時年最大值數列跨事件樣本	17
圖 3-3 伽瑪分布資料繪於常態機率圖	26
圖 3-4 常見分布的 L-moment ratio diagram (Hosking, 1990)	28
圖 3-5 頻率曲線示意圖(a) DDF 與(b) IDF	30
圖 4-1、三測站於各 MIT 下所得到的事件數	36
圖 4-2 事件最大值數列法示意圖	37
圖 4-3 事件最大值列法流程	37
圖 5-1 降雨頻率分析 44 年模擬結果	44
圖 5-2 10000 次模擬之 LMRD	44
圖 5-3 不同年事件數下年最大值之 LMRD	46
圖 5-4 不同事件發生率下年最大值之 LMRD	47
圖 6-1 雨量站、流量站與水位站位置	50
圖 6-2 五堵站事件最大值數列機率適合度檢定結果	56
圖 6-3 頭汴坑站事件最大值數列機率適合度檢定結果	57
圖 6-4 嘉義站事件最大值數列機率適合度檢定結果	58
圖 6-5 年最大值數列機率適合度檢定結果	59
圖 6-6 Poisson-PT3 與 NB-PT3 24 小時年最大值降雨累積機率分布函數	60
圖 6-7 三測站之降雨深度-延時-頻率曲線(DDF)	64
圖 7-1 3 小時降雨量累積分布函數	66
圖 7-2 三測站各類型事件延時長條圖	69
圖 7-3 SCS 單位歷線示意圖	72
圖 7-4 Q_p, EMS 等於 Q_p, AMS 、累積降雨不相等之樣本。	74

圖 7-5 Q_p, EMS 大於 Q_p, AMS 之樣本.....	74
圖 7-6 Q_p, EMS 小於 Q_p, AMS 之累積降雨樣本.....	75
圖 7-7 拔靴法示意圖.....	80
圖 7-8 100 年重現期 1000 次 95% 信賴區間模擬 (percentile), (a)AMS (b)EMS.	81
圖 7-9 Percentile Bootstrap (a)AMS、(b)EMS、(c)颱風事件數、(d)梅雨事件數、 (e)颱風事件降雨量、(f)梅雨事件降雨量累積分布.....	82
圖 7-10 五堵站百分比信賴區間估計結果.....	83
圖 7-11 頭汴坑站百分比信賴區間估計結果.....	84
圖 7-12 嘉義站百分比信賴區間估計結果.....	85



表目錄



表 1-1 朴子溪與高屏溪莫拉克颱風降雨頻率分析結果.....	2
表 3-1 國內降雨頻率分析常用統計分布.....	18
表 3-2 經驗累積分布公式.....	26
表 4-1 不同 MIT 下事件到達間隔 CV 計算.....	35
表 4-2 不同 MIT 下事件到達間隔針對指數分布的 KS-test 結果 (p-value).....	35
表 4-3 降雨事件分類規則.....	36
表 5-1 模擬使用參數.....	41
表 5-2 降雨頻率分析模擬結果 (單位：百分比).....	43
表 5-3 10000 次降雨模擬之 AMS 於 LMRD 拒絕率 (單位：百分比).....	43
表 5-4 AMS 於 Gumbel LMRD 之拒絕率 (單位：百分比).....	45
表 6-1 測站資訊.....	51
表 6-2 三測站 MIT=4 hr 下各類型降雨事件切割結果.....	51
表 6-3 降雨事件發生次數之卜瓦松與負二項分布機率適合度檢 (p-value).....	54
表 6-4 各測站雨量偏態係數.....	55
表 6-5 年最大值數列降雨類型.....	61
表 6-6 五堵站頻率分析結果.....	62
表 6-7 頭汴坑站頻率分析結果.....	62
表 6-8 嘉義站頻率分析結果.....	63
表 7-1 新增測站的基本資訊.....	66
表 7-2 流量站資訊.....	72
表 7-3 不同集流時間下各測站降雨和流量/水位之相關係數.....	72
表 7-4 三測站單位歷線參數與計算結果.....	73
表 7-5 三測站之單位歷線.....	73

表 7-6 各延時下 AMS 與 EMS 尖峰流量比較.....	74
表 7-7 長延時 Q_p , AMS 大於 Q_p , EMS 之樣本及與其對應之 Q_p , EMS 降雨延時 (粗體表尖峰流量相等).....	76
表 7-8 AMS 與 EMS 拔靴法信賴區間之理論值覆蓋率	80



第一章 前言



1.1 研究動機與目的

臺灣一年四季的降雨分布不均，不同類型降雨事件的降雨量亦有所不同，例如颱風和梅雨容易帶來過量的降雨，造成淹水、土石流等災害，尤其是近年來強降雨颱風事件的發生更加頻繁，因此，如何評估特定降雨量的發生機率以及其對於水利工程的風險將是必要的挑戰。降雨頻率分析是以歷史的雨量資料為基礎、推估特定降雨事件平均發生率的方法，在水利工程的暴雨設計中為決定該特定事件總雨量的必要步驟，其精準度會直接影響工程的成本以及可以承受的風險；此外頻率分析也可以用來估計災害事件的發生頻率，以評估特定事件的平均發生時間。

傳統上使用年最大值數列法進行頻率分析。由每年的雨量記錄中選取特定延時下的最大降雨量，進而推估所得年最大值數列 (Annual Maximum Series, AMS) 的機率分布，並以此計算不同重現期距的降雨量。在年最大值數列法中，由極端值理論推導而得的廣義極端值分布是最常使用的分布之一。年最大值數列法提供頻率分析一套簡單快速的估計流程，然而，其每年僅考慮一筆資料的方式，除了無法充分利用降雨紀錄之外，也使估計容易受到極端值的影響。以 2009 年的莫拉克颱風為例，根據水利署「莫拉克颱風暴雨量及洪流量分析」報告，莫拉克颱風的累積雨量在多處雨量站中的重現期均超過 2000 年(表 1-1)，顯示其災害之嚴重，但是，莫拉克颱風的降雨量明顯為歷年降雨紀錄中的離群值，在採用觀測紀錄長度平均約十五到二十五年、每年只取一次樣本的情況下，應用 AMS 的結果具有相當高的不確定性。此外，年最大值數列法的樣本選擇並沒有降雨事件的概念，除了會因為選取複數事件而使降雨量被高估，缺少樣本數概念的抽樣過程也無法符合極端值理論的假設。最後，臺灣的降雨來自於各種類型的降雨事件如颱風、梅雨、對流雨或者鋒面雨等，這些降雨的統計特性均不相同，若將這些樣本一起估計，將會造成誤差。

為了解決上述問題，本研究在考慮降雨事件和降雨類型的情形下，提出事件最大值數列法作為年最大值數列法之替代。除了使用模擬以及實際資料分析與年最大值數列法比較設計降雨之結果，也針對兩種方法所選取樣本造成之尖峰流量與設計降雨的信賴區間進行探討。



表 1-1 朴子溪與高屏溪莫拉克颱風降雨頻率分析結果

朴子溪	樟腦寮(2)	嘉義縣竹崎鄉	837.0	71	1346.0	>2000	1777.0	>2000	1816.0	34
朴子溪	沙坑	嘉義縣竹崎鄉	572.0	61	703.0	81	822.0	204	839.0	29
朴子溪	鶯鼓	嘉義縣東石鄉	412.5	>2000	489.5	>2000	516.0	>2000	521.5	43
朴子溪	樟腦寮	嘉義縣梅山鄉	653.0	21	1202.0	>2000	1595.0	>2000	1631.0	34
朴子溪	魚寮	嘉義縣太保市	460.5	>2000	551.0	>2000	582.0	>2000	585.5	43
朴子溪	朴子	嘉義縣朴子市	452.5	>2000	536.5	>2000	560.5	>2000	561.0	43
高屏溪	屏東(5)	屏東縣屏東市	667.0	141	886.0	124	947.0	159	959.0	38
	美濃(2)	高雄縣美濃鎮	507.0	>2000	749.0	>2000	828.0	>2000	871.0	19
	屏東	屏東縣屏東市	666.0	140	906.0	143	974.5	197	990.0	38
	溪埔	高雄縣大樹鄉	729.5	271	994.5	265	1057.5	378	1076.5	38
	旗山	高雄縣旗山鎮	621.0	>2000	813.0	>2000	854.5	>2000	881.0	15
	尾寮山	屏東縣三地門	1414.5	>2000	2215.5	>2000	2564.0	>2000	2701.0	21
	甲仙	高雄縣甲仙鄉	1077.5	>2000	1601.0	>2000	1856.0	>2000	1916.0	25
	古夏	屏東縣三地門鄉	683.5	>2000	946.0	>2000	1061.5	>2000	1127.0	25
	美濃	高雄縣美濃鎮	633.5	>2000	878	>2000	955.5	>2000	989.5	15
	里港	屏東縣里港鄉	710.5	>2000	955.5	>2000	1018	>2000	1039.5	15
	上德文	屏東縣三地門鄉	1185.5	>2000	1968.0	>2000	2194.5	>2000	2255.0	25
	新園	屏東縣鹽埔鄉	578.0	148	757.5	>2000	806.5	565	830.5	25
	月眉	高雄縣杉林鄉	744.0	>2000	1081.0	>2000	1205.0	>2000	1246.5	19
	吉東	高雄縣美濃鎮	547.5	>2000	728.0	>2000	789.0	>2000	820.5	19
	大津	高雄縣六龜鄉	738.5	>2000	1072.0	>2000	1241.0	>2000	1314.0	21

水利署 (2009)



1.2 研究架構與流程

事件最大值數列法的年最大值分布是由事件數分布和雨量分布組合而成的混合分布，藉由在分析時考慮所有降雨事件，可以避免年最大值數列法樣本數過少的缺點。為了瞭解兩種方法在估計上的差異，本研究使用卜瓦松分布 (Poisson distribution) 與伽瑪分布 (gamma distribution) 模擬颱風和梅雨的降雨紀錄，比較年最大值數列使用皮爾遜第三型分布、廣義極端值分布以及事件最大值數列 (Event Maximum Series, EMS) 使用卜瓦松與皮爾遜第三型分布 (Pearson Type III distribution, PT3) 之混合分布推估而得的回歸水準 (return level) 的偏誤與均方根誤差，並探討年最大值數列收斂至廣義極端值分布之情形。此外，本研究分別由臺灣北、中、南地區選定一個雨量站 (五堵、頭汙坑、嘉義)，針對年最大值數列使用皮爾遜第三型分布，事件最大值數列則使用負二項與皮爾遜第三型分布之混合分布進行實際資料分析。降雨量的參數推估與適合度檢定使用 Hosking 與 Wallis (1997) 所提出的線性動差法 (L-moment method) 和線性動差比圖 (L-Moment Ratio Diagram, LMRD) 決定，事件數目的分布則由最大概似推估 (Maximum Likelihood Estimator, MLE) 以及 Kolmogorov–Smirno 檢定 (Kolmogorov–Smirno Test, KS Test) 完成。除此之外，為了瞭解兩種方法所估計之年最大值降雨的關係是否會對應至尖峰流量，兩種方法所選取的年最大值降雨樣本將以三角形合成單位歷線計算尖峰流量並比較大小。最後，百分比拔靴法 (percentile bootstrap) 以及修正偏誤加速法 (Bias-Corrected and accelerated bootstrap, BCa) 則被用於計算兩種方法的回歸水準之信賴區間。

本文章節編排如下。頻率分析方法的文獻回顧與整理撰於第二章；第三章為傳統降雨頻率分析的方法介紹；在第四章將介紹年最大值混合分布以及事件最大值數列法；兩種方法的模擬比較於第五章進行討論；第六章為真實資料分析並於第七章進行討論；最後總結於第八章。



第二章 文獻回顧



2.1 極端值理論與頻率分析

降雨頻率分析藉由過去的雨量觀測紀錄推估測站在固定延時下的降雨量機率分布，並以此分布估計雨量的重現期距或者回歸水準。一般而言，頻率分析所關心的多為極端值事件，這些事件發生的頻率比較低，卻可能因為其極端值特性而造成重大影響，因此，極端值理論經常被應用於頻率分析來了解極端值事件的發生規律。

2.1.1 年最大值數列法

Fisher and Tippett (1928) 針對樣本之最大與最小順序統計量提出它們的漸進分布，並且被 Gnedenko (1943) 進一步發展出極端值理論，極端值理論指出，如果統計分布的最大順序統計量在經過適當的標準化之後不會變成退化分布，則其分布會收斂至廣義極端值分布，並且稱這些統計分布位於 Domain of Attraction 之中。極端值理論應用於頻率分析的研究可以見 Gumbel (1958) 的 block maxima 模型，藉由將紀錄資料以固定的間隔切割成複數個 blocks，則在樣本的分布相同並且獨立的假設下，當 block 中的樣本數足夠大時，每一個 block 中的最大值便可以使用廣義極端值分布逼近。如果 block 的時間單位為年，這些最大值樣本即稱為年最大值數列。傳統年最大值數列法藉由移動窗口 (moving window) 選取每年最大的樣本並假設其為年最大值，使用廣義極端值分布估計之。年最大值數列法的機率分布並不一定使用廣義極端值分布，經濟部水利署的水文設計應用手冊 (2001) 指出過去臺灣常用的統計分布除了廣義極端值分布與 Gumbel 分布以外，還有皮爾遜第三型分布、對數皮爾遜第三型分布以及三參數對數常態分布，該手冊針對全台降雨測站進行各分布的頻率分析，最後使用理論降雨量與觀測降雨量計算誤差指標來決定各測站在分析時所建議使用的分布。因為年最大值數列法操作簡單並且直觀，至今仍有許多人使用年最大值數列法進行分析，Svenssonc and Jones (2010) 比較九個國家採用的降雨頻率分析的方法，指出年最大值數列法是最常被使用的方法，並且廣義極端值分布是最常被使用的分布。雖然年最大值數列法因為其簡便的操作廣為眾人使用，許多研究也指出它的不足之處——年最大值數列法在篩選過程中只選取每年最大值的做法並無法充分利用所有資料 (Bačová and Onderka 2010)，一般而

言，進行重現期為 T 年的頻率分析時，資料的長度最好要大於等於 T (Hosking and Wallis 1997)，然而年最大值數列每年僅取一筆資料的做法，極容易發生資料長度不足的情況。Madsen et al. (1997) 指出在資料中年第二大的數值可能會高於其他年之年最大值，因此如果將之捨去則有可能造成估計值低估，並且在乾燥地區某些年的水文量（淹水、降雨）有可能非常低，將這些年的年最大值以相同的權重納入年最大值數列將有可能影響極端值的推估。

2.1.2 部分延時數列法

為了彌補年最大值數列法的缺點，部分延時數列法被提出。藉由選取紀錄中所有符合特定標準的資料為部分延時數列 (Partial Duration Series, PDS)，部分延時數列法可以納入更多資料進行推估。在部分延時數列法中，超越門檻峰值法是最具代表性的方法，因此在本文中部分延時數列皆指超越門檻峰值數列。經由決定閾值並選取高於閾值的峰值為超越門檻峰值數列 (Peak Over Threshold, POT)，數列的分布可以使用事件數分布——降雨量分布的混合分布進行描述。事件數與降雨量的分布可以有多種選擇，最常使用的組合為卜瓦松——廣義柏拉圖分布 (Poisson - generalized Pareto distribution)。然而卜瓦松分布具有期望值與變異數相等的特性，有時與資料的實際情況不相符，因此其他的離散分布例如二項分布、負二項分布也有可能被使用。Bhunya et al. (2013) 藉由模擬以及泰勒展開式比較負二項分布與卜瓦松分布於部分延時數列以及年最大值數列的回歸水準之變異數，指出當峰值的平均數目接近其變異數時，卜瓦松分布的表現稍佳，而在其他情形負二項分布皆更有效率。實際資料分析中負二項分布的表現在重現期小於 300 年時皆比較好，這是因為雙參數的負二項分布可以提供模型的變異數更多彈性。Bezak et al. (2014) 使用二項分布、負二項分布與卜瓦松分布配合指數分布以及卜瓦松——廣義柏拉圖分布進行估計並與年最大值數列進行比較，結果顯示，卜瓦松模型對於資料的配適優於負二項分布與二項分布，並且也優於年最大值數列法。在超越門檻峰值法中，降雨量的分布可以視為事件降雨量的分布高於閾值之條件分布，因此超越門檻峰值法的分布可以由事件降雨量或者年最大值數列的分布推導而得。以服從 Gumbel 分布的事件降雨量為例，當選取高於閾值的降雨為超越門檻峰值數列時，數列的分布會服從指數分布；而當超越門檻峰值數列服從廣義柏拉圖分布時，所對應事件降雨量則為廣義極端值分布（推導見 Coles et al. 2001）。在過去有許多文獻使用模擬比



較年最大值數列法與部分延時數列法的結果。Cunnane (1973) 比較部分延時數列使用卜瓦松——指數分布以及年最大值數列使用 Gumbel 分布估計之效率 (efficiency)，得到當重現期小於 10 年或者當數列的資料長度大於等於紀錄長度的 1.65 倍時，部分延時數列的表現才會優於年最大值數列。Wang (1991) 比較部分延時數列使用卜瓦松——廣義柏拉圖分布與年最大值數列以廣義極端值分布進行估計的結果，指出部分延時數列與年最大值數列具有相近的效率。Madsen et al. (1997) 在進行同樣比較後則提出當降雨量的形狀參數 (shape parameter) 為負值時適合採用部分延時數列，反之則為年最大值數列。Ferreira and De Haan (2015) 結合上述以及其他文獻，認為年最大值數列法與部分延時數列法的表現相近，甚至年最大值數列法的效率更高。眾多文獻在實際資料分析中皆建議使用部分延時數列法進行頻率分析 (Bezak et al. 2014; Vicente and Beguería 2003; Nagy et al. 2017; Urban et al. 2018)。其中 Bačová and Onderka (2010) 和 Nagy et al. (2017) 認為部分延時數列的優點在於可以針對更短的紀錄年限進行分析。綜合以上考量，部分延時數列法在頻率分析的表現優於年最大值數列法。

Pham et al. (2013) 認為部分延時數列無法普及是因為其模型的複雜性，除此之外，部分延時數列法中決定閾值的方法也有待解決。Bačová and Onderka (2010)、Bezak et al. (2014) 皆提及部分延時數列法對於使用者而言最困難部分即為閾值的決定方式。閾值越低則可以獲得越多的資料，但過低的閾值將無法維持峰值間的獨立性，甚至在使用廣義柏拉圖分布時無法確保雨量分布收斂至廣義極端值分布。為了決定適當的閾值以達成資料間彼此獨立以及盡可能地選取資料兩項目標，不同的閾值決定方法被眾多研究提出使用。Mean Residual Life plot (MRL plot) 法藉由廣義柏拉圖分佈超越閾值的平均值和閾值的線性關係，將部份延時數列在不同閾值下之超越閾值的平均值對閾值作圖，可以找出兩者開始呈現線性關係的閾值作為最終選擇 (Coles et al. 2001; Nagy et al. 2017)。此外，在不同閾值下使用 dispersion test 檢定超越峰值數目是否符合卜瓦松分布 (Cunnane 1979) 以及檢定超越峰值降雨量的參數變化 (Coles et al. 2001; Pham et al. 2013) 也可以決定最終所使用的閾值。Lang et al. (1999) 對部分延時數列法進行回顧，並針對 MRL plot 法與 dispersion test 進行討論，指出在實際應用時應該綜合考慮這些標準進行選擇。Anagnostopoulou and Tolika (2012) 比較 MRL plot 法、dispersion test 和檢驗廣義柏

拉圖分布分布參數穩定度的 threshold test，最後由數列長度以及為降雨量和閾值比所設定的標準得到應該使用由三種方法所得的閾值平均作為最終選擇的結論。由此可知閾值的決定沒有最佳方法，甚至需要依賴使用者的主觀認定，而基於統計檢定的方法可以檢視資料是否符合模型假設，卻不會考慮實際的物理現象，因此儘管這些方法可以提供使資料符合模型結果，卻無法提供令人滿意的最佳解釋。

2.1.3 事件頻率分析

Jun et al. (2018) 指出許多研究都建議頻率分析應該以事件作為分析基礎。藉由從降雨紀錄中切割出降雨事件，事件頻率分析可以避免部分延時數列中決定閾值的困難。此外，在充分利用所有的降雨紀錄之餘，所得到的事件也能提供如事件降雨量、事件延時、降雨強度、事件間隔時間等降雨過程的資訊，作為後續分析與降雨過程模擬使用。過去的研究指出傳統頻率分析方法所使用的設計延時過於人工、無法表現真正的自然現象 (Kao and Govindaraju 2007; Lee et al. 2010)，由於在選取樣本時不會考慮事件的真實延時，傳統頻率分析的短延時樣本可能來自長延時事件，而長延時樣本可能覆蓋複數個短延時事件，造成誤差。Yoo et al. (2015) 比較以雙變數指數分布的事件頻率分析結果以及使用 Gumbel 分布的傳統年最大值頻率分析結果，指出在低延時 (1 hr) 時事件頻率分析的結果明顯低於傳統頻率分析，而在高延時 (24 hr) 時差別則比較小，此結果顯示在該地的降雨事件中短延時降雨的發生頻率比較低，因此傳統頻率分析的低延時抽樣多來自高延時降雨事件，造成不同的估計結果。事件頻率分析包含單變數以及多變數兩種；單變數事件頻率分析藉由在考慮目標水文量 (例如降雨) 的分布時只推估其邊際分布 (marginal distribution) 來計算回歸水準。Jun et al. (2018) 針對降雨事件設置閾值並分別使用廣義柏拉圖分布和廣義卜瓦松分布 (generalized Poisson distribution) 配適事件降雨量分布和年事件發生數，以此獲得基於事件的年最大值混合分布來進行頻率分析並與年最大值數列法進行比較，結果顯示事件頻率分析將計算出更低並且更符合真實情形的回歸水準。

在降雨過程中會同時產生降雨量、事件延時、降雨強度、雨型等眾多變數，因此許多研究會考慮兩個以上的變數進行分析，是為多變數事件頻率分析。多變數事件頻率分析的困難在於如何描述變數之間的結構，除了使用聯合分布 (joint distribution) 例如雙變數指數分布 (bivariate exponential distribution) (Park et al. 2013;



Yoo et al. 2015)、Gumbel 混合模型 (Gumbel mixed model) (Lee et al. 2010)、雙變數極端值分布 (bivariate extreme value distribution) (Yue and Rasmussen 2002) 進行描述之外，耦合 (copula) 也是經常被使用的方法。Sklar (1959) 指出一個聯合分布可以分別以其邊際分布的累積分布函數 (Cumulative Distribution Function, CDF) 和其相關性結構 (dependence structure) 進行描述 (Ariff et al. 2012; Kao and Govindaraju 2007)，因此耦合方法藉由使用邊際分布為均勻分布的聯合分布來描述變數間的關係，並且聯合分布中變數間的相關性結構可以依需求而決定不同耦合使用。耦合提供使用者更多的選擇來描述變數的邊際分布以及變數之間的相關性，然而如何選擇也成為應用上的一項挑戰。

De Michele and Salvadori (2003) 為最早應用耦合於水文分析中的研究之一，在其後的眾多應用與文獻之中，Nelsen (2006) 提供了對於耦合的完整理論敘述 (Ariff et al. 2012; Kao and Govindaraju 2007)。Ariff et al. (2012) 針對馬來西亞降雨資料使用耦合建立降雨強度—延時—頻率曲線 (Intensity–Duration–Frequency curve, IDF curve) 並與經驗公式比較，考量到降雨強度與延時呈現負相關，他們使用 Farlie Gumbel Morgenstern (FGM), Gaussian, Ali–Mikhail Haq (AMH) 與 Frank 等四種耦合進行分析，並以赤池信息量準則 (Akaike's Information Criteria, AIC) 作為選擇最適合模型的標準，線性動差比圖則用於選取各降雨變數的邊際分布，結果顯示雖然兩種方法的結果差異不大，耦合在計算和後續分析上仍優於經驗公式。Kao and Govindaraju (2007) 基於 Frank 耦合通過 KS 檢定以及在變數相關性的良好配適結果，建議以其分析降雨量、延時與尖峰強度間的相關性，並考慮卡方檢定、KS 檢定與過去研究的結果，建議使用 Gumbel 與對數常態分布作為邊際分布。

雖然多變數頻率分析有充分的資訊可以提供，然而其分析過程中包含事件切割、變數與分布甚至耦合的決定，步驟相對複雜。Park et al. (2013) 指出雙變數頻率分析所需要面對的多項挑戰，首先在分析所有降雨事件時無法考慮如每年事件數或者降雨季節性的問題，因此使用年最大值降雨事件進行分析仍然是比較適合的方法。其次，在多變數頻率分析中重現期的定義不只有一種，因此分析的結果在解釋與應用上會比較模糊。最後，多維度的資料將使得頻率分析的分析過程更加複雜，增加分析難度。



2.2 降雨事件切割與過度離散

2.2.1 降雨事件切割

在頻率分析的過程中，年最大值數列法可以使用移動窗口選取每年的年最大值作為年最大值數列，部份延時數列的資料選取取決於閾值的設定，事件頻率分析則必須由降雨紀錄切割出互相獨立之降雨事件作為分析資料。獨立降雨事件的切割並沒有固定的方法 (Hou 2010)，最直接的做法是將降雨紀錄中的連續降雨視為一場降雨事件，一旦降雨量為 0 即視前後時刻的降雨為二個獨立事件 (Peters and Christensen 2002)。這種決定方法簡單迅速，然而現實中降雨事件具有間歇性，以颱風為例，即使測站的降雨因為颱風的雲帶散布或者地形等因素而停止一、二個小時，下一刻的降雨依然與颱風以及前一刻的降雨量相關，應該視為同一場降雨事件，因此若使用連續降雨作為切割標準，則有可能切出不獨立的降雨事件。基於上述的情形，事件之間應該具有最低事件間隔時間 (Minimum Inter-event Time, MIT) 作為切割標準，藉由將間隔時間未大於 MIT 的降雨紀錄視為同一事件，可以確保切割出的降雨事件彼此獨立。越長的 MIT 所獲得的事件越能保證事件之間彼此獨立，然而也會增加將獨立事件視為同一場事件的風險，因此 MIT 的決定對於事件頻率分析十分重要。

決定 MIT 的方法包括統計方法以及主觀認定。統計方法主要為 Grace and Eagleson (1967) 所提出的等級相關係數法 (rank correlation method) 以及 Restrepo and Eagleson (1982) 所提出的指數法 (exponential method)。等級相關係數法藉由計算並檢定不同延遲下降雨強度時間數列之等級相關係數，找出第一個使數列明顯不相關的延遲時間作為 MIT。在指數法中，假設降雨事件的發生為卜瓦松過程 (Poisson process)，則事件到達間隔 (inter arrival time)——即事件延時與事件間隔時間的和，會服從指數分布；並且在延時遠小於事件間隔的假設下可以直接使用事件間隔時間替代。指數法藉由指數分布標準差與平均數相等的特性，比較不同 MIT 下事件間隔時間的樣本標準差與樣本平均數之比值 (Coefficient of Variance, CV) 何者等於 1，可以決定最終所使用的 MIT。Bonta and Rao (1998) 比較指數法和等級相關係數法之結果，基於指數法可以提供比較長的 MIT 以及多數無降雨時間符合指數分布的結果，建議使用指數法決定 MIT。在前人研究中也指出指數分布對

於降雨、延時與事件間隔時間而言是合理且適當的假設 (Guo and .Adams 1998; Bonta 2004)，因此指數法至今也仍然被許多人所使用 (Wang et al. 2019; Hanel and Máca 2014; Bonta 2004; Iadanza et al. 2016; Sordo-Ward et al. 2016)。除了上述兩種統計方法，Medina-Cobo et al. (2016)亦應用碎形理論，在不同的 MIT 設置中挑選出讓降雨強度最符合理論的間隔時間。

Dunkerley (2008) 回顧不同研究決定 MIT 的方法，指出 4 到 8 小時為廣泛使用的間隔時間；此外，也有部分研究以植生或者土壤乾燥時間等作為依據 (Marin et al. 2000; Bracken et al. 2008)，這些標準可以額外結合降雨強度以及延時作為限制，然而因為所考慮的因素都不相同，方法間難以進行有意義的比較。除了考慮集水區特性，Haile et al. (2011) 考量對流雨的自然特性而將 MIT 定為三十分鐘。許恩菁 (1999) 以 3 到 6 小時的 MIT 進行事件切割，比較不同設計延時下所找到的各年年最大值降雨事件的實際延時平均值變化，得到 3 小時將會明顯低估降雨延時、4 小時與 5 小時接近、而 6 小時間隔將高出一般集流時間並且會造成事件延時過長的結論，因此決定以 4 小時作為最短間隔。針對都市地區，Joo et al. (2014) 使用 SWMM 計算集水區逕流後，將逕流未重疊的降雨視為獨立事件並以此定義 MIT，因此產生的降雨事件在逕流分析上並不會互相影響。Adams and Papa (2000) 藉由檢視不同 MIT 設置下的年平均事件數，選出不再明顯影響事件數的 MIT 作為最終使用。後者被 Liu et al. (2014) 進一步應用於美國維吉尼亞州。

MIT 的決定方法眾多，Guo and Adams (1998) 和 Chin et al. (2016) 指出 MIT 的決定必須視研究目的而定，本研究應用 MIT 於事件頻率分析提供暴雨設計的基礎，因此將以集流時間與事件數作為判斷標準切割事件。

2.2.2 過度離散

在單變數事件頻率分析中，為了描述年最大值分布，必須同時估計雨量分布以及事件發生數的分布。Jun et al. (2018) 指出在以事件進行頻率分析的研究中，少有人檢視事件的發生機率而皆專注於描述變數之間的關係，這是因為多數事件頻率分析採用多變數的方法進行分析。一般而言，降雨事件的發生次數可以使用卜瓦松分布來描述 (Todorovic and Vujica 1969; Bonta and Rao 1988; Jun et al. 2017)，即降雨事件為一卜瓦松過程。然而受到延時、群聚 (clustering) 以及一些未知特性的影響，有時在給定的模型之下，會發生資料的變異數大於模型所給定之變異數的情形，



稱為過度離散 (overdispersion)。在不同文獻之中過度離散特指資料的變異數大於其平均數的情形，因為事件的發生經常以卜瓦松分布描述，並且卜瓦松分布具有相同的平均數與變異數；本研究以後者作為過度離散的定義。各類型的資料如水文 (Dzupire et al. 2018; Katz and Parlange 1998)、生態 (Mcmahon et al. 2017; Harrison 2015; Lindén and Mäntyniemi 2011)、醫藥 (Plan 2014)、金融 (Cameron and Trivedi 1996) 都會發生過度離散的情況，在此情況下若使用卜瓦松分布進行配適，容易使資料被適合度檢定拒絕並在估計時產生誤差。在處理過度離散的資料時，負二項分布 (negative binomial distribution) 是最早也是最常被使用的分布之一。負二項分布的變異數大於期望值，因此可以用於描述過度離散的資料，除了降雨事件的發生，也被用於描述 POT 方法中超越閾值的尖峰數 (Zucchini and Adamson 1989; Bhunya et al 2013; Bezak et al 2014)。Eastoe and Tawn (2010) 指出使用負二項分布的問題包括它無法在理論上如同卜瓦松分布一般給予解釋、無法輕易地延伸模擬整個發生過程以及假設年最大值與事件數互相獨立，因此應用隨機效應模式 (random effect model) 描述超越峰值數目，然而在缺少解釋變數的情況下，模式依然會回歸負二項分布。

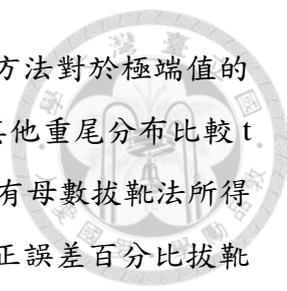
廣義卜瓦松分布在 1973 年被 Consul 與 Jain 所提出，在第二參數分別為 0、正與負的情況下，廣義卜瓦松分布可以對應卜瓦松分布、過度離散與欠離散 (underdispersion)，因此被 Jun. et al (2018) 用來描述降雨事件數。因為負二項分布與廣義卜瓦松分布都具有兩個參數，Joe 與 Zhu (2005) 針對過度離散的情形，藉由固定第一、二階動差使兩分布的參數一對一對應，比較兩者的機率質量函數 (Probability Mass Function, PMF) 與偏態系數 (skewness)。結果發現兩者並無明顯差異，不過負二項分布在 0 有比較高的質量，而廣義卜瓦松分布則比負二項分布更具有重尾 (heavy tail) 性質，並且偏態系數更高，因此負二項分布與廣義卜瓦松分布分別適用於資料中有大量 0 與資料為重尾時。上述的兩種分布均為卜瓦松混合分布 (Poisson mixture) 的一種 (Joe and Zhu 2005; Karlis and Xekalaki 2005)，代表它們都是使用一隨機變數作為參數的卜瓦松分布。卜瓦松混合分布可以處理不同情況下的過度離散問題，更多的混合分布可以參考 Karlis and Xekalaki (2005)。有時過度離散的原因來自於大量的 0 出現在紀錄之中，例如降雨紀錄中有降雨的時間會明顯少於沒降雨的時間，此時可以使用零膨脹模型 (zero-inflated model) 進

行模擬。零膨脹模型藉由與不同的分布結合可以形成如零膨脹卜瓦松、零膨脹負二項等模型，以負二項分布為例，藉由將發生次數為 0 的機率設為 $\pi + (1 - \pi)P_{NB}(N = 0)$ 、其他次數設為 $(1 - \pi)P_{NB}(N = n)$ ，因此模型中發生 0 的機率將被提升，而模型的變異數也會大於期望值。

2.3 不確定性分析與拔靴法

不確定性的定義有許多種，例如對於特定的因子、參數或者模型缺乏理解 (US EPA 1997) 或者是一種對於觀測資料與模型輸出準確度的一種量測 (Al Mamoon and Rahman 2014)。基本上，不確定性為對於模型以及模擬對象的不可掌握而造成的誤差，它的來源一般可以分為三種：——資料誤差、模型誤差與其他誤差，資料誤差來自於輸入資料在收集過程中所產生的誤差，包含抽樣誤差、測量誤差等；模型誤差則為分析過程中與模型有關的誤差，例如因為模型在簡化、選擇的過程中而產生的誤差，或者是參數估計的誤差；第三種誤差的來源可能是分析不完善、描述性誤差 (US EPA 1997)，或者是量化誤差 (Xu, Booij and Tong 2010)。降低與衡量不確定性的方法有很多種，Al Mamoon and Rahman (2014) 列舉眾多應用於設計降雨的不確定性分析方法，並針對蒙地卡羅法 (Monte Carlo method) 以及拔靴法進行回顧。這兩種方法簡單並且直觀，雖然必須依靠大量計算模擬產生資料再進行分析，隨著電腦計算能力的提升，越來越廣為眾人使用。

蒙地卡羅法藉由決定樣本之機率分布、模擬來自相同母體分布的新樣本，再由這些新樣本計算目標統計量來掌握統計量的不確定性。拔靴法由 Efron (1979) 所提出，藉由從假設的母體分布 (有母數拔靴法) 或者經驗分布 (無母數拔靴法) 中抽取拔靴法樣本並計算目標統計量來獲得目標統計量的樣本分布，以之計算信賴區間評估抽樣不確定性，除了被用來評估統計參數的抽樣誤差之外，也被廣泛應用於水文量的信賴區間建立。一般而言，建立信賴區間的拔靴法樣本數目建議為 1000 (Efron and Tibshirani 1994)。Zucchini and Adamson (1989) 針對降雨的部分延時數列使用拔靴法百分比拔靴法建立信賴區間，其抽樣過程包含抽取年事件數以及降雨樣本兩部分，並且由蒙地卡羅模擬發現 200 次抽樣即足夠建立信賴區間。Sane et al. (2018) 重複 1000 次拔靴法抽樣建立 IDF 曲線參數的信賴區間，並以此探討不同時間尺度下的降雨關係。由於需要以拔靴法評估不確定性的水文學通常為極端



值——例如頻率分析的結果，因此必須評估不同信賴區間建立方法對於極端值的表現。Kysely (2008) and Kysely (2010) 針對廣義極端值分布與其他重尾分布比較 t 分布拔靴法、百分比拔靴法、加速並修正誤差百分比拔靴法以及有母數拔靴法所得到的信賴區間覆蓋率，結果顯示在中小樣本數下雖然加速並修正誤差百分比拔靴法的表現最佳，無母數拔靴法的信賴區間會傾向於低估不確定性，因此建議使用有母數拔靴法甚至是其他數值分析方法建立信賴區間，這些結果在 Panagoulia et al. (2014) 針對平穩 (stationary) 與非平穩 (non stationary) 的降雨數列模擬中被進一步確認。Schendel and Thongwichian (2015) 使用 Kysely (2008) 所比較的拔靴法以及 Profile likelihood method、Test Inversion Bootstrap (TIB) 於淹水頻率分析，在其模擬比較中，TIB 的信賴區間覆蓋率表現明顯優於其他方法的估計，雖然該研究所使用的 TIB 必須符合母體為廣義極端值分布的假設，在 block maxima 的模型假設下 TIB 仍被建議用於建立信賴區間。除此之外，Schendel and Thongwichian (2017) 針對部分延時數列使用百分比拔靴法、Profile likelihood method 與 TIB 建立流量的信賴區間，針對卜瓦松分布與廣義柏拉圖分布的模擬結果顯示百分比拔靴法將低估信賴區間的長度，TIB 在重現期大於 100 年時表現最佳，此外雖然仍然低估，Profile likelihood method 的雙邊信賴區間在重現期小於 50 年時表現會優於 TIB。

在考量資料選取標準、現實中的降雨發生過程以及實作複雜度等因素，本研究提出事件最大值數列法作為頻率分析方法。事件最大值數列法以單變數事件頻率分析為基礎，使用最低間隔時間從降雨紀錄切割出颱風、對流雨、梅雨以及鋒面雨等四種類型降雨事件，再以事件降雨量分布與每年的事件數分布計算年最大值之混合分布並推估給定重現期下的回歸水準。模擬結果顯示事件最大值數列法較傳統年最大值數列法之結果有較低的均方根誤差，此外使用廣義極端值分布所得到的推估值可能因為事件數不足而產生更大的誤差。最後，以兩種方法針對臺灣五堵、頭汴坑與嘉義雨量站進行降雨頻率分析和尖鋒流量分析，再以拔靴法建立回歸水準的信賴區間並比較結果。

第三章 年最大值數列法



3.1 降雨頻率分析

在降雨事件中，事件的持續時間稱為延時，並且固定延時之下的事件降雨量可以視為一隨機變數，因此不同延時的事件降雨量都有其相應的機率密度函數可以描述其降雨量的發生機率。在降雨頻率分析中所使用的延時稱為設計延時 (design duration)，藉由決定設計延時並收集相應的雨量資料，工程師可以決定特定延時下的事件降雨量之機率分布，進而推估特定累積機率或者超越機率所對應之設計降雨量，作為工程設計與風險評估的參考。

假設特定延時下，測站每年最大的降雨量為一隨機變數 X ，稱為年最大值降雨，其中 X 超過降雨量 x 的機率為 p ，並將之定義為成功事件，則每一年所發生的年最大值降雨事件可以視為一次成功機率為 p 的白努力隨機試驗。其中，直到第一次成功發生所需要經過的試驗次數——即 X 第一次超過 x 所需要經過的試驗次數，會服從成功機率為 p 的幾何分布，平均達到成功所需要的試驗次數為幾何分布之期望值 $1/p$ ，因為試驗的間隔時間為年，事件成功發生所需要的平均時間為 $1/p$ 年，稱為重現期或者回歸年。因為 X 為最大降雨量，重現期也代表測站所遭遇的所有降雨降雨量超過 x 的平均時間，重現期所對應的 x 則稱為回歸水準，即設計暴雨中的設計降雨量。

在以頻率分析決定設計暴雨的降雨量之後，必須決定設計雨型才能將設計雨量分配至事件的各個小時。設計雨型為設計暴雨在各小時的降雨比例，其長度等於設計延時。藉由頻率分析與設計雨型，可以模擬特定重現期與設計延時下的暴雨事件並進行工程上的分析。

3.2 年最大值降雨數列法

為了推估年最大值降雨量的分布，不同的抽樣方法與模型被提出，年最大值數列法作為傳統的樣本選取方法，因為其快速簡單的執行過程而廣為應用於水文頻率分析之中。

在給定的設計延時下，藉由移動窗口從每年的降雨紀錄中抽取最大的累積降雨量，可以得到年最大值數列並推估其分布。以小時記錄為例，假設降雨紀錄的時

間總長度為 N 年、每年的降雨紀錄長度為 n_i 小時, $i=1, \dots, N$, $x_{i,k}$ 為第 i 年第 k 小時的雨量, 則在設計延時 D 小時下的年最大值數列 $\{a_i\}$ 可以表示為:

$$\{a_i \mid a_i = \max_{k \in \{1, \dots, n_i - D + 1\}} \left(\sum_k^{k+D-1} x_{i,k} \right)\}_{i=1, \dots, N} \quad (3-1)$$

再以統計分布配適之。因為數列的抽樣時間單位為年, 因此年最大值降雨量分布的超越機率之倒數可以直接對應雨量以年為單位的重現期。圖 3-1 為降雨紀錄長度 $n_i=10$ 小時、設計延時 $D=2$ 小時、紀錄長度 $N=10$ 年的年最大值數列法之選取範例。

年最大值數列法的樣本選取方便並且計算過程快速, 然而其資料選取的過程中並不會考慮降雨事件, 因此當移動窗口可以有更高的累積雨量時, 年最大值數列會產生跨降雨事件的樣本。在設計兩型中降雨發生於整場事件, 然而跨事件的樣本的降雨量之間具有沒降雨的間隔, 因此此類樣本將使設計暴雨高估雨量。圖 3-2 為嘉義雨量站 24 小時年最大值數列在 1969 年的樣本, 由於單一事件的累積雨量均小於跨事件累積雨量, 此樣本涵蓋了兩場降雨事件, 事件之間隔時間會使流量無法累積, 然而在設計暴雨中, 降雨量將被分配至每一小時, 因此兩場事件所造成的降雨量會被視為單一事件的降雨, 造成高估。

3.3 機率分布

3.3.1 常見分布

根據經濟部水利署水文設計應用手冊 (2001), 國內經常用於降雨頻率分析的統計分布包括廣義極端值分布與 Gumbel 分布以外, 還有皮爾遜第三型分布、對數皮爾遜第三型分布以及三參數對數常態分布, 它們的累積分布函數 $F(x)$ 與其機率密度函數 $f(x)$ (Probability Density Function, PDF) 列於表 3-1。

本研究使用皮爾遜第三型分布以及廣義極端值分布針對年最大值數列進行分析。皮爾遜第三型分布的機率密度函數為:

$$\frac{1}{\theta \Gamma(\beta)} \left(\frac{x - \gamma}{\theta} \right)^{\beta-1} e^{-\left(\frac{x - \gamma}{\theta} \right)} \quad (3-2)$$

其中 $\theta > 0$, $-\infty < \gamma < \infty$, $-\infty < \beta < \infty$ 。此分布實為伽瑪分布加上一平移 (位置參數, location parameter) 而得到的分布, 因此具有位置、尺度與形狀三項參數。除了

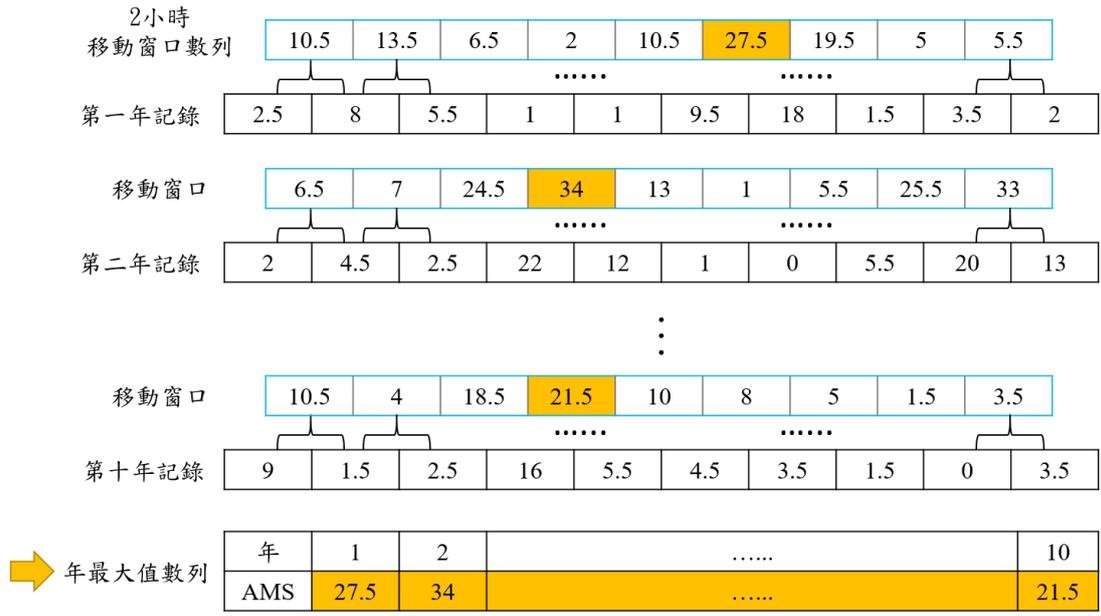
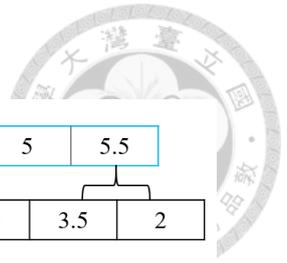


圖 3-1 年最大值數列法示意圖

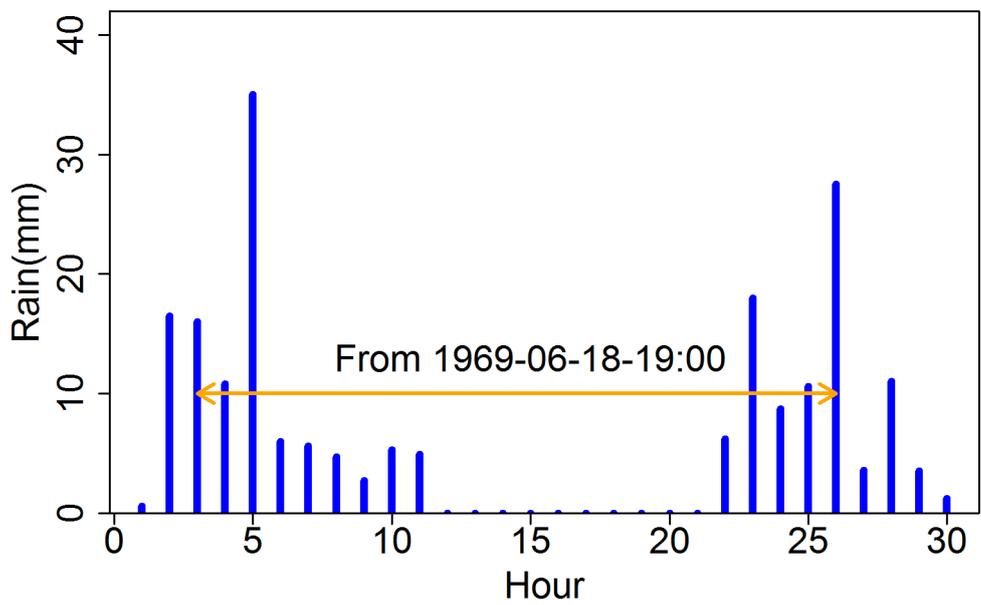
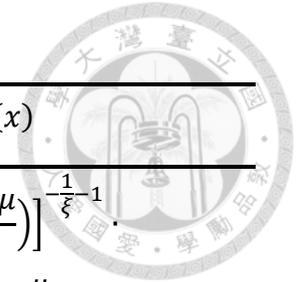


圖 3-2 嘉義雨量站 1969 年 24 小時年最大值數列跨事件樣本

表 3-1 國內降雨頻率分析常用統計分布



Distribution	Parameters	CDF $F(x)$	PDF $f(x)$
廣義極端值 分布	$1 + \xi ((m - \mu) / \sigma) > 0$ $-\infty < \mu < \infty, \sigma > 0, -\infty < \xi < \infty$	$\exp\{-[1 + \xi(\frac{x - \mu}{\sigma})]^{-1/\xi}\}$	$\frac{1}{\sigma} [1 + \xi(\frac{x - \mu}{\sigma})]^{\frac{1}{\xi}-1} \cdot \exp\{-[1 + \xi(\frac{x - \mu}{\sigma})]^{-1/\xi}\}$
Gumbel 分 布	$a > 0, -\infty < b < \infty$	$\exp\left\{-\exp\left[-\left(\frac{x - b}{a}\right)\right]\right\}$	$\frac{1}{a} \exp\left\{-\left[\left(\frac{x - b}{a}\right) + \exp\left(-\left(\frac{x - b}{a}\right)\right)\right]\right\}$
皮爾遜第三 型分布	$\theta > 0, -\infty < \gamma < \infty, -\infty < \beta < \infty$	$\frac{1}{\theta \Gamma(\beta)} \int_0^x \left(\frac{u - \gamma}{\theta}\right)^{\beta-1} e^{-\frac{u - \gamma}{\theta}} du$	$\frac{1}{\theta \Gamma(\beta)} \left(\frac{x - \gamma}{\theta}\right)^{\beta-1} e^{-\frac{x - \gamma}{\theta}}$
對數皮爾遜 第三型分布	$\theta > 0, -\infty < \gamma < \infty, -\infty < \beta < \infty$	$\frac{1}{\theta \Gamma(\beta)} \int_0^x \frac{1}{u} \left(\frac{\ln(u) - \gamma}{\theta}\right)^{\beta-1} e^{-\frac{\ln(u) - \gamma}{\theta}} du$	$\frac{1}{\theta x \Gamma(\beta)} \left(\frac{\ln(x) - \gamma}{\theta}\right)^{\beta-1} e^{-\frac{\ln(x) - \gamma}{\theta}}$
三參數對數 常態分布	$\sigma > 0, -\infty < a < \infty, -\infty < \mu < \infty$	$\int_a^x \frac{1}{(u - 1)\sigma\sqrt{2\pi}} e^{-\frac{[\ln(u - a) - \mu]^2}{2\sigma^2}} du$	$\frac{1}{(x - 1)\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x - a) - \mu]^2}{2\sigma^2}}$

臺大生物環境系統工程學系.(2001)



各參數的意義比較容易解釋之外，三個參數對於資料配適的自由度也比較高。Wu et al. (2011) 針對皮爾遜第三型分布建立其線性動差比圖之 95% 接受域 (acceptance region)，可以提供使用線性動差進行估計與檢定的基礎。

3.3.2 極端值理論與廣義極端值分布

廣義極端值分布是年最大值數列最常使用的分布之一，其應用是基於極端值理論而來。Fisher and Tippett 在 1928 年針對樣本的最大與最小順序統計量推導出它們的漸進分布，而後 Gnedenko (1943) 進一步發展出極端值理論，提供最大順序統計量逼近至廣義極端值分布的基礎，並且被 Gumbel (1958) 應用於其所提出的 block maxima 模型中。由 Coles et al. (2001)，令 X_1, \dots, X_n 為 n 個獨立並且服從相同分布 F_X 的隨機變數，則最大順序統計量 M_n 與其累積分布函數可以定義為：

$$M_n = \max\{X_1, \dots, X_n\} \quad (3-3)$$

$$\begin{aligned} F_{M_n}(m) &= P(M_n \leq m) = P(X_1 \leq m, \dots, X_n \leq m) \\ &= P(X_1 \leq m) * \dots * P(X_n \leq m) = F_X(m)^n \end{aligned} \quad (3-4)$$

由極端值理論可知，如果存在常數數列 $\{a_n > 0\}$ 、 $\{b_n\}$ 滿足：

$$P\left(\frac{M_n - b_n}{a_n} \leq m\right) \rightarrow G(m) \text{ as } n \rightarrow \infty \quad (3-5)$$

並且 G 為非退化分布，則 G 必為下列三種分布之一：

$$\text{I: } G(m) = \exp\left\{-\exp\left[-\left(\frac{m-b}{a}\right)\right]\right\}, \quad -\infty < m < \infty \quad (3-6)$$

$$\text{II: } G(m) = \begin{cases} 0, & m \leq b \\ \exp\left\{-\left(\frac{m-b}{a}\right)^{-\alpha}\right\}, & m > b \end{cases} \quad (3-7)$$

$$\text{III: } G(m) = \begin{cases} \exp\left\{-\left[-\left(\frac{m-b}{a}\right)^\alpha\right]\right\}, & m < b \\ 1, & m \geq b \end{cases} \quad (3-8)$$

其中 $a > 0, b$ ，並且 $\alpha > 0$

三種分布分別為 Gumbel, Fréchet 和 Weibull 分布，並且均屬於 GEV 分布：

$$G(m) = \exp\left\{-\left[1 + \xi\left(\frac{m-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (3-9)$$

其中， $1 + \xi\left(\frac{m-\mu}{\sigma}\right) > 0, -\infty < \mu < \infty, \sigma > 0, -\infty < \xi < \infty$ ，

當 ξ 等於 0 時， $G(m)$ 為 Gumbel 分布， ξ 大於 0 時為 Fréchet 分布， ξ 小於 0 時為

Weibull 分布，將 $G(m)$ 對 m 微分可以得到廣義極端值分布的機率密度函數：

$$g(m) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{m - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi} - 1} * \exp \left\{ - \left[1 + \xi \left(\frac{m - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (3-10)$$

在 block maxima 模型中，以年為 block 單位進行抽樣，因此年最大值即為每一 block 的最大順序統計量，當每年的樣本數足夠大時將會逼近廣義極端值分布，此為傳統年最大值數列被認為可以使用廣義極端值分布逼近的原因。然而，年最大值數列與極端值理論的假設實際上並不相符，在降雨事件數目上，最大順序統計量必須在每年的降雨事件數 n 趨近於無限時才能逼近 GEV 分布，在現實中只要事件數目足夠大即可成立，然而以颱風為例，每年登陸臺灣的颱風數不會超過十個，樣本數量並無法保證降雨分布收斂；除此之外，由於現實中的降雨事件數有限，當每年的降雨事件數 n 並不固定時，年最大值數列會由來自不同樣本數的最大順序統計量組成，因此數列並非具有相同分布的隨機樣本，無法以單一分布進行估計。更進一步地說，年最大值數列是以移動窗口針對降雨紀錄進行資料選取，選取的資料可能涵蓋兩場以上的降雨事件，因此年最大值數列的抽樣並不是從一群樣本之中選取最大值，而是沒有考慮降雨事件，也沒有樣本數 (block size) 概念的抽樣；即使以一天或者一小時作為樣本的單位 (block size=365 or 8760)，降雨的連續性也會使樣本之間彼此不獨立。因此，年最大值數列中並不是從獨立且相同分布的樣本中選取而來，無法直接套用至 block maxima 模型之中。最後，在考量降雨事件的情形下臺灣的降雨事件可以大致區分為梅雨、颱風雨、對流雨以及鋒面雨，每一種降雨類型的成因、特性皆不相同，所以事件降雨量分布也不會相同。在 GEV 的假設中所有樣本皆來自相同且獨立的分布，由於每年的年最大值降雨可能來自任一甚至複數類型的降雨，因此即使每年的降雨事件數目一致，不同類型的年最大值降雨量分布仍然會逼近至不同的 GEV 分布，造成年最大值數列資料分布的不一致。

3.4 參數推估

3.4.1 動差法

動差法 (Method of Moments, MOM) 的發展比多數參數推估方法都還要早，其估計過程也比較簡單快速。假設 X 為一隨機變數分布為 $f(x; \theta_1, \dots, \theta_k)$ ， x_1, \dots, x_n 為取自 X 的隨機樣本，則動差估計式可以由前 k 階的樣本動差去估計母體動差而得。定義樣本動差：



$$m_p = \frac{1}{n} \sum_{i=1}^n x_i^p, p = 1, \dots, k \quad (3-11)$$

母體動差則為：

$$\mu'_p = E[X^p], p = 1, \dots, k \quad (3-12)$$

其中，母體動差也同時會是 $\theta_1, \dots, \theta_k$ 的函數，因此藉由 k 個動差的等式：

$$m_p = \mu'_p = \mu'_p(\theta_1, \dots, \theta_k), p = 1, \dots, k \quad (3-13)$$

可以求得 $\theta_1, \dots, \theta_k$ 的估計式。

動差估計的優點在於計算簡單並且快速。然而並不是所有分布的參數都可以使用動差法進行估計，並且動差估計缺乏其他估計法所擁有的例如一致性、有效性等統計性質，因此動差法比較常被用於估計參數的初始猜測值。

3.4.2 最大概似法

最大概似估計是最廣為使用的估計之一。假設 $\mathbf{x} = x_1, \dots, x_n$ 為取自 $f(x; \theta)$ 、 $\theta = \theta_1, \dots, \theta_k$ 的隨機樣本，則可以定義概似函數 (likelihood function):

$$L(\theta_1, \dots, \theta_k | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta_1, \dots, \theta_k) \quad (3-14)$$

因為樣本之間彼此獨立，聯合機率函數可以拆開成為個別機率函數相乘：

$$f(x_1, \dots, x_n | \theta_1, \dots, \theta_k) = \prod_{i=1}^n f(x_i | \theta_1, \dots, \theta_k) \quad (3-15)$$

並且可以定義最大概似估計：

$$\hat{\theta}(\mathbf{x}) = \underset{\theta(\mathbf{x})}{\operatorname{argmax}} L(\theta_1, \dots, \theta_k | x_1, \dots, x_n) \quad (3-16)$$

在給定 θ 的情況下，概似函數其實就是 x 的密度（或者機率）乘積，因此藉由最大化概似函數，可以計算有最大概似取出樣本 \mathbf{x} 的參數 $\hat{\theta}(\mathbf{x})$ 。一般而言最大概似法的求解可以使用微分計算，當無法以微分直接求解時，即必須依靠數值方法求解。最大概似法具有許多良好的性質，例如隨著樣本增加而趨近無偏估計的一致性 (consistency)、漸進常態 (asymptotic normality)、漸進有效性 (asymptotic efficiency) 以及 MLE 在經過函數轉換後依然是轉換後參數之 MLE 的不變性 (invariance property)，然而 MLE 的精準度很容易受到樣本大小的影響，當樣本數比較低時必須謹慎使用。

3.4.3 線性動差法

Hosking and Wallis (1997) 介紹線性動差法並將之應用於區域頻率分析之中。

線性動差由 Greenwood et al. (1979) 所提出的機率加權動差 (Probability Weighted Moment, PWM) 線性組合而得，因此在機率加權動差所得到的估計結果也可以使用線性動差表示。針對分布為 $F(\cdot)$ 的隨機變數 X ，機率加權動差可以定義為：

$$M_{p,r,s} = E[X^p \{F(X)\}^r \{1 - F(X)\}^s] \quad (3-17)$$

其中 $\alpha_s = M_{1,0,s}$ 與 $\beta_r = M_{1,r,0}$ 為機率加權動差中經常使用的兩種特例。除了機率加權動差之外，線性動差也可以表示為順序統計量的期望值之線性組合。讓 $X_{i:n}$ 表示 n 個 i.i.d 隨機變數 X 中的第 i 順序統計量，則可以定義第 r 階線性動差 λ_r ：

$$\lambda_r \equiv r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E[X_{r-k:r}], r = 1, 2, \dots \quad (3-18)$$

順序統計量的期望值為：

$$E[X_{j:r}] = \frac{r!}{(j-1)!(r-j)!} \int x \{F(x)\}^{j-1} \{1 - F(x)\}^{r-j} dF(x) \quad (3-19)$$

因此第 r 階線性動差可以表示為：

$$\lambda_r = \int_0^1 x(F) P_{r-1}^*(F) dF, r = 1, 2, \dots \quad (3-20)$$

其中，

$$P_r^* = \sum_{k=0}^r p_{r,k}^* F^k \quad (3-21)$$

$$p_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \quad (3-22)$$

使用上述公式可以將線性動差分別以機率加權動差 (在此使用 β_r) 和順序統計量表示，在此列出第一到四階動差：

$$\lambda_1 = \beta_0 = E[X_{1:1}] \quad (3-23)$$

$$\lambda_2 = 2\beta_1 - \beta_0 = \frac{1}{2} E[X_{2:2} - X_{1:2}] \quad (3-24)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 = \frac{1}{3} E[X_{3:3} - 2X_{2:3} + X_{1:3}] \quad (3-25)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 = \frac{1}{4} E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}] \quad (3-26)$$

順序統計量為線性動差提供更直觀的解釋意義。與傳統中心動差相似，一階線性動差代表資料的中心、二階動差代表資料的尺度，並定義線性動差比 τ_r ：



$$\tau_r = \frac{\lambda_r}{\lambda_2}, r = 1, 2, 3, \dots \quad (3-27)$$

其中 λ_1 、 λ_2 和線性偏態係數 τ_3 (L-skewness) 以及線性峰態係數 τ_4 (L-kurtosis) 經常應用於機率分布之參數估計。線性動差可以使用機率加權動差或者 Plotting Position Estimator 進行估計，針對樣本數為 n 的資料，第 r 階樣本機率加權動差的不偏估計式為：

$$b_r = n^{-1} \sum_{i=1}^n \frac{(i-1)(i-2) \dots (i-r)}{(n-1)(n-2) \dots (n-r)} x_{i:n} \quad (3-28)$$

則第 r 階樣本線性動差 l_r 與線性動差比 t_r 可以表示為：

$$l_r = \sum_{k=0}^{r-1} p_{r-1,k}^* b_k, r = 1, 2, \dots, n \quad (3-29)$$

$$t_r = \frac{l_r}{l_2} \quad (3-30)$$

Plotting Position Estimator 為 $F(X_{i:n})$ 的估計式，此估計式並沒有固定的公式，Hosking (1990) 指出合理的估計式包括 $p_{i:n} \equiv (i + \gamma)/(n + \delta)$, $\delta > \gamma > -1$ ，其中 $p_{i:n} = (i - 0.35)/n$ 對於廣義極端值分布估計的結果表現良好。使用 Plotting Position Estimator 估計的線性動差 $\tilde{\lambda}_r$ 與線性動差比 \tilde{t}_r 為：

$$\tilde{\lambda}_r = \frac{1}{n} \sum_{j=1}^n P_{r-1}^*(p_{j:n}) x_{j:n}, r = 1, 2, \dots, n \quad (3-31)$$

$$\tilde{t}_r = \frac{\tilde{\lambda}_r}{\tilde{\lambda}_2} \quad (3-32)$$

兩種線性動差的估計方法雖然是有偏估計量，卻都具有一致性——在樣本數量為中等乃至大數量時會趨近於無偏估計，然而 Hosking and Wallis (1997) 指出 Plotting Position Estimator 可能產生不合理的估計、不具有尺度不變性 (scale invariant) 並且一般在線性動差比的估計會高於機率加權動差，因此機率加權動差法更適合用於線性動差與線性動差比的估計。Hosking (1985) 比較最大概似法、機率加權動差以及 Sextiles 方法對於 GEV 分布的估計情形，結果顯示因為具有一致性，當樣本數中等或者偏大 (≥ 50) 時，線性動差法與 MLE 的估計結果相近；當樣本數比較小時，雖然線性動差法的估計量偏誤略大於 MLE，其估計量的標準差卻顯著低於 MLE，因此線性動差法比 MLE 更適合用於估計 GEV 分布的參數。



3.5 機率適合度檢定

3.5.1 機率圖

機率圖 (probability plot) 藉由將資料點與理論分布作圖，可以由視覺判斷資料與假設分布是否相符。假設 X 是分布為 $F(\cdot)$ 的隨機變數，並且 G 為經過標準化後的 X 的累積分布函數：

$$F(x) = G\left(\frac{x - \mu}{\sigma}\right) = G(z) \quad (3-33)$$

則 Z 為 X 經過標準化後的隨機變數：

$$Z = G^{-1}(F(X)) = \frac{x - \mu}{\sigma} \quad (3-34)$$

由上式可以知道當分布 F 和分布 G 所對應的是隨機變數標準化前後的相同分布時，藉由將 X 對 Z 作圖，甚至無須標準化資料也可以獲得線性關係。在實作中通常使用經驗累積分布函數作為 F ，理論分布則為 G ，因此可以藉由資料點是否明顯偏離線性來進行檢定。累積經驗函數 F 的計算方式不只一種，令 $x_{(1)}, \dots, x_{(n)}$ 為隨機樣本 x_1, \dots, x_n 由小到大排序後的結果，Cunnane (1978) 整理過去的經驗累積分布公式並統一為：

$$F(x_{(i)}) = \frac{i - a}{n + 1 - 2a} \quad (3-35)$$

不同的 a 將對應不同的公式 (表 3-2)。圖 3-3 為以伽瑪分布的資料繪於常態分布機率圖之結果 (未標準化)，因為資料明顯偏離對角線，所以並不符合常態分布。

3.5.2 Kolmogorov-Smirnov 檢定

Kolmogorov-Smirnov 檢定藉由經驗分布函數與理論函數的最大距離計算檢定統計量，可以對任意理論分布進行檢定。假設 X 為一連續型隨機變數， x_1, \dots, x_n 為 X 之隨機樣本， $F(\cdot)$ 為欲檢定之理論分布， $F_n(\cdot)$ 為使用表 3-2 之 California formula 計算之經驗分布函數。定義：

$$D_n = \sup_{-\infty < x < \infty} |F(x) - F_n(x)| \quad (3-36)$$

除此之外，定義：

$$K = \sup_{t \in (0,1)} |B(t)| \quad (3-37)$$

其中 $B(t)$ 為一布朗橋 (brownian bridge)， K 之分布為 Kolmogorov 分布：

$$P(K \leq x) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2} = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{\infty} e^{-\frac{(2i-1)^2 \pi^2}{8x^2}} \quad (3-38)$$

當樣本的分布等於假設之理論分布時， $\sqrt{n}D_n$ 的分布會趨近於 Kolmogorov 分布：

$$\sqrt{n}D_n \xrightarrow{d} \sup_t |B(F(t))| \text{ as } n \rightarrow \infty \quad (3-39)$$

\xrightarrow{d} 表示分布收斂。因此任何分布均可以使用檢定量 $\sqrt{n}D_n$ 與 Kolmogorov 分布計算 p 值。 $\sqrt{n}D_n$ 的分布除了以 Kolmogorov 分布逼近之外，也可以藉由中央極限定理以常態分佈逼近。在虛無假設下（即樣本分布為 $F(\cdot)$ ），針對任一固定 x ：

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x) \quad (3-40)$$

其中 x_i 為獨立並且服從 $F(\cdot)$ 的隨機變數，因此 $I(x_i \leq x)$ 可視為成功機率為 $F(x)$ 的白努力隨機變數， $nF_n(x)$ 則為服從二項分布， $F_n(x)$ 之期望值與變異數為：

$$E[F_n(x)] = F(x) \quad (3-41)$$

$$\text{Var}(F_n(x)) = \frac{1}{n} F(x)[1 - F(x)] \quad (3-42)$$

由中央極限定理， $\sqrt{n}D_n = \sqrt{n}|F(x) - F_n(x)|$ 的分布將逼近常態分佈：

$$\sqrt{n}D_n \xrightarrow{d} N(0, F(x)[1 - F(x)]) \quad (3-43)$$

因此 $\sqrt{n}D_n$ 之 p 值或者假設檢定之臨界值 (critical value) 可以由常態分布獲得。

3.5.3 線性動差比圖

線性偏態係數 τ_3 以及線性峰態係數 τ_4 除了用於參數估計之外，也可以用於進行機率適合度檢定。傳統水文變數的適合度檢定步驟通常包含資料分布的假設、參數推估以及檢定量計算，因此分布的選擇與參數推估方法對於結果是否顯著相當重要。另一種檢定方法是使用動差比圖 (D'Agostino and Stephens 1986)，動差為各個分布所獨有的資訊，因此可以作為適合度檢定的依據，藉由將理論分布與樣本的峰態係數 (kurtosis) 對偏態係數作圖並比較距離或者建立接受域，可以判斷樣本是否來自該理論分布。動差比圖計算簡便並且不需要預先假設資料的分布，可以降低選擇造成誤差的可能性。除了偏態係數與峰態係數，Tsukatani and Shigemitsu (1980) 指出由於無法準確估計四階動差，建議使用變異數與偏態係數對皮爾遜類型的分布進行比較。Hosking 與 Wallis (1997) 指出雖然樣本偏態係數

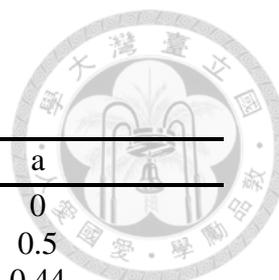


表 3-2 經驗累積分布公式

名稱	公式	a
Weibull	$i / (n + 1)$	0
Hazen	$(i - 0.5) / n$	0.5
Gringorten	$(i - 0.44) / (n + 0.12)$	0.44
Blom	$(i - 0.375) / (n + 0.25)$	0.375
the California formula	i / n	-

Cunnane (1978)

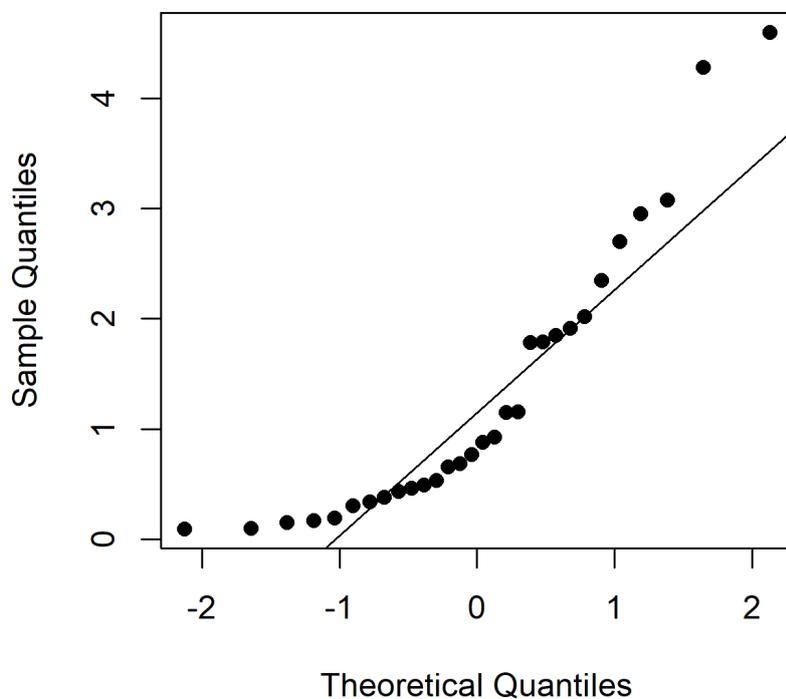


圖 3-3 伽瑪分布資料繪於常態機率圖

與樣本峰態係數會逼近常態分布，但在樣本數目中中等偏小時逼近的效果並不佳；Liou et al. (2008) 提及常態分布的樣本偏態係數與峰態係數的散布形狀會隨著樣本數增加而改變，因此難以使用數學式描述；這些情形造成動差比圖在建立接受域的過程中將面對許多困難。Hosking and Wallis (1997) 提出線性動差比圖 (圖 3-4) 作為機率適合度檢定的方法，樣本線性偏態係數以及樣本線性峰態係數的估計偏誤除了遠低於樣本偏態係數與樣本峰態係數之外，在一些理論分布中 (Ex. Gumbel) 它們的散布情形也接近雙變數常態分布、線性峰態係數可以比峰態係數提供更多的資訊，提供適合度檢定更準確的結果。在線性動差比圖中，每一個雙參數分布都為一個點、三參數分布則為線。Liou et al. (2008) 以及 Wu et al. (2011) 藉由模擬建立常態分布、Gumbel 分布與皮爾遜第三型分布在 LMRD 的 95% 接受域，此接受域會隨著樣本數的多寡而變化，藉由觀察樣本繪於 LMRD 的點是否落於機率分布之接受域，可以決定是否拒絕樣本來自理論分布的假設。

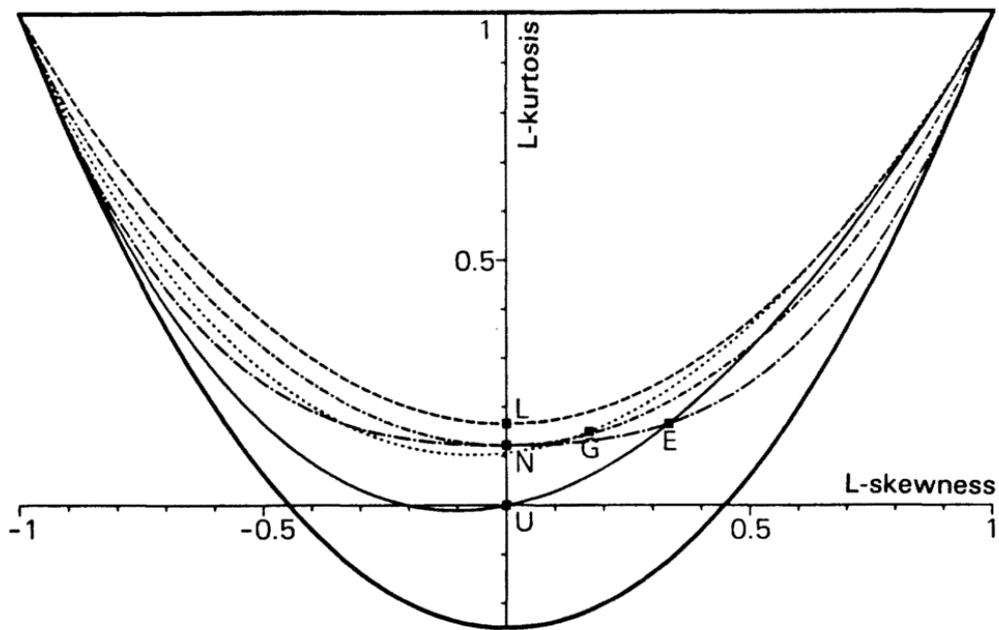
本研究參考前述方法，計算樣本之線性動差並使用 LMRD 進行適合度檢定與參數估計。卜瓦松分布與負二項分布之參數則以最大概似法推估，並使用 Conover (1972) 所提供的修正方法，以 K-S test 進行檢定。

3.6 降雨強度—延時—頻率曲線

在獲得不同設計延時以及不同重現期下的設計降雨後，一般會使用降雨強度—延時—頻率曲線或者深度—延時—頻率曲線 (Depth-Duration-Frequency curve, DDF curve) 呈現 (圖 3-5)。此處降雨深度即設計降雨量，延時則為設計降雨所對應之設計延時，降雨強度為單位時間內的平均降雨深度，可以由降雨深度除以延時獲得。兩種曲線分別代表在不同的重現期下，降雨強度/降雨深度與延時的關係，因為強度與深度可以經由延時互相轉換，因此兩種曲線所提供的資訊是相等的。一般而言降雨和延時並非線性關係，隨著事件延時增加，降雨深度會隨之增加，降雨強度則會降低，然而增加量與降低量將隨著延時增加而遞減。在獲得不同重現期下延時與降雨強度的資料點後，可以使用降雨強度—延時公式予以分析。國內常用的暴雨公式包括：

$$(1) \text{ Talbot 公式: } I = \frac{a}{d+b} \quad (3-44)$$

$$(2) \text{ Sherman 公式: } I = \frac{a}{d^b} \quad (3-45)$$



Distributions: U uniform, E exponential, G Gumbel, L logistic, N Normal,
 — generalized Pareto, generalized extreme value, ---- generalized logistic,
 -·-·- lognormal, - - - gamma, — bound for all distributions

圖 3-4 常見分布的 L-moment ratio diagram (Hosking, 1990)

$$(3) \text{ Horner 公式: } I = \frac{a}{(d+b)^c} \quad (3-46)$$

其中， I 為降雨強度(mm/hr)， a 、 b 、 c 為參數， d 為延時(min)。

由公式可知，Talbot 與 Sherman 公式分別為 Horner 公式 $c = 1$ 以及 $b = 0$ 的特例，因此以 Horner 公式最廣為使用。公式參數的計算可以藉由找出使無因次均方誤差最小的參數：

$$(a, b, c) = \underset{(a, b, c)}{\operatorname{argmin}} \left(\frac{I - \hat{I}}{I} \right)^2 \quad (3-47)$$

\hat{I} 是由公式計算之降雨強度。如果上述目標式所得到的 IDF 曲線發生相交的情形，則可以使用均方誤差最小作為目標式計算：

$$(a, b, c) = \underset{a, b, c}{\operatorname{argmin}} (I - \hat{I})^2 \quad (3-48)$$

在 IDF 曲線中，資料點的延時是使用移動窗口取樣時所設定的設計延時，而非降雨事件的真實延時，因此 IDF 中的資料分布並不是降雨事件的雨量——延時雙變數散布圖，而是針對水利工程設計所繪製的曲線，這點必須特別注意

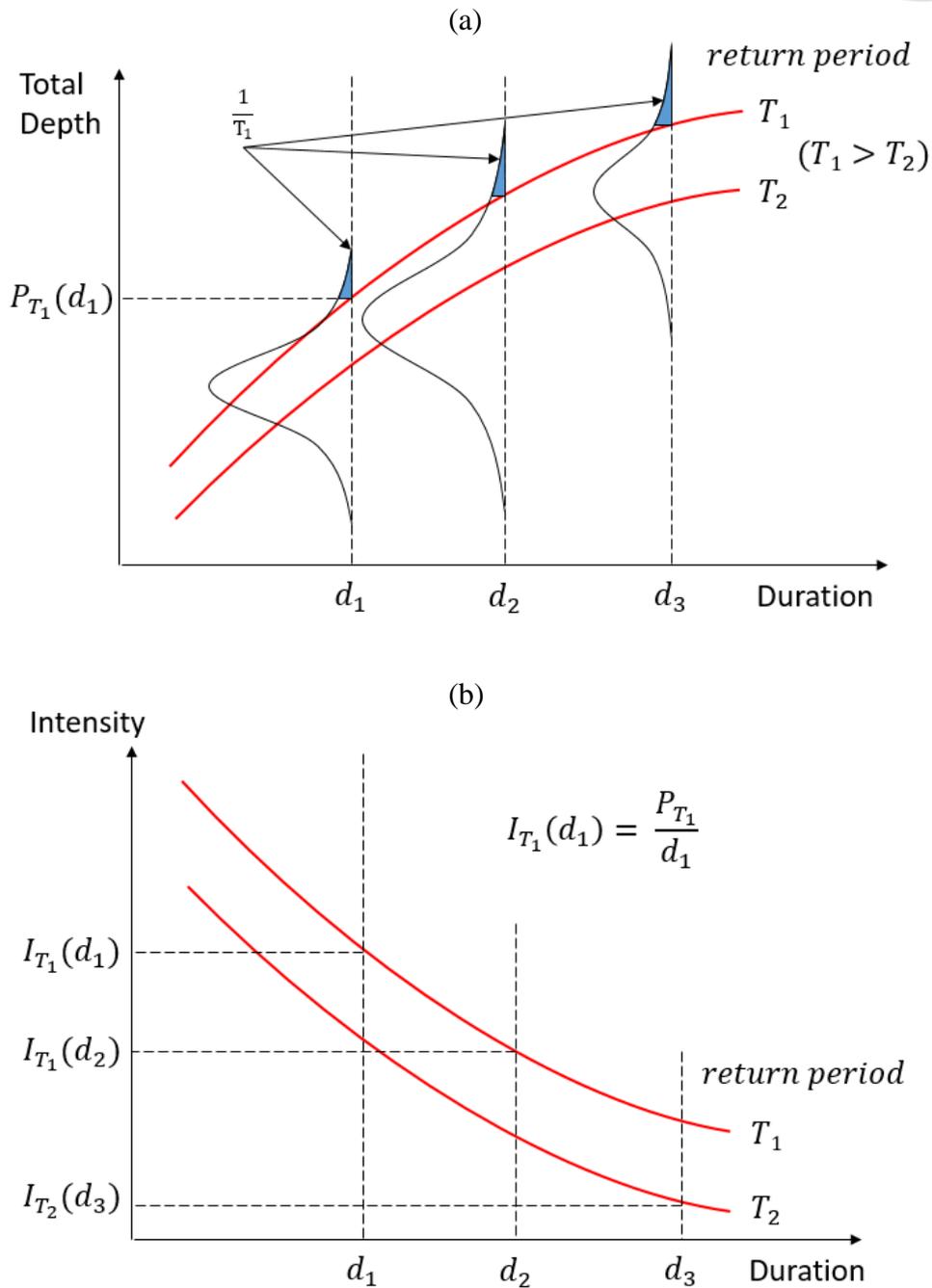


圖 3-5 頻率曲線示意圖(a) DDF 與(b) IDF



第四章 事件最大值數列法

為了處理年最大值數列法在樣本的數量與選取上所遭遇的問題，本研究提出事件最大值數列法作為傳統頻率分析的替代方法。事件最大值數列法將臺灣的降雨分為颱風、梅雨、對流與和鋒面雨等四種類型，當每年的降雨事件數固定時各類型年最大值降雨量分布為一混合分布，並且可以獲得年最大值雨量之邊際分布；而考慮各類型降雨下的年最大值量降雨分布將會是由各類型年最大值降雨分布組成的混合分布。分析結果顯示事件最大值數列法比傳統年最大值數列法更適合應用於暴雨設計。

4.1 年最大值與混合分布

由於年最大值數列法並未考慮降雨事件的影響，因此當其樣本涵蓋兩場以上的降雨事件時，將造成雨量的高估。而極端值理論的假設則顯示在進行頻率分析時必須考量每年的降雨事件發生數。因此頻率分析中的年最大值應該從降雨事件出發來考慮。

以設計延時 24 小時為例，假設每一年的颱風事件數為 n ，颱風事件 24 小時的累積降雨量 X 服從降雨分布 $F_X(\cdot)$ ，則颱風事件的年最大值降雨量 Y_n 為 n 個 X 的最大順序統計量，可以參考式(3-2)、(3-3) 的表達：

$$Y_n = \max\{X_1, X_2, \dots, X_{N-1}, X_n\} \quad (4-1)$$

$$F_{Y_n}(y|N = n) = [F_X(y)]^n \quad (4-2)$$

因為每年的降雨事件數 N 為一隨機變數，因此 Y_n 的分布為固定年事件數 N 為 n 下的條件分布。除此之外，事件數與年最大值降雨的聯合分布是由事件數分布與事件累積雨量分布所組成的混合分布：

$$P(Y_n \leq y, N = n) = P(N = n) \cdot [F_X(y)]^n \quad (4-3)$$

並可以推得颱風 24 小時年最大值降雨的邊際分布：

$$F_{Y_n}(y) = \sum_{n=0}^{\infty} P(N = n) \cdot [F_X(y)]^n \quad (4-4)$$

其他類型的降雨事件——梅雨、對流與以及鋒面雨的 24 小時年最大值降雨量分布也能使用同樣的方法獲得。由於降雨事件彼此獨立，各類型的年最大值降雨量也互相獨立，真正的年最大值降雨量 Y 實為各類型年最大值降雨中的最大值：

$$Y = \max\{Y_1, Y_2, Y_3, Y_4\} \quad (4-5)$$

$Y_i, i=1, 2, 3, 4$ 分別對應颱風、梅雨、對流雨、鋒面雨的 24 小時年最大值降雨，因此 Z 的累積機率分布可以表示為各個類型年最大值降雨分布的混合分布：

$$F_Y(y) = F_{Y_1}(y) \cdot F_{Y_2}(y) \cdot F_{Y_3}(y) \cdot F_{Y_4}(y) \quad (4-6)$$

藉由此混合分布可以找到對應不同重現期的回歸水準。

4.2 事件發生次數分布

在事件年最大值分布的建立過程中，除了特定延時下的事件降雨量分布必須決定之外，每年的降雨事件數分布也需要由使用者決定。年降雨事件數 N 為一離散型隨機變數，本研究將針對各類型降雨使用卜瓦松分布與負二項分布進行模擬並比較。

4.2.1 卜瓦松分布

卜瓦松分布是最常被用於描述事件發生次數的分布，可以視為二項分布的試驗次數趨近於無限的極限情形。其機率質量函數為：

$$P(N = n) = \frac{e^{-\lambda} \lambda^n}{n!}, n = 0, 1, 2, \dots \quad (4-7)$$

其中 $\lambda > 0$ 為事件的平均發生次數，並且 N 的期望值和變異數均為 λ 。除了由二項分布逼近，卜瓦松分布也可以解釋為在固定時間下，平均發生率為常數的事件發生次數分布，因此 λ 可以改以平均發生率 μ 乘以時間間隔 t 表示。在將年事件數分布為參數 λ 的卜瓦松分布假設下，颱風事件年最大值雨量的累積分布函數經由重新參數化可以改寫為：

$$\begin{aligned} F_{Y_n}(y; \lambda) &= \sum_{n=0}^{\infty} P_{pois}(N = n) \cdot F_X(y)^n \\ &= \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} \cdot F_X(y)^n \\ &= e^{-\lambda[1-F_X(y)]} \end{aligned} \quad (4-8)$$

4.2.2 負二項分布

負二項分布的機率質量函數為：



$$P_{NB}(N = n) = \binom{n+r-1}{n} p^n (1-p)^r, n = 0, 1, 2, \dots \quad (4-9)$$

其中 $r > 0, 0 < p < 1$ ， N 的期望值為 $r(1-p)/p$ ，變異數為 $r(1-p)/p^2$ 。並且負二項分布可以藉由假設卜瓦松分布的平均發生率 μ 服從尺度參數為 $1/\beta$ 、形狀參數為 α 的伽瑪分布(即位置參數為 0 的皮爾遜第三型分布)推導而得：

$$\begin{aligned} P_{NB}(N = n) &= \int_0^{\infty} P(N = n|\mu) \cdot f(\mu) \\ &= \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^y}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta \mu} d\mu \\ &= \binom{\alpha+n-1}{n} \left(\frac{t}{\beta+t}\right)^n \left(1 - \frac{t}{\beta+t}\right)^\alpha \end{aligned} \quad (4-10)$$

因此卜瓦松分布實為負二項分布在平均發生率為常數時的特例。在負二項分布的假設下，年最大值雨量的累積分布函數為：

$$\begin{aligned} F_{Y_n}(y; r, p) &= \sum_{n=0}^{\infty} P_{NB}(N = n) \cdot F_X(y)^n \\ &= \sum_{n=0}^{\infty} \binom{n+r-1}{n} p^n (1-p)^r \cdot F_X(y)^n \\ &= \frac{(1-p)^r}{(1-pG(z))^r} \end{aligned} \quad (4-11)$$

4.3 事件最大值數列法

4.3.1 降雨事件切割

在使用 EMS 方法估計年最大值降雨混合分布以前，必須先切割出互相獨立的降雨事件。Powell et al. (2007) 使用指數法計算間隔時間小於 12 小時的 CV 時，發現少有 MIT 的 CV 會接近 1，針對本研究的資料所設定的 1 至 8 小時 MIT 中亦有部分類型的降雨有類似情形，CV 在 8 小時之中不斷下降而沒有靠近 1 的最低值或者 CV 不高於 1(表 4-1)，而 8 小時以上的 MIT 將明顯過高；除此之外，事件到達間隔服從指數分布的假設在颱風以外的降雨事件均為 K.S. test 所拒絕(表 4-2)。考慮臺灣的天氣特性，本研究將降雨事件依照月份分為對應梅雨的春季、對應颱風雨和對流雨的夏季以及對應鋒面雨的冬季降雨。由於對流雨必定於白天發生，將夏季延時 12 小時以下的降雨事件歸類至對流雨、延時大於 12 小時的事件則歸類為颱風雨，並根據中央氣象局颱風資料庫所公布之有發警報颱風列表，將與警報時間重



疊的降雨歸類為颱風事件。此外，為了避免儀器和紀錄過程所產生的誤差和降低低強度降雨事件的影響，本研究將滿足以下任一條件——延時小於兩小時、降雨量 0.5mm 或者降雨強度小於 2mm/hr 的事件移除。四種降雨類型的分類規則列於表 4-3。

各類型降雨的 MIT 選擇範圍設定在 1 到 8 小時之間，從圖 4-1 可以發現當 MIT 大於 4 小時時的降雨事件數將不再明顯受到 MIT 的增加影響。除此之外，由於事件切割的最終目標為提供事件最大值頻率分析進行樣本篩選與後續的逕流分析，過長的事件間隔時間將因為缺少降雨而難以被涵蓋在窗口之中，即使被選取為樣本，考慮到臺灣集水區坡陡流急、集流時間短的特性，此一類型的樣本會因為長時間間隔而在逕流分析中產生雙峰，並不適合作為設計單一暴雨事件的樣本。因此針對各類型降雨，本研究將以 4 小時作為 MIT 進行事件切割。

4.3.2 事件最大值數列

以事件最大值雨量進行頻率分析必須先估計不同類型降雨事件的雨量分布和事件年平均發生率來計算事件的一年最大值降雨量分布，再以這些分布計算年最大值降雨量的最終分布。以颱風事件為例，假設降雨紀錄長達 N 年並且已經完成事件切割，每年的颱風事件數為 $n_i, i=1 \dots N$ ，則第 i 年中第 j 次的颱風事件可以表示為 $X_{i,j} = \{x_{i,j,1}, \dots, x_{i,j,d_{i,j}}\}$ ， $i = 1 \dots N, j = 1 \dots n_i$ ， $x_{i,j,k}$ 代表第 i 年第 j 次颱風事件中第 k 小時的降雨紀錄， $d_{i,j}$ 為該場事件的降雨延時。給定設計延時 D 小時，藉由移動窗口可以選出該事件的最大降雨量 $P_{i,j}$ ：

$$P_{i,j} = \begin{cases} \max_{k \in \{1, \dots, d_{i,j}-D+1\}} \left(\sum_k^{k+D-1} x_{i,j,k} \right) & \text{if } D \leq d_{i,j} \\ \sum_1^{d_{i,j}} x_{i,j,k} & \text{if } D > d_{i,j} \end{cases} \quad (4-12)$$

$$i = 1 \dots N, j = 1 \dots n_i$$

$\{P_{i,j}\}_{i=1 \dots N, j=1 \dots n_i}$ 即為颱風事件的事件最大值數列。利用事件最大值數列，可以推估颱風事件年發生次數的分布以及在設計延時 D 小時下的事件最大降雨量分布，進而推導颱風年最大降雨量的混合分布(4-4)。梅雨、對流雨以及鋒面雨的年最大值降雨分布可以使用同樣過程計算而得，最後，經由(4-6)獲得最終所使用的年最大值降雨分布，以此推估設計延時 D 小時下、不同重現期所對應的回歸水準。



表 4-1 不同 MIT 下事件到達間隔 CV 計算

MIT	01B030				01F680				467480			
	對流	颱風	梅雨	鋒面	對流	颱風	梅雨	鋒面	對流	颱風	梅雨	鋒面
1	1.63	1.25	1.55	1.47	1.65	0.82	1.69	1.82	1.45	1.15	1.63	1.65
2	1.34	1.13	1.36	1.26	1.46	0.98	1.48	1.64	1.35	1	1.46	1.49
3	1.24	0.97	1.31	1.26	1.4	0.95	1.38	1.53	1.27	1.02	1.38	1.4
4	1.16	0.98	1.27	1.2	1.36	0.91	1.29	1.4	1.21	0.98	1.29	1.39
5	1.11	0.96	1.22	1.2	1.27	0.83	1.24	1.34	1.2	0.97	1.2	1.34
6	1.14	1.01	1.19	1.14	1.25	0.84	1.21	1.31	1.17	0.96	1.16	1.31
7	1.1	0.99	1.15	1.13	1.24	0.83	1.18	1.22	1.16	0.93	1.11	1.28
8	1.1	1.02	1.15	1.08	1.23	0.87	1.17	1.13	1.15	0.92	1.09	1.26

表 4-2 不同 MIT 下事件到達間隔針對指數分布的 KS-test 結果 (p-value)

MIT	01B030				01F680				467480			
	對流	颱風	梅雨	鋒面	對流	颱風	梅雨	鋒面	對流	颱風	梅雨	鋒面
1	0	0.01	0	0	0	0.52	0	0	0	0	0	0
2	0	0.4	0	0	0	0.17	0	0	0	0.01	0	0
3	0	0.81	0.01	0	0	0.35	0	0	0	0	0	0
4	0	0.82	0.05	0.1	0	0.79	0	0	0	0.04	0	0
5	0	0.69	0.13	0.14	0	0.76	0	0	0	0.14	0	0
6	0.02	0.81	0.13	0.49	0	0.83	0	0	0	0.12	0	0
7	0.08	0.7	0.13	0.26	0	0.85	0	0.02	0	0.22	0	0
8	0.15	0.76	0.06	0.77	0	0.84	0	0.24	0	0.32	0	0



表 4-3 降雨事件分類規則

降雨類型	月份	延時(hr)	降雨強度
對流雨	7~10月	$12 \geq dr > 1$	2mm/hr 以上
鋒面雨	1~4月、11~12月	$dr > 1$	2mm/hr 以上
梅雨	5~6月	$dr > 1$	2mm/hr 以上
颱風	7~10月	$dr > 12$	2mm/hr 以上

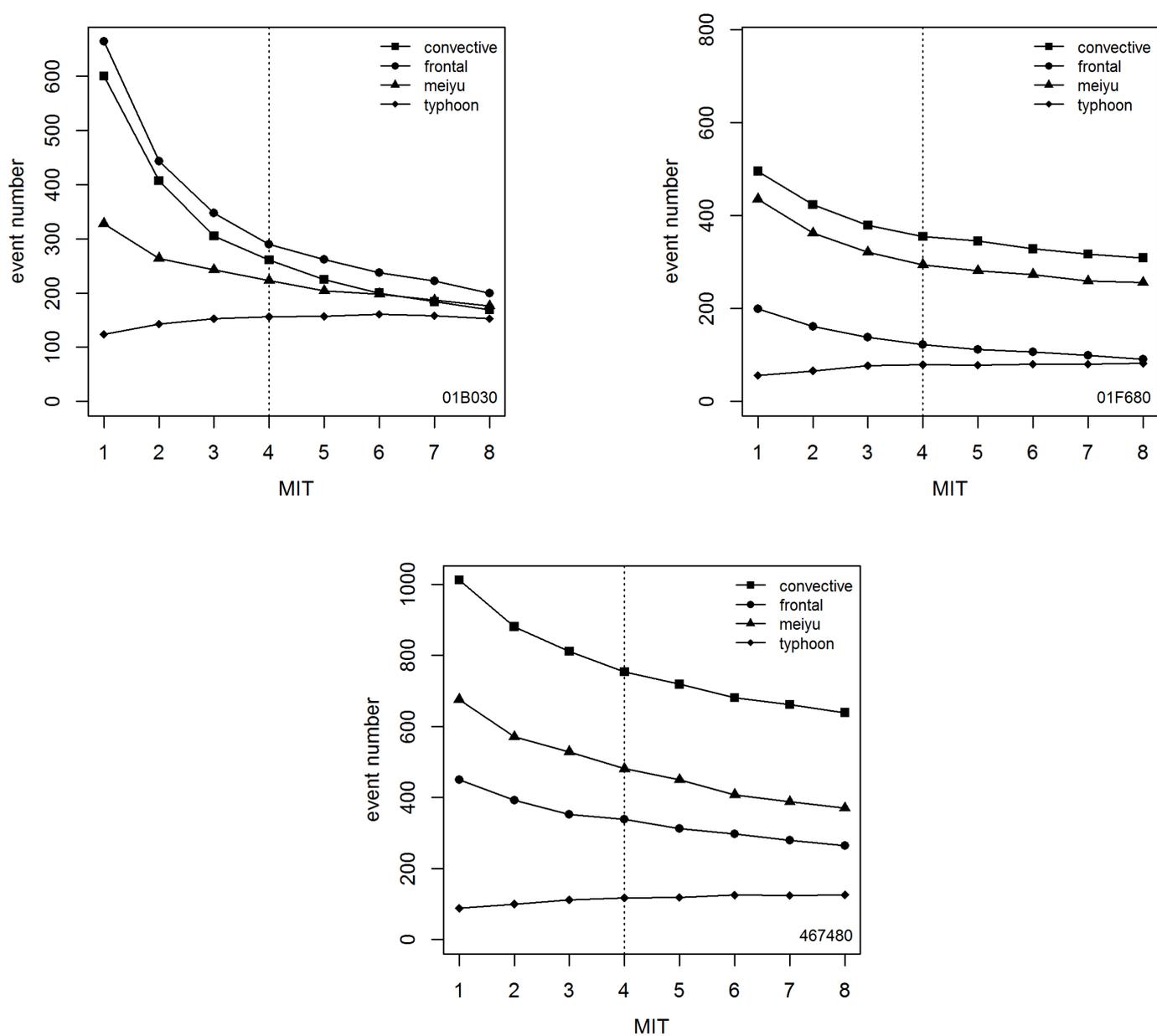


圖 4-1、三測站於各 MIT 下所得到的事件數

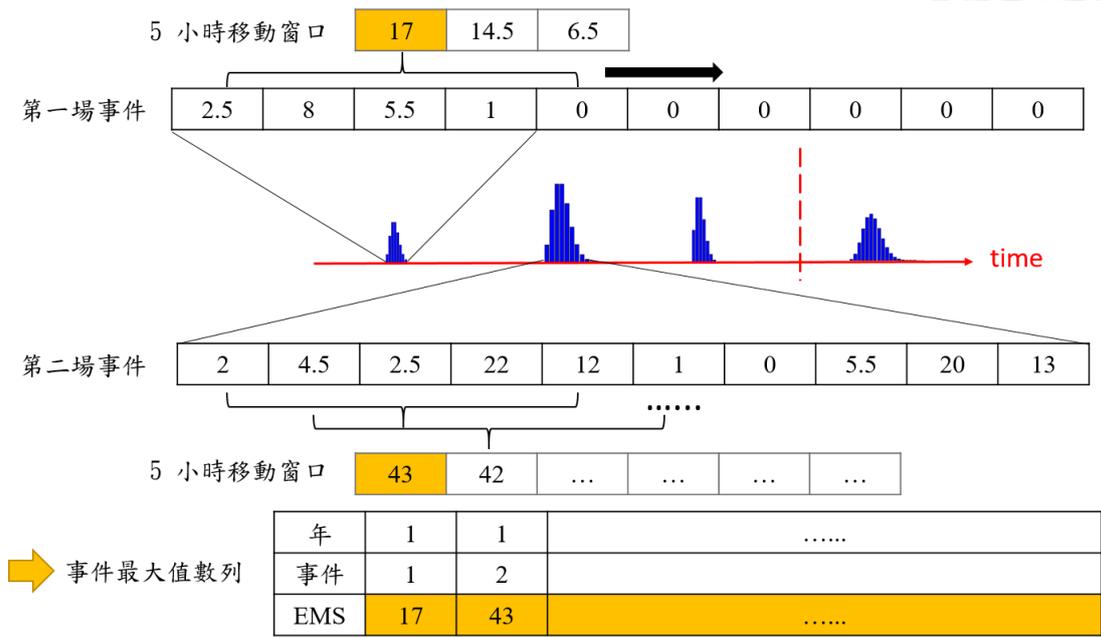
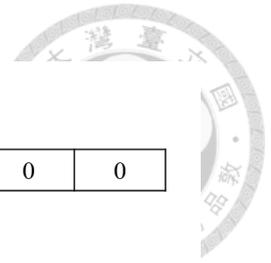


圖 4-2 事件最大值數列法示意圖

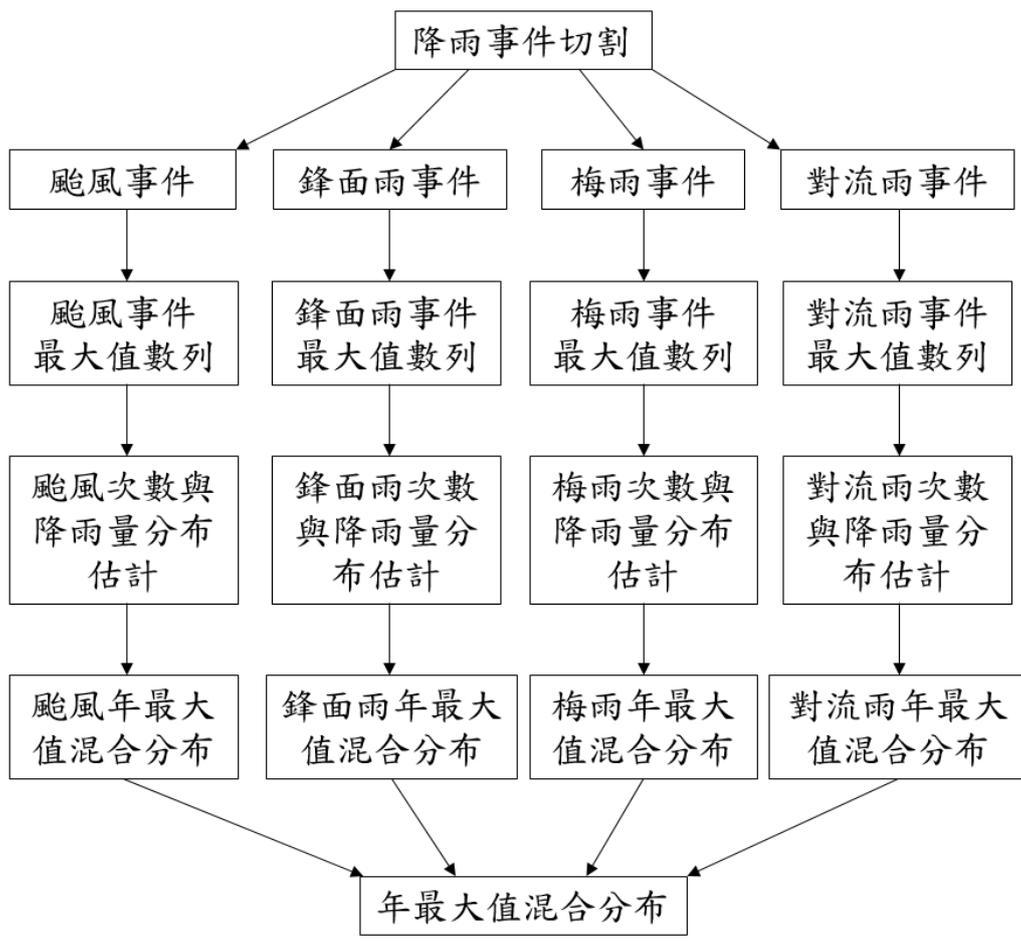


圖 4-3 事件最大值列法流程





第五章 蒙地卡羅模擬

為了比較 EMS 與 AMS 的推估結果，本研究以蒙地卡羅法模擬降雨每年降雨事件並比較年最大值數列法與事件最大值數列法的回歸水準估計結果。降雨事件類型使用颱風與梅雨作為代表，結果顯示在偏誤與均方根誤差的比較上事件最大值數列的表現結果較佳。

5.1 降雨資料生成與頻率分析模擬

颱風和梅雨的降雨紀錄可以經由模擬每年的事件數與事件降雨量而得，在此事件數以最廣為使用的卜瓦松分布進行模擬，降雨量則以伽瑪分布進行模擬。假設每一年的颱風事件數與梅雨事件數各自服從不同平均發生率的卜瓦松分布，則在第 i 年可以從分布中產生颱風事件數 n_{1i} 與梅雨事件數 n_{2i} 作為當年的降雨事件數；除此之外，假設在固定延時下的颱風降雨量 X_1 與梅雨降雨量 X_2 各自服從一伽瑪分布（位置參數 γ 為 0 的皮爾遜第三型分布），藉由產生彼此獨立的 n_{1i} 個颱風降雨量以及 n_{2i} 個梅雨降雨量，可以模擬第 i 年的降雨紀錄。卜瓦松分布與伽瑪分布的參數由嘉義雨量站的 24 小時降雨紀錄以線性動差法推估而得，並列於表 5-1。降雨紀錄的長度 N 設定為嘉義站的紀錄長度 44 年以及自記雨量站常見的紀錄長度 20 年進行模擬。

在產生降雨紀錄後便能以此紀錄進行年最大值數列與事件最大值數列頻率分析，由於降雨量來自於固定延時，因此不需要再以移動窗口選取。與實際資料不同的部分在於年最大值數列是由每年的颱風、梅雨事件數列取得，因此不存在跨事件的樣本。根據事件最大值數列的年最大值分布推導可以知道模擬資料的年最大值分布為：

$$\begin{aligned} F_Y(y) &= F_{Y_1}(y) \cdot F_{Y_2}(y) \\ &= e^{-\lambda_1[1-F_{X_1}(y)]} \cdot e^{-\lambda_2[1-F_{X_2}(y)]} \end{aligned} \quad (5-1)$$

由此分布可以針對特定重現期推得其理論回歸水準，並與每一次的模擬推估值進行比較。頻率分析的模擬過程總共重複一萬次，並在最後計算兩種方法的推估值之偏誤與均方根誤差，兩種指標的定義如下：

$$\text{Bias} = E[\hat{\theta} - \theta] \quad (5-2)$$



$$\text{RMSE} = \left(E \left[(\hat{\theta} - \theta)^2 \right] \right)^{1/2} = (\text{Var}(\hat{\theta}) + \text{Bias}^2)^{1/2} \quad (5-3)$$

其中期望值的部分可以以一萬次模擬平均計算。每一次模擬的步驟如下：

從 $i = 1$ 開始，

- (1) 進入第 i 年的降雨紀錄模擬。
- (2) 分別由對應的卜瓦松分布產生第 i 年的颱風事件數 n_{1i} 與梅雨事件數 n_{2i} 。
- (3) 由對應的伽瑪分布產生 n_{1i} 個颱風事件最大雨量值 $\{X_{i,1}, \dots, X_{i,n_{1i}}\}$ 與 n_{2i} 個梅雨事件最大雨量值 $\{Y_{i,1}, \dots, Y_{i,n_{2i}}\}$ 。
- (4) 選取 $a_i = \max\{X_{i,1}, \dots, X_{i,n_{1i}}, Y_{i,1}, \dots, Y_{i,n_{2i}}\}$ 作為第 i 年之年最大值。
- (5) 重複流程(1)-(4)，使 i 從 1 遞增至 N ，產生事件最大值數列與年最大值數列。
- (6) 針對重現期=5、10、25、50、100、200 年進行事件最大值與年最大值頻率分析，計算回歸水準與誤差。事件最大值數列以伽瑪分布估計雨量分布並結合卜瓦松分布推估年最大雨量之混合分布 (EMS)；年最大值數列則以廣義極端值分布和皮爾遜第三型分布進行估計 (AMS-GEV 和 AMS-PT3)。

5.2 頻率分析結果討論

模擬後所得到的偏誤以及均方根誤差以百分比的形式呈現於圖 5-1 和表 5-2。不論在 44 年還是 20 年的模擬，隨著重現期增加，三種方法的偏誤與均方根誤差皆上升，這是因為各累積分布函數的尾端上升比較緩慢，因此在高重現期下些許的誤差也會使回歸水準產生明顯偏誤。在本模擬中年最大值數列來自於該年獨立降雨量中的最大值，因此較諸實際資料中的年最大值數列更符合 block maxima 的假設，然而結果中 AMS-GEV 的推估具有最大的偏誤與均方根誤差，可能的原因包括年最大值數列的低樣本數和每年不固定的事件數使年最大值數列無法收斂至廣義極端值分布。AMS-PT3 與 EMS 的估計接近於無偏誤，然而隨著重現期上升，AMS-PT3 的均方根誤差逐漸高於較 EMS，同時 AMS-GEV 維持最高的均方根誤差，顯示 AMS 方法的推估結果受到樣本數的限制，而具有較高的不確定性。當記錄長度由 44 年減少至 20 年時，各方法的偏誤些微增加，主要的變化為均方根誤差變為原本的 1.5 倍，三種方法受到記錄長度的影響相近，而 EMS 均為最適合的方法。

5.3 極端值理論模擬

在模擬過程中所產生的 10000 次 44 年年最大值數列可以使用 LMRD 進行皮



爾遜第三型分布與 Gumbel 分布的適合度檢定並計算拒絕率，而所有的年最大值可以用以估計年最大值分布的理論線性動差比。因為兩分布的接受域針對正確分布的拒絕機率為 5%，年最大值數列的拒絕率可以作為理論線性動差比之外衡量年最大值分布是否接近假設分布的標準。在此，年最大值數列分為颱風事件的年最大值數列、梅雨事件的年最大值數列以及考量兩種類型降雨下的年最大值數列 (AMS)。檢定結果列於表 5-3 與圖 5-2，紅點是使用一萬次模擬所得到的所有資料估計的理論線性動差比，橘點則為一萬次模擬的線性動差比，左圓為常態分佈與其接受域，右圓則為 Gumbel 分布與其接受域，綠與黑曲線是皮爾遜第三型分布與其接受域。對於 Gumbel 分布，颱風和梅雨的年最大值數列分布拒絕率都很低，其中梅雨的理論線性動差比幾乎與 Gumbel 分布重疊，顯示年平均事件數為 11 的梅雨已經收斂至廣義極端值分布，然而 AMS 的拒絕率明顯高於理論值 (35%)，顯示其未能很好地逼近至 Gumbel。無論是颱風、梅雨或者是 AMS 的降雨量均來自於伽瑪分布，因此在皮爾遜第三型分布的適合度檢定拒絕率仍然接近 5%，然而梅雨事件因為年平均發生率比較高，因此其年最大值分布的線性動差比已經偏離皮爾遜第三型分布，拒絕率也比其他兩種年最大值數列高。為了更進一步探討年最大值數列的收斂情形，本研究分別在固定年事件數 (1~11 場事件/年) 以及年事件數服從卜瓦松分布 ($\lambda=1\sim 11$) 的情況下，分別模擬各類型年最大值數列，並以 LMRD 進行 Gumbel 分布的適合度檢定，計算各個情形下年最大值數列被拒絕的次數以及年最大值分布的理論動差比。

5.4 極端值模擬結果討論

檢定結果可見表 5-4、圖 5-3 與圖 5-4。在固定年事件數下，颱風與梅雨的年最大值符合 block maxima 的假設，因此隨著年事件數上升，颱風與梅雨的線性動差比由皮爾遜第三型分布朝 Gumbel 分布靠近。AMS 之線性動差比雖然在最初偏離皮爾遜第三型分布，其趨勢也與颱風、梅雨相似；AMS 可以視為颱風年最大值與梅雨年最大值的混合數列，因此當兩種年最大值趨近極端值分布時，AMS 也會趨近兩種極端值分布的混合數列，而不易被檢定拒絕。其中年事件數等於 1 時因為颱風的降雨普遍大於梅雨，因此年最大值數列的資料主要來自於颱風，拒絕率也與颱風相近，此情形將隨著事件數上升而變得不明顯。三種數列的拒絕率會隨著年事

件數的上升而下降，大約在年事件數為 6 到 8 時到達 10% 拒絕率，然而之後下降的幅度會逐漸趨緩，可能與事件數不夠大有關。

當年事件數為卜瓦松隨機變數時，各個年最大值數列的線性動差比同樣隨著平均發生率上升而逼近 Gumbel 分布，並且很明顯地不屬於皮爾遜第三型分布。在 $\lambda=1$ 時隨機事件數的拒絕率會高於固定事件數等於 1 的拒絕率，然而從 $\lambda=2$ 開始拒絕率的下降速度就高於固定事件數的下降速度，並維持較低的拒絕率，可能的原因是卜瓦松分布提供了更高階的順序統計量加入年最大值數列的機會，因此數列可以更快逼近極端值分布，也因此 λ 上升時拒絕率無法持續降低。以 $\lambda=3$ 的年最大值數列為例，雖然 n 小於 3 的機率為 0.4——即有將近 50% 的年最大值降雨來自於事件數等於 1 或者 2 的年份，仍然有 30% 的降雨來自事件數大於 3 的年份，使得此數列比固定 n 等於 3 的年最大值數列更接近 Gumbel 分布。相對於颱風與梅雨的年最大值數列拒絕率由 $\lambda=1$ 開始快速下降，AMS 的拒絕率在 $\lambda=1$ 時比前兩者低，其後卻下降地比較緩慢並高於其他數列，顯示在隨機事件數的情形下，分布的混合對於年最大值在廣義極端值分布的逼近有更顯著的影響。



表 5-1 模擬使用參數

	Gamma's Scale θ	Gamma's Shape β	Poisson's Rate λ
颱風	105.10	1.24	2.68
梅雨	34.94	1.07	10.93

表 5-2 降雨頻率分析模擬結果 (單位：百分比)

重現期		5	10	25	50	100	200	
理論雨量(mm)		306.89	389.85	494.36	571.42	647.57	723.16	
N=20	Bias	AMS-PT3	-0.1	-0.07	0.22	0.51	0.73	0.81
		AMS-GEV	-3.05	-3.48	-1.13	2.38	7.15	13.12
		EMS	0.18	0.22	0.26	0.4	0.53	0.56
	RMSE	AMS-PT3	14.52	16.72	20.5	23.26	25.74	27.92
		AMS-GEV	14.47	15.56	19.02	23.97	31.34	41.3
		EMS	13.95	14.95	15.92	16.54	17.07	17.51
N=44	Bias	AMS-PT3	-0.11	-0.12	0.10	0.23	0.29	0.30
		AMS-GEV	-2.98	-3.11	-0.63	2.58	6.73	11.75
		EMS	0.05	0.09	0.16	0.22	0.27	0.31
	RMSE	AMS-PT3	9.81	11.25	13.65	15.36	16.90	18.25
		AMS-GEV	10.01	10.73	12.89	16.31	21.46	28.29
		EMS	9.42	10.02	10.61	10.98	11.30	11.58

表 5-3 10000 次降雨模擬之 AMS 於 LMRD 拒絕率 (單位：百分比)

	颱風 ($\lambda=2.68$)	梅雨 ($\lambda=10.93$)	AMS
Gumbel 分布	6.08	7.52	35.04
PT3 分布	5.63	8.19	5.27

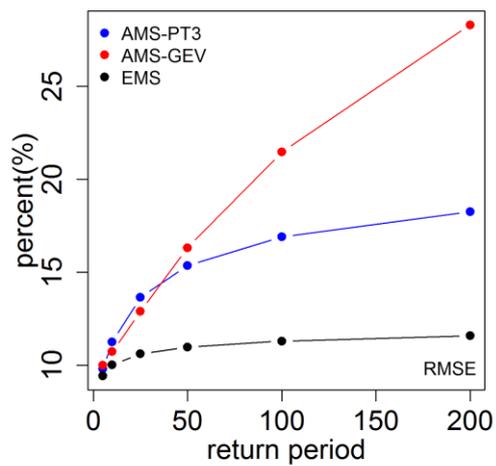
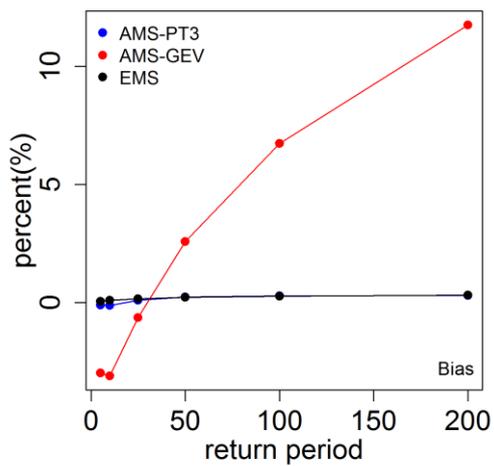


圖 5-1 降雨頻率分析 44 年模擬結果

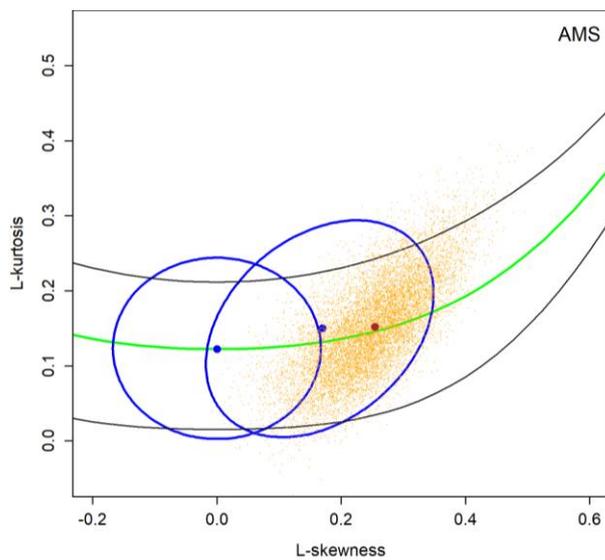
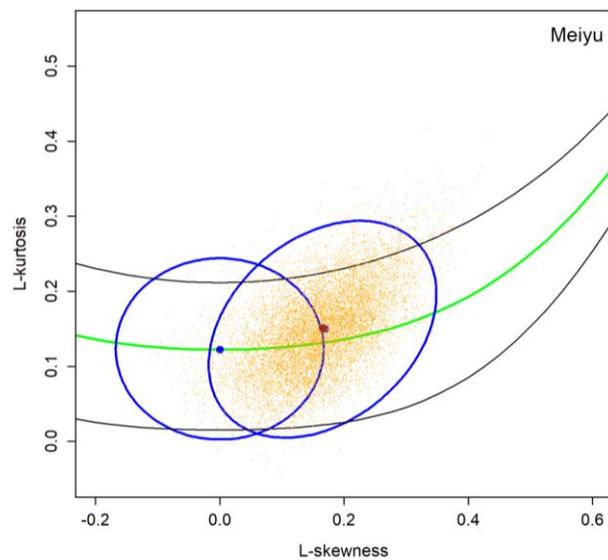
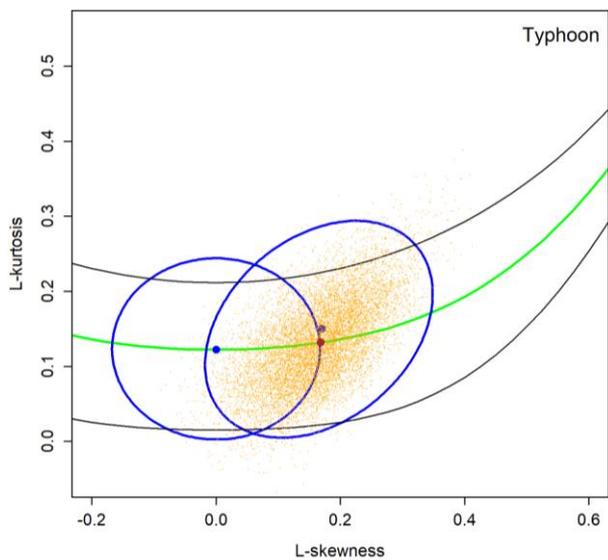


圖 5-2 10000 次模擬之 LMRD



表 5-4 AMS 於 Gumbel LMRD 之拒絕率 (單位：百分比)

類型 年事件數/ λ	固定年事件數			隨機年事件數		
	颱風	梅雨	AMS	颱風	梅雨	AMS
1	63.83	75.93	63.53	97.94	97.96	78.18
2	24	32.71	32.32	19.82	19.40	23.75
3	15.36	19.94	20.4	5.84	5.75	12.65
4	11.77	14.85	15.86	6.62	5.60	9.40
5	10.43	12.5	12.77	6.36	6.61	8.15
6	9.28	11.24	10.8	6.65	7.35	7.51
7	8.92	10.42	10.38	6.08	6.33	6.96
8	8.8	9.74	8.95	7.29	6.51	7.28
9	8.47	9.46	9.06	6.69	6.52	7.40
10	7.77	9.18	8.53	6.70	6.78	7.34
11	7.91	8.86	8.35	6.66	7.10	6.78

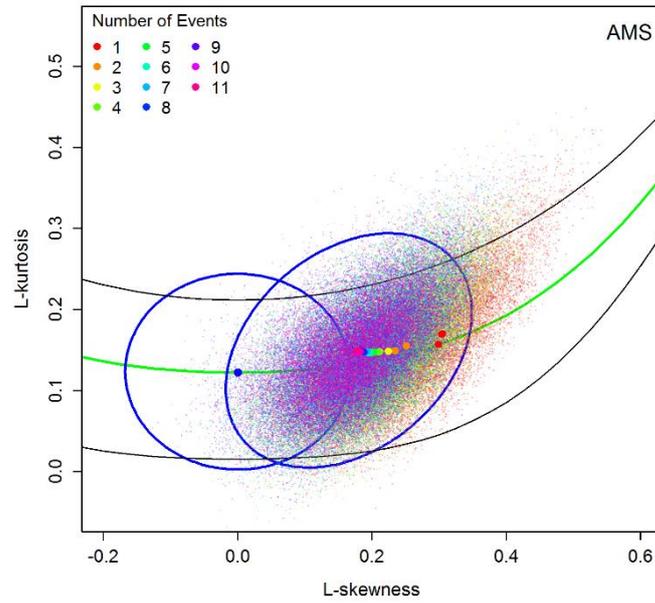
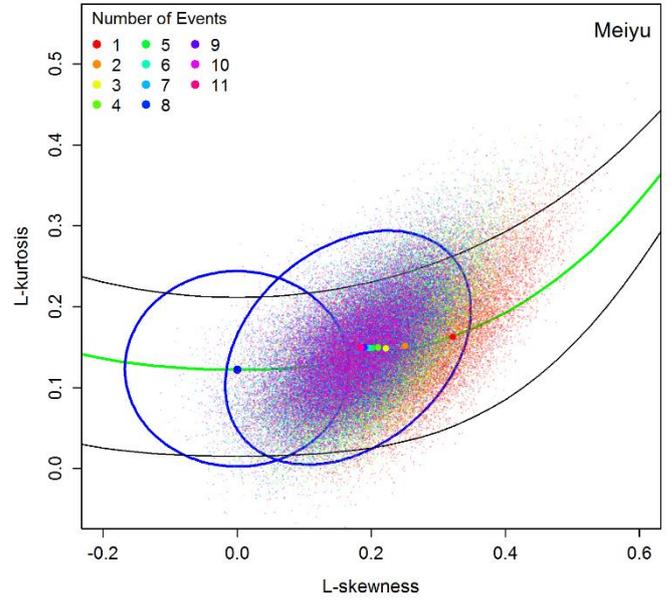
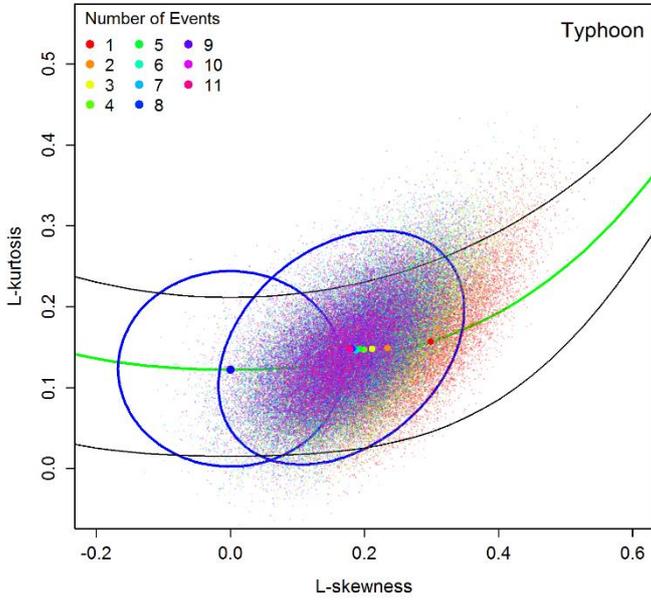


圖 5-3 不同年事件數下年最大值之 LMRD

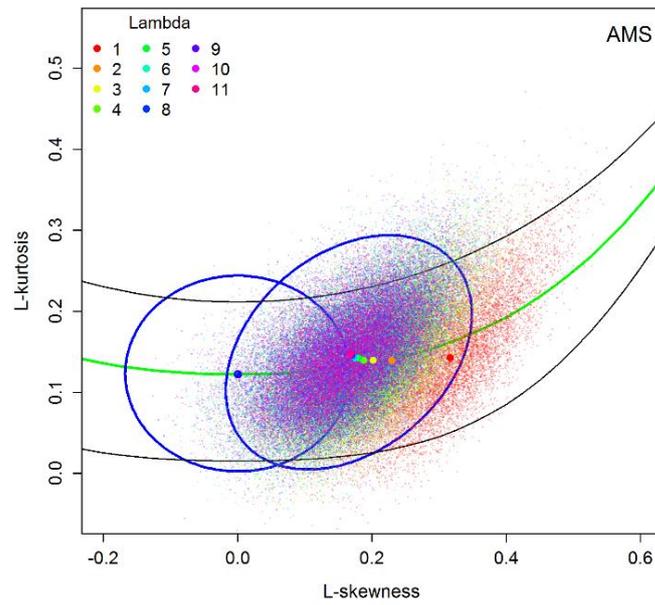
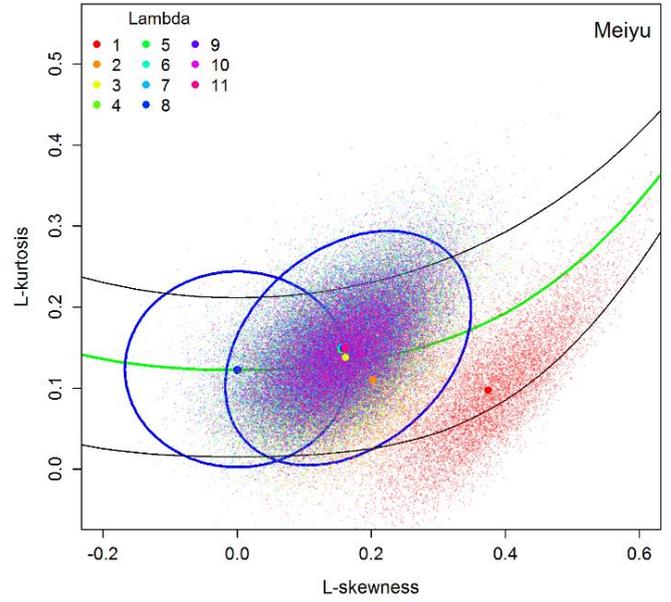
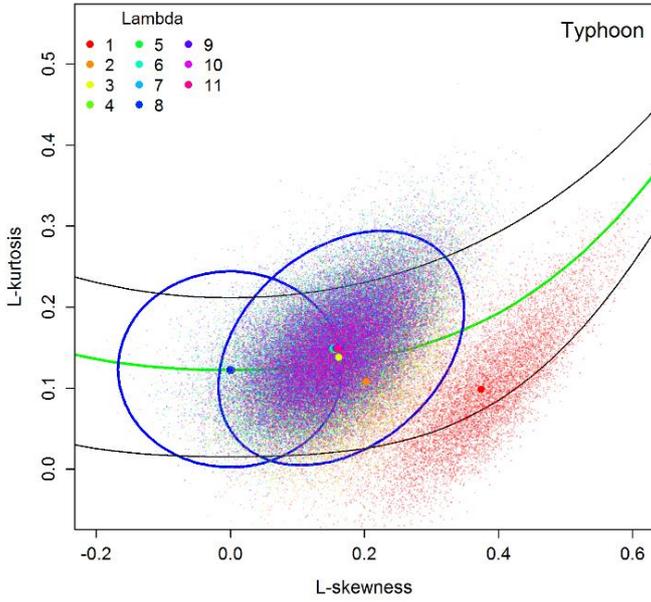


圖 5-4 不同事件發生率下年最大值之 LMRD



第六章 真實資料分析



在以模擬結果確認 EMS 的適用性之後，本研究選定臺灣本島北、中、南各一降雨測站進行 AMS 與 EMS 的頻率分析結果比較。

6.1 研究測站

北、中、南所對應的三個測站分別為五堵、頭汴坑與嘉義雨量站，測站資訊與位置可見於表 6-1 以及圖 6-1。三個測站的記錄長度均大於 30 年，五堵站長度較其他兩站稍短。

6.2 事件切割結果

根據前述之規則，本研究針對三個雨量站之降雨紀錄以 4 小時作為分隔的標準進行降雨事件切割，其結果列於表 6-2。切割的結果基本上與臺灣氣候情形相符，在北部，事件平均降雨量和事件平均延時由大到小排列依序為颱風、鋒面雨、梅雨、對流雨。中部與南部因為冬季受鋒面以及東北季風的影響較小，平均降雨量和平均延時的順序為颱風、梅雨、對流雨、鋒面雨。除此之外，降雨事件的發生率排序則接近相反。當降雨事件的延時越長，降雨時間越多，事件累積降雨量也會越高，因此降雨量和延時多呈現正相關。夏季降雨以 12 小時為界分為颱風和對流雨，颱風作為極端降雨的主要來源，具有長延時、高降雨量的特徵，並且發生率也是四種降雨類型中最低的；與之相對的對流雨具有短延時、低降雨量的特徵，然而其發生次數高，並且降雨強度並不一定會低於其他類型的降雨事件；梅雨和鋒面雨的影響來自鋒面與季風，因此會受到南北差異影響而有不同關係。受到氣溫和西南季風的影響，對流雨和梅雨的平均發生率以南部為最高；冬季的鋒面雨降雨主要發生於北部，在南部的鋒面雨事件數雖然大於中部，其平均延時與降雨量均為最低，因此為冬季枯水期。

6.3 降雨頻率分析

在獲得各類型降雨事件之後，可以以移動窗口獲得年最大值數列和事件最大值數列，並針對降雨量以及其發生次數進行參數估計與機率適合度檢定。使用 KS test 對降雨事件數進行適合度檢定的結果列於表 6-3，在五堵與頭汴坑測站除了颱風

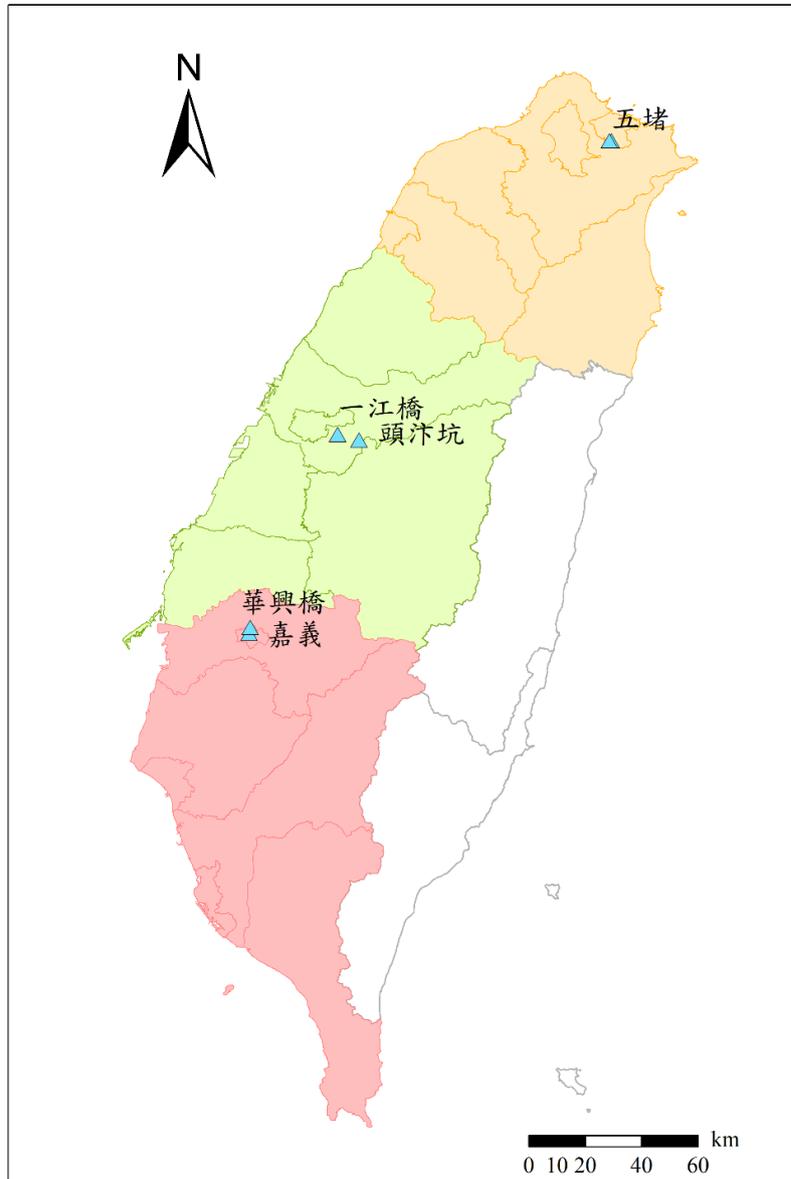


圖 6-1 雨量站、流量站與水位站位置



表 6-1 測站資訊

站名	站號	所屬單位	流域	X 座標 (TM2)	Y 座標 (TM2)	資料蒐集年限	紀錄 年限
五堵	01B030	水利署	淡水河	320278	2774708	1970、1974- 1975、1983-2011	32
頭汴坑	01F680	水利署	烏溪	230799	2667784	1970-2011	42
嘉義	467480	氣象局	朴子溪	191727	2598683	1969-2012	44

表 6-2 三測站 MIT=4 hr 下各類型降雨事件切割結果

測站	類型	總事 件數	平均發生率 (次/年)	事件 平均延時(hr)	事件平均 降雨量(mm)
01B030	對流雨	261	8.15	4.49	28.49
	鋒面雨	290	9.06	18.04	59.16
	梅雨	223	6.96	8.71	45.45
	颱風	156	4.87	33.80	194.19
01F680	對流雨	355	8.45	4.17	38.22
	鋒面雨	122	2.9	8.55	33.35
	梅雨	294	7	9.2	60
	颱風	79	1.88	26.15	217.24
467480	對流雨	754	17.14	3.91	29.36
	鋒面雨	339	7.7	5.09	21.57
	梅雨	481	10.93	6.77	38.71
	颱風	118	2.68	20.5	142.18

以外的降雨事件發生次數均拒絕卜瓦松分布的假設，而不拒絕負二項分布；從 CV 的結果 (表 4-1) 也可以看出，只有颱風的 CV 會從大於 1 降低至小於 1，其他類型的降雨 CV 都維持在大於 1 的情形，因此除了颱風可以使用卜瓦松過程描述之外，其他類型的降雨都具有過度離散的情形。發生過度離散的原因可能來自降雨事件的群聚以及其延時不固定的影響；由於鋒面系統的持續時間比較長，並且可以持續產生雨胞，因此降雨事件的發生在時間上將具有相關性。此外，降雨事件的數量為來自一固定時間內的抽樣，當事件延時無法被忽略或者長度為隨機時，事件發生次數的變異數將受到影響，因此負二項分布更適合用於描述各類型降雨事件的年發生次數。

Wu et al. (2011) 針對皮爾遜第三型分布所建立的 LMRD 95% 接受域適用於樣本數介於 20 到 300、偏態係數小於 3 的樣本。除了嘉義站的對流雨和梅雨事件數過高之外，其餘事件的樣本數尚在合理範圍中。表 6-4 為各測站不同延時下的雨量偏態係數，偏態係數大於 3 的資料包括五堵站鋒面雨的低延時降雨、頭汴坑站的梅雨和低延時的年最大值數列以及嘉義站的低延時鋒面雨和高延時梅雨，這些測站資料的檢定結果並不一定準確。降雨分布的適合度檢定以 LMRD 的形式呈現 (圖 6-2 到圖 6-5)。除了颱風以外，三個測站的所有事件降雨分布的 Gumbel 分布假設均被拒絕。皮爾遜第三型分布假設的拒絕次數並不高，主要發生於低延時鋒面雨 (五堵站)、高延時梅雨 (南部測站) 還有少部分的颱風數列，除了颱風以外這些降雨的偏態係數均大於 3，因此拒絕結果僅作為參考。被拒絕的原因是數列中存在極端值降雨，若將之移除則樣本分布不會被拒絕，考慮到目前使用較為嚴格的 95% 接受域，並且偏態係數不符合使用前提，本研究仍然以皮爾遜第三型分布作為事件降雨量分布。年最大值數列在北部高延時與中部低延時的雨量均拒絕皮爾遜第三型分布假設，樣本數較少使得各分布的接受域比事件最大值數列的各類型降雨更大，並且使年最大值數列受到極端值的影響更顯著，因此在移除數列中的極端值後，各延時下的拒絕率將下降。雖然沒有進入 Gumbel 分布的接受域，考量到廣義極端值分布的線性動差比為一曲線，即使不符合極端值理論假設，年最大值數列仍然有可能因為保有極端值特性而接近廣義極端值分布。

圖 6-6 為三個測站依卜瓦松——皮爾遜第三型分布模型 (Poisson-PT3) 與負二項——皮爾遜第三型分布模型 (NB-PT3) 模型由 24 小時事件最大值數列估計而得

之累積機率函數。圖標中 Con, Typ, Mei, Fron 與 EMS 分別代表對流雨、颱風、梅雨、鋒面雨和考慮所有事件的年最大值分布，NB 與 Pois 則代表使用的事件數分布為負二項分布還是卜瓦松分布。從圖中可以獲得兩項結論——首先，由五堵、頭汴坑測站的對流雨、梅雨和鋒面雨的年最大值分佈可以看出，儘管卜瓦松分布被拒絕，隨著累積機率上升，Poisson-PT3 的累積機率分布函數逐漸與 NB-PT3 重合，顯示在重現期足夠高的時候，累積機率接近 1、不同次數下的最大順序統計量相對接近，兩種模型的估計結果會幾乎相同，因此後續分析以使用能處理過度離散的 NB-PT3 模型為主。其次，隨著累積機率上升，可以發現颱風降雨事件的年最大值分布和最終的年最大值分布幾乎重合，顯示對於此三測站，颱風事件的年最大降雨量足以代表整年的年最大雨量，這與臺灣極端降雨事件多來自颱風的事實相符；由年最大值數列的降雨類型（表 6-5）也可以發現類似情況——年最大值數列的樣本多來自颱風事件。針對五堵、頭汴坑與嘉義測站，以設計延時 1、2、3、6、12、24、48 和 72 小時對重現期 5、10、25、50、100、200 年進行 EMS 以及 AMS 頻率分析的結果列於表 6-6 到表 6-8，圖 6-7 為其 DDF，並於下一章進行討論。



表 6-3 降雨事件發生次數之卜瓦松與負二項分布機率適合度檢 (p-value)

	分布	對流雨	鋒面雨	梅雨	颱風
01B030	卜瓦松	0.01	0.00	0.01	0.93
	負二項	0.23	0.08	0.48	0.99
01F680	卜瓦松	0	0	0	0.73
	負二項	0.25	0.44	0.62	0.87
467480	卜瓦松	0.69	0.16	0.62	0.82
	負二項	0.78	0.46	1	0.94



表 6-4 各測站雨量偏態係數

	延時	對流雨	鋒面雨	梅雨	颱風	AMS
01B030	1	1.59	4.09	1.83	1.47	1.02
	2	1.86	3.82	1.88	1.87	1.47
	3	2.04	4.22	1.9	1.65	1.17
	6	2.44	3.16	2.01	1.74	1.25
	12	2.93	2.4	1.64	2.67	2.16
	24	2.93	2.31	1.54	3.28	2.41
	48	2.93	2.26	2.12	3.55	2.54
	72	2.93	2.39	2.15	3.33	2.47
01F680	1	1.66	2.69	8.07	1.16	4.31
	2	2.55	1.7	10.29	0.99	4.72
	3	3.33	1.54	9.57	1.02	4.19
	6	3.92	2	8.95	1.23	3.28
	12	3.78	1.69	8.29	1.18	2.38
	24	3.78	1.63	7.01	1.23	1.67
	48	3.78	2.08	6.45	1.37	1.45
	72	3.78	2.12	6.4	1.37	1.29
467480	1	2.12	5.92	1.81	2.67	1.69
	2	1.97	4.49	1.98	3.22	2.37
	3	1.93	3.98	2.16	2.78	2.28
	6	1.82	2.88	2.41	1.97	1.76
	12	1.78	2.65	2.58	2.22	2.11
	24	1.78	2.63	3.55	2.55	2.17
	48	1.78	2.93	4.78	2.92	2.28
	72	1.78	2.94	5.27	2.9	2.03

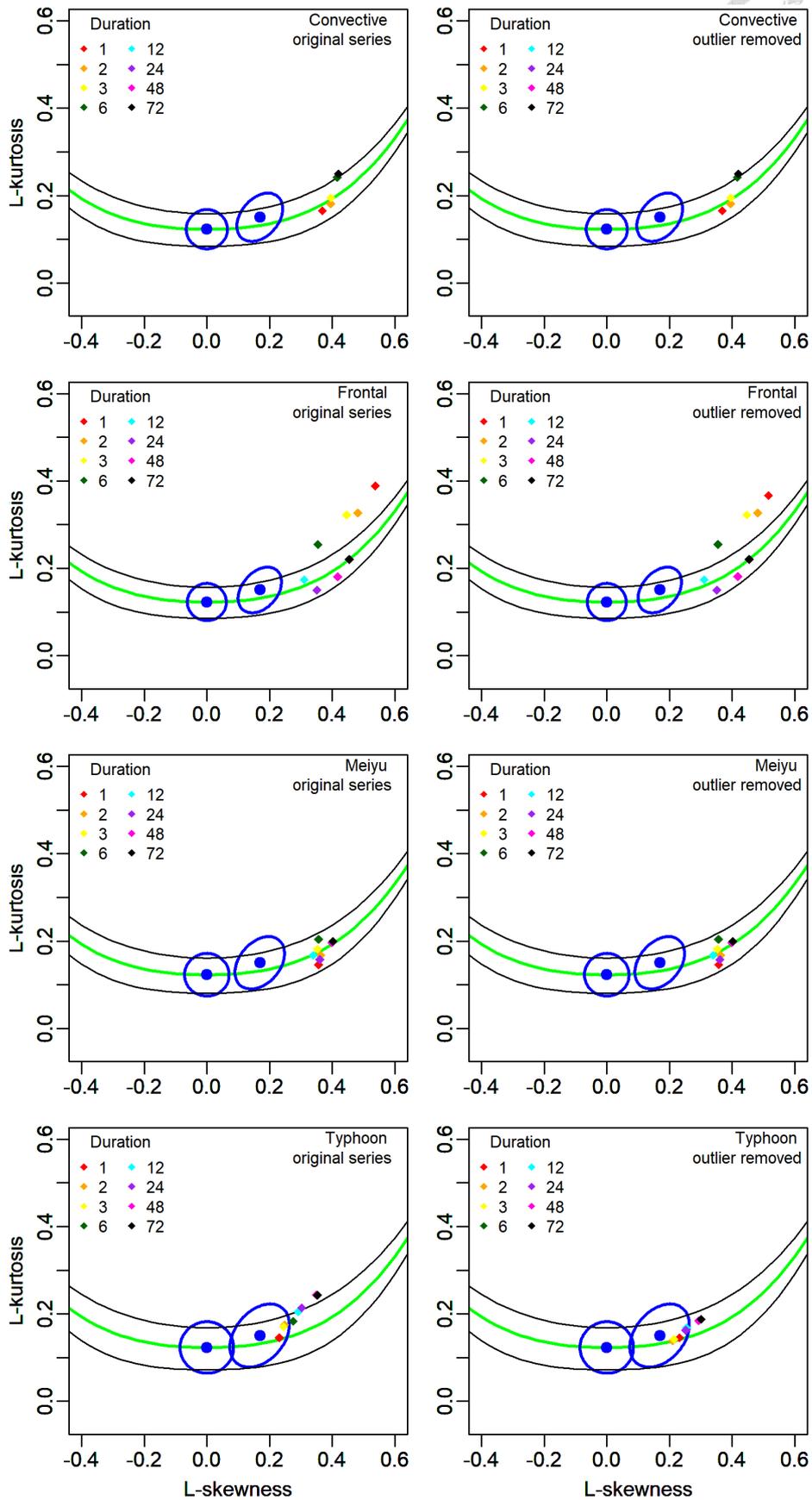


圖 6-2 五堵站事件最大值數列機率適合度檢定結果

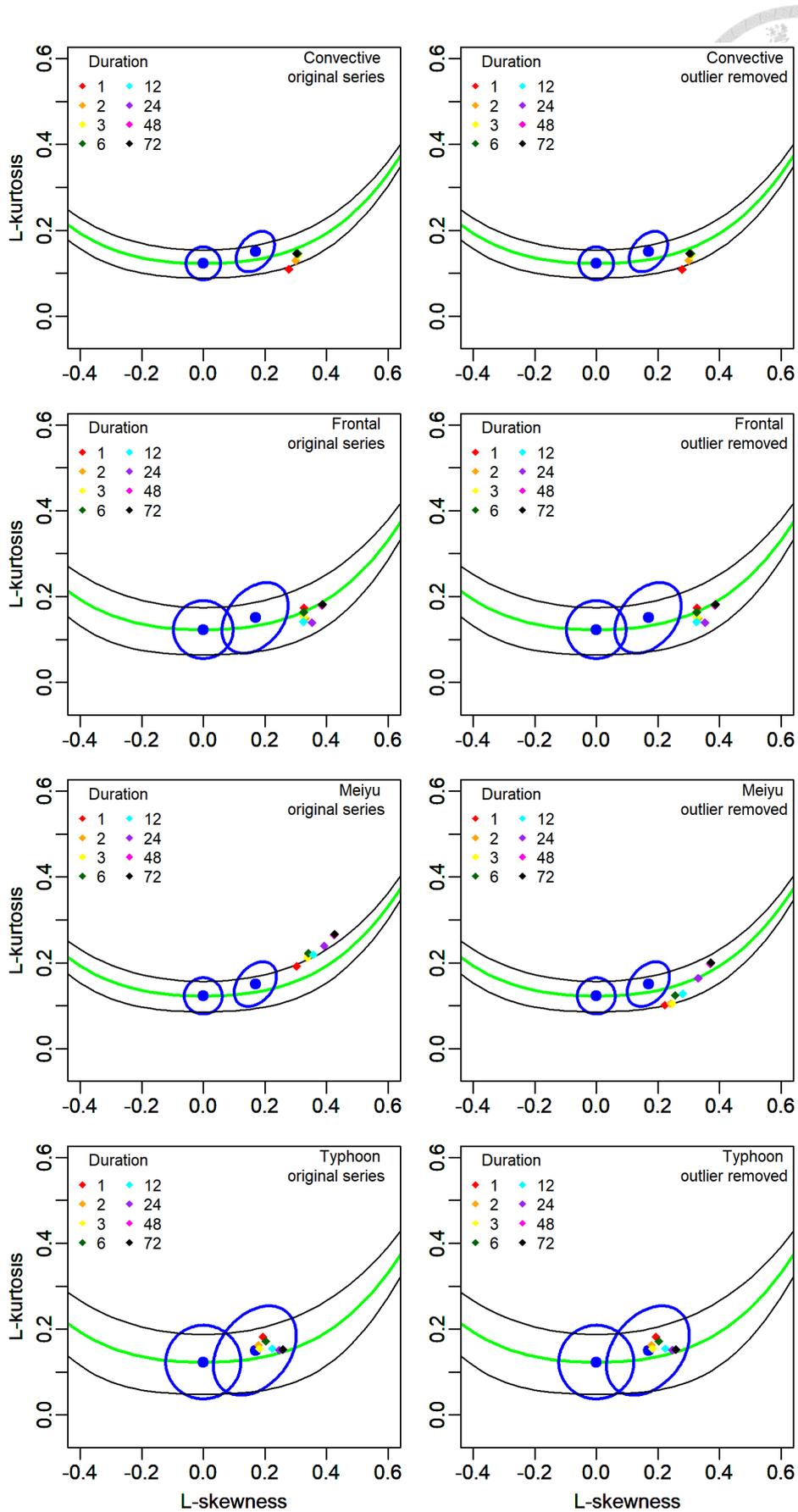


圖 6-3 頭汴坑站事件最大值數列機率適合度檢定結果

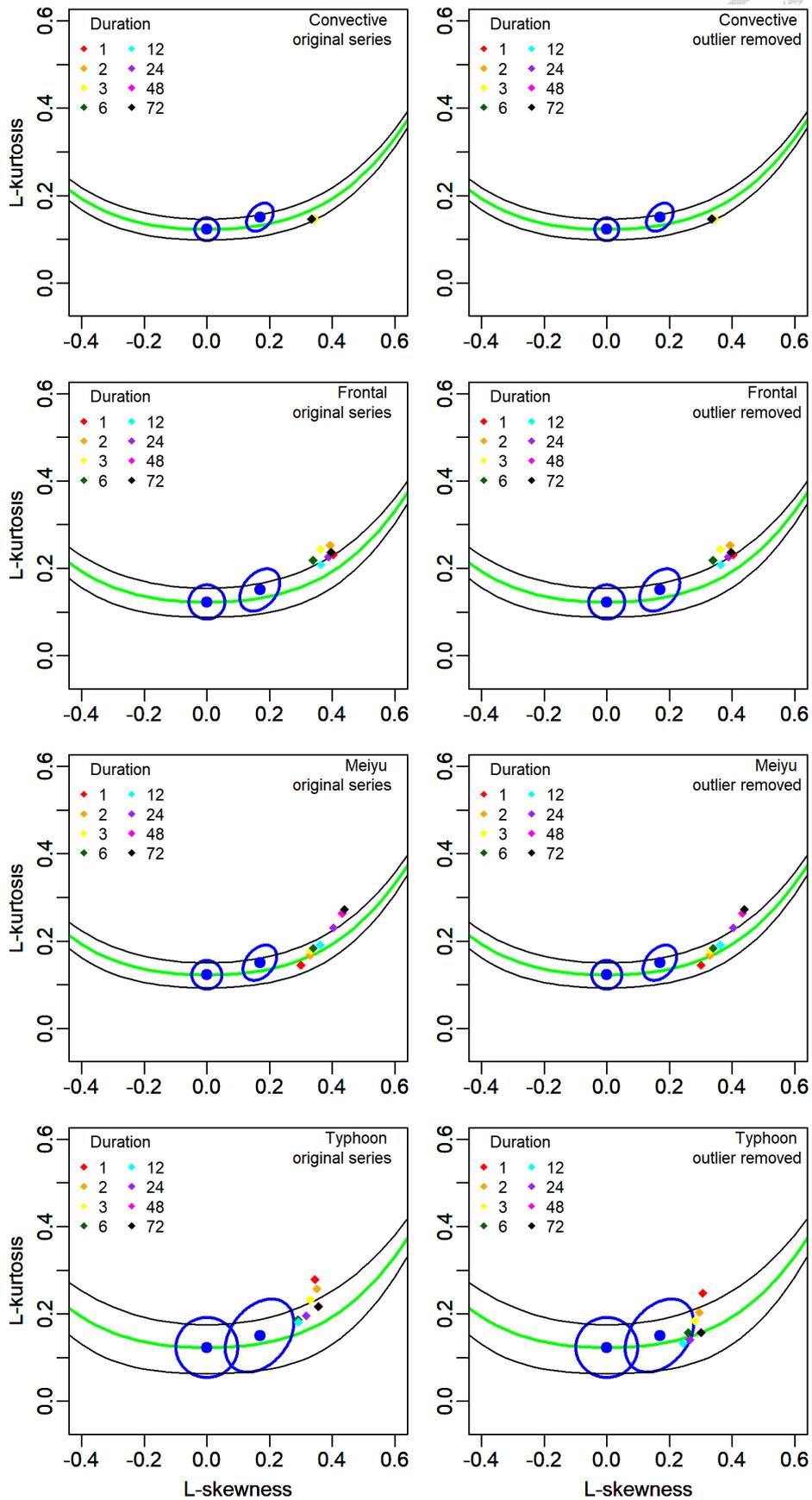


圖 6-4 嘉義站事件最大值數列機率適合度檢定結果

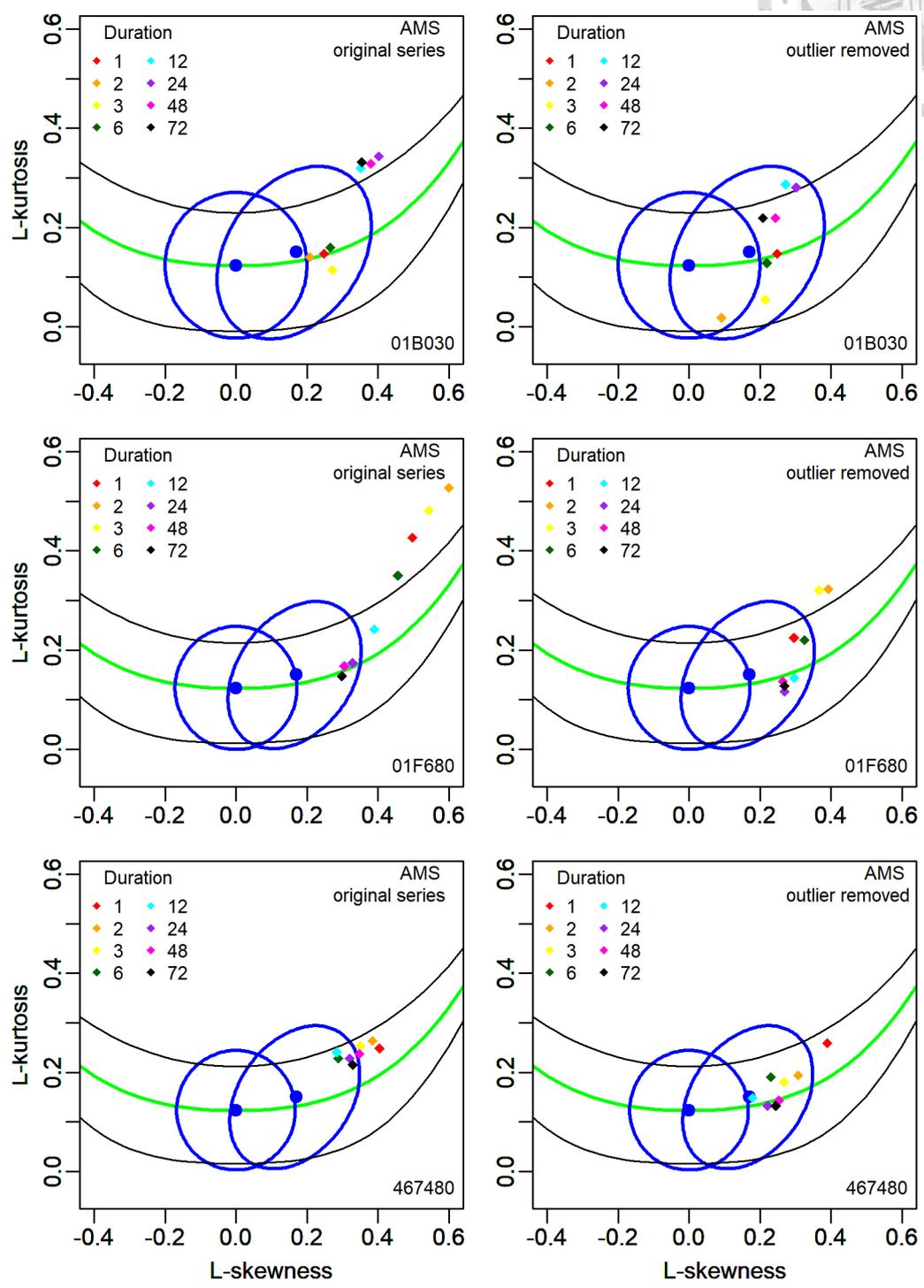
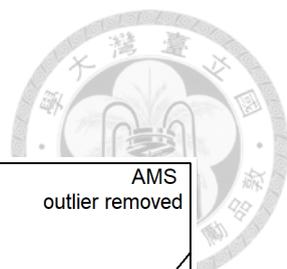


圖 6-5 年最大值數列機率適合度檢定結果

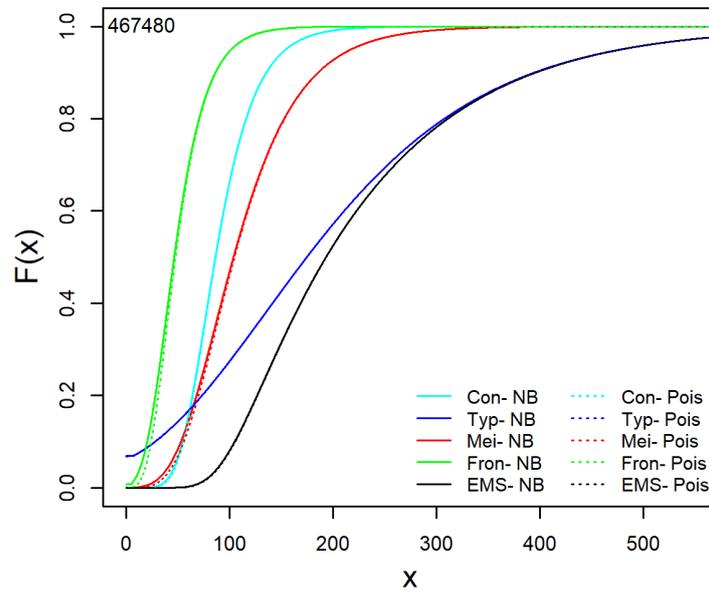
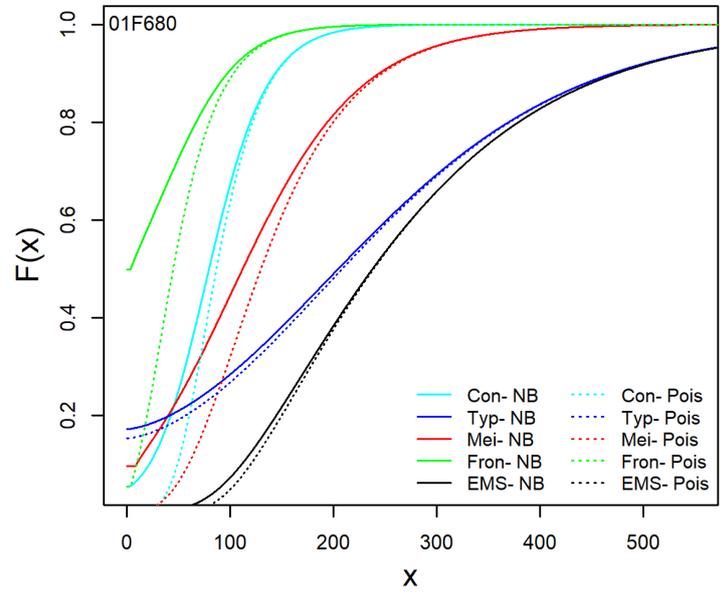
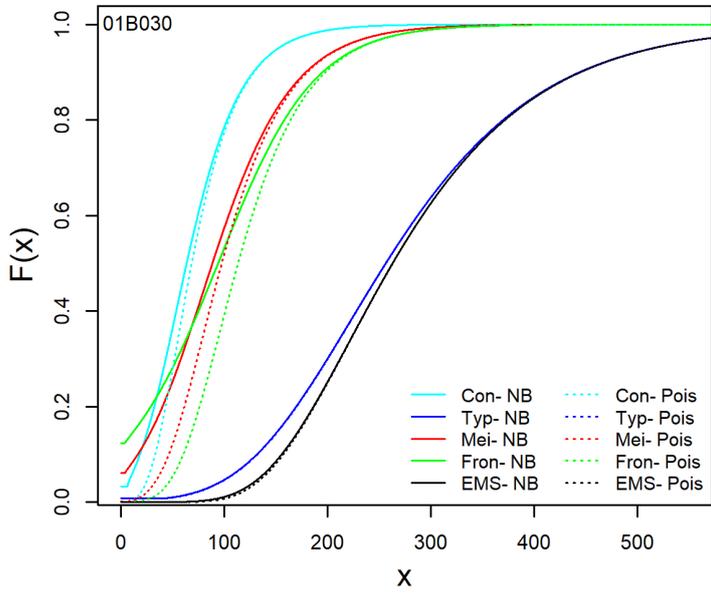


圖 6-6 Poisson-PT3 與 NB-PT3 24 小時年最大值降雨累積機率分布函數

表 6-5 年最大值數列降雨類型
(括號內為涵蓋 2 次以上降雨事件的樣本數，對流雨和颱風合併計算)

	設計延時	對流雨	鋒面雨	梅雨	颱風
01B030 (Total 32 years)	1	0	5 (0)	8 (0)	19 (0)
	2	0	6 (1)	8 (0)	18 (0)
	3	0	6 (2)	10 (0)	16 (0)
	6	1	5 (1)	4 (1)	22(1)
	12	0	3 (0)	2 (1)	27 (5)
	24	0	3 (0)	3 (2)	26 (7)
	48	0	1 (1)	2 (2)	29 (17)
	72	0	4 (3)	2 (3)	26 (23)
01F680 (Total 42 years)	1	1	1 (0)	7 (0)	33 (0)
	2	1	0 (0)	9 (0)	32 (0)
	3	1	0 (0)	5 (0)	36 (3)
	6	0	1 (0)	10 (0)	31 (8)
	12	0	1 (1)	13 (2)	28 (5)
	24	1	2 (1)	13 (4)	26 (11)
	48	1	2 (2)	14 (13)	25 (22)
	72	1	3 (3)	15 (15)	23 (24)
467480 (Total 44 years)	1	1	1 (0)	9 (0)	33 (0)
	2	1	2 (0)	15 (0)	26 (0)
	3	1	1 (0)	14 (0)	28 (1)
	6	0	1 (0)	14 (0)	29 (4)
	12	0	0 (0)	14 (3)	30 (6)
	24	0	1 (0)	16 (10)	27 (14)
	48	0	1 (0)	14 (2)	29 (25)
	72	1	2 (2)	13(12)	28 (29)

表 6-6 五堵站頻率分析結果

		AMS						EMS						
		RT	5	10	25	50	100	200	5	10	25	50	100	200
Original	1	78	93	112	126	140	154	77	89	104	115	126	138	
	2	112	132	157	175	193	210	112	129	150	166	181	197	
	3	138	167	203	230	256	283	137	157	183	203	222	241	
	6	196	236	288	326	364	401	196	230	273	304	336	367	
	12	281	357	458	536	613	690	279	332	399	448	497	546	
	24	370	492	659	789	920	1052	372	445	537	606	674	741	
	48	493	660	886	1060	1235	1411	486	593	730	833	936	1038	
	72	557	733	968	1147	1327	1507	516	630	776	886	996	1105	
Removing the largest value	1	78	93	112	126	140	154	76	87	101	112	123	133	
	2	105	119	135	146	156	166	108	123	142	156	171	185	
	3	131	154	182	202	222	242	131	150	174	191	208	224	
	6	186	220	261	291	320	349	188	219	257	285	313	341	
	12	262	318	390	443	495	547	261	306	363	404	445	485	
	24	339	422	530	611	691	771	341	400	474	529	582	635	
	48	445	549	680	777	871	964	438	523	630	709	788	866	
	72	503	611	745	841	935	1028	469	561	676	762	848	933	

表 6-7 頭汴坑站頻率分析結果

		AMS						EMS						
		RT	5	10	25	50	100	200	5	10	25	50	100	200
Original	1	83	110	150	182	215	248	80	93	108	119	129	140	
	2	118	169	249	315	385	457	120	139	161	178	195	211	
	3	148	210	302	377	455	535	145	169	197	218	238	259	
	6	218	302	422	517	613	711	204	242	291	327	362	397	
	12	300	409	558	672	788	904	288	354	436	496	554	612	
	24	389	515	681	806	931	1056	378	474	594	682	768	853	
	48	460	600	782	918	1054	1189	429	542	684	788	891	993	
	72	495	640	827	967	1106	1245	430	543	685	790	892	994	
Removing the largest value	1	76	90	108	121	134	147	77	89	104	115	126	136	
	2	109	132	164	188	213	238	114	132	154	170	186	202	
	3	135	167	210	243	276	310	138	161	189	209	230	250	
	6	198	251	320	373	425	477	195	235	285	322	358	394	
	12	276	354	454	530	604	678	281	349	433	494	553	612	
	24	364	464	592	687	780	873	372	471	593	681	768	853	
	48	435	552	701	812	921	1029	422	538	682	787	891	992	
	72	470	596	756	874	991	1107	423	539	683	788	892	994	



表 6-8 嘉義站頻率分析結果

		AMS						EMS						
		RT	5	10	25	50	100	200	5	10	25	50	100	200
Original	1	78	100	129	151	174	197	77	89	106	119	133	147	
	2	113	143	184	215	247	279	111	131	159	180	203	226	
	3	131	163	206	239	272	305	130	154	187	213	240	267	
	6	175	215	267	306	345	383	172	208	256	293	330	367	
	12	244	306	387	446	506	564	236	294	368	423	478	533	
	24	332	428	555	651	746	841	311	396	504	585	665	745	
	48	407	526	684	805	925	1046	351	456	592	694	796	897	
	72	444	564	722	842	961	1081	353	458	595	697	799	901	
Removing the largest value	1	75	94	120	139	159	179	74	86	101	112	123	135	
	2	106	129	158	180	202	224	107	124	145	162	179	196	
	3	124	148	179	201	224	246	125	145	171	192	212	233	
	6	167	199	240	270	299	328	165	196	237	268	299	331	
	12	228	272	326	364	401	438	220	267	327	371	415	458	
	24	309	378	464	527	588	648	286	354	441	504	567	630	
	48	380	465	573	653	731	808	320	403	510	589	668	746	
	72	418	507	619	701	782	862	322	405	513	592	671	750	

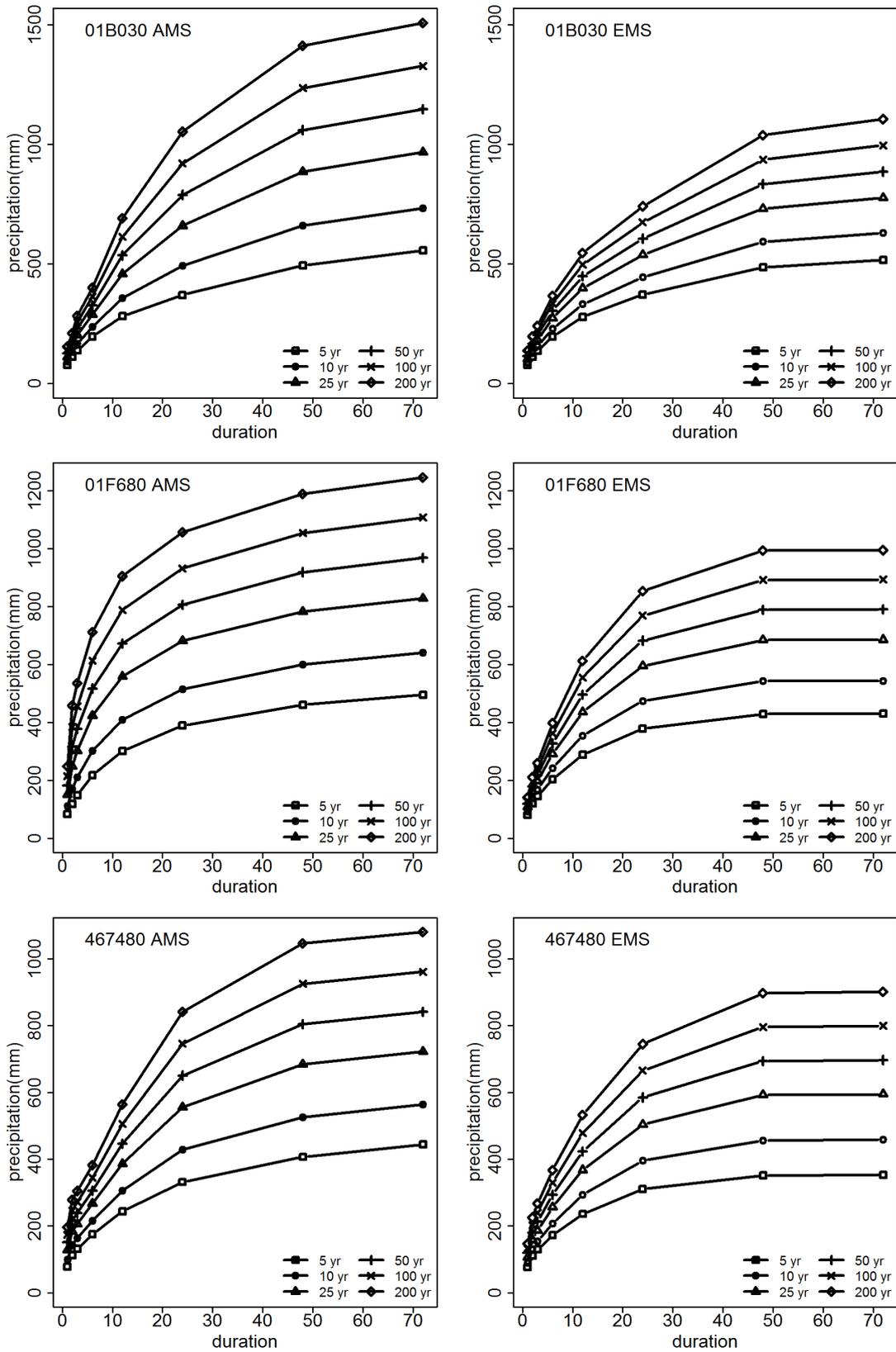
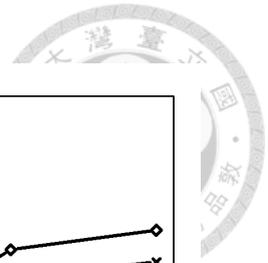


圖 6-7 三測站之降雨深度-延時-頻率曲線(DDF)

第七章 結果與討論



7.1 設計降雨討論

根據表 6-6，圖 6-7 的降雨頻率分析結果，可以發現在三個測站 EMS 的估計結果均小於 AMS，並且差距隨著設計延時以及重現期增加而增加。EMS 傾向較低估計的原因來自於事件的切割與分類，由於事件最大值數列的降雨量不會跨事件累積，在各設計延時下 EMS 所得到的降雨量必然小於等於 AMS。此外，每一年 EMS 的年最大值樣本有可能來自於不同類型的降雨，因此即使每年的 EMS 年最大值與 AMS 相等，也有可能被用於估計不同類型的降雨分布，而各類型降雨事件中的低雨量樣本將導致在相同重現期下，事件降雨量的回歸水準低於 AMS 的回歸水準，使 EMS 的估計將傾向小於 AMS 的估計。然而 EMS 的估計結果並不必然會低於 AMS 的結果，本研究額外選取花蓮立山與高雄楠溪雨量站的 3 小時降雨紀錄作為範例說明。兩測站的基本資料列於表 7-1，圖 7-1 為颱風事件降雨量 (Typ)、颱風年最大值分布 (Typ-Ams) 以及由 EMS 和 AMS 估計之年最大值雨量分布，灰點為颱風事件最大值數列與年最大值數列的經驗累積分布函數。可以看出颱風事件降雨量分布以及 AMS 分布與其資料點的配適結果良好，並且因為抽樣方法的影響，事件降雨量分布位於 AMS 分布的左方，代表在相同重現期之下，由事件降雨量分布所推估的回歸水準都會低於由 AMS 分布所推估的。然而，當颱風事件降雨量分布經過卜瓦松或者負二項分布與最大順序統計量轉換成為颱風年最大值混合分布時，混合分布將因為最大順序統計量的加權而向右移動，在立山與楠溪兩側站的範例中，颱風年最大值混合分布在高重現期下的回歸水準大於 AMS 的回歸水準。此外，因為 EMS 年最大值分布接近颱風年最大值分布，EMS 的估計結果將高於 AMS 的估計結果，因此 EMS 的估計結果是否會低於 AMS 的估計結果需視測站資料而定。

AMS 的樣本數與紀錄年限相等、EMS 的樣本數則為各類型事件的發生次數，因此 EMS 的樣本數將遠多於 AMS，每一筆樣本對於 EMS 估計所具有的影響力也比較小——即 EMS 比較不容易受到極端值的影響。藉由將 AMS 與 EMS 數列中的極端值移除一筆並進行適合度檢定，可以比較兩種方法受極端值影響的情形。圖 6-2 到圖 6-5 右圖為移除最大值後重新計算的線性動差比，因為只移除一筆資料，



表 7-1 新增測站的基本資訊

站名	站號	所屬單位	流域	X 座標 (TM2)	Y 座標 (TM2)	資料蒐集年限	紀錄年限
立山	C1Z040	水利署	秀姑巒溪	283444	2593399	1997-2019	23
楠溪	C1V460	氣象局	濁水溪	239018	2592604	1993-2017	25

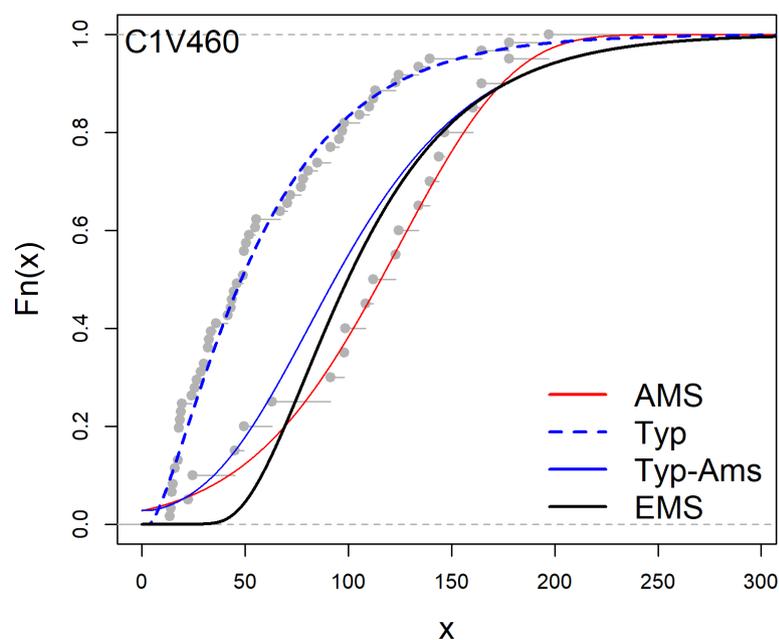
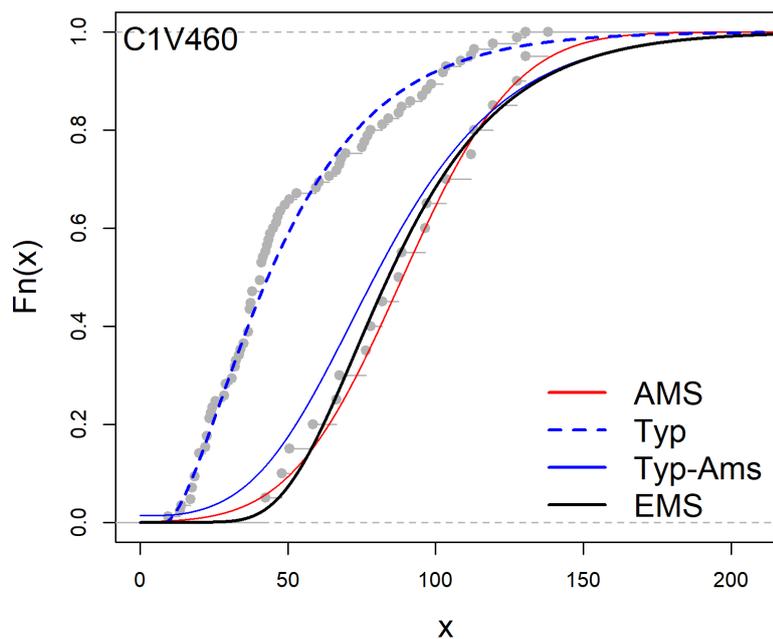


圖 7-1 3 小時降雨量累積分布函數

在 EMS 中僅會有一種類型的線性動差比產生變化，五堵、頭汴坑、嘉義被移除的最大值降雨分別來自颱風、梅雨和颱風。在將極端值移除後多數樣本均進入皮爾遜第三型分布的接受域，其中 AMS 的變化特別明顯，顯示單一極端值可以輕易影響由 AMS 所推得的分布形狀。使用移除極端值後的數列進行頻率分析的結果列於表 6-6 到表 6-8，與沒有移除極端值的數列相比，各重現期與設計延時下的回歸水準明顯降低，降低的幅度基本上隨著設計延時與重現期上升而增加，AMS 的變化較 EMS 為劇烈。在五堵站中，EMS 與 AMS 在長延時的線性動差比有比較明顯的變化，AMS 受極端值影響更大，其長延時線性動差比在移除極端值之後進入皮爾遜第三型分布的接受域，並且回歸水準的差最高可達將近 500 mm (72 小時、200 年)。在頭汴坑站，EMS 的線性動差比在移除極端值後更接近皮爾遜第三型分布，然而回歸水準估計值幾乎沒有變化，原因在於其極端值來自梅雨分布；在本研究中三個測站的年最大值混合分布與颱風事件降雨分布在高累積機率下幾乎相等，因此雖然梅雨事件的降雨分布受到極端值移除的影響，在颱風降雨分布不受影響的情況下，年最大值混合分布的估計也不會有明顯變化。然而，頭汴坑站的 AMS 線性動差比依然受到顯著影響，各延時與重現期的回歸水準明顯下降，其中設計延時 1、2、3、6 小時的線性動差比在極端值移除後進入皮爾遜第三型的接受域，因此其回歸水準的變化甚至大於高延時下的變化。嘉義站的 EMS 與 AMS 線性動差比變化並不明顯，依然維持前述之關係。

相較於 AMS，EMS 的回歸水準在相同重現期的長延時 (48 至 72 小時) 下的變動比較小，這是因為基於事件的樣本選取方法不會涵蓋複數事件，從圖 7-2 可以看出，在長延時上升趨緩的原因在於所有測站各類型事件的延時真正超過 48 小時的數量極少，因此即使設計延時增加，由 EMS 所選取的事件降雨量也不會上升，使 72 小時的事件降雨分布接近 48 小時事件降雨分布。此情形與現實中很少有降雨事件的延時長達 72 小時的情況相符，也顯示 72 小時的設計延時並不現實。另一方面，當設計延時大於事件延時的時候，AMS 會將鄰近事件一併納入考量，因此長延時下的 AMS 樣本將包含兩個以上的降雨事件，造成 AMS 在長設計延時下設計雨量持續上升，從表 6-5 可以發現將近所有 72 小時的 AMS 樣本都涵蓋 2 場以上的降雨事件，然而，這並不符合設計暴雨時單一事件的假設目標，以 AMS 進行設計很容易高估設計降雨量並產生誤差。



7.2 尖峰流量分析

在使用頻率分析產生設計雨量、決定雨型並生成設計暴雨後便可以進行逕流分析和尖峰流量計算，當樣本的累積雨量越高，後續的尖峰流量也會越高。在頻率分析結果中 AMS 方法的頻率分析結果傾向高於 EMS 結果，因此理論上其設計暴雨所造成的尖峰流量也將高於 EMS 方法的尖峰流量。然而考慮到 AMS 樣本選取的跨事件情形，這些 AMS 樣本在現實中的尖峰流量是否真的高於 EMS 樣本卻有待討論。由於測量困難，現行的河川流量主要藉由觀測河川水位再由水位——流量率定曲線獲得，然而河川的渠道、坡度等特性會隨著沖蝕、淤積而改變，因此率定曲線也會不斷變化，使流量無法準確估計。本節使用美國土壤保育局 (Soil Conservation Service, SCS；現改名為 Natural Resources Conservation Service, NRCS) 所發展之三角形單位歷線，針對三個雨量站中 AMS 與 EMS 的各年年最大值降雨進行逕流分析並比較尖峰流量 $Q_{p,AMS}$ 與 $Q_{p,EMS}$ 之差異。結果顯示，儘管 AMS 可能挑選出累積雨量比較大的樣本，這些樣本的尖峰流量皆可以被 EMS 的樣本所涵蓋，甚至會發生 AMS 樣本的尖峰流量比較低的情形。

7.2.1 SCS 三角形單位歷線

在設計暴雨中進行頻率分析並設計雨型後，經過降雨——逕流演算可以獲得集水區出口之流量。逕流的演算方法中以單位歷線最被廣為使用。如果集水區具備足夠的雨量與流量資料，可以直接計算集水區之單位歷線；如果缺乏紀錄，合成單位歷線 (synthesized unit hydrograph) 為經常使用的方法，其中以 SCS 三角形單位歷線最為普遍 (鄭克聲、陳葦庭、葉彙中 1999)。單位歷線的定義為在集水區內均勻降下一單位強度固定的超滲降雨所形成的直接逕流歷線。SCS 三角形單位歷線的計算方式可以參考圖 7-3，在設定到達尖峰流量前所經過的流量占總流量的 37.5% 的情況下，直接逕流的基期 t_b 會等於到達尖峰流量所需時間 t_p 的 2.67 倍，並且，尖峰流量 Q_p (cms) 可以表示為：

$$Q_p = \frac{484 * Area}{t_p} \quad (7-1)$$

$Area$ 為集水區面積(平方公里)。時間的單位均為小時，並且可以使用下式獲得：

$$t_l = 0.6 * t_c \quad (7-2)$$

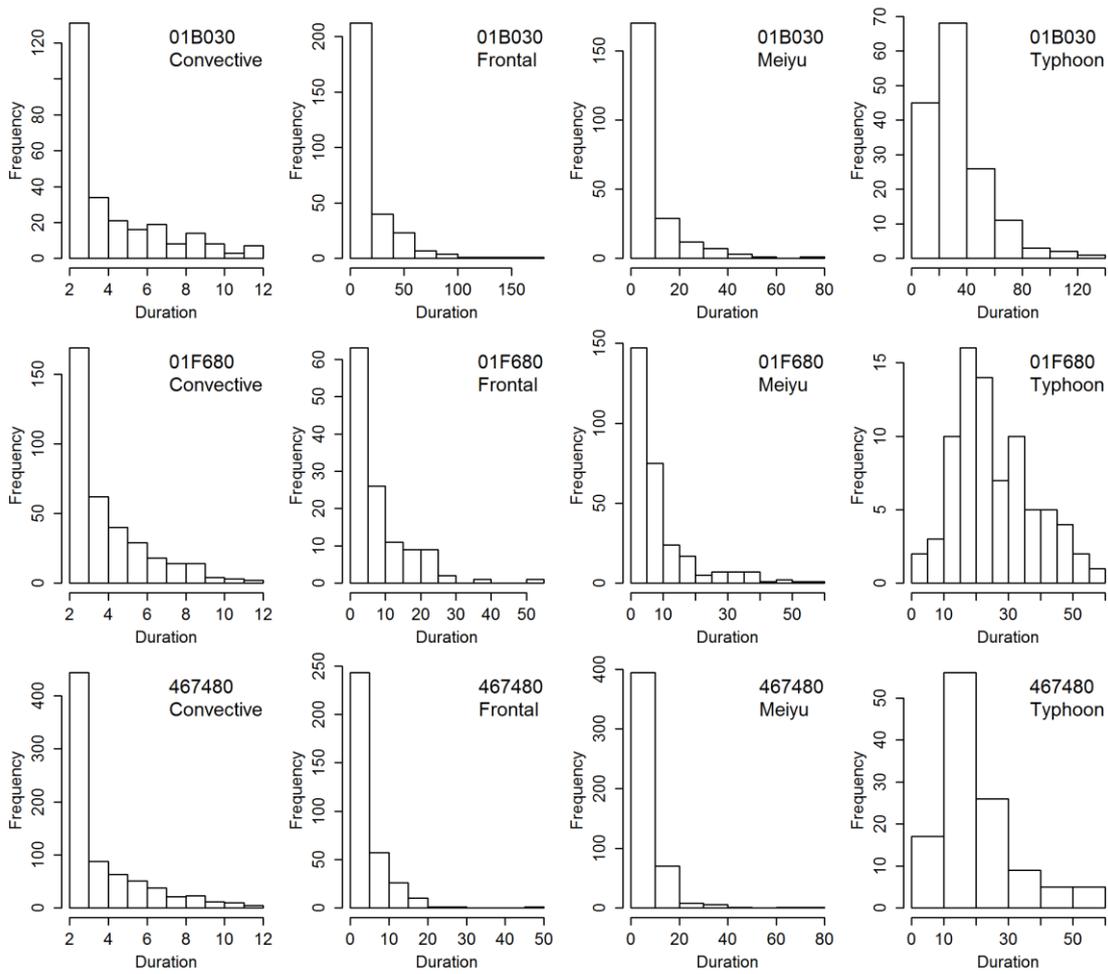
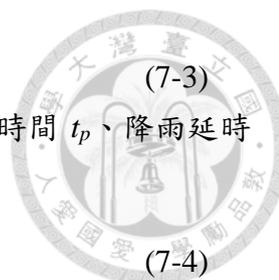


圖 7-2 三測站各類型事件延時長條圖



$$t_p = 0.5t_r + t_l \quad (7-3)$$

其中 t_c 為集水區集流時間。鄭克聲等人 (1999) 指出，尖峰到達時間 t_p 、降雨延時 t_r 和集流時間 t_c 需滿足以下關係式：

$$t_c \geq t_p > t_r \quad (7-4)$$

並且降雨延時 t_r 和延遲時間 t_l 須滿足：

$$t_r \leq 0.29t_l \quad (7-5)$$

在利用上式獲得無因次單位歷線之後，針對設計暴雨隨時間之雨量變化，可以使用摺合積分 (convolution integral) 計算集水區出口之直接逕流歷線與尖峰流量。

7.2.2 研究測站

根據雨量站的位置，本研究選取鄰近之流量站或者水位站作為集水區出口，並使用 SCS 三角形單位歷線計算 AMS 與 EMS 所選取之降雨事件樣本的尖峰流量 $Q_{p,AMS}$ 與 $Q_{p,EMS}$ 。三個測站的位置可見圖 6-1，測站資訊與參數列於表 7-2 和表 7-3，其中集水區的集流時間 t_c 若以集流時間公式 Kirpich、Rziha 或者加州公路局等計算會有高估的現象，在此以 1 至 10 小時作為延遲時間計算降雨與流量(如無流量資料則以水位資料代替)的相關係數 (表 7-3)，並以相關係數最高的延遲時間作為集水區集流時間，其中，頭汴坑站最低的相關係數發生於集流時間為 1 小時的時候，為滿足式(7-4)，本研究以 2 小時作為集流時間，單位歷線計算結果列於表 7-5。

7.2.3 尖峰流量比較

每一年計算而得的尖峰流量包含三種比較結果—— $Q_{p,AMS}$ 大於、小於或者等於 $Q_{p,EMS}$ ，前面兩種情形的發生次數可見表 7-6。相等的結果代表產生尖峰流量的降雨為同樣的降雨事件，此時當累積雨量也相等時代表兩種方法的窗口選取到單一事件內完全相同的紀錄，因此毋須討論。然而當窗口的累積雨量不同時，則代表 AMS 方法的窗口涵蓋兩場以上的降雨事件，以圖 7-4 為例，雖然在 48 小時窗口下 $Q_{p,AMS}$ 和 $Q_{p,EMS}$ 相等，AMS 窗口很明顯地因為涵蓋四場降雨事件而有更高的累積雨量，因此此樣本將增加 AMS 頻率分析的估計結果，然而從現實的尖峰流量來看此增長並不會影響流量，因此此 AMS 樣本將會使頻率分析結果高估雨量。當兩種方法的尖峰流量不同時，必定是因為 AMS 方法所選取的降雨數列涵蓋兩個以上的



降雨事件而造成差異。 $Q_{p,AMS}$ 小於 $Q_{p,EMS}$ 的情形顯示 AMS 所選取的樣本累積雨量比 EMS 從單一事件所選取的樣本累積雨量還多，複數事件的時間間隔卻導致雨量無法累積在單一尖峰流量、而使其尖峰流量低於 EMS 數列的尖峰流量 (圖 7-5)。這一類型的樣本一般發生於長延時數列，會使頻率分析的雨量估計結果上升，並且在暴雨事件設計時產生高估的尖峰流量。 $Q_{p,EMS}$ 低於 $Q_{p,AMS}$ 的情形同樣易發生於長延時樣本之中，此時的 AMS 樣本是由複數降雨事件所組成，因此其中產生 $Q_{p,AMS}$ 的事件在短延時事件中必定已經被比較過，在低延時的 EMS 樣本之中會存在累積雨量大於等於該短延時事件的樣本。當設計延時增加，EMS 轉而選擇其他具有更高累積雨量的長延時事件，而 AMS 則可以藉由選取複數事件繼續涵蓋該強降雨事件造成兩種方法的尖峰流量差異。換言之，在此情形下即使 $Q_{p,EMS}$ 低於 $Q_{p,AMS}$ 也不會造成低估，因為該尖峰流量在短延時的 EMS 樣本中已經出現過。以圖 7-6 為例，在圖 7-6 的 24 小時樣本中 $Q_{p,AMS}$ 高於 $Q_{p,EMS}$ ，然而由圖 7-6 的 12 小時樣本可以發現 24 小時的尖峰流量在 12 小時時已經在兩數列中出現過，隨著設計延時增加，EMS 轉而選取擁有更多累積雨量的事件，AMS 則藉由納入第二場事件繼續涵蓋原降雨事件，此時與第一種情形 ($Q_{p,AMS}$ 等於 $Q_{p,EMS}$ 、累積雨量不相等) 的情況相同，AMS 樣本中增加的降雨量既不來自相同降雨事件、對於尖峰流量也不會產生影響，因此將造成雨量高估。表 7-7 列出所有尖峰流量小於 AMS 的 EMS 樣本，經過比對可以發現雖然這些樣本的尖峰流量低於 $Q_{p,AMS}$ ，在同一年的低延時樣本中卻已經出現過相同甚至更高的尖峰流量，因此 EMS 抽樣並不會造成低估。

7.3 不確定性分析

7.3.1 拔靴法與信賴區間

拔靴法由 Efron 於 1979 年提出，是一種藉由重複抽樣來評估不確定性的方法，可以分為有母數拔靴法以及無母數拔靴法。有母數拔靴法類似於蒙地卡羅法，在知道母體分布的情況下，經由參數估計決定母體分布，再從估計的分布中產生新的樣本並估計目標統計量的信賴區間；無母數拔靴法以取後放回的方式從樣本中重複抽樣 (即由經驗分布 $F_n(x)$ 產生樣本) 並估計統計量與信賴區間。在無母數拔靴法中，經由重複抽樣產生的新樣本稱為拔靴法樣本 (bootstrap sample)，當抽取的樣本能夠代表母體時，樣本與母體之間的抽樣不確定性可以藉由拔靴法樣本與樣本之間

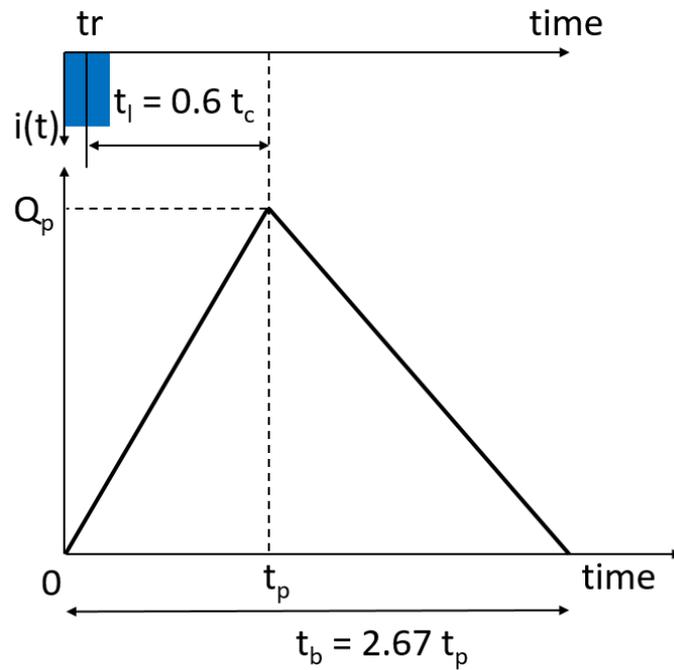


圖 7-3 SCS 單位歷線示意圖

表 7-2 流量站資訊

測站 編號	流量站/ 水位站	對應雨 量站	流域	X 坐標 (TM97)	Y 坐標 (TM97)	流域面積 (km ³)	資料蒐集年 限
1140H058	五堵 (流量站)	五堵	基隆河	319559.1	2774673.7	204.41	1966~1994, 1996~2002, 2004~2017
1430H051	一江橋 (水位站)	頭汴坑	頭汴坑溪	223184.3	2669800.2	96.38	2006-2017
1550H013	華興橋 (流量站)	嘉義	朴子溪	192217.7	2600836.9	170.64	2001,2003~ 2015

表 7-3 不同集流時間下各測站降雨和流量/水位之相關係數

Hr	1	2	3	4	5	6	7	8	9	10
1140H058	0.523	0.572	0.603	0.616	0.613	0.602	0.587	0.571	0.554	0.538
1430H051	0.184	0.174	0.162	0.152	0.144	0.139	0.133	0.129	0.125	0.120
1550H013	0.584	0.585	0.548	0.503	0.461	0.424	0.399	0.380	0.363	0.349

表 7-4 三測站單位歷線參數與計算結果

	t_c (hr)	t_r (hr)	t_p (hr)	t_l (hr)	t_b (hr)	Q_p (cms)
五堵	4	1.00	2.90	2.40	7.74	146.82
一江橋	2	1.00	1.70	1.20	4.54	118.09
華興橋	2	1.00	1.70	1.20	4.54	209.08

表 7-5 三測站之單位歷線

Time(hr)	五堵 UH(t,T)	一江橋 UH(t,T)	華興橋 UH(t,T)
0	0	0	0.00
0.5	25.31	34.73	61.50
1	50.62	69.47	122.99
1.5	75.94	104.20	184.49
2	101.25	105.61	186.99
2.5	126.57	84.82	150.17
3	143.79	64.02	113.34
3.5	128.63	43.22	76.52
4	113.47	22.42	39.70
4.5	98.31	1.62	2.87
5	83.15	0.00	0.00
5.5	67.99		
6	52.84		
6.5	37.68		
7	22.52		
7.5	7.36		
8	0.00		

表 7-6 各延時下 AMS 與 EMS 尖峰流量比較

設計延時(hr)		1	2	3	6	12	24	48	72	記錄長
1140H058	$Q_{p,AMS}$ Higher	0	0	0	0	0	0	1	6	32
	$Q_{p,EMS}$ Higher	0	0	0	1	0	1	4	3	32
1430H051	$Q_{p,AMS}$ Higher	0	0	0	0	0	1	3	4	42
	$Q_{p,EMS}$ Higher	0	0	0	0	0	3	1	2	42
1550H013	$Q_{p,AMS}$ Higher	0	0	0	0	0	3	8	9	44
	$Q_{p,EMS}$ Higher	0	0	0	0	1	6	4	9	44

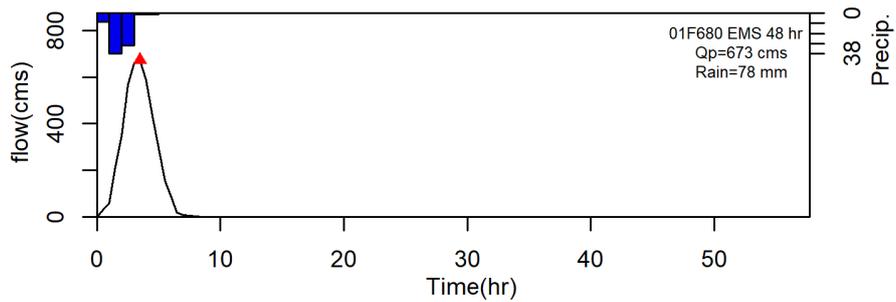
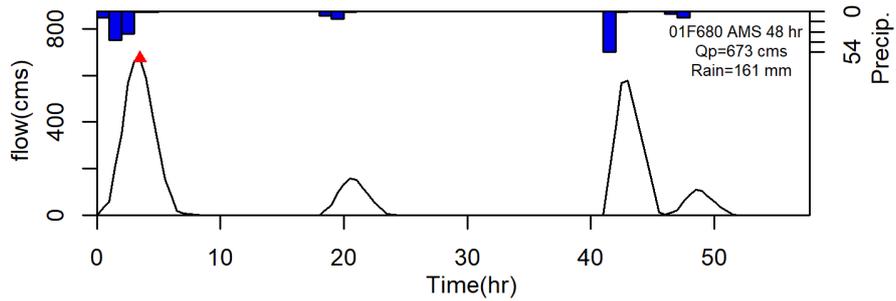


圖 7-4 $Q_{p,EMS}$ 等於 $Q_{p,AMS}$ 、累積降雨不相等之樣本。

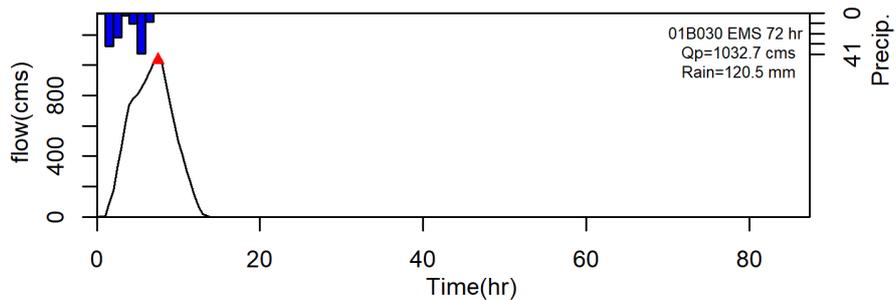
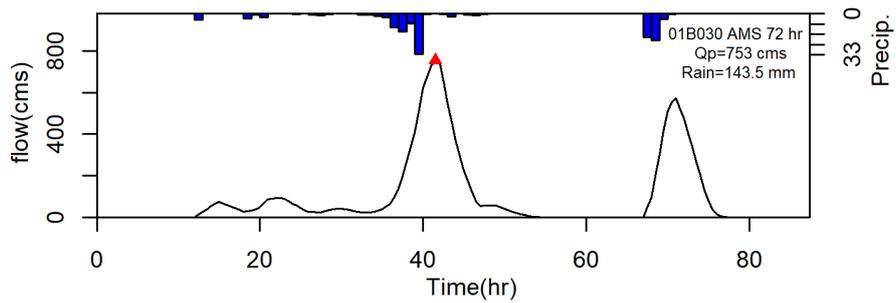


圖 7-5 $Q_{p,EMS}$ 大於 $Q_{p,AMS}$ 之樣本

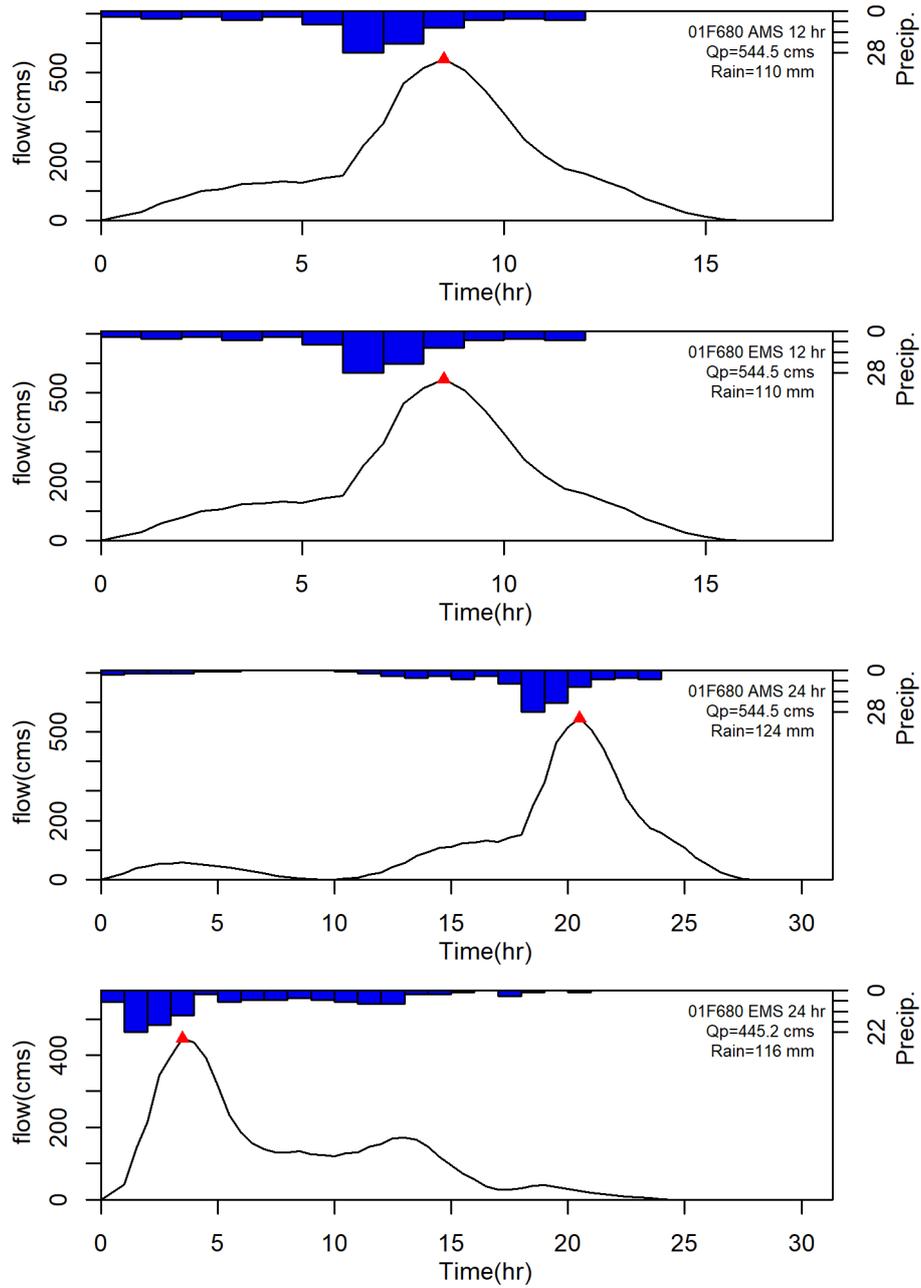


圖 7-6 $Q_{p,EMS}$ 小於 $Q_{p,AMS}$ 之累積降雨樣本



表 7-7 長延時 $Q_{p,AMS}$ 大於 $Q_{p,EMS}$ 之樣本及與其對應之 $Q_{p,EMS}$ 降雨延時 (粗體表尖峰流量相等)

1140H058			1430H051			1550H013			1550H013		
Duration	Year	Matched EMS duration	Duration	Year	Matched EMS design duration	Duration	Year	Matched EMS duration	Duration	Year	Matched EMS duration
48	1992	2, 3, 6, 12, 24	24	1998	12	24	1988	12	48	2012	1, 2, 3, 6, 12
72	1988	3, 6, 12, 24	48	1974	3, 6, 12	24	1997	1, 2, 3, 6, 12, 48, 72	72	1982	2, 3, 6
72	1990	6, 12	48	1987	12	24	2010	1, 2, 3	72	1983	3, 6, 12, 24
72	1992	2, 3, 6, 12, 24	48	1991	12, 24	48	1976	2, 3, 6, 12, 24	72	1984	2, 3, 6, 12, 24(higher)
72	1995	1, 2, 3	72	1974	3, 6, 12	48	1983	3, 6, 12, 24	72	1988	12
72	2000	2, 12, 24, 48	72	1987	12	48	1988	12	72	1992	2
72	2011	12	72	1991	12, 24	48	1992	2	72	1999	1, 2, 3, 6, 24(higher)
			72	2005	1, 2, 3, 6	48	1999	1, 2, 3, 6, 24(higher)	72	2007	3, 6, 24
			72	2007	1,2,3,6(higher)	48	2007	3, 6, 24	72	2010	1, 2, 3
						48	2010	1, 2, 3	72	2012	1, 2, 3, 6, 12

的抽樣不確定性來估計，並且用以估計信賴區間。本研究使用無母數拔靴法建立回歸水準之信賴區間，在不假設母體分布的情況下，經由重複抽樣從資料中提取資訊以評估抽樣不確定性。拔靴法中常見的信賴區間估計方法包括標準常態拔靴法 (standard normal bootstrap)、t 分布拔靴法 (t bootstrap)、百分位拔靴法以及修正偏誤加速法。前兩者藉由假設模型輸出值近似常態分佈，分別以常態分佈與 t 分布計算信賴區間；百分比拔靴法與 BCa 法則是以拔靴法抽樣分布的分位數作為信賴區間的上下界，其中 BCa 方法特別針對具偏態與偏誤的統計量給予調整。針對一組 *i.i.d* 的隨機樣本 X_1, \dots, X_N ，可以計算目標估計 $\hat{\theta}$ (在本研究中為回歸水準)，此外，對樣本進行取出放回的重複抽樣可以獲得拔靴法樣本 X_1^*, \dots, X_N^* 並獲得估計 $\hat{\theta}^*$ ，重複 B 次則可以計算 $\hat{\theta}_{(1)}^*, \dots, \hat{\theta}_{(B)}^*$ ，再以 $\hat{\theta}^*$ 建立拔靴法樣本分布 $F_{\hat{\theta}^*}$ 。若將 $\hat{\theta}$ 視為母體理論值、 $\hat{\theta}^*$ 為抽取樣本，則 $F_{\hat{\theta}^*}$ 即為抽樣分布；因此拔靴法樣本分布 $F_{\hat{\theta}^*}$ 可以視為真正的抽樣分布 $F_{\hat{\theta}}$ 的估計量 $\hat{F}_{\hat{\theta}}$ ，再由此 $\hat{F}_{\hat{\theta}}$ 估計信賴區間。

本研究使用百分比拔靴法與 BCa 法作為信賴區間的建立方法，百分位拔靴法所建立的 $100 \cdot (1-\alpha)\%$ 信賴區間上下界 $[\hat{\theta}_L, \hat{\theta}_U]$ 可以由下式計算：

$$[\hat{\theta}_L, \hat{\theta}_U] = [\hat{F}_{\hat{\theta}^*, \frac{\alpha}{2}}, \hat{F}_{\hat{\theta}^*, 1-\frac{\alpha}{2}}] \quad (7-6)$$

即由 $\hat{\theta}^*$ 所估計的抽樣分布 $F_{\hat{\theta}^*}$ 的 $\frac{\alpha}{2}$ 與 $1-\frac{\alpha}{2}$ 分位數。

BCa 方法所選取的百分位數受到校正偏誤的 z_0 與校正偏態的 a 兩個因子所調整，可以在多數情形下獲得比百分位拔靴法更好的結果。針對 $100 \cdot (1-\alpha)\%$ 的信賴區間，可以計算上下界所使用的累積機率 α_1, α_2 ：

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 + \hat{a}(\hat{z}_0 + z_{\alpha/2})} \right) \quad (7-7)$$

$$\alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 + \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \right) \quad (7-8)$$

其中 $z_\alpha = \Phi^{-1}(\alpha)$, $\Phi(\cdot)$ 為標準常態分布之累積機率函數，此外 \hat{z}_0 與 \hat{a} 定義為：

$$\hat{z}_0 = \Phi^{-1} \left(\frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_b < \hat{\theta}) \right) \quad (7-9)$$

$$\hat{a} = \frac{\sum_{i=1}^n (\bar{\theta}_{(i)} - \theta_{(i)})^3}{6 \left(\sum_{i=1}^n (\bar{\theta}_{(i)} - \theta_{(i)})^2 \right)^{3/2}} \quad (7-10)$$



BCa 的信賴區間可以由下式計算：

$$[\hat{\theta}_L, \hat{\theta}_U] = [\hat{F}_{\hat{\theta}, \alpha_1}, \hat{F}_{\hat{\theta}, \alpha_2}] \quad (7-11)$$

7.3.2 年最大值數列抽樣

令 a_i , $i = 1 \dots N$ 為第 i 年的年最大值降雨 ($a_i = \max\{x_{i,j}\}$)，則年最大值數列可以使用

拔靴法建立信賴區間：

- (1) 由 a_1, \dots, a_N 抽取年最大值事件樣本 $a_1^* \dots a_N^*$
- (2) 使用皮爾遜第三型分布估計回歸水準 \hat{x}
- (3) 重複上述步驟 B 次，並計算 $100 \cdot (1-\alpha)\%$ 信賴區間的上下界。

重複抽樣次數 B 並沒有標準，Zucchini and Adamson (1989) 中使用模擬發現 200 次的抽樣即可對部分延時數列建立符合需求的信賴區間。在此設定 B 為 1000 以確保產生足夠的拔靴法樣本。

7.3.3 事件最大值數列抽樣

針對事件最大值數列使用拔靴法進行信賴區間的建立可以參考 Zucchini and Adamson (1989) 中對部分延時數列使用拔靴法建立信賴區間的過程。以颱風雨梅雨兩種事件為例，讓 N 代表降雨紀錄長度、 n_{1i}, n_{2i} , $i = 1 \dots N$ 分別代表第 i 年的颱風以及梅雨事件數， $x_{i,j}$, $i = 1 \dots N$, $j = 1 \dots n_{1i}$ 是特定延時下第 i 年第 j 個颱風事件的降雨量， $y_{i,j}$, $i = 1 \dots N$, $j = 1 \dots n_{2i}$ 為梅雨降雨量，則事件最大值數列的 $100 \cdot (1-\alpha)\%$ 信賴區間建立過程如下：

- (1) 針對每一年的事件發生數進行抽樣，得到颱風事件數的拔靴法樣本 $n_{1,1}^* \dots n_{1,N}^*$ 以及梅雨事件數的拔靴法樣本 $n_{2,1}^* \dots n_{2,N}^*$ 。
- (2) 針對每一年的事件數，從特定延時下的事件降雨量之中抽取對應數量的颱風降雨量 $x_{1,1}^* \dots x_{N,n_{1N}}^*$ 與梅雨降雨量 $y_{1,1}^* \dots y_{N,n_{2N}}^*$ 。
- (3) 由每一年的事件估計颱風、梅雨事件的年最大值混合分布參數，並計算相應之回歸水準 \hat{x} 。
- (4) 重複上述步驟 B 次，並計算 $100 \cdot (1-\alpha)\%$ 信賴區間的上下界。

7.3.4 信賴區間模擬

信賴區間的定義為有具有特定機率涵蓋理論值的區間，95% 信賴區間即代表該



區間有 95% 的機率涵蓋理論值，藉由模擬重複產生樣本與信賴區間，可以驗證前述方法是否符合預期並能與 AMS 之信賴區間進行比較。其過程如下：

- (1) 參考 4.1 的頻率分析模擬產生降雨資料。
- (2) 由上述步驟建立事件最大值數列以及年最大值數列之 95% 和 90% 信賴區間。
- (3) 重複模擬 1000 次並計算信賴區間對於理論值的覆蓋率。

模擬結果列於表 7-8，基本上無論是百分位法還是 BCa 法在 AMS 與 EMS 方法中皆無法達到理論覆蓋率，BCa 信賴區間在 EMS 法中的覆蓋率有提升，在 AMS 方法中的覆蓋率卻下降。圖 7-8 是 100 年重現期的 1000 次百分比信賴區間模擬結果，綠色為未涵蓋理論值的信賴區間，未覆蓋的信賴區間傾向低於理論值，從圖 7-9 拔靴法過程中的抽樣分布中可以看出，各分布在低與高累積機率時容易產生偏差，因此當重現期設定為 2 年（累積機率 0.5）時，未經調整的百分位信賴區間的覆蓋率都相當準確，Bca 方法的結果也相近。本研究中雖然並未採用廣義極端值分布進行頻率分析，AMS 仍然有低樣本數、重尾的特性，因此 AMS 的信賴區間覆蓋率與 Kysely (2008) 具有類似的結果；此外，隨著重現期上升，AMS 信賴區間的覆蓋率也隨之下降，而 EMS 則相對穩定。整體而言，EMS 的信賴區間在拔靴法下較 AMS 為穩定與準確。

7.3.5 真實資料設計降雨之信賴區間

針對三個測站不同設計延時與重現期下所計算的百分比信賴區間結果繪於圖 7-10。兩種方法的信賴區間間隔都隨重現期與延時增加而增加，從真實資料以及模擬都可以看出 AMS 的信賴區間比 EMS 還要寬，特別是在上界的部分，25 年重現期的信賴區間上界已經超過 200 年重現期的估計值，顯示 AMS 的信賴區間的上界估計值具有比較高的不確定性；EMS 具有比較短的信賴區間，此外重疊的情形也比較不明顯，可以提供使用者更好的不確定性分析依據。

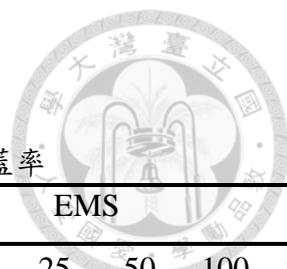


表 7-8 AMS 與 EMS 拔靴法信賴區間之理論值覆蓋率

		AMS							EMS						
重現期		2	5	10	25	50	100	200	2	5	10	25	50	100	200
Percenti	90%	91.1	85.9	82.1	78.2	77.4	76.7	76.8	88.3	86.3	85.8	85.1	84.9	84.7	84.4
	95%	95.1	93.2	89.9	87.1	85.7	85.1	84.9	95.5	93.9	92.9	92.1	91.8	91.7	91.7
BCa	90%	90.2	86.3	83.2	79.2	78.5	78.3	77.8	88.6	87.8	87.9	88.0	88.0	87.9	88.0
	95%	94.6	92.2	88.0	85.1	84.1	83.9	83.3	95.0	95.1	94.5	94.6	94.2	94.1	94.0

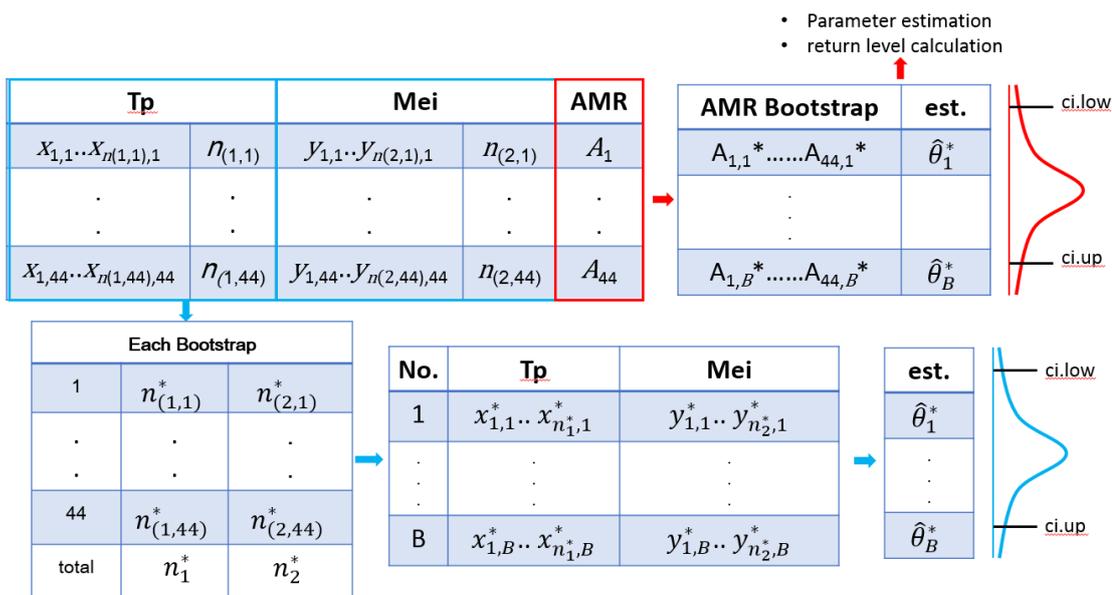


圖 7-7 拔靴法示意圖

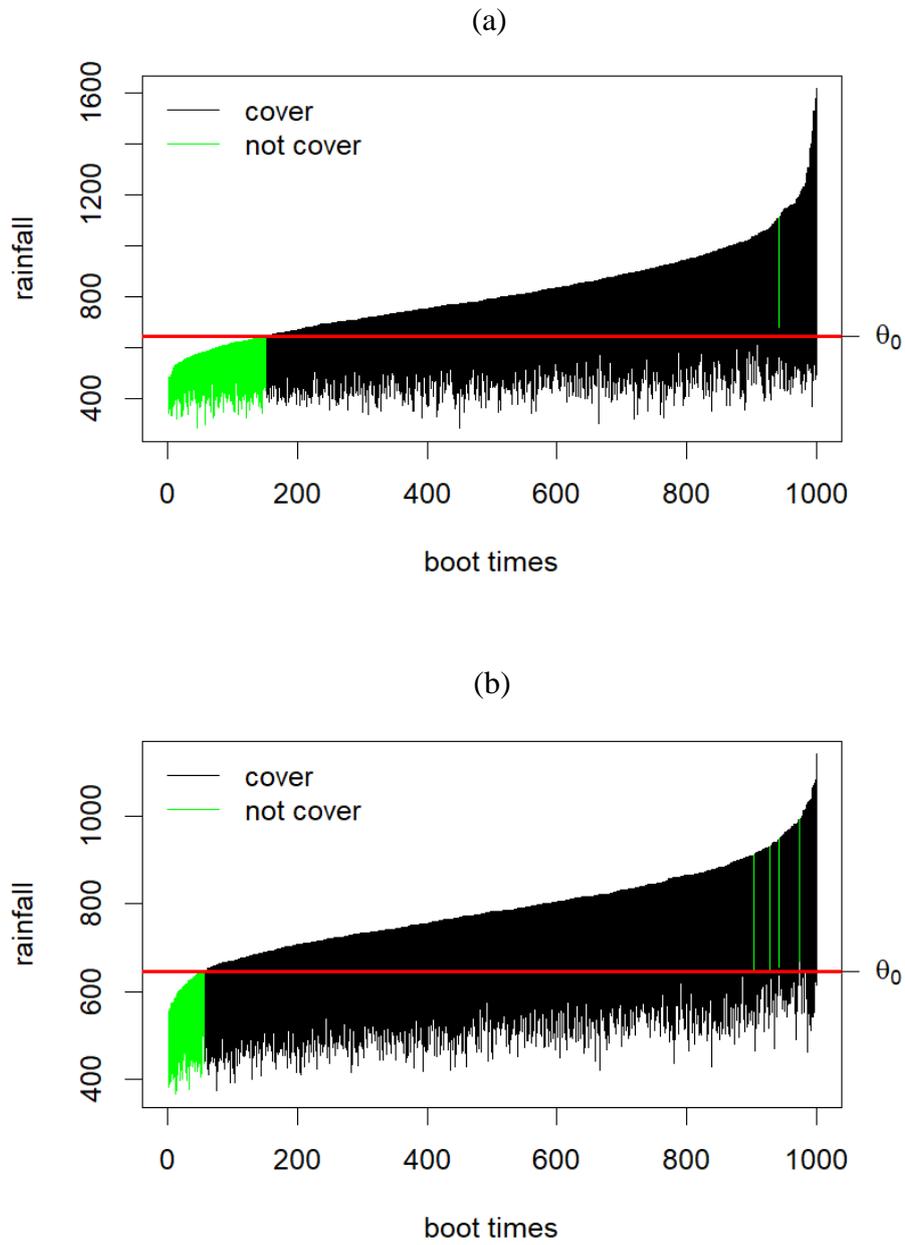
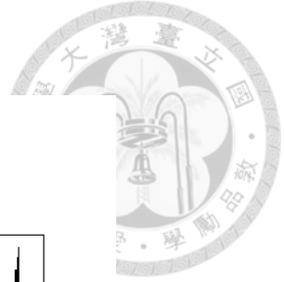


圖 7-8 100 年重現期 1000 次 95%信賴區間模擬 (percentile)，(a)AMS (b)EMS

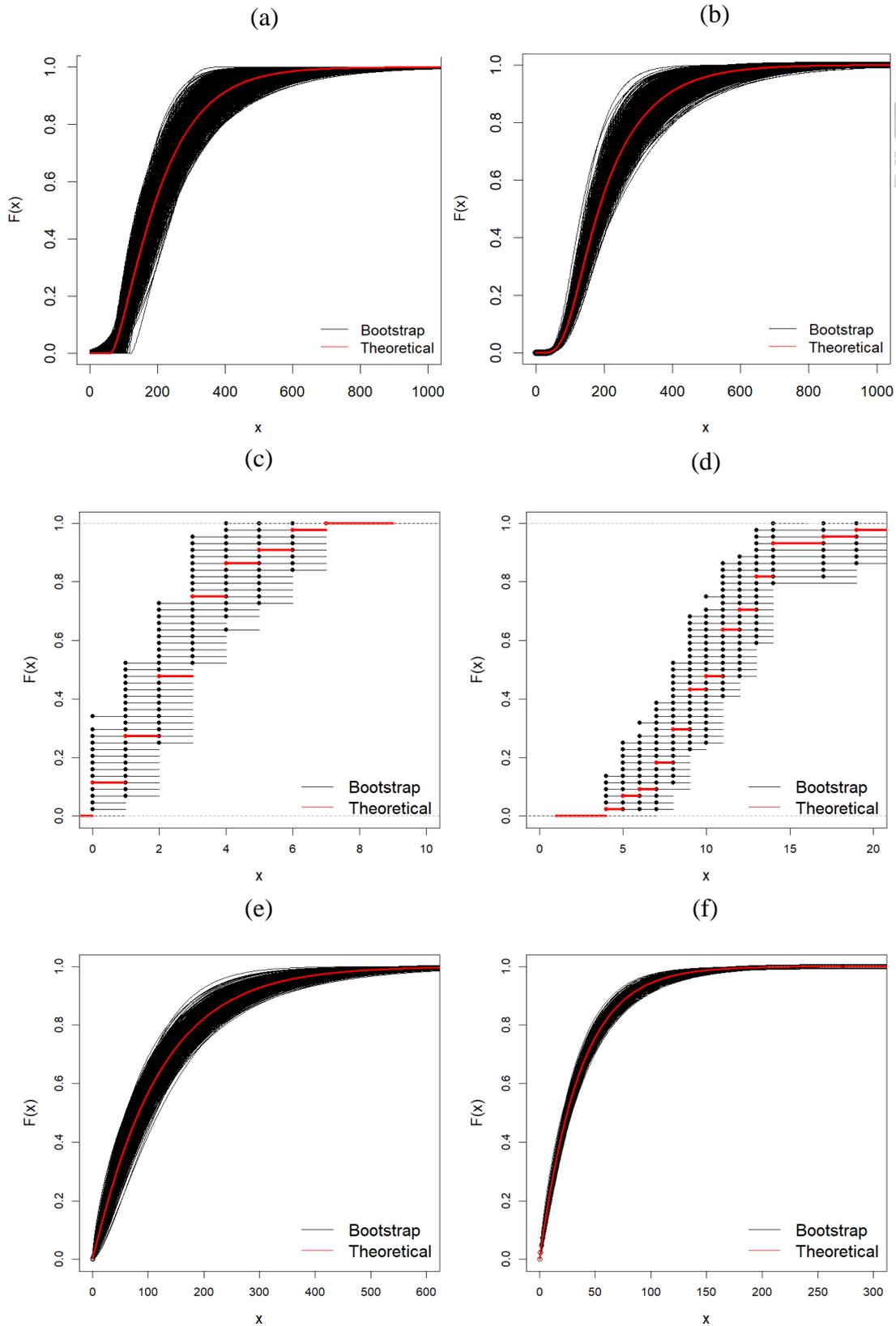


圖 7-9 Percentile Bootstrap (a) AMS、(b)EMS、(c)颱風事件數、(d)梅雨事件數、(e)颱風事件降雨量、(f)梅雨事件降雨量累積分布

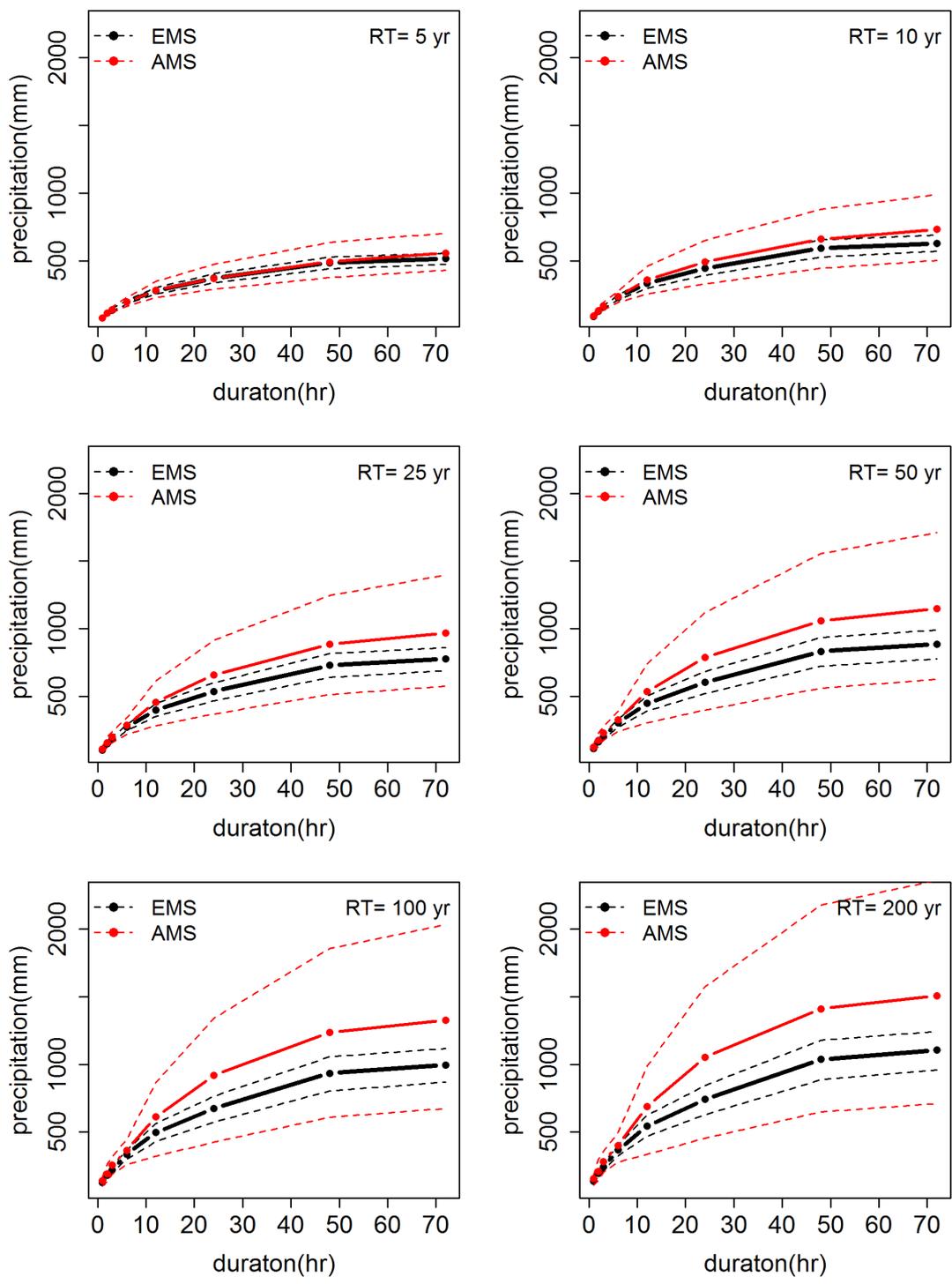
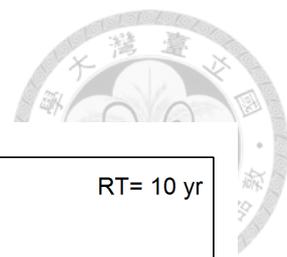


圖 7-10 五堵站百分比信賴區間估計結果

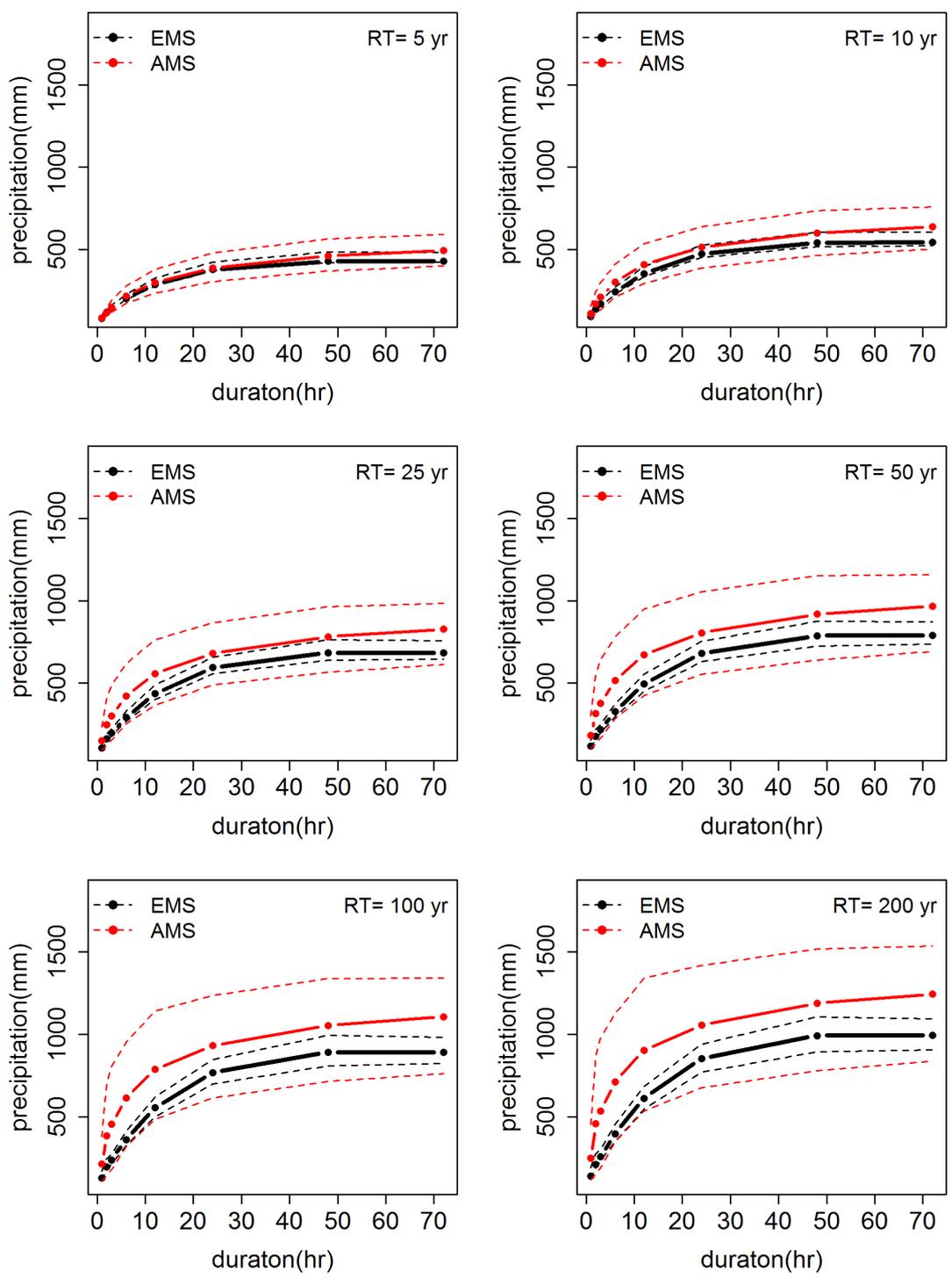


圖 7-11 頭汴坑站百分比信賴區間估計結果

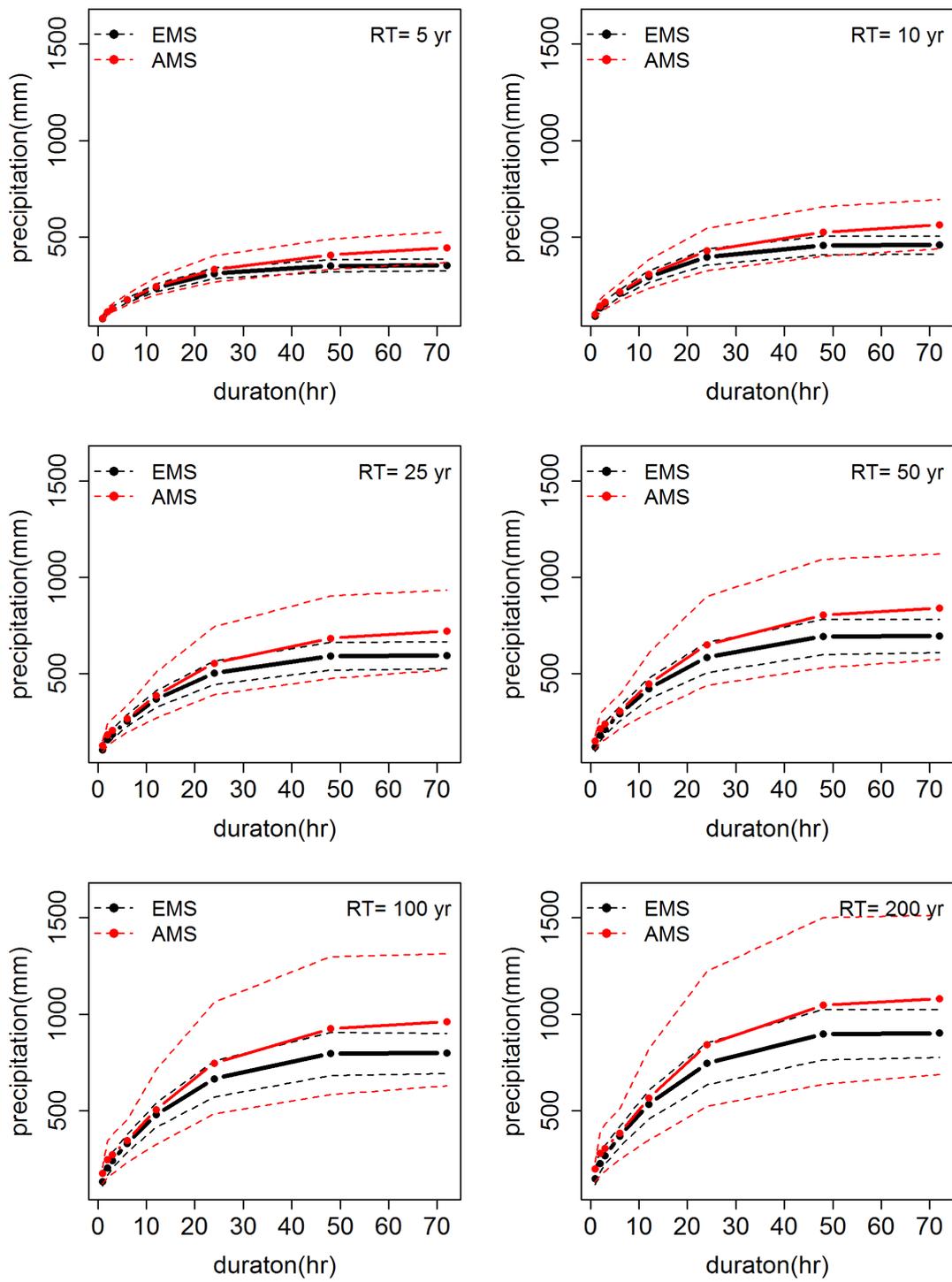


圖 7-12 嘉義站百分比信賴區間估計結果



第八章 結論



本研究提出事件最大值數列法作為傳統年最大值數列法之替代。在事件最大值數列法中，降雨紀錄使用最小事件間格時間切割為四種降雨事件——颱風、梅雨、對流雨和鋒面雨，因此年最大值的分布為此四種類型降雨的年事件數分布與雨量分布之混合分布。在本研究中，年最大值數列法的年事件數採用負二項分布、事件降雨量而則採用皮爾遜第三型分布進行分析，除了以序率模擬和年最大值數列法比較雨量估計結果、並探討年最大值於極端值分布之收斂以外，也針對臺灣北、中、南各一測站進行實際資料分析。最後分別使用三角形合成單位歷線和無母數拔靴法計算尖峰流量和信賴區間並進行比較，得到的結論如下：

- (1) 模擬結果顯示，年最大值數列法受限於樣本數目，所估計回歸水準的均方根誤差會高於事件最大值數列法。除此之外，與極端值理論假設不符合的情形將使廣義極端值分布的推估有更高的偏誤與均方根誤差。在廣義極端值分布的收斂方面，隨機年事件數會使得年最大值分布在低發生率時偏離廣義極端值分布，而後更快速得接近逼近廣義極端值分布。除此之外，當年事件數為隨機時，年最大值是選取自單一分布還是多種分布將影響收斂至廣義極端值分布的速度。
- (2) 在實際資料分析中，EMS 高重現期下的年最大值分布與颱風年最大值分布相當接近，顯示颱風降雨可以代表年最大值降雨。AMS 的估計普遍高於 EMS 的估計，原因包括 EMS 考慮到降雨事件而不會選取跨事件樣本，以及 EMS 的各年年最大值樣本分散於四種類型的降雨之中，與其他低雨量事件一起估計。然而，順序統計量的加權有可能讓 EMS 的估計值高於 AMS。除此之外，取樣不跨事件也使 EMS 的估計在長延時下增加趨緩
- (3) 在移除極端值的影響中，兩種方法的適合度檢定都會受到極端值影響，並且 AMS 受極端值影響的程度高於 EMS，EMS 的極端值如果發生在颱風對於設計降雨會有比較明顯的影響。移除極端值後，AMS 的估計結果會更接近甚至低於 EMS 的估計結果，如果針對擁有莫拉克颱風紀錄的南部雨量站進行移除極端值的頻率分析，因為缺少莫拉克颱風的極端降雨紀錄，AMS 的估計結果也有機會小於等於 EMS 之結果，顯示 AMS 受到單一年度的極端降雨影響，很容易產生

高估或者低估的結果。

- (4) 針對 AMS 與 EMS 的年最大值樣本使用合成單位歷線計算尖峰流量，結果顯示在長延時下 AMS 所選取之樣本容易涵蓋兩場以上的降雨事件並造成兩種結果。第一種為 AMS 樣本雖然因為跨事件選取而有高雨量，其尖峰流量卻低於 EMS 之尖峰流量，因此雨量實際上被高估。第二種則是 AMS 因為跨事件取樣而可以在長延時樣本中繼續選取短延時強降雨的降雨樣本，因此長延時下的尖峰流量將高於 EMS 的尖峰流量；然而該強降雨樣本會出現於 EMS 的短延時樣本之中，因此在工程設計中並不會產生影響。尖峰流量分析顯示 AMS 所得到的雨量樣本所產生之尖峰流量並不會高於 EMS 雨量樣本的尖峰流量而發生高估雨量的情形。
- (5) 本研究使用百分比拔靴法與 BCa 法針對回歸水準建置信賴區間。模擬結果顯示以拔靴法建立 EMS 信賴區間覆蓋率會優於 AMS 的區間，並且隨著重現期增加 EMS 信賴區間也比較穩定。然而兩種方法都無法達到理論覆蓋率，未來仍有改進空間。

參考文獻

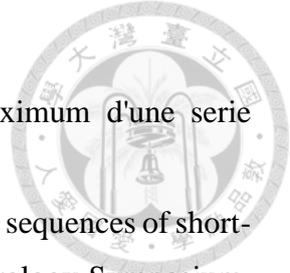


- Adams BJ, Papa F (2000). *Urban Stormwater Management Planning with Analytical Probabilistic Models*. Wiley, New York.
- Al Mamoon A, Rahman A (2014). Uncertainty in design rainfall estimation: A review. *J. Hydrol. Environ. Res*, 2(1):65-75.
- Anagnostopoulou C, Tolika K (2012). Extreme precipitation in Europe: statistical threshold selection based on climatological criteria. *Theoretical and Applied Climatology*, 107(3-4):479-489.
- Ariff NM, Jemain AA, Ibrahim K, Zin WW (2012). IDF relationships using bivariate copula for storm events in Peninsular Malaysia. *Journal of Hydrology*, 470:158-171.
- Bačová-Mitková V, Onderka M (2010). Analysis of extreme hydrological events on the Danube using the peak over threshold method. *Journal of Hydrology and Hydromechanics*, 58(2):88-101.
- Bezak N, Brilly M, Šraj M (2014). Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal*, 59(5):959-977.
- Bhunya PK, Berndtsson R, Jain SK, Kumar R (2013). Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS). *Journal of Hydrology*, 497:121-132.
- Bonta JV (2004). Stochastic simulation of storm occurrence, depth, duration, and within-storm intensities. *Transactions of the ASAE*, 47(5):1573.
- Bonta JV, Rao AR (1988). Factors affecting the identification of independent storm events. *Journal of Hydrology*, 98(3-4):275-293.
- Bracken LJ, Cox NJ, Shannon J (2008). The relationship between rainfall inputs and flood generation in south-east Spain. *Hydrological Processes*, 22(5):683-696.
- Cameron AC, Trivedi PK (1996). Count data models for financial data. *Handbook of statistics*, 14(12):363-391.
- Chin RJ, Lai SH, Chang KB, Jaafar WZW, Othman F (2016). Relationship between minimum inter-event time and the number of rainfall events in Peninsular Malaysia. *Weather*, 71(9):213-218.
- Coles S, Bawa J, Trenner L, Dorazio P (2001). *An introduction to statistical*



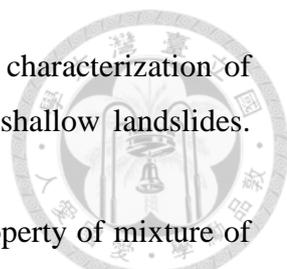
modeling of extreme values. Springer, London.

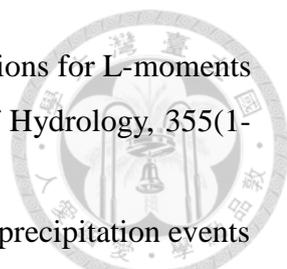
- Conover WJ (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, 67(339):591-596.
- Consul PC, Jain GC (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4):791-799.
- Cunnane C (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18(3-4):257-271.
- Cunnane C (1978). Unbiased plotting positions—a review. *Journal of Hydrology*, 37(3-4):205-222.
- Cunnane C (1979). A note on the Poisson assumption in partial duration series models. *Water Resources Research*, 15(2):489-494.
- D'Agostino RB (1986). *Goodness-of-fit-techniques*. Marcel Dekker, New York.
- De Michele C, Salvadori G (2003). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres*, 108(D2):1-11.
- Dunkerley D (2008). Identifying individual rain events from pluviograph records:a review with analysis of data from an Australian dryland site. *Hydrological Processes*, 22(26):5024-5036.
- Dzipire NC, Ngare P, Odongo L (2018). A poisson-gamma model for zero inflated rainfall data. *Journal of Probability and Statistics*, 2018(10):1-13.
- Eastoe EF, Tawn JA (2010). Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water resources research*, 46(2):W02510. doi:[10.1029/2009WR007757](https://doi.org/10.1029/2009WR007757).
- Efron B (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics*, 7(1):1-26.
- Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Chapman & Hall, Boca Raton, FL.
- Ferreira A, De Haan L (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of statistics*, 43(1):276-298.
- Fisher RA, Tippett LHC (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the*

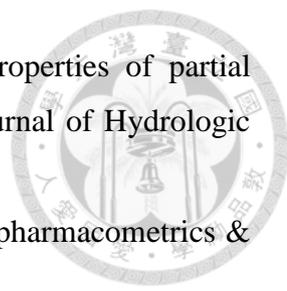


Cambridge Philosophical Society 24(2):180-190.

- Gnedenko B (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 44(3):423-453.
- Grace RA, Eagleson PS (1967). A model for generating synthetic sequences of short-time-interval rainfall depths. *Proceedings of International Hydrology Symposium, Fort Collins, Colorado*, 1:268-276
- Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water resources research*, 15(5):1049-1054.
- Gumbel EJ (1958). *Statistics of Extremes*, Columbia Univ. Press, New York.
- Guo Y, Adams BJ (1998). Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume. *Water Resources Research*, 34(12):3421-3431.
- Haile AT, Rientjes TH, Habib E, Jetten V, Gebremichael M (2011). Rain event properties at the source of the Blue Nile River. *Hydrology and Earth System Sciences*, 15(3):1023-1034.
- Hanel M, Máca P (2014). Spatial variability and interdependence of rain event characteristics in the Czech Republic. *Hydrological processes*, 28(6):2929-2944.
- Harrison XA (2015). A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology evolution. *PeerJ*, 3:e1114.
- Hosking JRM (1990). L-moment: analysis and estimation of distributions using linear-combinations of order-statistics. *Journal of the Royal Statistical Society. Series B: Methodological*, 52(1):105-124.
- Hosking JRM and Wallis JR (1997). *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, Cambridge
- Hosking JRM, Wallis JR, Wood EF (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251-261
- Hou JC (2010). 雙變數迦瑪分布頻率因子序率模擬法及其應用於評估氣候變遷對年颱風降雨量影響之研究. 臺灣大學生物環境系統工程學研究所學位論文

- 
- Iadanza C, Trigila A, Napolitano F (2016). Identification and characterization of rainfall events responsible for triggering of debris flows and shallow landslides. *Journal of Hydrology*, 541:230-245
 - Joe H, Zhu R (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(2):219-229
 - Joo J, Lee J, Kim J, Jun H, Jo D (2014). Inter-event time definition setting procedure for urban drainage systems. *Water*, 6(1):45-58
 - Jun C, Qin X, Gan TY, Tung YK, De Michele C (2017). Bivariate frequency analysis of rainfall intensity and duration for urban stormwater infrastructure design. *Journal of Hydrology*, 553:374-383
 - Jun C, Qin X, Tung YK, De Michele C (2018). Storm event-based frequency analysis method. *Hydrology Research*, 49(3):700-710
 - Kao S, Govindaraju R (2007) A bivariate frequency analysis of extreme rainfall with implications for design. *J Geophys Res* 112. doi:[10.1029/2007JD008522](https://doi.org/10.1029/2007JD008522)
 - Karlis D, Xekalaki E (2005). Mixed poisson distributions. *International Statistical Review*, 73(1):35-58
 - Katz RW, Parlange MB (1998). Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, 11(4):591-601
 - Kysely J (2008). A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *Journal of Applied Meteorology and Climatology*, 47(12):3236-3251
 - Kysely J (2010). Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theoretical and applied climatology*, 101(3-4):345-361
 - Lang M, Ouarda TBMJ, Bobée B (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225(3-4):103-117
 - Lee CH, Kim TW, Chung G, Choi M, Yoo C (2010). Application of bivariate frequency analysis to the derivation of rainfall–frequency curves. *Stochastic Environmental Research and Risk Assessment*, 24(3):389-397
 - Lindén A, Mäntyniemi S (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414-1421

- 
- Liou JJ, Wu YC, Cheng KS (2008). Establishing acceptance regions for L-moments based goodness-of-fit tests by stochastic simulation. *Journal of Hydrology*, 355(1-4):49-62
 - Liu, J., Sample, D. J., Zhang, H. (2014). Frequency analysis for precipitation events and dry durations of Virginia. *Environmental Modeling & Assessment*, 19(3):167-178
 - Madsen H, Rasmussen PF, Rosbjerg D (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water resources research*, 33(4):747-757
 - Marin CT, Bouten W, Sevink J (2000). Gross rainfall and its partitioning into throughfall, stemflow and evaporation of intercepted water in four forest ecosystems in western Amazonia. *Journal of Hydrology*, 237(1-2):40-57
 - McMahon BJ, Purvis G, Sheridan H, Siriwardena GM, Parnell AC (2017). A novel method for quantifying overdispersion in count data and its application to farmland birds. *Ibis*, 159(2):406-414
 - Medina-Cobo MT, García-Marín AP, Estévez J, Ayuso-Muñoz JL (2016). The identification of an appropriate Minimum Inter-event Time (MIT) based on multifractal characterization of rainfall data series. *Hydrological Processes*, 30(19):3507-3517
 - Nagy BK, Mohssen M, Hughey KFD (2017). Flood frequency analysis for a braided river catchment in New Zealand: Comparing annual maximum and partial duration series with varying record lengths. *Journal of Hydrology*, 547:365-374
 - Nelsen RB (2006). *An introduction to copulas*. Springer Science & Business Media, New York
 - Panagoulia D, Economou P, Caroni C (2014). Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change. *Environmetrics*, 25(1):29-43
 - Park M, Yoo C, Kim H, Jun C (2013). Bivariate frequency analysis of annual maximum rainfall event series in Seoul, Korea. *Journal of Hydrologic Engineering*, 19(6):1080-1088
 - Peters O, Christensen K (2002). Rain: Relaxations in the sky. *Physical Review E*, 66(3):036120

- 
- Pham HX, Shamseldin AY, Melville B (2013). Statistical properties of partial duration series: Case study of North Island, New Zealand. *Journal of Hydrologic Engineering*, 19(4):807-815.
 - Plan EL (2014). Modeling and simulation of count data. *CPT: pharmacometrics & systems pharmacology*, 3(8):1-12.
 - Powell DN, Khan AA, Aziz NM, Raiford JP (2007). Dimensionless rainfall patterns for South Carolina. *Journal of Hydrologic Engineering*, 12(1):130-133.
 - Restrepo PJ, Eagleson PS (1982). Identification of independent rainstorms. *J. Hydrol.*, 55:303-319.
 - Sane Y et al (2018). Intensity–duration–frequency (IDF) rainfall curves in Senegal. *Natural Hazards and Earth System Sciences*, 18(7):1849-1866.
 - Schendel T, Thongwichian R (2015). Flood frequency analysis: Confidence interval estimation by test inversion bootstrapping. *Advances in water resources*, 83:1-9.
 - Schendel T, Thongwichian R (2017). Confidence intervals for return levels for the peaks-over-threshold approach. *Advances in water resources*, 99:53-59.
 - Sklar M (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229-231.
 - Sordo-Ward A, Bianucci P, Garrote L, Granados A (2016). The influence of the annual number of storms on the derivation of the flood frequency curve through event-based simulation. *Water*, 8(8):335.
 - Svensson C, Jones DA (2010). Review of rainfall frequency estimation methods. *Journal of Flood Risk Management*, 3(4):296-313.
 - Todorovic P, Vujica M (1969). *Stochastic Process of Precipitation*. Hydrology papers (Colorado State University), no. 35
 - Tsukatani T, Shigemitsu K (1980). Simplified Pearson distributions applied to air pollutant concentration. *Atmospheric Environment* (1967), 14(2):245-253.
 - US Environmental Protection Agency (1997). *Guiding principles for Monte Carlo analysis*. USEPA Risk Assessment Forum, Washington, DC
 - Vandenberghe S, Verhoest NEC, Onof C, De Baets B (2011). A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall. *Water Resources Research*, 47(7):197-203.
 - Vicente-Serrano SM, Beguería-Portugués S (2003). Estimating extreme dry-spell

risk in the middle Ebro valley (northeastern Spain): a comparative analysis of partial duration series with a general Pareto distribution and annual maxima series with a Gumbel distribution. *International Journal of Climatology*, 23(9):1103-1118.

- Urban S, Wang Y, McBean EA, Binns A, Gharabaghi B (2018). Evaluation of Stormwater Infrastructure Design Storms Developed Using Partial Duration and Annual Maximum Series Models. *Journal of Hydrologic Engineering*, 23(12):04018051.
- Wang QJ (1991). The POT model described by the tugenalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, 129(1-4):263-280.
- Wang W, Yin S, Xie Y, Nearing MA (2019). Minimum Inter-Event Times for Rainfall in the Eastern Monsoon Region of China. *Transactions of The ASABE* 62(1):9-18
- Wu YC, Liou JJ, Su YF, Cheng KS (2011). Establishing acceptance regions for L-moments based goodness-of-fit tests for the Pearson type III distribution. *Stochastic environmental research and risk assessment*, 26(6):873-885.
- Xu YP, Booij MJ, Tong YB (2010). Uncertainty analysis in statistical modeling of extreme hydrological events. *Stochastic environmental research and risk assessment*, 24(5):567-578.
- Yoo C, Park C, Jun C (2015). Evaluation of the concept of critical rainfall duration by bivariate frequency analysis of annual maximum independent rainfall event series in Seoul, Korea. *Journal of Hydrologic Engineering*, 21(1):05015016.
- Yue S, Rasmussen P (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14):2881-2898.
- 許恩菁 (1999). 設計暴雨雨型序率模式之研究. 臺灣大學生物環境系統工程學研究所學位論文
- 經濟部水利署. (2009). 莫拉克颱風暴雨量及洪流量分析.
- 臺大生物環境系統工程學系. (2001). 水文設計應用手冊, 第一版. 經濟部水資源局.
- 鄭克聲, 陳葦庭, 葉彙中. (1999). 坡地開發滯留池之水文設計探討. *臺灣水利*, 47(4):41-57.